



DETECTORES DE NOVIDADES E CLASSIFICADORES ESPECIALIZADOS EM SISTEMAS DE SONAR PASSIVO

Victor Hugo da Silva Muniz

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: João Baptista de Oliveira e Souza
Filho

Rio de Janeiro
Agosto de 2019

DETECTORES DE NOVIDADES E CLASSIFICADORES ESPECIALIZADOS
EM SISTEMAS DE SONAR PASSIVO

Victor Hugo da Silva Muniz

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. João Baptista de Oliveira e Souza Filho, D.Sc.

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. Michel Pompeu Tcheou, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
AGOSTO DE 2019

Muniz, Victor Hugo da Silva

Detectores de Novidades e Classificadores Especializados em Sistemas de Sonar Passivo/Victor Hugo da Silva Muniz. – Rio de Janeiro: UFRJ/COPPE, 2019.

XVIII, 78 p.: il.; 29, 7cm.

Orientador: João Baptista de Oliveira e Souza Filho

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2019.

Referências Bibliográficas: p. 72 – 76.

1. Sonar passivo. 2. Sistema de apoio à decisão. 3. Aprendizagem de máquina. I. Souza Filho, João Baptista de Oliveira e. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*“Carpe diem quam minimum
credula postero.”*

*(Aproveita o dia e confia o
mínimo possível no amanhã.)*

(Horácio, Odes I, 11.8)

Agradecimentos

A Deus, pela inspiração que me fez seguir adiante naqueles momentos em que deixei de acreditar em mim. Afinal, sem Fé, não há esperança e a vida perde completamente o sentido.

À COPPE/UFRJ, pela oportunidade de cursar este Mestrado nesta Instituição que é uma referência de ensino.

Ao Instituto de Pesquisas da Marinha (IPqM), pela disponibilização da base de dados utilizada neste trabalho.

Ao meu orientador e amigo de longa data, Prof. João Baptista, pela confiança em meu potencial, suporte constante e conselhos. Sou grato por você sempre tentar extrair o melhor de mim, por mais que eu relute às vezes.

A Prof^a. Mariane e ao Prof. Michel, pela participação na banca e pelas correções propostas.

Ao Prof. Luiz Wagner, pela gentileza em revisar de forma minuciosa esta Dissertação.

Aos meus pais, Carmen Lucia e Hugo, os quais investiram na minha educação desde cedo para que eu me tornasse um profissional bem-sucedido. Em especial, a minha mãe, sempre a meu lado durante os maus momentos, preocupada com o meu bem-estar.

Às minhas amigas, Carla e Patrícia, que contribuíram para tornar esses últimos anos especiais e mais leves. Nunca achei que pudesse ter a sorte de conhecer duas pessoas tão iluminadas em um só lugar. Nossos momentos de descontração e até os de desespero vão ficar para sempre registrados neste capítulo da minha história.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

DETECTORES DE NOVIDADES E CLASSIFICADORES ESPECIALIZADOS EM SISTEMAS DE SONAR PASSIVO

Victor Hugo da Silva Muniz

Agosto/2019

Orientador: João Baptista de Oliveira e Souza Filho

Programa: Engenharia Elétrica

Em submarinos, cabe aos operadores de sonar a tarefa de identificar e classificar contatos de sonar passivo, de forma que possíveis ameaças sejam detectadas. A automatização deste processo é extremamente relevante, visto que facilita o trabalho do profissional desta área, ao exigir um menor esforço físico e mental durante a vigilância. A proposta deste estudo é investigar a eficiência de modelos especializados na constituição de tal sistema, visando a derivar um mecanismo que detecte de forma eficaz navios desconhecidos, bem como identifique corretamente os rótulos daqueles conhecidos. Três níveis de especialização foram considerados: não-especializado, especializado por classes, e especializado por navios, assumindo as seguintes técnicas para a construção do sistema: Análise de Componentes Principais (PCA), Análise de Componentes Principais por *Kernel* (KPCA), Máquinas de Vetor-Suporte de Uma Classe (OCSVM), Modelos de Mistura de Gaussianas (GMM), k -vizinhos mais próximos (kNN), k -vizinhos mais próximos esparsos (s-kNN) e *Local Outlier Factor* (LOF). Experimentos conduzidos com dados reais adquiridos em uma raia acústica mostraram um melhor desempenho dos modelos especializados em navios, que atingiram uma taxa de detecção de novidades de 83,4%, conjugada com uma taxa média de reconhecimento de classes conhecidas de 90,5%. Em relação especificamente à tarefa de classificação das classes conhecidas, 98,7% são corretamente rotuladas.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SPECIALIZED NOVELTY DETECTORS AND CLASSIFIERS IN PASSIVE SONAR SYSTEMS

Victor Hugo da Silva Muniz

August/2019

Advisor: João Baptista de Oliveira e Souza Filho

Department: Electrical Engineering

In submarines, sonar operators have the task of identifying and classifying passive sonar contacts, so that possible threats are detected. The automation of this process is extremely relevant, since it facilitates the work of the professional of this area, requiring less physical and mental efforts during the surveillance. The proposal of this study is to investigate the efficiency of specialized models in the constitution of such a system, aiming to derive a mechanism that effectively detects unknown ships, as well as correctly identifies the labels of those already known. Three levels of specialization were considered: non-specialized, specialized in classes, and specialized in ships, assuming the following techniques for the construction of the system: Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), One-Class Support-Vector Machines (OCSVM), Gaussian Mixture Models (GMM), k -Nearest Neighbors (kNN), sparse k -Nearest Neighbors (s-kNN) and Local Outlier Factor (LOF). Experiments conducted with real data acquired on an acoustic lane showed a better performance of the models specialized in ships, which reached a novelty detection rate of 83.4%, conjugated with an average recognition rate of known classes of 90.5%. Regarding specifically the task of classifying the known classes, 98.7% are correctly labeled.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
Lista de Símbolos	xv
Lista de Abreviaturas	xvii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	1
1.3 Metodologia	2
1.4 Contribuições	2
1.5 Organização	3
2 Automatização em Sonar Passivo	4
2.1 Sistemas Automáticos de Classificação	6
2.1.1 Classificação Automática de Contatos	7
2.1.2 Detecção de Novidades	7
2.2 Decomposição do Problema Multiclasse	8
2.3 Técnicas para a Detecção de Novidades	9
2.3.1 Análise de Componentes Principais	10
2.3.2 Análise de Componentes Principais por Kernel	13
2.3.3 Máquinas de Vetor-Suporte de Uma Classe	16
2.3.4 Modelos de Mistura de Gaussianas	19
2.3.5 k -vizinhos mais próximos	21
2.3.6 k -vizinhos mais próximos esparso	23
2.3.7 <i>Local Outlier Factor</i>	26
3 Sistema Proposto	30
3.1 Níveis de Especialização	30
3.1.1 Global	31

3.1.2	Especializado em Classes	32
3.1.3	Especializado em Navios	33
3.2	Topologias	33
3.2.1	Unânime	34
3.2.2	Competitiva	35
3.2.3	Hierárquica	36
3.3	Avaliação de Desempenho	38
3.3.1	Métodos de Reamostragem	38
3.3.2	Índices de Desempenho	42
3.4	Sintonia de Hiperparâmetros	45
4	Resultados	48
4.1	Base de Dados	48
4.2	Particionamento dos Dados	50
4.3	Hiperparâmetros escolhidos	51
4.4	Sistema Detector de Novidades	52
4.4.1	PCA	52
4.4.2	KPCA	54
4.4.3	OCSVM	56
4.4.4	GMM	58
4.4.5	kNN	59
4.4.6	s-kNN	61
4.4.7	LOF	64
4.4.8	Comparativo das Técnicas	65
4.5	Classificação de Classes de Contatos	67
4.6	Discussão	68
5	Conclusões e Trabalhos Futuros	70
	Referências Bibliográficas	72
A	Produção Científica	77

Lista de Figuras

2.1	Sistema de Sonar Ativo com operação monoestática.	4
2.2	Sistema de Sonar Ativo com operação biestática.	5
2.3	Sistema de Sonar Passivo.	6
2.4	Exemplo que ilustra um conjunto de dados bidimensional arbitrário \mathcal{X} (a) projetado em um subespaço unidimensional, representado pela reta u , correspondente à direção de maior variância dos dados (b). . .	11
2.5	Erro de reconstrução.	12
2.6	Exemplo ilustrando a técnica KPCA.	14
2.7	Exemplo que ilustra um conjunto de dados bidimensional arbitrário \mathcal{X} (a), bem como este mesmo conjunto quando projetado em um espaço de maior dimensão (b).	17
2.8	Exemplo contendo: (a) um conjunto de dados bidimensional arbitrário \mathcal{X} ; (b) os mesmos dados projetados em um espaço de maior dimensão.	18
2.9	Exemplo ilustrando uma mistura de Gaussianas arbitrária para $\mathbf{x} \in \mathbb{R}^2$. Em (a), são exibidos os contornos de probabilidade constante para cada componente da mistura. Em (b), o esboço da superfície contendo a densidade final.	21
2.10	Exemplo ilustrando a proximidade entre um conjunto de dados de treino e um dado de teste pertencente a classe conhecida (a) ou não (b).	22
2.11	Técnica k NN para a constituição de um classificador.	23
2.12	Exemplo ilustrando a diferença entre os conjuntos de treinamento utilizados pelas técnicas k NN (a) e s - k NN (b).	24
2.13	Ilustração do conceito de densidade local para $k = 3$	27
2.14	Cardinalidade de N_k para diferentes valores de k	27
2.15	Distância de alcançabilidade $da_k(\mathbf{x}, \mathbf{y})$ para $k = 4$ em 2 casos hipotéticos.	28

3.1	Exemplo que ilustra um conjunto de dados bidimensional arbitrário \mathcal{X} de 2 classes distintas separados através de uma abordagem especializada: em classes (a) e em subclasses (b).	31
3.2	Detector Global ou não-especializado.	32
3.3	Detectors Especializados em Classes.	32
3.4	Detectors Especializados em Navios.	33
3.5	Topologia Tradicional.	34
3.6	Topologia Unânime.	35
3.7	Topologia Competitiva.	36
3.8	Topologia Hierárquica.	37
3.9	Representação do processo de amostragem.	38
3.10	Ilustração do método Validação Cruzada via <i>k-fold</i> para $k = 10$ em suas duas primeiras rodadas.	39
3.11	Ilustração da dinâmica dos conjuntos de Treinamento, Validação e Teste no processo de desenvolvimento e avaliação de um modelo.	40
3.12	Ilustração do método <i>k-fold</i> aninhado para $k^{(1)} = k^{(2)} = 10$	41
3.13	Exemplo ilustrando três CODs em um problema arbitrário.	44
3.14	Representação dos subconjuntos do conjunto de validação utilizados para o cálculo da $AUC_{i,m}$	47
4.1	Espectrogramas referentes a quatro cenários de operação.	50
4.2	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica PCA.	53
4.3	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica PCA.	53
4.4	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica KPCA.	55
4.5	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica KPCA.	56
4.6	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica OCSVM.	57
4.7	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica OCSVM.	58
4.8	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica GMM.	59
4.9	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica GMM.	59
4.10	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica kNN.	60

4.11	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica kNN.	61
4.12	Percentual médio de amostras para os diferentes níveis de especialização considerando a faixa de ϵ investigada para a técnica s-kNN.	62
4.13	Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica s-kNN ($\epsilon = 0,4$).	63
4.14	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica s-kNN.	64
4.15	Diagrama de caixas da área sob a COD para diferentes topologias e especializações considerando a técnica LOF.	65
4.16	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando a técnica LOF.	65
4.17	Diagrama de caixa da área sob a COD para a melhor combinação de topologia e especialização identificada para cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).	66
4.18	Diagrama de diferença crítica obtido na comparação de <i>ranks</i> médios das medidas AUC considerando cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).	66
4.19	COD para a melhor combinação de topologia e especialização identificada para cada técnica.	67
4.20	COD para a técnica kNN sob o enfoque EN-HI.	69

Lista de Tabelas

3.1	Hiperparâmetros presentes em cada uma das técnicas.	46
4.1	Distribuição do quantitativo de corridas (NC) e de espectros (NE) para as diferentes classes e navios.	49
4.2	Faixas de valores de hiperparâmetros investigados para cada técnica adotada como geradora do escore de novidade.	51
4.3	Percentual de variância explicada (Var_e) para os detectores baseados na técnica PCA.	52
4.4	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica PCA.	53
4.5	Percentual de variância explicada (Var_e) para os detectores baseados na técnica KPCA.	54
4.6	Largura do <i>kernel</i> (γ) para os detectores baseados na técnica KPCA.	54
4.7	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica KPCA.	55
4.8	Largura do <i>kernel</i> (K_s) para os detectores baseados na técnica OCSVM.	56
4.9	Penalidade associada às variáveis de folga (ν) para os detectores baseados na técnica OCSVM.	57
4.10	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica OCSVM.	58
4.11	Número de distribuições base da mistura Gaussiana (m) para os detectores baseados na técnica GMM.	58
4.12	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica GMM.	59
4.13	Número de vizinhos (k) para os detectores baseados na técnica kNN.	60
4.14	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica kNN.	61
4.15	Mediana da AUC para as diferentes topologias e especializações considerando a faixa de ϵ investigada para a técnica s-kNN.	62
4.16	Número de vizinhos (k) para os detectores baseados na técnica s-kNN ($\epsilon = 0,4$).	63

4.17	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica s-kNN.	64
4.18	Número de vizinhos (k) para os detectores baseados na técnica LOF.	64
4.19	<i>Rank</i> médio para as diferentes topologias e especializações considerando a técnica LOF.	65
4.20	<i>Rank</i> médio para a melhor topologia e especialização de cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).	67
4.21	TU e TK para a melhor topologia e especialização de cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).	68
4.22	TA médio para cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF) de acordo com o nível de especialização empregado (EC ou EN) para as topologias UN/CO.	68
4.23	Resultados descritos na literatura quanto à detecção de novidades.	69
4.24	Resultados descritos na literatura quanto à classificação de contatos conhecidos em rótulos de classe.	69

Lista de Símbolos

Q_j	Quantitativo de navios da j -ésima classe conhecida, p. 42
R_j	Quantitativo de navios da j -ésima classe desconhecida, p. 43
TA	Acurácia média de acertos, p. 45
TA_{C_j}	Acurácia média de acertos no reconhecimento da j -ésima classe, p. 45
$TA_{N_i^j}$	Acurácia média de acertos no reconhecimento do i -ésimo navio da j -ésima classe, p. 45
TK	Taxa média de detecção de eventos conhecidos, p. 43
TK_{C_j}	Taxa média de detecção da j -ésima classe conhecida, p. 42
$TK_{N_i^j}$	Taxa média de detecção do i -ésimo navio da j -ésima classe conhecida, p. 42
TU	Taxa média de detecção de novidades, p. 43
TU_{C_j}	Taxa média de detecção da j -ésima classe desconhecida, p. 43
$TU_{N_i^j}$	Taxa média de detecção do i -ésimo navio da j -ésima classe desconhecida, p. 43
\mathcal{F}	<i>Feature space</i> ou espaço de características, p. 13
\mathcal{K}	Conjunto contendo os contatos conhecidos, p. 50
\mathcal{K}_N^j	Conjunto contendo os navios da j -ésima classe conhecida, p. 42
\mathcal{K}_C	Conjunto contendo os contatos conhecidos rotulados por classe, p. 42
\mathcal{K}_N	Conjunto contendo os contatos conhecidos rotulados por navio, p. 42

\mathcal{U}	Conjunto contendo os contatos desconhecidos, p. 50
\mathcal{U}_C^j	Conjunto contendo os navios da j -ésima classe desconhecida, p. 43
\mathcal{U}_C	Conjunto contendo os contatos desconhecidos rotulados por classe, p. 43
\mathcal{U}_N	Conjunto contendo os contatos desconhecidos rotulados por navio, p. 51

Lista de Abreviaturas

CM-MLP	<i>Class-Modular Multi-layer Perceptron</i> ou Perceptron Multicamadas de Classe Modular, p. 7
COD	Curva de Operação do Detector, p. 44
CO	(Topologia) Competitiva, p. 36
CP	Curvas Principais, p. 7
DC	Diferença Crítica, p. 52
DEMON	<i>DEModulation On Noise</i> , p. 5
EC	(Nível de) Especialização por Classes, p. 32
EM	Maximização da Esperança, p. 20
EN	(Nível de) Especialização por Navios, p. 33
FA	<i>Factor Analysis</i> ou Análise Fatorial, p. 8
GL	(Nível de) Especialização Global ou não-especializado, p. 31
GMM	<i>Gaussian Mixture Models</i> ou Modelos de Mistura de Gaussianas, p. 2
HI	(Topologia) Hierárquica, p. 37
IIQ	Intervalo Interquartil, p. 53
IPqM	Instituto de Pesquisas da Marinha, p. 48
KPCA	<i>Kernel Principal Component Analysis</i> ou Análise de Componentes Principais por <i>Kernel</i> , p. 2
LOFAR	<i>LOw Frequency Analysis and Recording</i> , p. 5
LOF	<i>Local Outlier Factor</i> , p. 2

NC	Número de Corridas, p. 49
NE	Número de Espectros, p. 49
NNd	<i>Nearest Neighbor density</i> , p. 3
OCSVM	<i>One-Class Support-Vector Machines</i> ou Máquinas de Vetor-Suporte de Uma Classe, p. 2
OMP	<i>Orthogonal Matching Pursuit</i> , p. 25
PCA	<i>Principal Component Analysis</i> ou Análise de Componentes Principais, p. 2
RF	<i>Random Forest</i> , p. 3
ROC	<i>Receiver Operating Characteristic</i> , p. 43
SAC	Sistemas Automáticos de Classificação, p. 1
SMO	<i>Sequential Minimal Optimization</i> , p. 19
SVM	Máquinas de Vetor-Suporte, p. 17
Sonar	<i>SOUND Navigation and Ranging</i> , p. 1
UN	(Topologia) Unânime, p. 35
kNN	<i>k-Nearest Neighbors</i> ou <i>k</i> -vizinhos mais próximos, p. 2
s-kNN	<i>Sparse k-Nearest Neighbors</i> ou <i>k</i> -vizinhos mais próximos esparsos, p. 2

Capítulo 1

Introdução

Sistemas de sonar passivo representam o principal mecanismo de defesa de submarinos, sendo utilizados para detectar o ruído emitido por outras embarcações ou animais marinhos. De posse de características tonais e espectrais deste sinal, os operadores de sonar, treinados especialmente para esta tarefa, identificam a natureza deste contato e avaliam se este representa alguma ameaça.

1.1 Motivação

Técnicas de Aprendizagem de Máquina, explorando dados representativos do problema, podem auxiliar o trabalho do operador de sonar, aliviando o esforço de vigilância exigido no desempenho desta função.

Todavia, a obtenção de bases de dados envolvendo uma ampla variedade de embarcações e cenários operativos é inviável, devido aos custos financeiros elevados e às dificuldades operacionais associadas a esta tarefa, bem como pela natureza sigilosa de tais informações.

Por esse motivo, sistemas automáticos de classificação (SAC) costumam ser gerados a partir de uma base de dados contendo um número restrito de classes. Portanto, é imprescindível dispor de ferramentas para a identificação daquelas classes não utilizadas durante o seu desenvolvimento. Em outras palavras, além de o sistema efetuar a classificação dos contatos, deve ser capaz de detectar novidades, alertando o operador de sonar para a tomada das contramedidas necessárias.

1.2 Objetivos

Como objetivo principal, pretende-se desenvolver um sistema que realize a detecção de novidades de forma eficiente, auxiliando o trabalho de decisão do operador de sonar. Destaca-se que, neste caso, as novidades são contatos pertencentes a classes

não contempladas durante o desenvolvimento do modelo.

Como objetivo secundário, o sistema também deve classificar corretamente os contatos conhecidos, de forma que os identifique através de seus rótulos de classe.

1.3 Metodologia

O sistema de apoio à decisão criado utilizou detectores especializados em navios para um reconhecimento de novidades mais eficaz, além de novos mecanismos para a fusão destes detectores. Para a construção deste sistema, diferentes técnicas clássicas foram avaliadas: Análise de Componentes Principais (PCA), Análise de Componentes Principais por *Kernel* (KPCA), Máquinas de Vetor-Suporte de Uma Classe (OCSVM), Modelos de Mistura de Gaussianas (GMM), k -vizinhos mais próximos (kNN) e *Local Outlier Factor* (LOF) [1]. Também foi proposta uma nova técnica que introduz mecanismos de esparsidade no algoritmo kNN, intitulada aqui como kNN esparso (s-kNN).

Ademais, ressalta-se que os ensaios executados utilizaram uma base de dados constituída pelo ruído irradiado por navios reais em corridas de prova na raia acústica da Marinha do Brasil, e contemplam um total de 28 navios pertencentes a 8 classes, submetidos a diferentes condições operacionais.

1.4 Contribuições

A contribuição principal deste trabalho reside em avaliar, sobre a base de dados em questão, se um maior nível de especialização para a geração dos modelos leva a um ganho na eficiência quanto à detecção de novidades. Adicionalmente, novas topologias para a integração dos detectores são investigadas, bem como o impacto da introdução de mecanismos de esparsidade visando à simplificação dos modelos. Em síntese, o sistema de apoio à decisão final permite um auxílio direto ao trabalho dos operadores de sonar, uma vez que contempla a dupla função de detectar novidades e identificar o rótulo de classes conhecidas, ao considerar um nível de especialização maior do que o empregado em trabalhos anteriores [2].

Em paralelo ao desenvolvimento desta Dissertação de Mestrado, três artigos científicos foram produzidos e submetidos a Simpósios e Congressos Nacionais em 2019, conforme detalhado no Apêndice A. O primeiro, aprovado no 14º Simpósio Brasileiro de Automação Inteligente, trata da construção de detectores de novidades especializados em navios, integrados por diferentes topologias e baseados nas técnicas PCA, GMM e kNN, que alcançou ganhos expressivos em relação a trabalhos anteriores.

O segundo artigo foi submetido ao XVI Congresso Brasileiro de Inteligência Computacional e está em fase de avaliação. Este também discute o desenvolvimento de um detector de novidades utilizando o algoritmo kNN, porém comparando este a outras técnicas, tais como o *Nearest Neighbor density* (NNd) [3] e o *Local Outlier Factor* (LOF) [4].

O último artigo também foi submetido ao XVI Congresso Brasileiro de Inteligência Computacional, e, assim como o anterior, também está em fase de avaliação. Este investiga o problema da classificação de navios por meio de uma estratégia hierárquica, construída com o auxílio das técnicas clássicas: *Random Forest* (RF) [5] e kNN.

1.5 Organização

O trabalho está organizado da forma descrita a seguir. O Capítulo 2 apresenta alguns princípios relativos à operação do sonar passivo, revisita alguns trabalhos associados e descreve os conceitos fundamentais e as técnicas utilizadas para o desenvolvimento dos modelos. No Capítulo 3 é descrita a metodologia adotada para a construção do sistema proposto, explorando diferentes combinações de níveis de especialização e topologias. Neste capítulo são também discutidas algumas medidas de desempenho de tal sistema, as quais servirão para a seleção dos hiperparâmetros e para a escolha da melhor abordagem ao problema. No Capítulo 4 é apresentada a base de dados e são discutidos os resultados. Por fim, no Capítulo 5, têm-se as conclusões e os trabalhos futuros.

Após as referências bibliográficas, o Apêndice A apresenta uma breve descrição dos artigos científicos desenvolvidos ao longo deste trabalho.

Capítulo 2

Automatização em Sonar Passivo

O Sonar, abreviação de *SOund Navigation and Ranging*, é uma tecnologia amplamente utilizada para a exploração e mapeamento do oceano, visto que as ondas acústicas sofrem uma atenuação significativamente menos expressiva em meio aquático do que as ondas de radar e de luz [6]. Ademais, é empregado em sistemas de comunicação, no desenvolvimento de cartas náuticas, na busca por artefatos no fundo do mar, e na identificação de animais marinhos, como cardumes, golfinhos e baleias. No contexto militar, visa à descoberta de ameaças submarinas para a navegação, tais como embarcações inimigas, minas e torpedos [7].

Existem dois grupos de sistemas de sonar: o ativo e o passivo. No caso ativo, o sistema é formado por transdutores que emitem pulsos acústicos na água. Com isso, um alvo por eles atingido produz um eco, isto é, reflexões de tais pulsos, os quais são coletados e analisados para a extração de informações, em especial a distância ao alvo. Esse processo está ilustrado na Figura 2.1, na qual o alvo em questão é um navio.

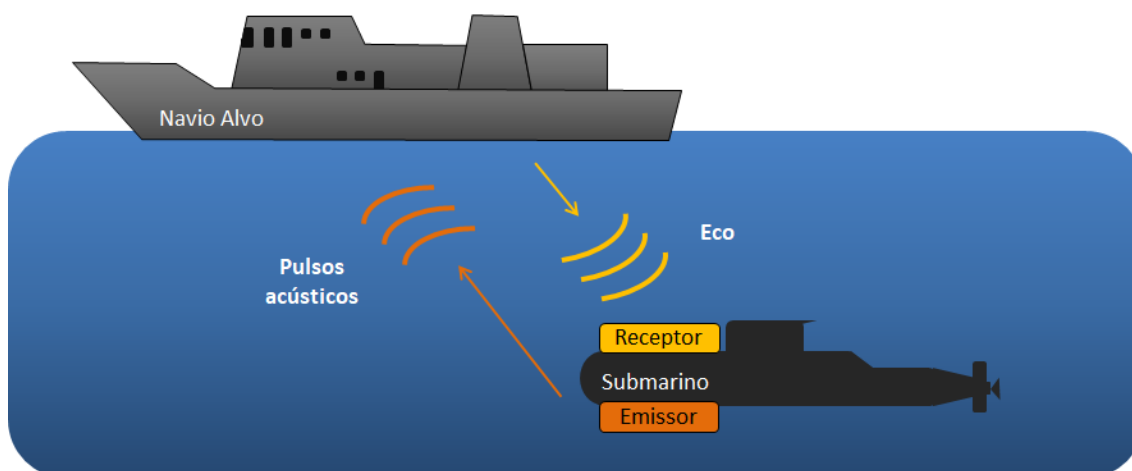


Figura 2.1: Sistema de Sonar Ativo com operação monoestática.

O sonar ativo pode ser classificado quanto à localização do emissor e do receptor [8]. Quando estes se encontram juntos, o sonar realiza uma operação monoestática, caso apresentado na Figura 2.1. Entretanto, quando o emissor e o receptor estão separados a uma distância consideravelmente maior do que as distâncias entre o emissor-alvo e o alvo-receptor, a operação é dita biestática, conforme ilustrado na Figura 2.2.

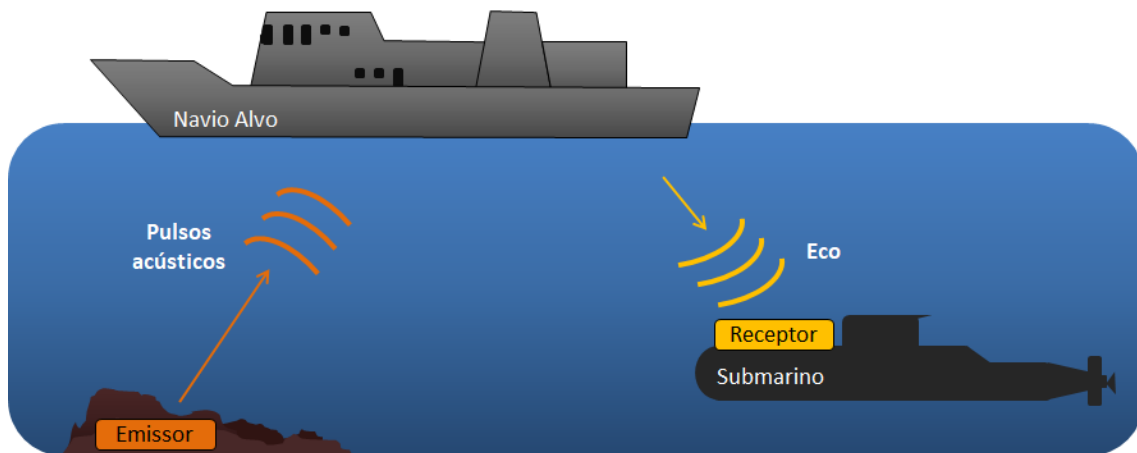


Figura 2.2: Sistema de Sonar Ativo com operação biestática.

Na modalidade passiva, o sistema de sonar apenas detecta as ondas acústicas que chegam aos hidrofones, portanto sem emitir qualquer sinal próprio. Assim, os sensores sonoros apenas captam as vibrações emitidas pelo contato, as quais são subsequentemente processadas visando à identificação da sua assinatura acústica. Dessa forma, o sonar passivo assume um papel vantajoso em embarcações de caráter militar que não podem ser descobertas, tais como os submarinos. A Figura 2.3 ilustra tal dinâmica, na qual há um navio alvo a ser identificado através de seu ruído.

O ruído captado por um sistema de sonar passivo pode ser oriundo de múltiplas fontes, entre elas: do sistema de propulsão do navio; de máquinas auxiliares, tais como geradores, bombas e equipamentos de ar condicionado; da hélice, e da passagem do fluxo irregular de água pelo navio em movimento, o qual é chamado de ruído hidrodinâmico [9].

A partir dos sinais recebidos e devidamente processados, operadores de sonar treinados utilizam as ferramentas de análise LOFAR (*LOw Frequency Analysis and Recording*), DEMON (*DEModulation On Noise*), e até a própria escuta de tais sinais para a identificação dos contatos, realizada por meio de suas características tonais e espectrais [10].

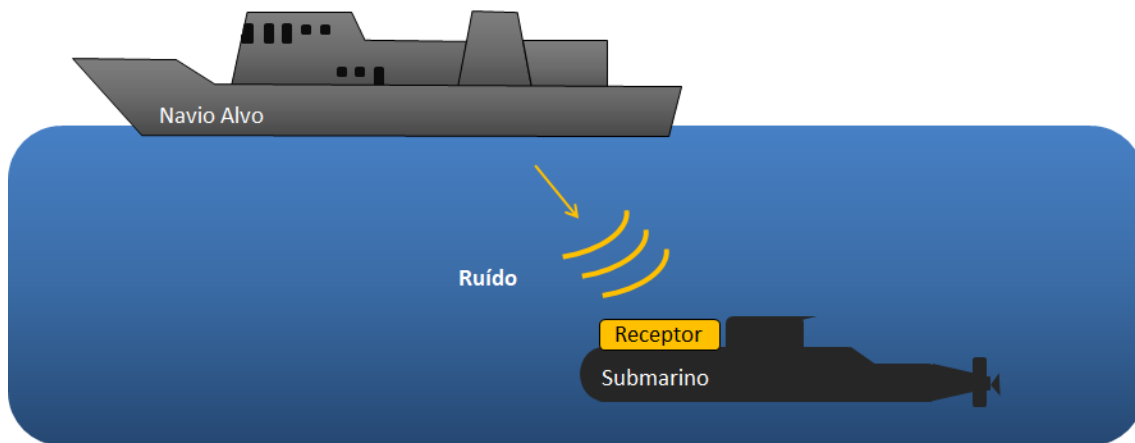


Figura 2.3: Sistema de Sonar Passivo.

2.1 Sistemas Automáticos de Classificação

Em operações militares, a atividade de classificação de possíveis ameaças em submarinos é nomeada como classificação de contatos, sendo realizada por meio da aquisição e processamento de sinais de sonar passivo. Um número expressivo de técnicas de Aprendizagem de Máquina vem sendo empregado para o desenvolvimento de sistemas de apoio à decisão. No contexto abordado, os mecanismos citados visam a auxiliar o operador de sonar, ao automatizar as tarefas de detecção e identificação dos contatos. Tal procedimento alivia o trabalho do profissional, visto que reduz a necessidade de constante vigilância humana, evitando um maior desgaste físico e mental do operador, o qual pode se dedicar mais intensivamente apenas a situações de alto risco.

Dada a existência de inúmeras embarcações, cada qual possuindo uma assinatura acústica própria, é plausível que o sistema se depare com uma embarcação desconhecida (ou novidade), isto é, que não faça parte do conjunto de características espectrais previamente conhecidas. Esta situação é comum, considerando-se ainda o fato de que algumas embarcações têm natureza sigilosa. Portanto, é imprescindível que o sistema gerado seja capaz de reconhecer novidades, assegurando que classificações errôneas não ocorram, as quais podem ter consequências graves.

As subseções seguintes sintetizam as principais abordagens utilizadas e os resultados obtidos por alguns trabalhos da literatura, considerando, primeiramente, o problema da classificação de contatos de sistemas de sonar passivo, para, em seguida, abordar aqueles dedicados à detecção de novidades.

2.1.1 Classificação Automática de Contatos

Os SAC vêm se mostrando bem sucedidos no auxílio ao processo de tomada de decisão do operador de sonar. Neste cenário, redes neurais especialistas podem ser empregadas para integrar tais sistemas. Em [11], foi proposto um sistema de classificação baseado em um Perceptron Multicamadas de Classe Modular (CM-MLP), o qual opera de forma flexível e em tempo real, alcançando uma acurácia de 84,4% de acertos de classificação. Tal análise foi baseada em sinais reais produzidos por 34 navios pertencentes a 8 classes de contato.

Redes neurais convolucionais também foram investigadas para tratar o problema de classificação [12], produzindo uma acurácia média de 95,5% para um conjunto de teste formado por 25 navios pertencentes a 8 classes.

Outra abordagem explorada na literatura foi a técnica das Curvas Principais (CP) [13], a qual se destaca por realizar uma expressiva redução de dimensionalidade para dados não-lineares. Dois trabalhos que realizaram ensaios com a mesma base de dados que esta Dissertação podem ser destacados: (i) em [14], foi desenvolvido um SAC para sinais de sonar passivo constituído por curvas principais hierárquicas. Nesse sistema, de caráter especialista, cada classe foi representada por um banco de curvas, e os eventos foram rotulados de acordo com a distância às curvas principais integrantes deste banco, atingindo uma eficiência média de classificação da ordem de 96,2%; (ii) Um classificador hierárquico baseado em CP foi proposto em [15], alcançando uma acurácia média de 91,4% quanto à identificação de navios. Neste sistema, primeiramente, é identificada a classe, de forma similar a [14]. Em sequência, o navio mais provável é definido através do valor do índice de projeção nesta curva, com base na avaliação dos critérios Limites Heurísticos, Limiares ótimos e Classificação *Fuzzy*, com resultados equivalentes.

Com base nos resultados supracitados, observa-se que a criação de modelos hierárquicos e especializados em classes estão associados a taxas de acertos de classificação expressivas, as quais podem ser ainda maiores caso ocorra uma maior especialização do sistema, dada diretamente no âmbito dos navios.

2.1.2 Detecção de Novidades

Dentro do contexto tratado, o problema da detecção de novidades se resume em identificar a presença de embarcações desconhecidas através dos sinais provenientes do sistema de sonar passivo. Cada embarcação apresenta uma assinatura acústica característica, a qual deve permitir sua distinção em relação a outras existentes. Portanto, o operador de sonar precisa ser alertado quando esta análise se depara com um sinal diferente do habitual, isto é, uma novidade.

Um sistema de reconhecimento para classes desconhecidas foi desenvolvido por

meio de componentes principais especializadas nas classes conhecidas em [16]. A partir do ruído irradiado por 25 navios pertencentes a 8 classes distintas, dentre as quais 4 classes foram tratadas como conhecidas, o sistema proposto apresentou uma eficiência média de detecção de novidades de 62,9%, ao considerar uma detecção de classes conhecidas de 70,0%.

Em [2], uma variedade de técnicas foi investigada para a constituição de um sistema de detecção de novidades explorando uma topologia hierárquica, considerando as seguintes técnicas: PCA, KPCA, kNN, GMM, CP e Análise Fatorial (FA). No sistema em questão, foram desenvolvidos detectores classe-especialistas, os quais foram avaliados em experimentos que utilizaram a mesma base de estudo deste trabalho. Ao se utilizar 4 classes conhecidas e 4 classes desconhecidas, obteve-se uma detecção de novidades de aproximadamente 80,0%, para 90,0% de detecção de classes conhecidas, através da técnica kNN.

Máquinas de Vetor-Suporte de Uma Classe também foram aplicadas para lidar com esta questão em [9]. Primeiramente, foi efetuado um pré-processamento para a redução dimensional dos sinais envolvidos, seguida de um encapsulamento das classes conhecidas através de um modelo OCSVM. As novidades assumidas eram os sinais que não foram encapsulados por este modelo. Assumindo uma taxa de detecção de novidades de 75,0%, obteve-se uma detecção de sinais conhecidos de 62,4%, considerando uma base de dados constituída por 4 classes de embarcações. Destaca-se que em seus experimentos foram avaliadas todas as combinações possíveis de classes, nas quais 3 foram tratadas como conhecidas, enquanto a restante assumida como novidade.

2.2 Decomposição do Problema Multiclasse

Uma vez que o sistema de sonar passivo trabalha com o reconhecimento de múltiplas classes de navios, deve-se buscar o tratamento mais adequado para abordar este problema.

Em um problema em que se opera com M classes distintas, uma possível abordagem é considerar que todas as classes são tratadas conjuntamente, logo tanto a extração dos atributos quanto a decisão caberão a um modelo único. Assim, dado um conjunto com M classes, segundo esta proposta, um único modelo atribuirá um dos M rótulos disponíveis para um dado de entrada.

Entretanto, uma abordagem alternativa é aquela onde o problema supracitado pode ser decomposto em uma série de problemas de classificação binária, em que cada um deles terá associado um modelo de decisão construído segundo um determinado algoritmo e com base em uma fração de dados representativa de 2 classes. Dois métodos comuns para se realizar esta decomposição são o *one-against-one* (um-

contra-um) e o *one-against-all* (um-contra-todos) [17].

No *one-against-one*, são criados subconjuntos de dados considerando todas as combinações 2 a 2 possíveis das M classes do problema original. Para cada uma das $M(M-1)/2$ combinações possíveis, é gerado um modelo de decisão. Assim, supondo um problema com 4 classes (A, B, C, D), 6 problemas binários seriam gerados (A \times B, A \times C, A \times D, B \times C, B \times D, C \times D), e, portanto, 6 modelos seriam criados.

Já pelo método *one-against-all*, um problema multiclasse de M classes é decomposto em M problemas de duas classes, sendo o conjunto de treinamento do modelo formado por exemplos de uma classe principal e por uma “superclasse” constituída pelos exemplos das classes restantes. Neste caso, ao se considerar o mesmo problema original com 4 classes (A, B, C, D), 4 problemas binários seriam gerados (A \times BCD, B \times ACD, C \times ABD, D \times ABC), e, assim, 4 modelos seriam utilizados.

Uma outra alternativa consiste em decompor o problema multiclasse original em M problemas de classe única, uma vertente nomeada de *one-class classification* [18] [19]. Neste caso, assume-se um conjunto de treinamento do modelo constituído apenas por exemplos de uma única classe específica por vez. Sob esta perspectiva, o problema de 4 classes (A, B, C, D) seria tratado como 4 problemas de classe única (A, B, C, D), e assim, 4 modelos de decisão seriam desenvolvidos. Esta alternativa possui como diferencial das anteriores, o fato de que o sistema de classificação obtido é escalável, dado que novas classes podem ser adicionadas ao problema sem qualquer impacto ao conhecimento previamente adquirido. Contudo, seu desempenho tende a ser inferior relativamente às outras abordagens, uma vez que os classificadores são construídos isoladamente, sem interação com qualquer outra classe. Esta abordagem será aplicada para a construção dos detectores de novidades especializados, a serem descritos na Seção 3.1.

2.3 Técnicas para a Detecção de Novidades

De modo generalizado, a detecção de novidades pode ser compreendida como a tarefa de reconhecer se dados de teste diferem sob algum aspecto de dados que estão disponíveis durante o treinamento de um modelo [20]. Assim, um exemplo não-contemplado no treinamento é testado via sua comparação com o modelo de normalidade gerado para o treinamento, de forma que é produzido um escore de dissimilaridade, o qual é comparado a um limiar de decisão. A partir do modelo, classifica-se o exemplo testado como novidade se esta medida superar um dado limiar. Ressalta-se que o modelo de normalidade supracitado é desenvolvido com base no conceito de *one-class classification*, isto é, o modelo explora apenas características relativas à classe definida como conhecida, a qual deve ser distinguida de qualquer outra classe indisponível na etapa de treinamento.

Ao longo das décadas, o número de técnicas propostas para a detecção de novidades na área de Aprendizagem de Máquina se mostrou crescente [20]. Neste trabalho foram avaliadas as mais comuns, baseadas em estratégias variadas: reconstrução, descrição do domínio, probabilidade e distância.

Na estratégia baseada em reconstrução, busca-se encontrar representações para o conjunto de dados da classe conhecida. Assim, quando um dado desconhecido é representado desta forma, este costuma apresentar um maior erro do que o obtido para um dado supostamente pertencente às classes conhecidas. A PCA e a KPCA são exemplos de métodos desta categoria ([21],[22]).

Entre métodos que consideram a descrição de domínio, o OCSVM se destaca [23], utilizando funções discriminativas para delimitar uma fronteira do espaço que defina uma região contendo a maior parte dos dados conhecidos. Assim, os dados situados fora desta região são assumidos como novidades.

Para a abordagem que faz uso da probabilidade, caso do GMM [24], busca-se estimar a função de densidade de probabilidade do conjunto definido pela classe conhecida. Dessa forma, dados pertencentes a regiões de baixa densidade de probabilidade tendem a ser classificados como novidades.

Por fim, a estratégia baseada em distâncias é aquela que faz uso de uma medida da distância entre os dados. Para tal, convencionou-se que dados conhecidos tendem a estar mais agrupados, logo estão mais próximos entre si do que exemplos não presentes no treinamento do modelo. Algoritmos clássicos dentro desta categoria são o LOF [25] e o kNN [26] - o qual também é muito utilizado em problemas de classificação ([27], [28]).

A seguir, as técnicas supracitadas serão descritas de forma detalhada.

2.3.1 Análise de Componentes Principais

A PCA (*Principal Component Analysis*) é uma técnica de redução dimensional que projeta dados de P dimensões em um subespaço ortogonal de dimensão $L \ll P$ [29]. Tais projeções definem novas variáveis descorrelacionadas e de máxima variância, representando direções ótimas para a descrição dos dados, as quais levam a um valor esperado mínimo do erro quadrático de reconstrução.

A Figura 2.4 exemplifica o conceito desta técnica para um conjunto de dados bi-dimensional arbitrário $\mathcal{X} = \{\mathbf{x} \mid \mathbf{x} = (x_1, x_2)\}$, que são projetados em um subespaço unidimensional correspondente à direção de maior variância.

Considerando um vetor de atributos $\mathbf{x} \in \mathbb{R}^P$ com média zero; um inteiro $L \ll P$, representando o quantitativo de componentes principais a expandir \mathbf{x} , e $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_L]^T$, $\mathbf{y} \in \mathbb{R}^L$, são determinadas as L componentes lineares descorrelacionadas de \mathbf{x} segundo:

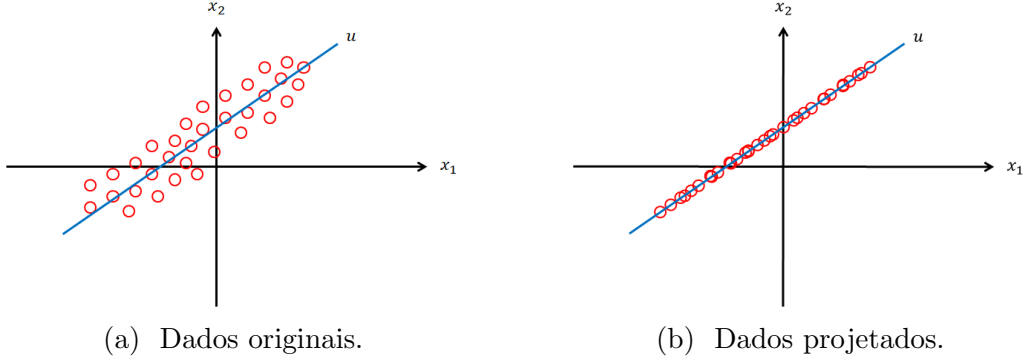


Figura 2.4: Exemplo que ilustra um conjunto de dados bidimensional arbitrário \mathcal{X} (a) projetado em um subespaço unidimensional, representado pela reta u , correspondente à direção de maior variância dos dados (b).

$$y_i = \mathbf{u}_i^T \mathbf{x} \in \mathbb{R}, \quad \mathbf{u}_i \in \mathbb{R}^P, \quad i = 1, 2, \dots, L, \quad (2.1)$$

onde as direções \mathbf{u}_i são definidas de forma que a variância de y_i seja maximizada, impondo-se também as seguintes restrições:

$$\mathbf{u}_i^T \mathbf{u}_i = 1, \quad \mathbf{u}_i^T \mathbf{u}_j = 0 \quad \text{se } i \neq j, \quad e \quad \text{Var}(y_1) \geq \text{Var}(y_2) \geq \dots \geq \text{Var}(y_L) > 0. \quad (2.2)$$

Pode-se mostrar [29] que as L componentes principais de \mathbf{x} (\mathbf{u}_i) podem ser calculadas a partir dos L autovetores de sua matriz de covariância $\Sigma_{\mathbf{x}} \doteq \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ associados aos seus L maiores autovalores. Tais autovetores serão os vetores-coluna que constituem a matriz de transformação $\mathbf{U}_L \in \mathbb{R}^{P \times L}$, tal que $\mathbf{U}_L = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L]$, definindo direções mutuamente ortogonais formadoras de uma base associada ao subespaço de máxima variância. Claramente, a projeção dos dados neste subespaço resulta no vetor:

$$\mathbf{y}_L = \mathbf{U}_L^T \mathbf{x}. \quad (2.3)$$

Quanto à reconstrução do vetor de atributos original, esta é definida por:

$$\hat{\mathbf{x}} = \mathbf{U}_L \mathbf{y}_L. \quad (2.4)$$

Assim, o erro quadrático de reconstrução do vetor de atributos (Figura 2.5), ou distância ao subespaço associado a \mathbf{U}_L , pode ser expresso como a diferença entre a sua magnitude original e a projetada, conforme se segue¹:

¹No desenvolvimento apresentado, considerar \mathbf{I} como a matriz Identidade.

$$\begin{aligned}
\|\mathbf{e}\|^2 &= \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \\
&= \|\mathbf{x} - \mathbf{U}_L \mathbf{y}_L\|^2 \\
&= (\mathbf{x} - \mathbf{U}_L \mathbf{y}_L)^T (\mathbf{x} - \mathbf{U}_L \mathbf{y}_L) \\
&= \underbrace{\mathbf{x}^T \mathbf{x}}_{\|\mathbf{x}\|^2} - \underbrace{\mathbf{x}^T \mathbf{U}_L}_{\mathbf{y}_L^T} \mathbf{y}_L - \mathbf{y}_L^T \underbrace{\mathbf{U}_L^T \mathbf{x}}_{\mathbf{y}_L} + \mathbf{y}_L^T \underbrace{\mathbf{U}_L^T \mathbf{U}_L}_{\mathbf{I}} \mathbf{y}_L \\
&= \|\mathbf{x}\|^2 - \mathbf{y}_L^T \mathbf{y}_L - \mathbf{y}_L^T \mathbf{y}_L + \mathbf{y}_L^T \mathbf{y}_L \\
&= \|\mathbf{x}\|^2 - \mathbf{y}_L^T \mathbf{y}_L \\
&= \|\mathbf{x}\|^2 - \|\mathbf{y}_L\|^2.
\end{aligned} \tag{2.5}$$

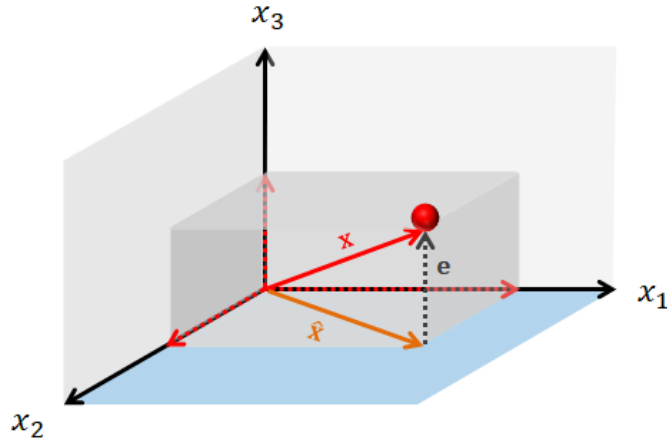


Figura 2.5: Erro de reconstrução.

Portanto, o erro de reconstrução será zero quando a matriz de transformação \mathbf{U}_L considerar $L = P$.

No contexto da detecção de novidades, o erro de reconstrução pode ser considerado como um escore de dissimilaridade de um dado de teste em relação ao conjunto que contempla a(s) classe(s) conhecida(s), visto que dados que pertençam a categorias desconhecidas tendem a não ser bem representados no subespaço definido por \mathbf{U}_L .

O Algoritmo 1 busca resumir essa abordagem em um pseudocódigo, considerando $\mathbf{x}_t \in \mathbb{R}^P$ como o exemplo de teste; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento com N exemplos conhecidos P -dimensionais; $L \in \mathbb{N}^*$, o número de componentes PCA; e $\mathbf{1}_N^T$, um vetor-linha contendo N componentes iguais a 1.

Fundamental para a construção do detector de novidades por esta técnica, a definição do parâmetro L está condicionada à proporção de variância explicada (Var_e), que pode ser expressa pela razão entre os somatórios dos L autovalores considerados na expansão PCA pelo somatório de todos os P autovalores [29]:

Algoritmo 1 Geração do escore de dissimilaridade baseado em PCA

```
1: função ESCORE-DISSIMILARIDADE-PCA( $\mathbf{x}_t, \mathbf{X}_{tr}, L$ )
2:   Executar apenas uma vez ou quando  $\mathbf{X}_{tr}$  ou  $L$  forem modificados:
3:      $\bar{\mathbf{x}}_{tr} \leftarrow \text{media}(\mathbf{X}_{tr})$ 
4:      $\tilde{\mathbf{X}}_{tr} \leftarrow \mathbf{X}_{tr} - \bar{\mathbf{x}}_{tr} \mathbf{1}_N^T$ 
5:      $\mathbf{U}_L \leftarrow \text{PCA}(\tilde{\mathbf{X}}_{tr}, L)$ 
6:   Executar sempre:
7:      $\tilde{\mathbf{x}}_t \leftarrow \mathbf{x}_t - \bar{\mathbf{x}}_{tr}$ 
8:      $\mathbf{y}_L \leftarrow \mathbf{U}_L^T \tilde{\mathbf{x}}_t$ 
9:      $s \leftarrow \|\tilde{\mathbf{x}}_t\|^2 - \|\mathbf{y}_L\|^2$ 
10:  devolve  $s$ 
11: fim função
```

$$\text{Var}_e = \frac{\sum_{k=1}^L \lambda_k}{\sum_{k=1}^P \lambda_k} \times 100 \text{ (\%)}, \quad (2.6)$$

de forma que quando $L = P$, a variância explicada é de 100%, visto o erro de reconstrução ser nulo.

2.3.2 Análise de Componentes Principais por Kernel

Ao passo em que a PCA é eficaz ao lidar com dados que descrevem problemas lineares, representando-os em espaços de atributos de menor dimensão, a Análise de Componentes Principais por *Kernel* (*Kernel Principal Component Analysis*, KPCA) é capaz de tratar problemas envolvendo dados com descrições não-lineares, conforme é ilustrado na Figura 2.6a [30]. Ao se considerar uma transformação não-linear ϕ que mapeia, de forma implícita, um ponto $\mathbf{x} \in \mathbb{R}^P$ em um espaço de alta dimensão \mathbb{R}^M , nomeado espaço de características \mathcal{F} (*feature space*), possivelmente infinito, obtém-se uma imagem $\phi(\mathbf{x})$ (Figura 2.6b). Em seguida, considera-se uma PCA aplicada sobre os dados mapeados neste novo espaço, o qual definirá um modelo com componentes principais não-lineares no espaço original.

Assim, considerando a existência de uma matriz de dados $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{P \times N}$, é necessário assumir as imagens destes dados no espaço dos atributos, isto é, $\{\phi(\mathbf{x}_i)\}_{i=1}^N$. Ademais, é conveniente que os dados mapeados estejam centralizados neste novo espaço.

Seja a amostra média em \mathcal{F} dada por:

$$\bar{\phi} = \frac{1}{N} \sum_{i=1}^N \phi(\mathbf{x}_i). \quad (2.7)$$

Logo, a matriz centralizada dos dados mapeados é definida por:

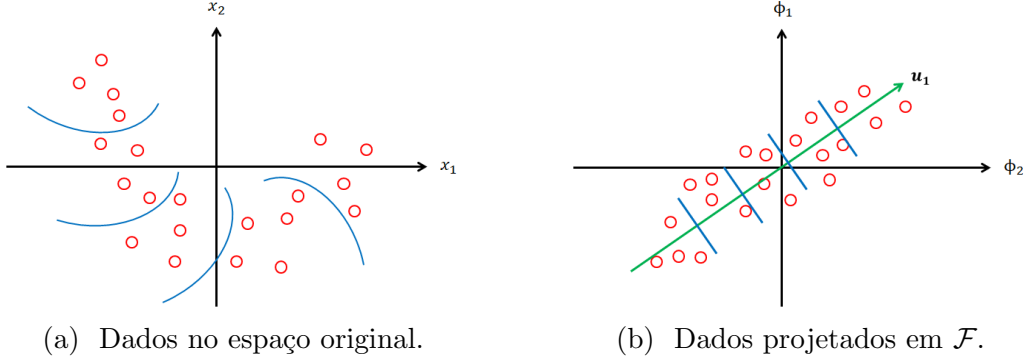


Figura 2.6: Ao realizar a PCA no espaço de características, obtêm-se as componentes principais, das quais a primeira é denotada pelo vetor \mathbf{u}_1 . As linhas em azul no espaço de características representando dados com mesmos valores de projeção linear sobre a primeira componente principal correspondem a projeções não-lineares no espaço original. Adaptação da Figura 12.16 de (Bishop, 2016).

$$\tilde{\Phi} \doteq \left[\underbrace{\phi(\mathbf{x}_1) - \bar{\phi}}_{\tilde{\phi}(\mathbf{x}_1)}, \underbrace{\phi(\mathbf{x}_2) - \bar{\phi}}_{\tilde{\phi}(\mathbf{x}_2)}, \dots, \underbrace{\phi(\mathbf{x}_N) - \bar{\phi}}_{\tilde{\phi}(\mathbf{x}_N)} \right] \in \mathbb{R}^{M \times N}. \quad (2.8)$$

Por fim, as componentes principais em \mathcal{F} podem ser extraídas a partir dos autovetores da matriz de covariância

$$\Sigma_{\tilde{\phi}(\mathbf{x})} \doteq \frac{1}{N} \tilde{\Phi} \tilde{\Phi}^T \in \mathbb{R}^{M \times M}. \quad (2.9)$$

Considerando o desafio de tratar um problema M -dimensional, onde o valor de M pode ser muito alto ou infinito, bem como os casos onde a função de mapeamento $\phi(\mathbf{x})$ é desconhecida, a KPCA permite a extração das componentes principais não-lineares diretamente dos dados no espaço original, através da chamada matriz de *kernel* $\mathcal{K} = [\kappa(\mathbf{x}_i, \mathbf{x}_j)] \in \mathbb{R}^{N \times N}$, explorando-se o truque de *kernel* [29]. Cada elemento desta matriz corresponde a uma função de *kernel*, definida pelo produto interno de pares de dados mapeados em \mathcal{F} :

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) \doteq \phi^T(\mathbf{x}_i) \phi(\mathbf{x}_j) \in \mathbb{R}. \quad (2.10)$$

Existem vários tipos de funções de *kernel*, porém neste trabalho a gaussiana - Eq. (2.11) - será a utilizada. Nela, o parâmetro γ está associado à largura do *kernel*.

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \quad (2.11)$$

Deve-se definir também o *kernel* centralizado:

$$\tilde{\kappa}(\mathbf{x}_i, \mathbf{x}_j) \doteq (\phi(\mathbf{x}_i) - \bar{\phi})^T (\phi(\mathbf{x}_j) - \bar{\phi}) \in \mathbb{R}, \quad (2.12)$$

o qual pode ser calculado com base nas funções de *kernel*, conforme a Eq. (2.13):

$$\tilde{\kappa}(\mathbf{x}_i, \mathbf{x}_j) = \kappa(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{N} \sum_{m=1}^N \kappa(\mathbf{x}_i, \mathbf{x}_m) - \frac{1}{N} \sum_{n=1}^N \kappa(\mathbf{x}_n, \mathbf{x}_j) + \frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N \kappa(\mathbf{x}_n, \mathbf{x}_m). \quad (2.13)$$

De forma matricial, a matriz de *kernel* centralizado $\tilde{\mathcal{K}} = \tilde{\Phi}^T \tilde{\Phi}$ pode ser calculada por [29]:

$$\tilde{\mathcal{K}} = \mathcal{J} \mathcal{K} \mathcal{J} \in \mathbb{R}^{N \times N}, \quad (2.14)$$

onde a matriz \mathcal{J} é dada por:

$$\mathcal{J} = \mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \in \mathbb{R}^{N \times N} \quad (2.15)$$

tendo o papel de matriz centralizadora, na qual $\mathbf{I} \in \mathbb{R}^{N \times N}$ é a matriz identidade, e $\mathbf{1} \in \mathbb{R}^N$ é um vetor-coluna com elementos unitários.

Após a obtenção da matriz de *kernel* centralizado $\tilde{\mathcal{K}}$, pode-se extrair os seus autovetores ($\mathbf{u}_i \in \mathbb{R}^N$) e autovalores ($\lambda_i \in \mathbb{R}$) normalizados. Destaca-se a necessidade subsequente de normalizar os autovetores, de forma que $\|\mathbf{u}_i\|^2 = \lambda_i^{-1}$.

Por fim, para um dado de teste $\mathbf{x}_t \in \mathbb{R}^P$, sua i -ésima componente principal de *kernel* é dada por:

$$y_i = \boldsymbol{\alpha}_i^T [\tilde{\kappa}(\mathbf{x}_1, \mathbf{x}_t) \tilde{\kappa}(\mathbf{x}_2, \mathbf{x}_t) \dots \tilde{\kappa}(\mathbf{x}_N, \mathbf{x}_t)]^T, \quad 1 \leq i \leq L \ll N, \quad (2.16)$$

onde $\boldsymbol{\alpha}_i = \mathbf{u}_i / \lambda_i$, logo $\mathbf{y}_L = [y_1 \ y_2 \ \dots \ y_L]^T$.

De forma similar a PCA, o parâmetro L está relacionado à proporção de variância explicada (Var_e), dada pela razão entre os somatórios dos L autovalores considerados na expansão KPCA por todos os N autovalores:

$$\text{Var}_e = \frac{\sum_{k=1}^L \lambda_k}{\sum_{k=1}^N \lambda_k} \times 100 \quad (\%), \quad (2.17)$$

de forma que quando $L = N$, a variância explicada é de 100%.

O erro quadrático de reconstrução associado ao dado de teste também é obtido de forma análoga ao da técnica PCA, porém, calculado de forma implícita no espaço de características \mathcal{F} , conforme a Eq. (2.18), na qual $\tilde{\phi}(\mathbf{x}_t)$ corresponde à centralização do dado mapeado $\phi(\mathbf{x}_t)$, e $\hat{\phi}(\mathbf{x}_t)$ é a suposta reconstrução do vetor $\phi(\mathbf{x}_t)$ neste mesmo espaço.

$$\|\mathbf{e}\|^2 = \|\tilde{\phi}(\mathbf{x}_t) - \hat{\phi}(\mathbf{x}_t)\|^2 = \|\tilde{\phi}(\mathbf{x}_t)\|^2 - \|\mathbf{y}_L\|^2 \quad (2.18)$$

O termo $\|\tilde{\phi}(\mathbf{x}_t)\|^2$ pode ser computado através da Eq. (2.19), e simplificado ao

considerar uma função de *kernel* gaussiana, resultando na Eq. (2.20), visto que: $\kappa(\mathbf{x}_t, \mathbf{x}_t) = 1$.

$$\|\tilde{\phi}(\mathbf{x}_t)\|^2 = \kappa(\mathbf{x}_t, \mathbf{x}_t) - \frac{2}{N} \sum_{n=1}^N \kappa(\mathbf{x}_t, \mathbf{x}_i) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \kappa(\mathbf{x}_i, \mathbf{x}_j) \quad (2.19)$$

$$\|\tilde{\phi}(\mathbf{x}_t)\|^2 = 1 - \frac{2}{N} \sum_{n=1}^N \kappa(\mathbf{x}_t, \mathbf{x}_i) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \kappa(\mathbf{x}_i, \mathbf{x}_j) \quad (2.20)$$

O Algoritmo 2 sintetiza os procedimentos supracitados em um pseudocódigo, considerando $\mathbf{x}_t \in \mathbb{R}^P$ como o exemplo de teste; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento com N exemplos conhecidos P -dimensionais; $L \in \mathbb{N}^*$, o número de componentes da expansão; e $\gamma \in \mathbb{R}$, a largura do *kernel* utilizado na Eq. (2.11).

Algoritmo 2 Geração do escore de dissimilaridade baseado em KPCA

- 1: **função** ESCORE-DISSIMILARIDADE-KPCA($\mathbf{x}_t, \mathbf{X}_{tr}, L, \gamma$)
 - 2: Executar apenas uma vez ou quando \mathbf{X}_{tr} ou L forem modificados:
 - 3: Obter a matriz de *kernel* $\mathcal{K} \in \mathbb{R}^{N \times N}$ para \mathbf{X}_{tr} pela Eq. (2.11).
 - 4: Centralizar \mathcal{K} com as Eqs. (2.14) e (2.15), resultando em $\tilde{\mathcal{K}} \in \mathbb{R}^{N \times N}$.
 - 5: Calcular os L autovetores $\mathbf{u}_i \in \mathbb{R}^N$ de $\tilde{\mathcal{K}}$, e normalizá-los, de forma que $\|\mathbf{u}_i\|^2 = \lambda_i^{-1}$.
 - 6: Executar sempre:
 - 7: Computar as L primeiras componentes principais de *kernel* $y_i \in \mathbb{R}$ de \mathbf{x}_t através da Eq. (2.16), gerando $\mathbf{y}_L \in \mathbb{R}^L$.
 - 8: Computar o termo $\|\tilde{\phi}(\mathbf{x}_t)\|^2$ através da Eq. (2.19) ou (2.20).
 - 9: $s \leftarrow \|\tilde{\phi}(\mathbf{x}_t)\|^2 - \|\mathbf{y}_L\|^2$
 - 10: **devolve** s
 - 11: **fim função**
-

Em resumo, a detecção de novidades baseada na técnica KPCA também utiliza o erro quadrático de reconstrução, porém no espaço das características, como um escore de dissimilaridade entre um dado de teste e o conjunto de dados conhecidos.

2.3.3 Máquinas de Vetor-Suporte de Uma Classe

Máquinas de Vetor-Suporte (*Support-Vector Machines*, SVM) são uma técnica para a construção de regressores e classificadores que se utilizam de dados selecionados no conjunto de treinamento para a constituição das funções de regressão/classificação correspondentes, daí o nome de vetores-suporte. Frequentemente, por meio do truque de *kernel*, as SVM derivam uma superfície de decisão não-linear, utilizando uma função de mapeamento não-linear implícita do conjunto de dados de treinamento em um espaço de maior dimensão. O objetivo desta transformação é que pontos de classes distintas, os quais não poderiam ser separados em seu espaço original através de uma superfície de decisão simples, tal como uma reta ou um hiperplano, passem

a ser separáveis por um hiperplano neste novo espaço, conhecido como espaço de características \mathcal{F} . Note que o mapeamento desta solução no espaço original usualmente gera uma curva ou superfície não-linear, permitindo melhores regressores e classificadores.

A Figura 2.7a ilustra este processo, no qual as duas classes representadas pelos círculos em azul e vermelho não são linearmente separáveis em \mathbb{R}^2 . Após o mapeamento em \mathbb{R}^3 , os pontos podem ser separados de forma linear, conforme a Figura 2.7b.

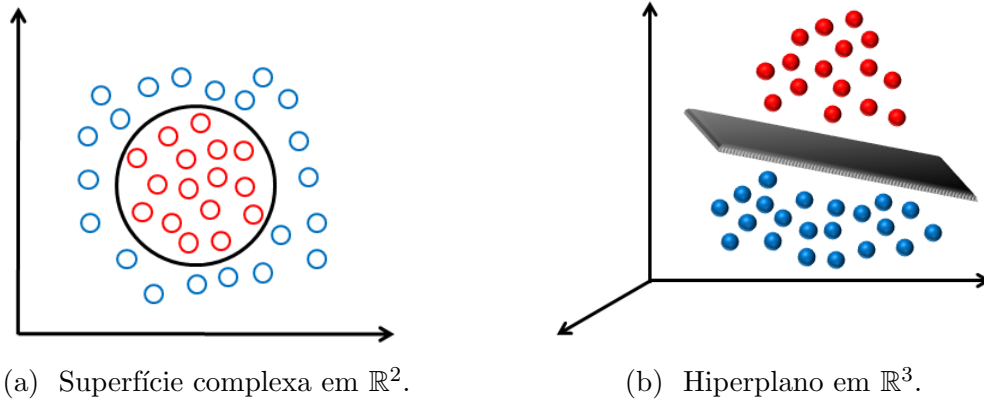


Figura 2.7: Claramente, percebe-se que as duas classes passam a ser separáveis por um hiperplano em \mathbb{R}^3 , o qual corresponde a uma curva não-linear em \mathbb{R}^2 .

Para o truque de *kernel* supracitado, será considerada a função de *kernel* do tipo gaussiana (2.21), de forma similar ao descrito na KPCA. Entretanto, neste caso, o parâmetro K_s está associado à largura do *kernel*:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{K_s^2}}. \quad (2.21)$$

A técnica OCSVM [31] é uma variante das técnicas de vetores-suporte. Tal modelo usa um hiperplano com máxima distância da origem, no entanto que defina um semi-espaço contendo a maior parte dos dados pertencentes à classe conhecida no espaço \mathcal{F} , conforme ilustrado na Figura 2.8.

Sejam os hiperplanos definidos por:

$$\mathbf{w}^T \phi(\mathbf{x}) = \pm \rho, \quad \mathbf{w} \in \mathcal{F}. \quad (2.22)$$

Da teoria de SVM usual, a minimização de $\|\mathbf{w}\|$ leva a maximização de ρ . O objetivo é obter \mathbf{w} , tal que $\|\mathbf{w}^T \phi(\mathbf{x}_i)\| > \rho$ para a maior parte dos $\mathbf{x}_i \in \mathbf{X}_{tr}$, que definem o conjunto de treinamento. Porém, com o intuito de prevenir a ocorrência de *overfitting*, visto que podem existir dados com ruído no conjunto de treinamento, opta-se por inserir variáveis de folga ξ_i para permitir que alguns dados se acomodem na região interior aos hiperplanos definidos pela Eq. (2.22), logo associados à região

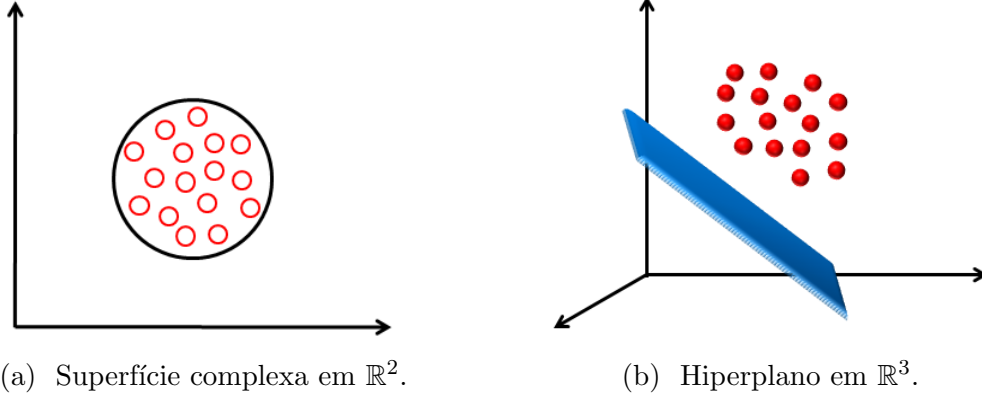


Figura 2.8: Exemplo contendo: (a) um conjunto de dados bidimensional arbitrário \mathcal{X} ; (b) os mesmos dados projetados em um espaço de maior dimensão.

definida por $\|\mathbf{w}^T \phi(\mathbf{x}_i)\| < \rho$. A solução deste problema recai na minimização da seguinte função objetivo quadrática:

$$\min_{(\mathbf{w}, \rho, \xi \in \mathbb{R}^N)} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu N} \sum_{i=1}^N \xi_i - \rho, \quad (2.23)$$

sujeita a:

$$\mathbf{w}^T \phi(\mathbf{x}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad \rho \geq 0, \quad 1 \leq i \leq N. \quad (2.24)$$

Destaca-se que quanto maior for o valor de ρ , maiores serão a região entre os hiperplanos e o número de dados contidos nela, isto é, maior o erro de treinamento, para uma direção específica \mathbf{w} . Já o parâmetro $\nu \in (0, 1]$ controla a influência do somatório relativo às variáveis de folga na função objetivo, portanto está relacionado à fração de dados na região $\|\mathbf{w}^T \phi(\mathbf{x}_i)\| < \rho$.

Ao se utilizar os multiplicadores de Lagrange $\alpha_i, \beta_i \geq 0$ [32], introduz-se o Lagrangiano:

$$L(\mathbf{w}, \xi, \rho, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu N} \sum_{i=1}^N \xi_i - \rho - \sum_{i=1}^N \alpha_i ((\mathbf{w}^T \phi(\mathbf{x}_i)) - \rho + \xi_i) - \sum_{i=1}^N \beta_i \xi_i. \quad (2.25)$$

Derivando-se a Eq. (2.25) com respeito às variáveis primais \mathbf{w}, ξ, ρ e igualando o resultado a zero, obtêm-se:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i \phi(\mathbf{x}_i), \quad (2.26)$$

$$\alpha_i = \frac{1}{\nu N} - \beta_i, \quad \sum_{i=1}^N \alpha_i = 1. \quad (2.27)$$

Convém definir que, na Eq. (2.26), todos os \mathbf{x}_i associados a $\alpha_i > 0$ são chamados de vetores-suporte. Aplicando-se o truque de *kernel*, a expansão em vetores-suporte transforma a função de decisão $f(\mathbf{x}) = \text{sinal}(\mathbf{w}^T \Phi(\mathbf{x}_i) - \rho)$ em²:

$$f(\mathbf{x}) = \text{sinal}\left(\sum_{i=1}^N \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) - \rho\right). \quad (2.28)$$

Ao substituir as Eqs. (2.26) e (2.27) na Eq. (2.25), obtém-se o problema dual expresso pela Eq. (2.29), o qual pode ser solucionado através do algoritmo de minimização quadrática *Sequential Minimal Optimization* (SMO) [33], segundo o qual a expressão é minimizada através de uma série de minimizações por dois pontos:

$$\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \kappa(\mathbf{x}_i, \mathbf{x}_j), \quad \text{sujeito a } 0 \leq \alpha_i \leq \frac{1}{\nu N}, \quad \sum_{i=1}^N \alpha_i = 1. \quad (2.29)$$

Pode-se mostrar [32] que no ponto ótimo as duas primeiras restrições de desigualdade presentes na Eq. (2.24) tornam-se igualdades, caso α_i e β_i sejam diferentes de zero, ou seja, $0 < \alpha_i < 1/(\nu N)$. Portanto, para $\alpha_i \neq 0$ e seu \mathbf{x}_i associado, ρ satisfaz:

$$\rho = (\mathbf{w}^T \Phi(\mathbf{x}_i)) = \sum_{j=1}^N \alpha_j \kappa(\mathbf{x}_j, \mathbf{x}_i). \quad (2.30)$$

O Algoritmo 3 exprime o cálculo do escore de dissimilaridade utilizado na detecção de novidades, o qual é equivalente ao argumento da função definida por $f(\mathbf{x})$. Consideram-se $\mathbf{x}_t \in \mathbb{R}^P$ um exemplo de teste; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento contendo N exemplos conhecidos P -dimensionais; e os parâmetros K_s , ν e ξ relativos ao algoritmo OCSVM.

2.3.4 Modelos de Mistura de Gaussianas

Os GMM (*Gaussian Mixture Models*) pertencem à categoria de modelos de variáveis latentes, nos quais é assumido que as variáveis observadas são correlacionadas porque são provenientes de uma “causa” comum desconhecida [1]. Mais especificamente, eles consistem em um modelo probabilístico que considera dados produzidos por um conjunto de M distribuições base gaussianas multivariadas.

²Para $z \geq 0$, temos $\text{sinal}(z) = 1$; e caso contrário, $\text{sinal}(z) = -1$.

Algoritmo 3 Geração do escore de dissimilaridade baseado em OCSVM

- 1: **função** ESCORE-DISSIMILARIDADE-OCSVM($\mathbf{x}_t, \mathbf{X}_{tr}, K_s, \nu, \xi$)
 - 2: Executar apenas uma vez ou quando $\mathbf{X}_{tr}, K_s, \nu$ ou ξ forem modificados:
 - 3: Aplicar SMO para solucionar o problema de minimização definido pela Eq. (2.29) e obter α .
 - 4: Calcular ρ através da Eq. (2.30).
 - 5: Executar sempre:
 - 6: $s \leftarrow \sum_{i=1}^N \alpha_i k(\mathbf{x}_{tr_i}, \mathbf{x}_t) - \rho$
 - 7: **devolve** s
 - 8: **fim função**
-

Assim, o modelo assume a forma descrita na Eq. (2.31), a qual representa a superposição de M densidades gaussianas $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, $1 \leq m \leq M$, as chamadas componentes da mistura, cada uma com média $\boldsymbol{\mu}_m \in \mathbb{R}^P$ e matriz de covariância $\boldsymbol{\Sigma}_m \in \mathbb{R}^{N \times N}$. Neste caso, o vetor $\mathbf{x} \in \mathbb{R}^P$ é assumido como uma amostra proveniente da mistura, cujos parâmetros são denotados por $\boldsymbol{\theta}$:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (2.31)$$

na qual:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_m|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_m)^T \boldsymbol{\Sigma}_m^{-1} (\mathbf{x} - \boldsymbol{\mu}_m)\right). \quad (2.32)$$

Ademais, os termos π_m ($0 \leq \pi_m \leq 1$) presentes na Eq. (2.31) representam os coeficientes da mistura, sujeitos a:

$$\sum_{m=1}^M \pi_m = 1, \quad (2.33)$$

e relacionados à probabilidade a priori de ocorrência de cada um dos elementos da mistura. Na Figura 2.9 são ilustradas as curvas de nível e a superfície correspondente a um modelo de misturas contendo três componentes com diferentes valores de π_m , $\boldsymbol{\Sigma}_m$ e $\boldsymbol{\mu}_m$.

Observa-se que para se definir os parâmetros da distribuição de mistura de gaussianas, é imprescindível obter os parâmetros π_m , $\boldsymbol{\mu}_m$ e $\boldsymbol{\Sigma}_m$ de cada componente. Para solucionar este problema, utiliza-se o Algoritmo de Maximização da Esperança (EM) [30], resumido no Algoritmo 4, no qual $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$ representa o conjunto de treinamento com N exemplos conhecidos P -dimensionais; e $M \in \mathbb{N}^*$, o número de componentes da mistura.

Neste trabalho, assumiu-se que as distribuições compartilham as matrizes de co-

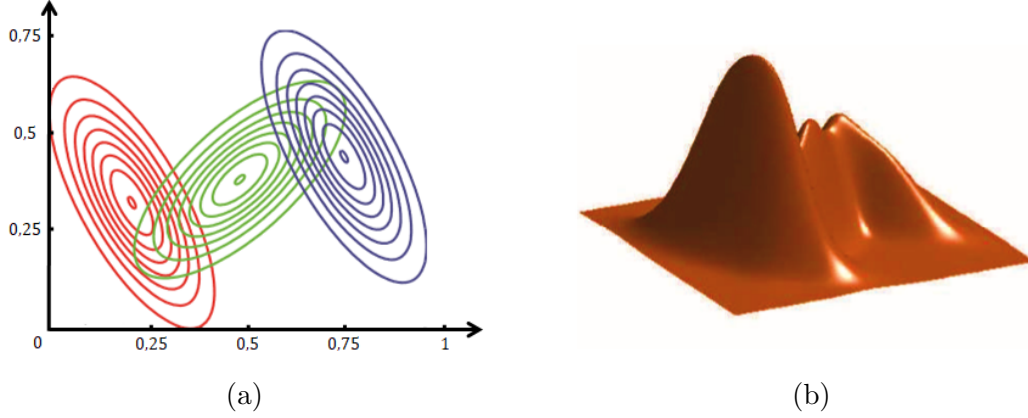


Figura 2.9: Adaptação da Figura 11.3 de (Murphy, 2012).

Algoritmo 4 Algoritmo EM para a obtenção dos parâmetros do GMM

```

1: função ALGORITMO-EM( $\mathbf{X}_{tr}, M$ )
2:   Inicializar aleatoriamente os parâmetros  $\pi_m$ ,  $\boldsymbol{\mu}_m$  e  $\boldsymbol{\Sigma}_m$ .
3:   enquanto não ocorrer a convergência dos parâmetros faça
4:     para  $m \leftarrow 1$  até  $M$  faça
5:       para  $i \leftarrow 1$  até  $N$  faça
6:          $r_{mi} \leftarrow \frac{\pi_m \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)}{\sum_{j=1}^M \pi_j \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$ 
7:       fim para
8:     fim para
9:     para  $m \leftarrow 1$  até  $M$  faça
10:       $N_m \leftarrow \sum_{i=1}^N r_{mi}$ 
11:       $\boldsymbol{\mu}_m \leftarrow \frac{1}{N_m} \sum_{i=1}^N r_{mi} \mathbf{x}_i$ 
12:       $\boldsymbol{\Sigma}_m \leftarrow \frac{1}{N_m} \sum_{i=1}^N r_{mi} (\mathbf{x}_i - \boldsymbol{\mu}_m)(\mathbf{x}_i - \boldsymbol{\mu}_m)^T$ 
13:       $\pi_m \leftarrow \frac{N_m}{N}$ 
14:    fim para
15:  fim enquanto
16: fim função

```

variância, ou seja, $\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}$. Logo, simplificações podem ser efetuadas no Algoritmo 4, discutidas em [34].

Para lidar com a tarefa de detecção de novidades, após se construir um modelo de misturas que contemple os dados conhecidos, pode-se utilizar o valor negativo do logaritmo da verossimilhança deste modelo para um dado sinal de teste ($-\log p(\mathbf{x}_t | \boldsymbol{\theta})$) como representante do escore de dissimilaridade (Algoritmo 5). Por conseguinte, quanto menor for o valor do escore, mais similar é o evento à classe conhecida.

2.3.5 k -vizinhos mais próximos

A técnica de aprendizagem dos k -vizinhos mais próximos é baseada em instâncias, explorando a proximidade de um dado de teste com seus vizinhos mais próximos. Tais vizinhos são identificados no conjunto de treinamento do modelo, correspon-

Algoritmo 5 Geração do escore de dissimilaridade baseado em GMM

- 1: **função** ESCORE-DISSIMILARIDADE-GMM($\mathbf{x}_t, \mathbf{X}_{tr}, M$)
 - 2: Executar apenas uma vez ou quando \mathbf{X}_{tr} ou M forem modificados:
 - 3: $[\pi_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}] \leftarrow EM(\mathbf{X}_{tr}, M)$
 - 4: Executar sempre:
 - 5: $s \leftarrow -\log \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma})$
 - 6: **devolve** s
 - 7: **fim função**
-

dendo, assim, à classe conhecida. Usualmente, tal proximidade é inferida por meio da distância Euclidiana.

A Figura 2.11 ilustra tal procedimento. O conjunto de treinamento é representado pelos círculos em vermelho, enquanto os dados de teste consistem nos símbolos em azul.

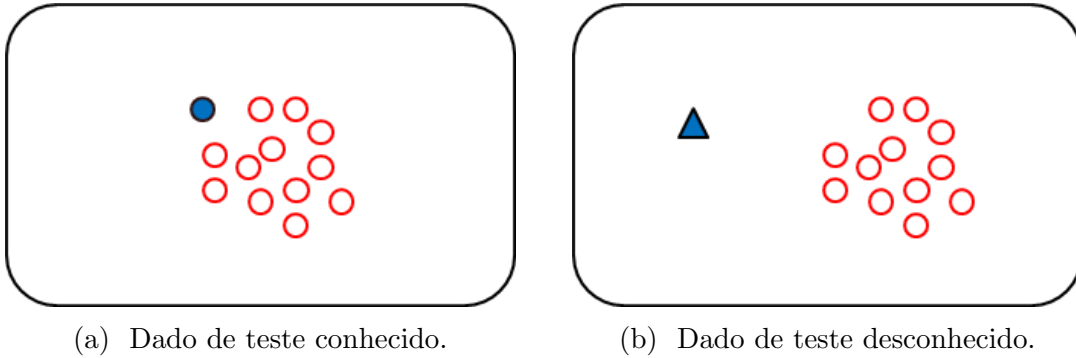


Figura 2.10: Exemplo ilustrando a proximidade entre um conjunto de dados de treino e um dado de teste pertencente a classe conhecida (a) ou não (b).

Com o intuito de aplicar esta técnica na tarefa de detecção de novidades, foi assumido o valor da mediana da distância Euclidiana entre um dado e seus k -vizinhos mais próximos como escore de dissimilaridade. Claramente, se tal distância for baixa, tal evento é similar à classe conhecida; se a distância for alta, o evento é dissimilar, representando, portanto, uma novidade. Ressalta-se que a opção pelo uso da mediana é justificado pelo seu potencial em mitigar possíveis efeitos de eventuais *outliers* presentes no conjunto de treinamento [2].

O Algoritmo 6 sintetiza este procedimento em um pseudocódigo, considerando $\mathbf{x}_t \in \mathbb{R}^P$ como um exemplo de teste; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento contendo N exemplos conhecidos $\mathbf{x}_{tr_i} \in \mathbb{R}^P$; e $k \in \mathbb{N}^*$, o número de vizinhos.

A técnica kNN também é amplamente utilizada para a construção de sistemas de classificação multiclasse. Para um dado sob avaliação, são identificados os k vizinhos mais próximos, sendo considerado o rótulo mais frequente dentre tais vizinhos para se definir o rótulo a ser atribuído pelo classificador. A Figura 2.11 ilustra este processo, onde o indicador estrela sinaliza um dado arbitrário a ser classificado.

Algoritmo 6 Geração do escore de dissimilaridade baseado em kNN

```
1: função ESCORE-DISSIMILARIDADE-KNN( $\mathbf{x}_t, \mathbf{X}_{tr}, k$ )
2:   para  $i \leftarrow 1$  até  $N$  faça
3:      $d_i \leftarrow \|\mathbf{x}_t - \mathbf{x}_{tr_i}\|^2$ 
4:   fim para
5:    $\mathbf{d} \leftarrow$  Armazenar os  $k$  menores valores de  $d_i$ .
6:    $s \leftarrow$  Calcular a mediana do vetor  $\mathbf{d}$ .
7:   devolve  $s$ 
8: fim função
```

Assumindo $k = 4$, claramente, o rótulo atribuído será Classe B.

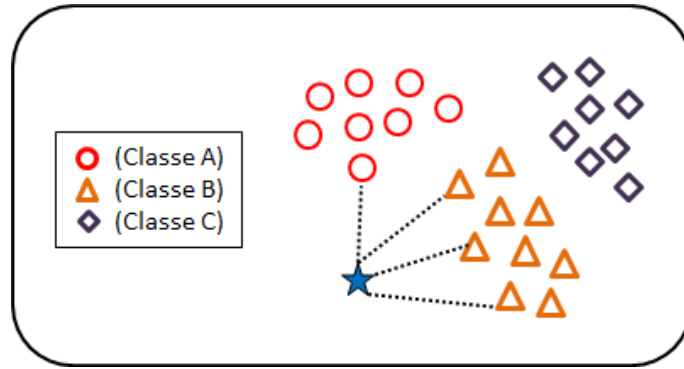


Figura 2.11: Técnica k NN para a constituição de um classificador.

O pseudocódigo referente a este sistema de classificação é apresentado no Algoritmo 7.

Algoritmo 7 Sistema de classificação baseado em k NN

```
1: função CLASSIFICADOR-KNN( $\mathbf{x}_t, \mathbf{X}_{tr}, k$ )
2:   para  $i \leftarrow 1$  até  $N$  faça
3:      $d_i \leftarrow \|\mathbf{x}_t - \mathbf{x}_{tr_i}\|^2$ 
4:   fim para
5:    $\mathcal{I} \leftarrow$  Armazenar os índices  $i$  associados aos  $k$  menores valores de  $d_i$ .
6:    $\mathbf{r} \leftarrow$  Armazenar os rótulos de  $\mathbf{x}_{tr_i}$ , para  $i \in \mathcal{I}$ .
7:    $r \leftarrow$  Calcular a moda do vetor  $\mathbf{r}$ .
8:   devolve  $r$ 
9: fim função
```

2.3.6 k -vizinhos mais próximos esparsos

O algoritmo kNN utiliza um conjunto de treinamento $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$ contendo N observações, as quais devem ser comparadas com um dado de teste \mathbf{x}_t quanto ao critério da distância Euclidiana. A ideia é obter um novo conjunto $\mathbf{X}'_{tr} \in \mathbb{R}^{P \times Q}$, em que $Q \ll N$, o qual reúna os dados mais relevantes de \mathbf{X}_{tr} para esta tarefa, de

forma que o custo computacional seja reduzido, sem prejuízo significativo à acurácia do sistema. Destaca-se que esta abordagem esparsa do kNN, nomeada de s-kNN, é uma proposta desta Dissertação.

A Figura 2.12 compara o kNN e o s-kNN em um problema hipotético. Nela, os círculos em vermelho representam o conjunto de treinamento de cada modelo, enquanto o círculo em azul é o dado de teste. Na Figura 2.12a, todas as N observações disponíveis compõem o conjunto \mathbf{X}_{tr} , enquanto na Figura 2.12b, $N/2$ observações foram selecionadas para integrar este conjunto. Pode-se observar que o conjunto esparsado realiza uma descrição da “fronteira” dos dados de forma razoavelmente similar ao conjunto original. No entanto, assumindo $k = 3$, por exemplo, diferentes dados são eleitos como os vizinhos mais próximos em cada técnica, implicando em um valor de escore de dissimilaridade diferente quanto à detecção de novidades. Assim, deve-se investigar o impacto da redução do quantitativo de amostras descritivas do conjunto, referidas aqui como dicionário ou base representativa da classe $\Phi \in \mathbb{R}^{N \times P}$ [35], no desempenho do modelo.

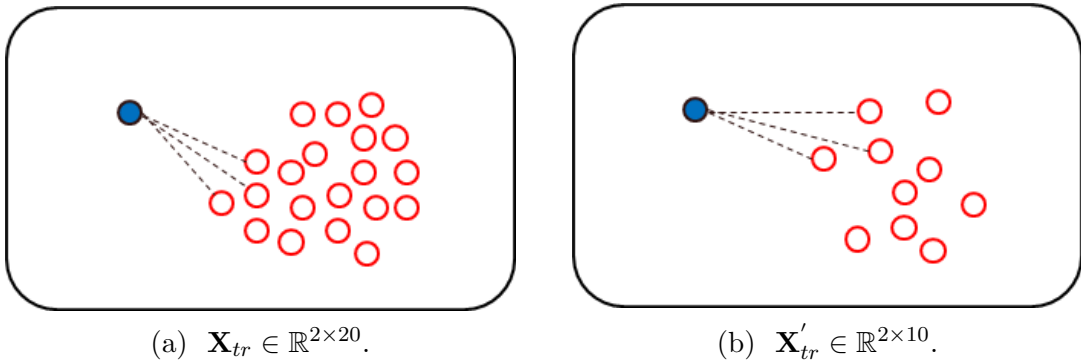


Figura 2.12: Exemplo ilustrando a diferença entre os conjuntos de treinamento utilizados pelas técnicas kNN (a) e s-kNN (b).

A representação de um sinal em uma base específica envolve definir um conjunto de coeficientes de expansão (α). Seja um dicionário composto por N vetores de norma unitária, referidos como átomos — $\phi_n \in \mathbb{R}^P$ — responsáveis pela definição de um subespaço para a expansão do sinal, associado a $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_N]$.

Ao assumir um dicionário completo³ e redundante⁴, há várias formas de representar um sinal \mathbf{x} como uma combinação dos átomos, conforme a Eq. (2.34):

$$\hat{\mathbf{x}} = \sum_{n=1}^N \alpha_n \phi_n. \quad (2.34)$$

Neste contexto, torna-se interessante obter a melhor forma de compor esta representação, que será alcançada quando o vetor de coeficientes for esparsado⁵. Para isso,

³Quando os átomos do dicionário abrangem todo o espaço do sinal.

⁴Quando os átomos do dicionário formam um conjunto linearmente dependente.

⁵Visto que existem infinitas soluções, deseja-se aquela que envolve o menor número de átomos,

deve-se minimizar a norma-zero⁶ de $\boldsymbol{\alpha}$ ($\|\boldsymbol{\alpha}\|_0$), ao passo que o erro de reconstrução do sinal seja mínimo, isto é, $E[\mathbf{x} - \hat{\mathbf{x}}] \rightarrow 0$.

A escolha dos átomos que constituem o dicionário pode ser efetuada pelo algoritmo *Orthogonal Matching Pursuit*⁷ (OMP) [35], o qual consiste em uma busca iterativa através de passos na direção de ótimos locais. Partindo de um conjunto vazio e $\boldsymbol{\alpha} = 0$, acrescenta-se um novo membro a este conjunto de forma que tal ação maximize a redução do erro residual (\mathbf{r}), definido como $\|\mathbf{e}\|_2$, onde: $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. Se o valor deste erro for menor que um limiar ϵ arbitrário, o procedimento termina.

Seja \mathcal{I} o conjunto que contém os índices das amostras de \mathbf{X}_{tr} presentes no conjunto de átomos escolhidos Φ . A seleção de um novo membro i será dada por:

$$i = \arg \max_{i \notin \mathcal{I}} | \langle \mathbf{x}_{tr_i}, \mathbf{r} \rangle |, \quad \text{para } \mathbf{x}_{tr_i} \in \mathbf{X}_{tr}. \quad (2.35)$$

Pode-se mostrar [35] que a cada iteração, o resíduo é ortogonal à aproximação do vetor \mathbf{x} fornecida pelo Φ atual. Com isso, o número de observações de \mathbf{X}_{tr} que sejam linearmente dependentes do resíduo é reduzido, pois assim que uma delas é selecionada para a constituição de Φ , as outras já estarão cobertas, logo não contribuirão significativamente para a redução do resíduo, de forma que não serão selecionadas. O resíduo pode ser determinado pela ortogonalização de Gram-Schmidt [36].

Uma vez que o resíduo deve ser reduzido a cada iteração k , pode-se reescrever o problema de minimização da seguinte forma:

$$\min_{\boldsymbol{\alpha}_{(k)}} \|\mathbf{r}_{(k-1)} - \Phi_{(k)} \boldsymbol{\alpha}_{(k)}\|^2, \quad (2.36)$$

o qual pode ser solucionado com base no conceito de matrizes pseudoinversas [37]:

$$\boldsymbol{\alpha}_{(k)} = (\Phi_{(k)}^T \Phi_{(k)})^{-1} \Phi_{(k)}^T \mathbf{r}_{(k-1)}. \quad (2.37)$$

Assim, o resíduo é atualizado através da Eq. (2.38), e uma nova iteração ocorre caso a norma Euclidiana do resíduo seja superior a um limiar pré-determinado.

$$\mathbf{r}_{(k)} = \mathbf{r}_{(k-1)} - \Phi_{(k)} (\Phi_{(k)}^T \Phi_{(k)})^{-1} \Phi_{(k)}^T \mathbf{r}_{(k-1)}. \quad (2.38)$$

O Algoritmo 8 apresenta as etapas descritas em pseudocódigo, considerando o vetor-coluna $\mathbf{x} \in \mathbb{R}^P$ como o exemplo a ser representado; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento contendo N observações $\mathbf{x}_{tr_i} \in \mathbb{R}^P$ candidatas a integrar o conjunto Φ ; e o limiar $\epsilon \in \mathbb{R}$ que baliza o critério de parada. Na linha 9, \mathbf{I} representa a matriz identidade.

por simplicidade.

⁶A norma-zero de um vetor é dada pelo número total de componentes não-nulas.

⁷O algoritmo foi escolhido devido a sua simplicidade e bons resultados.

Algoritmo 8 Seleção das amostras para o conjunto Φ através de OMP

```
1: função SELECAO-AMOSTRAS-OMP( $\mathbf{x}, \mathbf{X}_{tr}, \epsilon$ )
2:    $\mathbf{r} = \mathbf{x}, \Phi = \emptyset$  e  $\mathcal{I} = \emptyset$ 
3:   enquanto ( $\|\mathbf{r}\|_2 \geq \epsilon$ ) faça
4:      $i \leftarrow \max_{i \notin \mathcal{I}} | \langle \mathbf{x}_{tr_i}, \mathbf{r} \rangle |$ 
5:      $\mathcal{I} \leftarrow \mathcal{I} \cup \{i\}$ 
6:      $\Phi \leftarrow [\Phi \ \mathbf{x}_{tr_i}]$ 
7:      $\mathbf{r} \leftarrow (\mathbf{I} - \Phi(\Phi^T \Phi)^{-1} \Phi^T) \mathbf{r}$ 
8:   fim enquanto
9:   devolve  $\Phi$ 
10: fim função
```

Suponha um conjunto \mathcal{X} que contenha Z amostras \mathbf{x}_i ($1 \leq i \leq Z$). A técnica s-kNN propõe que cada \mathbf{x}_i seja expandido individualmente considerando as amostras do conjunto de treinamento \mathbf{X}_{tr} como um conjunto de possíveis membros de Φ_i , através da técnica OMP. Por sua vez, a união dos Z conjuntos Φ_i forma o conjunto Φ que será utilizado pelo algoritmo kNN. Em outras palavras, apenas as amostras do conjunto de treinamento que nunca foram selecionadas serão descartadas do conjunto final. Esta abordagem é diferente da usual, a qual é realizada em batelada [38], porém se mostrou efetiva.

Para a detecção de novidades, neste caso, o escore de dissimilaridade será calculado conforme descrito no Algoritmo 6 da técnica kNN, tão somente substituindo o conjunto de treinamento pelo conjunto esparsos Φ .

2.3.7 Local Outlier Factor

A técnica LOF se baseia no conceito de densidade local, a qual é calculada através da distância entre os dados [4]. O procedimento consiste em comparar a densidade local de um dado de teste com as densidades locais de seus k -vizinhos mais próximos pertencentes ao conjunto de treinamento, de forma que é gerado um valor de escore que indica a chance de este dado ser espúrio (um *outlier*). Assim, quando tais densidades são suficientemente distintas, trata-se o dado como uma novidade. Destaca-se que a técnica opera em um âmbito local, visto que apenas uma vizinhança restrita é levada em consideração nesta análise.

A Figura 2.13 exemplifica o conceito da densidade local de um dado. Nela, há um conjunto de dados em vermelho agrupados, representando a classe conhecida, e um dado de teste em azul simbolizando uma novidade. Ao considerar 3 vizinhos mais próximos ($k = 3$), observa-se que o raio do círculo tracejado centrado no dado de teste é maior do que aquele centrado em um dos dados conhecidos. Logo, a densidade local do dado de teste é menor do que aquela associada ao dado em vermelho.

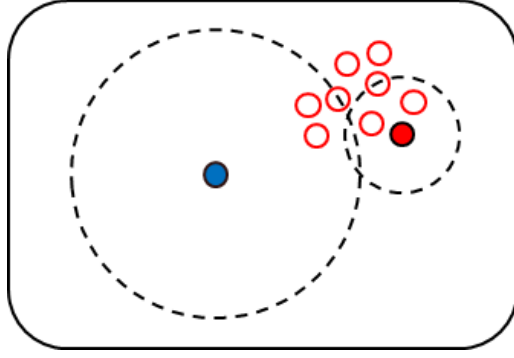


Figura 2.13: Ilustração do conceito de densidade local para $k = 3$.

Sejam: (i) $\text{dist}(\mathbf{x}, \mathbf{y}) \in \mathbb{R}$ a distância Euclidiana entre os dados $\mathbf{x}, \mathbf{y} \in \mathbb{R}^P$; (ii) $\text{dist}_k(\mathbf{x}) \in \mathbb{R}$ a distância Euclidiana entre um dado \mathbf{x} e o seu k -ésimo vizinho ($k \in \mathbb{N}^*$); (iii) $N_k(\mathbf{x})$ o conjunto que contém todos os vizinhos com distância menor ou igual a $\text{dist}_k(\mathbf{x})$, cuja cardinalidade ($\#N_k(\mathbf{x})$) pode ser maior do que o valor de k . Tal situação ocorre quando existem vizinhos locais posicionados a uma mesma distância de \mathbf{x} , conforme ilustrado na Figura 2.14. Neste exemplo hipotético, o dado no centro do conjunto apresenta a mesma distância: para o seu segundo e terceiro vizinhos (Figura 2.14a); e para o seu quarto, quinto e sexto vizinhos (Figura 2.14b). Desta forma, pela análise da Figura 2.14a, observa-se que para $k = \{2, 3\}$, $\#N_k = 3$; e quando $k = \{4, 5, 6\}$, $\#N_k = 6$, segundo a Figura 2.14b.

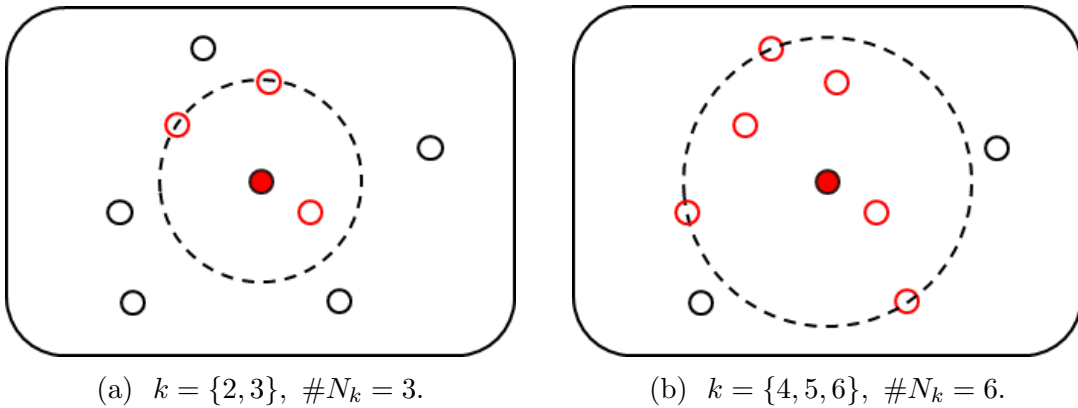


Figura 2.14: Cardinalidade de N_k para diferentes valores de k .

De posse dos valores anteriores, pode-se calcular a distância de alcançabilidade de \mathbf{x} a partir de \mathbf{y} , dada por:

$$\text{da}_k(\mathbf{x}, \mathbf{y}) = \max\{\text{dist}_k(\mathbf{y}), \text{dist}(\mathbf{x}, \mathbf{y})\}. \quad (2.39)$$

A Figura 2.15 exemplifica os 2 casos que podem definir o valor da distância de alcançabilidade, ao considerar $k = 4$: (i) $\text{dist}_k(\mathbf{y}) > \text{dist}(\mathbf{x}, \mathbf{y})$ (Figura 2.15a); (ii) $\text{dist}(\mathbf{x}, \mathbf{y}) > \text{dist}_k(\mathbf{y})$ (Figura 2.15b).

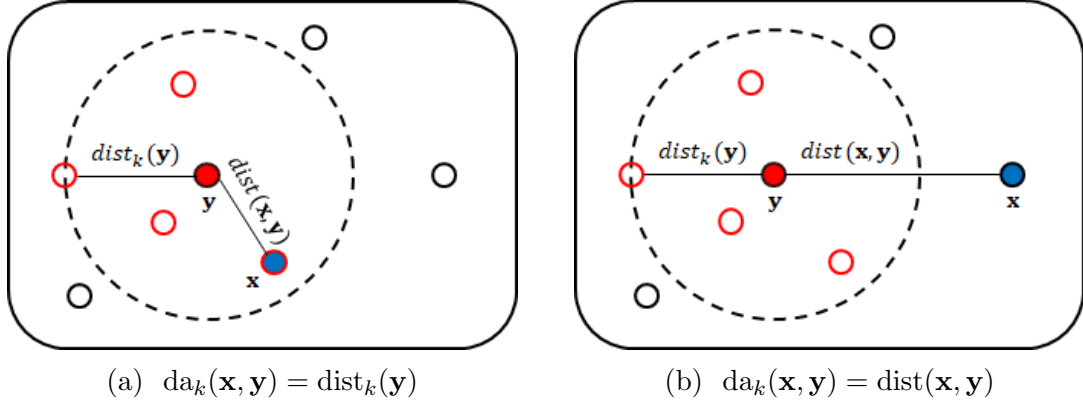


Figura 2.15: Distância de alcançabilidade $da_k(\mathbf{x}, \mathbf{y})$ para $k = 4$ em 2 casos hipotéticos.

Outro conceito fundamental é o da densidade local de alcançabilidade, expresso pela Eq. (2.40), a qual indica o quão longe se deve percorrer, a partir de \mathbf{x} , a fim de alcançar o próximo dado ou o conjunto de dados. Assim, quanto maior for a distância aos vizinhos próximos, mais esparsa é a região em que o dado está localizado, logo menos densa. Tal valor é obtido através do inverso da média da distância de alcançabilidade de \mathbf{x} a partir de seus vizinhos.

$$\text{dal}_k(\mathbf{x}) = 1 / \left(\frac{\sum_{\mathbf{y} \in N_k(\mathbf{x})} da_k(\mathbf{x}, \mathbf{y})}{\#N_k(\mathbf{x})} \right) \quad (2.40)$$

Por fim, pode-se obter o valor do escore $\text{LOF}_k(\mathbf{x})$ através da comparação das densidades de alcançabilidade locais de \mathbf{x} com seus vizinhos:

$$\text{LOF}_k(\mathbf{x}) = \frac{1}{\#N_k(\mathbf{x})} \sum_{\mathbf{y} \in N_k(\mathbf{x})} \frac{\text{dal}_k(\mathbf{y})}{\text{dal}_k(\mathbf{x})} \quad (2.41)$$

Quanto mais próximo de 1 é o escore, mais similares são as densidades locais de \mathbf{x} e de seus vizinhos, sinalizando que o dado pertence à classe conhecida. Entretanto, valores significativamente maiores que 1 indicam novidades. Dessa forma, assume-se que essa medida será o escore de dissimilaridade necessário para o modelo de detecção de novidades, sintetizado no Algoritmo 9. Considera-se $\mathbf{x}_i \in \mathbb{R}^P$ como o exemplo de teste; $\mathbf{X}_{tr} \in \mathbb{R}^{P \times N}$, o conjunto de treinamento contendo N exemplos conhecidos $\mathbf{x}_{tr_i} \in \mathbb{R}^P$; e $k \in \mathbb{N}^*$, o número de vizinhos.

Algoritmo 9 Geração do escore de dissimilaridade baseado em LOF

```
1: função ESCORE-DISSIMILARIDADE-LOF( $\mathbf{x}_t, \mathbf{X}_{tr}, k$ )
2:   Executar apenas uma vez ou quando  $\mathbf{X}_{tr}$  ou  $k$  forem modificados:
3:   para  $i \leftarrow 1$  até  $N$  faça
4:     para  $j \leftarrow 1$  até  $N$  faça
5:        $\text{dist}(\mathbf{x}_{tr_i}, \mathbf{x}_{tr_j}) \leftarrow \|\mathbf{x}_{tr_i} - \mathbf{x}_{tr_j}\|^2$ , para  $i \neq j$ .
6:     fim para
7:      $[\text{dist}_k(\mathbf{x}_{tr_i}), N_k(\mathbf{x}_{tr_i})] \leftarrow$  Armazenar os  $k$  menores valores de  $\text{dist}(\mathbf{x}_{tr_i}, \mathbf{X}_{tr})$ 
      e os exemplos de  $\mathbf{X}_{tr}$  associados, desconsiderando o próprio  $\mathbf{x}_{tr_i}$ .
8:   fim para
9:   para  $i \leftarrow 1$  até  $N$  faça
10:    para  $j \leftarrow 1$  até  $N$  faça
11:       $\text{da}_k(\mathbf{x}_{tr_i}, \mathbf{x}_{tr_j}) \leftarrow \max\{\text{dist}_k(\mathbf{x}_{tr_j}), \text{dist}(\mathbf{x}_{tr_i}, \mathbf{x}_{tr_j})\}$ , para  $i \neq j$ .
12:    fim para
13:     $\text{dal}_k(\mathbf{x}_{tr_i}) \leftarrow 1 / \left( \frac{\sum_{\mathbf{x}_{tr_j} \in N_k(\mathbf{x}_{tr_i})} \text{da}_k(\mathbf{x}_{tr_i}, \mathbf{x}_{tr_j})}{\#N_k(\mathbf{x}_{tr_i})} \right)$ .
14:  fim para
15:
16:  Executar sempre:
17:  para  $i \leftarrow 1$  até  $N$  faça
18:     $\text{dist}(\mathbf{x}_t, \mathbf{x}_{tr_i}) \leftarrow \|\mathbf{x}_t - \mathbf{x}_{tr_i}\|^2$ 
19:     $\text{da}_k(\mathbf{x}_t, \mathbf{x}_{tr_i}) \leftarrow \max\{\text{dist}_k(\mathbf{x}_{tr_i}), \text{dist}(\mathbf{x}_t, \mathbf{x}_{tr_i})\}$ 
20:  fim para
21:   $N_k(\mathbf{x}_t) \leftarrow$  Armazenar os  $k$  exemplos de  $\mathbf{X}_{tr}$  que acarretam os menores valores
  de  $\text{dist}(\mathbf{x}_t, \mathbf{X}_{tr})$ .
22:   $\text{dal}_k(\mathbf{x}_t) \leftarrow 1 / \left( \frac{\sum_{\mathbf{x}_{tr_i} \in N_k(\mathbf{x}_t)} \text{da}_k(\mathbf{x}_t, \mathbf{x}_{tr_i})}{\#N_k(\mathbf{x}_t)} \right)$ 
23:   $s \leftarrow \frac{1}{\#N_k(\mathbf{x}_t)} \sum_{\mathbf{x}_{tr_i} \in N_k(\mathbf{x}_t)} \frac{\text{dal}_k(\mathbf{x}_{tr_i})}{\text{dal}_k(\mathbf{x}_t)}$ 
24:  devolve  $s$ 
25: fim função
```

Capítulo 3

Sistema Proposto

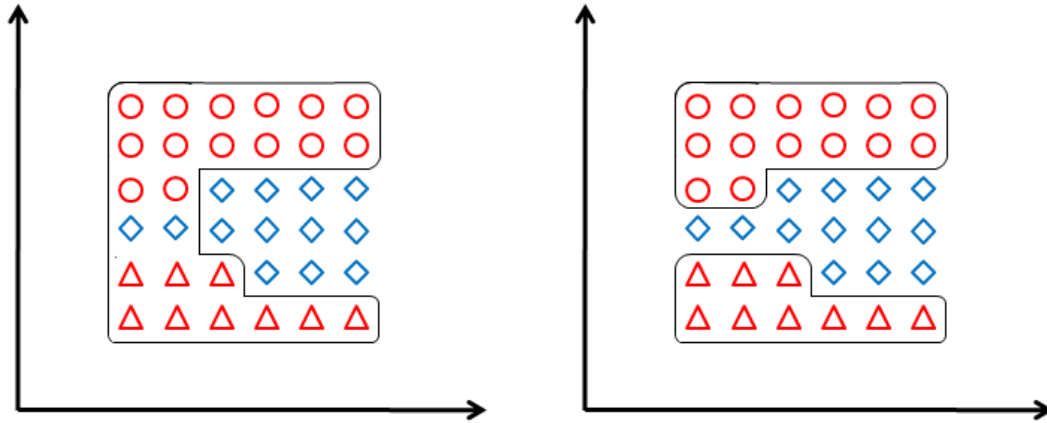
O propósito deste capítulo é descrever os principais conceitos relativos ao sistema detector de novidades e ao classificador de contatos desenvolvidos. Inicialmente, a questão da especialização dos detectores será discutida, bem como as ferramentas para a sua integração. Em sequência, serão descritos os mecanismos e os indicadores para a análise de desempenho de tais sistemas.

3.1 Níveis de Especialização

Ao se operar com múltiplas classes conhecidas, os detectores podem ser não-especializados ou especializados. O primeiro enfoque trata os dados como um conjunto único, desconsiderando qualquer informação sobre as diferentes classes presentes no desenvolvimento dos modelos, enquanto o último se utiliza desta informação para gerar modelos específicos para cada classe. Tal enfoque explorando a especialização tem se mostrado promissor, conforme relatado em publicações recentes ([2], [39], [40], [41] e [42]).

Suponha um modelo que busque separar os dados pertencentes às classes vermelha e azul, ilustrados na Figura 3.1. Observe que a classe vermelha é composta por dados de 2 subclasses: círculo e triângulo. Para uma abordagem especializada em classes, obtém-se o separador exibido na Figura 3.1a, que apresenta alguns erros de classificação ao tentar encapsular todos os dados da classe vermelha. Entretanto, ao se adotar um maior nível de especialização, neste caso, uma abordagem especializada em subclasses, pode-se obter a separação apresentada na Figura 3.1b, a qual é constituída por 2 separadores que encapsulam as subclasses vermelhas, isolando totalmente a classe azul.

O aumento no nível de especialização leva ao desenvolvimento de um maior número de modelos, contribuindo para reduzir a heterogeneidade dos dados a serem modelados. Desta forma, superfícies de encapsulamento mais simples podem ser adotadas. Ao se utilizar uma especialização pode-se ter vários modelos operando em



(a) Modelo especializado em classes. (b) Modelo especializado em subclasses.

Figura 3.1: Observa-se a presença de erros de classificação no caso (a).

paralelo ou uma divisão do espaço de dados com a conseguinte atuação de modelos especializados.

Os detectores de novidades aqui analisados consideraram três níveis de especialização: Global, Especializado em Classes, e Especializado em Navios, os quais são apresentados a seguir.

Para os exemplos descritos neste capítulo, a interpretação dos rótulos utilizados é a que se segue: a letra maiúscula designa a classe do navio, enquanto a combinação de letra e número refere-se a um navio em particular. Portanto, o rótulo B2 se refere ao navio 2 pertencente à classe B.

3.1.1 Global

O nível Global (GL), também chamado de não-especializado, considera apenas um único detector responsável por identificar se um contato é proveniente de uma dentre todas as classes conhecidas. Em operação, este produz um escore de dissimilaridade considerando uma dada técnica de detecção de novidades (discutidas na Seção 2.3) entre o contato sob avaliação e uma “superclasse” formada por todas as classes conhecidas. Se tal medida for igual ou superior a um determinado limiar, tal contato é rotulado como novidade.

A Figura 3.2 ilustra um detector GL, o qual atua sobre 4 navios hipotéticos (A1, A2, B1, B2) que formam uma “superclasse”. Neste caso, o detector trabalha com diferentes navios de variadas classes, portanto, frequentemente heterogêneas. Com isso, a sua eficiência tende a se degradar à medida em que se aumenta o número de classes conhecidas consideradas.

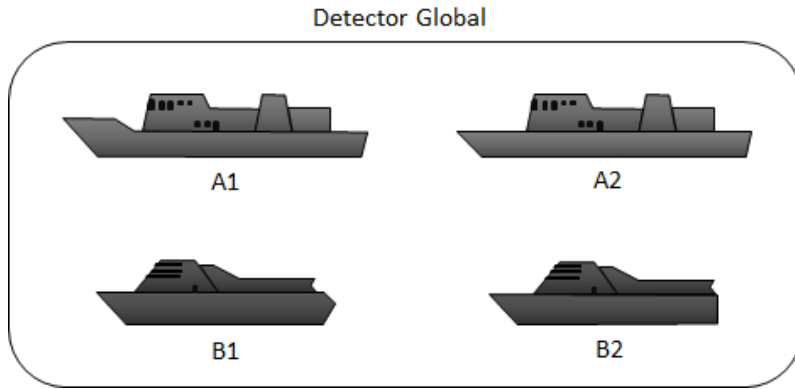


Figura 3.2: Detector Global ou não-especializado.

3.1.2 Especializado em Classes

O nível Especializado em Classes (EC) é constituído pela fusão de detectores classe-especialistas, isto é, ele considera a existência de vários detectores, cada qual especializado em uma determinada classe. Como cada detector produz um escore de dissimilaridade, é necessário que o sistema apresente uma estratégia para a integração das saídas dos modelos (a ser apresentada na Seção 3.2).

A Figura 3.3 exhibe dois detectores EC, que atuam independentemente para as classes hipotéticas (A, B). Aqui, a introdução da especialização reduz a heterogeneidade das informações com as quais os detectores devem lidar, portanto melhores resultados de detecção são esperados.

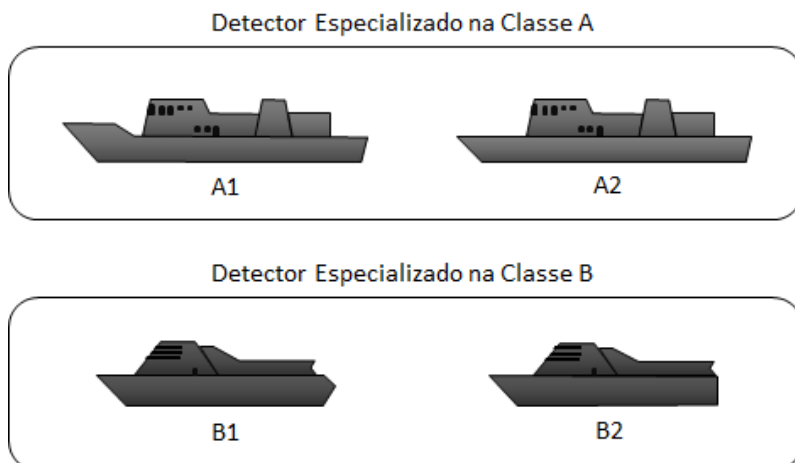


Figura 3.3: Detectores Especializados em Classes.

3.1.3 Especializado em Navios

O nível Especializado em Navios (EN) é similar ao EC, embora considere modelos navio-especialistas. Tal abordagem é originalmente proposta neste trabalho.

A Figura 3.4 mostra quatro detectores EN que atuam independentemente para os navios hipotéticos (A1, A2, B1, B2). Espera-se que o maior nível de especialização leve a melhores resultados relativamente ao nível EC, visto a maior homogeneidade das subclasses consideradas, porém ao custo de um maior número de detectores requeridos.

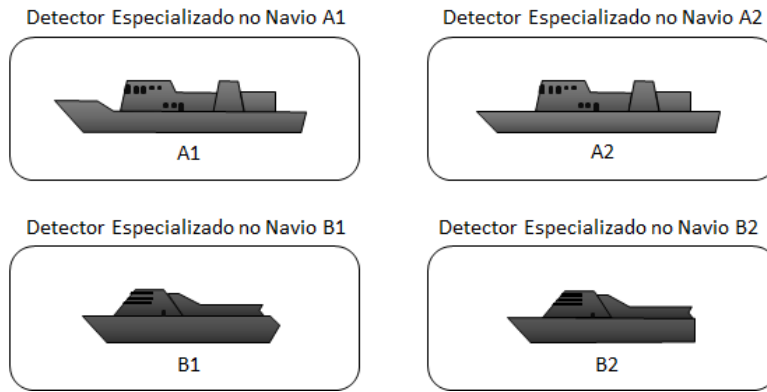


Figura 3.4: Detectores Especializados em Navios.

3.2 Topologias

Tradicionalmente, a sistemática exibida na Figura 3.5 é aquela mais utilizada para a realização das tarefas de detecção de novidades (em vermelho) e de classificação dos contatos (em azul), ou seja, tais sistemas operam de forma independente.

O sinal advindo do sistema de sonar passivo, já devidamente processado, pode ativar um sistema de classificação, gerando um rótulo de classe $r_C \in \{A, B, \dots\}$ ou um rótulo de navio $r_N \in \{A1, A2, \dots, B1, \dots\}$. Por outro lado, caso o sinal pré-processado atue no detector de novidades, uma variável contínua s é produzida no bloco gerador do score. Tal valor segue para um bloco decisor, que o compara com um limiar interno λ e, por fim, gera a variável binária $d \in \{0, 1\}$, onde 0 indica um sinal conhecido ($s < \lambda$); e 1, uma novidade ($s \geq \lambda$).

Observa-se que a Topologia Tradicional é a única que permite uma classificação e detecção de novidades simultânea quando se opera com o detector GL, visto que o detector de novidades não é capaz de discernir as classes conhecidas entre si, apenas o conjunto destas com relação às desconhecidas. Entretanto, ao se operar com múltiplos detectores (EC ou EN), é possível propor topologias, no espírito de *one-class classifiers* [19], que realizam simultaneamente a detecção de novidades e

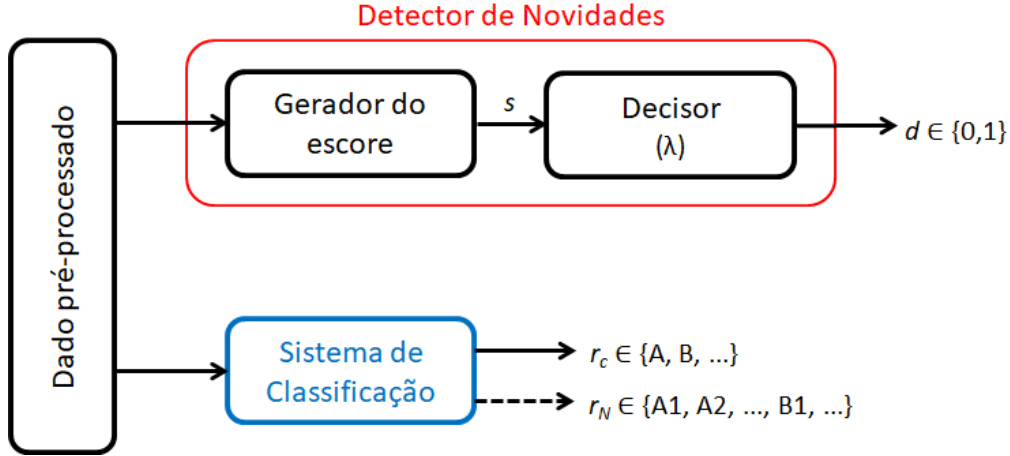


Figura 3.5: Topologia Tradicional.

a classificação, sendo a última um subproduto da primeira, conforme ficará mais claro a seguir. Cabe destacar que este será o foco do trabalho, tendo em vista que o desenvolvimento de classificadores com ou sem especialização em classes, segundo metodologias tradicionais ou de decomposição do tipo um-contra-todos já foi realizado em outros trabalhos ([11], [43], [44], [45], [46]).

Nesta linha, três propostas serão avaliadas para este problema: a Unânime, a Competitiva e a Hierárquica, descritas a seguir.

3.2.1 Unânime

A Figura 3.6 exhibe o diagrama em blocos da Topologia Unânime (UN). Em vermelho, os blocos do sistema que são utilizados na tarefa de detecção de novidades; em azul, os blocos designados para a classificação dos contatos.

Cada bloco gerador do escore produz uma variável contínua s_i ($1 \leq i \leq N$) relacionada à dissimilaridade entre o sinal de teste (\mathbf{x}_t) e a classe ou navio associados ao detector. Tal variável alimenta o bloco decisor i , que possui um limiar interno λ_i , e produz como saída a variável binária $d_i = 1$, se $s_i \geq \lambda_i$; e $d_i = 0$, caso contrário. Por fim, é considerado um E-lógico sobre todos os d_i 's. Logo, um contato só é considerado novidade ($d = 1$) com a concordância unânime entre todos os detectores ($\forall i: d_i = 1$); ao passo que um contato será conhecido ($d = 0$), se ao menos um detector assim o definir ($\exists i: d_i = 0$).

A classificação do contato é de responsabilidade do bloco seletor, o qual elege o índice $i \in \{1, 2, \dots, N\}$ correspondente ao valor mínimo de s_i dentre o conjunto de escores $S = \{s_1, s_2, \dots, s_N\}$. Em seguida, a variável i é introduzida no bloco interpretador para a geração do rótulo final de classe (r_C) ou de navio (r_N). Em outras palavras, a função deste último bloco é interpretar a variável i em termos

de um rótulo com significado real. Observa-se que os N blocos geradores de escore participam destas duas tarefas.

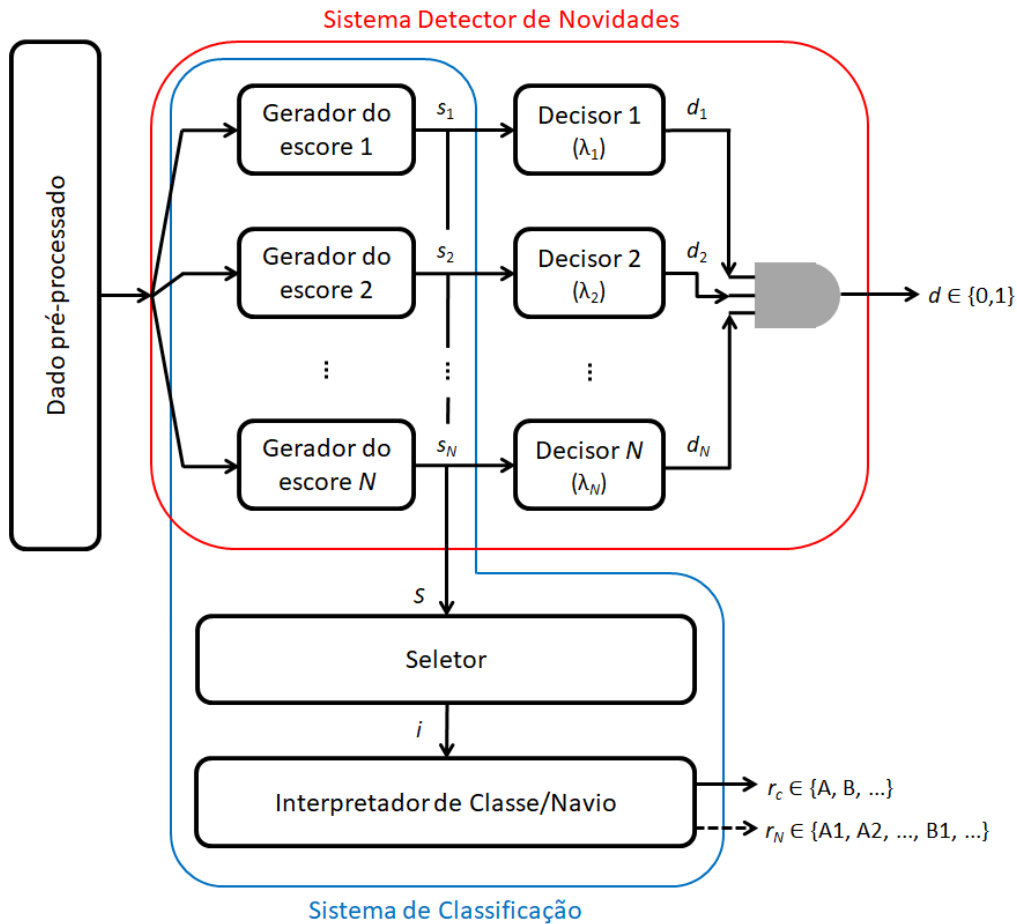


Figura 3.6: Topologia Unânime.

3.2.2 Competitiva

O diagrama em blocos da Topologia Competitiva (CO) é apresentado na Figura 3.7. Destacados em vermelho estão os blocos do sistema presentes na tarefa de detecção de novidades; em azul, os blocos destinados à classificação dos contatos. Ressalta-se a presença dos N geradores de escore e do bloco seletor nas duas tarefas.

Nesta topologia, o dado pré-processado é avaliado por todos os blocos geradores de escore. Entretanto, há a presença do bloco seletor, contendo um mecanismo de escolha que opta apenas pelo escore de menor dissimilaridade, portanto associado à classe ou ao navio de maior verossimilhança com o dado sob avaliação. Por seqüência, esta saída s_i ($1 \leq i \leq N$) é aplicada ao bloco decisor i correspondente, o qual apresenta um limiar interno λ_i . Quando $s_i < \lambda_i$, temos $d_i = 0$; e quando $s_i \geq \lambda_i$, temos $d_i = 1$. Conseqüentemente, a variável de decisão final será $d = d_i$.

Neste caso, a classificação do contato é mais direta, visto que o bloco seletor também está presente na etapa da detecção de novidades. De forma similar à Topologia UN, a variável discreta i pode ser aplicada no bloco interpretador para a obtenção do rótulo de classe (r_C) ou de navio (r_N).

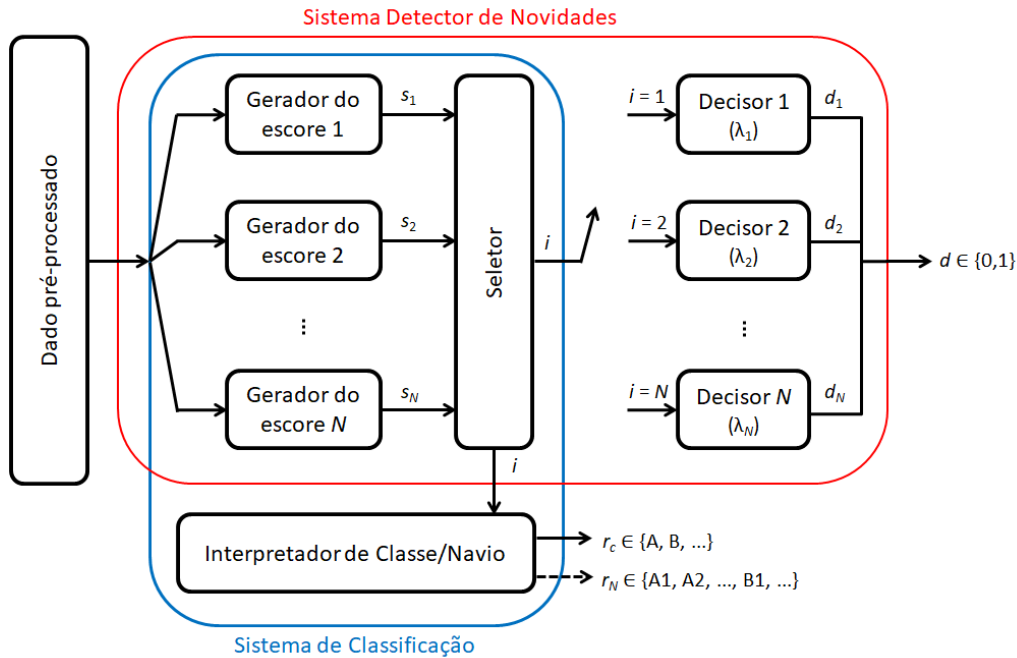


Figura 3.7: Topologia Competitiva.

Uma vantagem da utilização da Topologia CO é o menor número requerido de comparações com limiares, dado a utilização de apenas um bloco decisor i do sistema, ao passo que na Topologia UN, todos os N blocos decisores são requeridos para se obter a decisão final. Ademais, a Topologia CO é menos sensível a regiões de confusão entre as classes, quando comparada a Topologia UN por fazer uso de descritores de classe mais específicos.

3.2.3 Hierárquica

A Figura 3.8 exibe o diagrama em blocos da Topologia Hierárquica (HI). Novamente, a caixa em cor vermelha demarca os blocos que atuam na tarefa da detecção de novidades e a caixa em cor azul, os blocos utilizados para a classificação de contatos.

Esta topologia opera em dois estágios. Inicialmente, o dado pré-processado é submetido a um bloco seletor que, especialmente para esta arquitetura, consiste em um sistema de classificação de alta-eficiência, cujo papel é rotular o dado como oriundo de uma dentre as classes ou navios conhecidos através de uma variável i ($1 \leq i \leq N$). Tal sistema consiste em um classificador multiclasse arbitrário,

usualmente constituído por um único modelo, de forma a considerar todas as classes/navios conhecidos simultaneamente, portanto inferindo melhor as fronteiras das classes conhecidas, aspecto útil para aquelas de maior confusão.

Em seguida, apenas o par gerador do escore i e decisor i é selecionado para produzir a saída do sistema. De forma similar às outras topologias, a saída s_i gerada será comparada a um limiar interno λ_i presente no decisor i , de forma que para $s_i \geq \lambda_i$, temos $d_i = 1$; e caso contrário, $d_i = 0$. Claramente, a variável de decisão final será dada por $d = d_i$.

Relativamente à classificação do contato, esta já é provida pelo bloco seletor, logo a variável discreta i pode ser facilmente encaminhada ao bloco interpretador para a rotulação do contato em classe (r_C) ou navio (r_N).

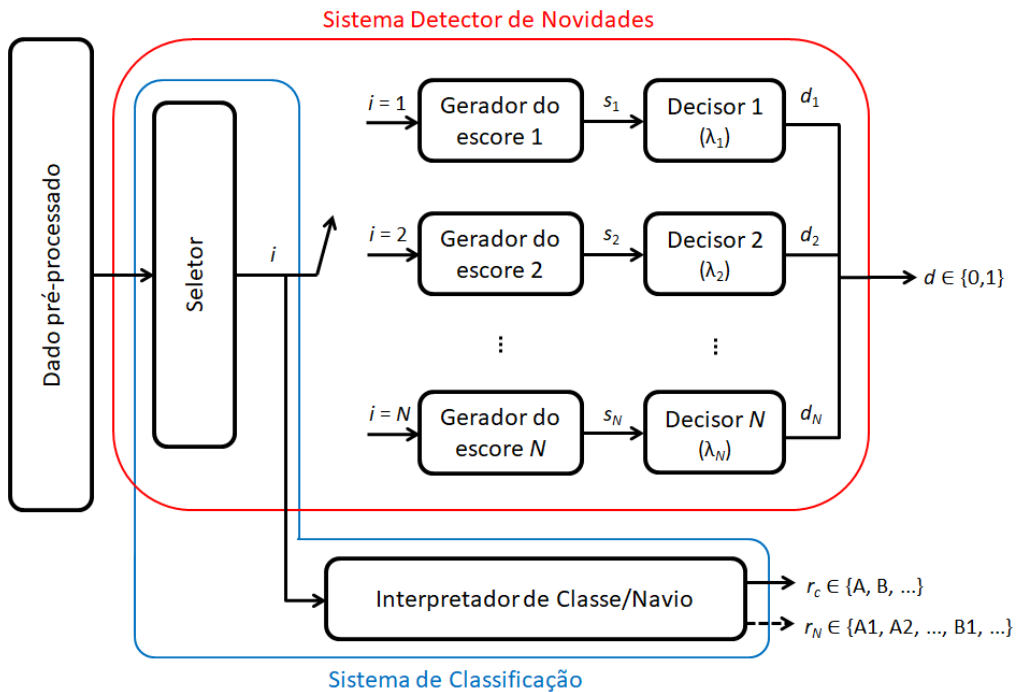


Figura 3.8: Topologia Hierárquica.

A Topologia HI possui como diferencial em relação às topologias anteriores o fato de que a tarefa de classificação de contatos não é influenciada pelas saídas dos geradores de escore existentes. Assim, o bloco seletor e os blocos geradores do escore podem ser construídos de forma independente, para, após, serem integrados ao sistema. Ademais, como ocorre na Topologia CO, apenas um bloco decisor i é utilizado para a decisão final. Note que caso o bloco seletor seja construído por *one-class classifiers*, o sistema será classe-escalável [18], isto é, novas classes podem ser adicionadas sem um possível maior impacto no conhecimento previamente adquirido.

3.3 Avaliação de Desempenho

A avaliação de desempenho é uma etapa imprescindível para a definição dos hiperparâmetros dos modelos e, por conseguinte, para sua comparação final quanto às eficiências obtidas. Para isto, faz-se necessário definir métodos de reamostragem dos dados e índices de desempenho adequados ao problema.

3.3.1 Métodos de Reamostragem

Dados do mundo real usualmente envolvem variáveis cuja distribuição é desconhecida. Procedimento comum é buscar estimar o comportamento da distribuição original através de amostras desta população.

Ao se desenvolver um dado modelo, utilizam-se amostras para o treinamento e subsequente avaliação. Portanto, um aspecto desejável é que o processo de amostragem empregado para a geração de conjuntos de treinamento e avaliação resulte em amostras que representem de forma eficaz a realidade, conforme ilustrado na Figura 3.9.

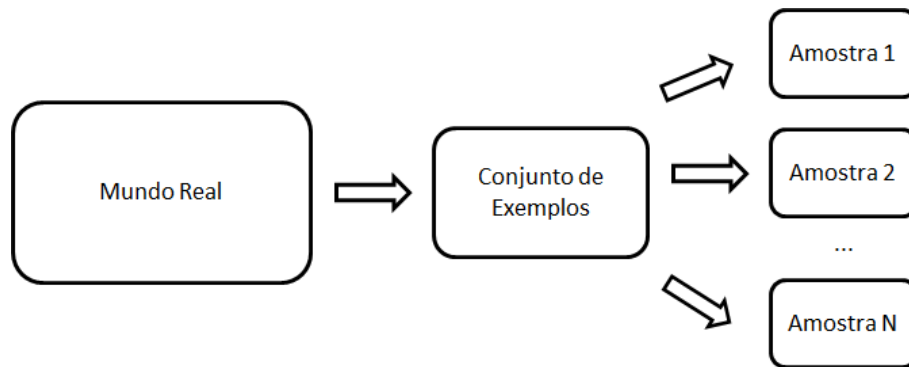


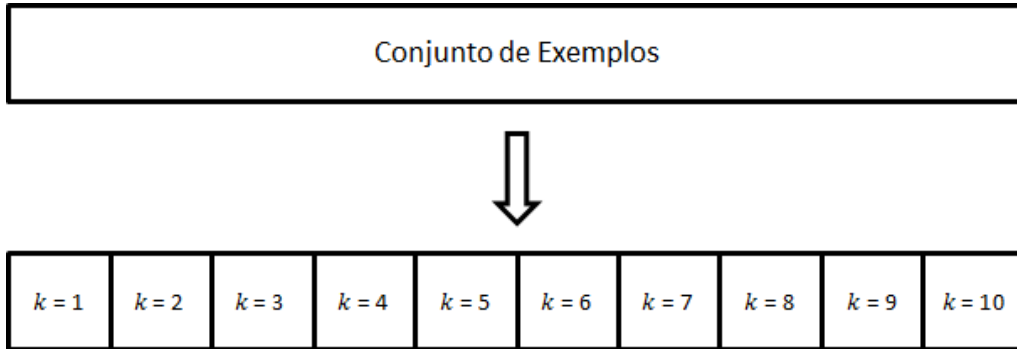
Figura 3.9: Representação do processo de amostragem.

Ao se efetuar repetidamente a retirada de amostras de um conjunto de exemplos para se criar uma amostra destinada ao desenvolvimento de um modelo, realiza-se o processo de reamostragem, o qual é fundamental para a obtenção de um modelo final ajustado. O papel dos métodos de reamostragem é permitir tal análise sem a necessidade de coletar novos dados.

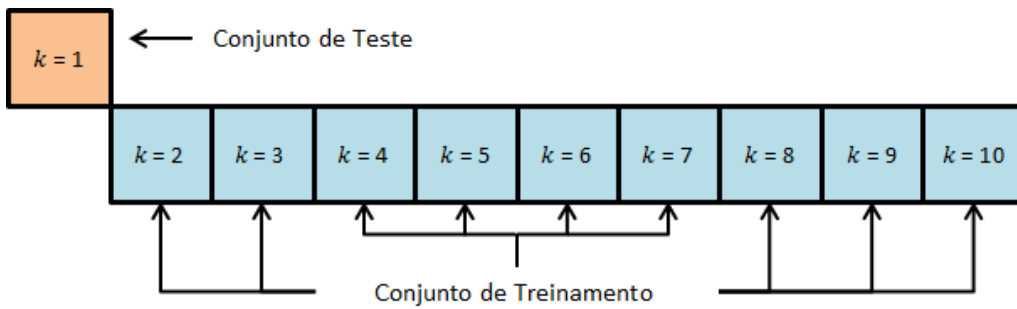
A Validação Cruzada via *k-fold* é um método de reamostragem que gera resultados com menor viés¹ quando comparado a outros métodos, como a divisão simples em treino e teste. Sua abordagem consiste em particionar, de forma aleatória, o conjunto de exemplos em *k* grupos ou *folds*, de tamanho aproximadamente igual. Em sequência, o primeiro *fold* é tratado como o conjunto de teste do modelo que será criado com base nos $(k - 1)$ *folds* restantes, que definem o conjunto de treinamento

¹Um viés elevado implica em um modelo com desempenho subestimado ou superestimado.

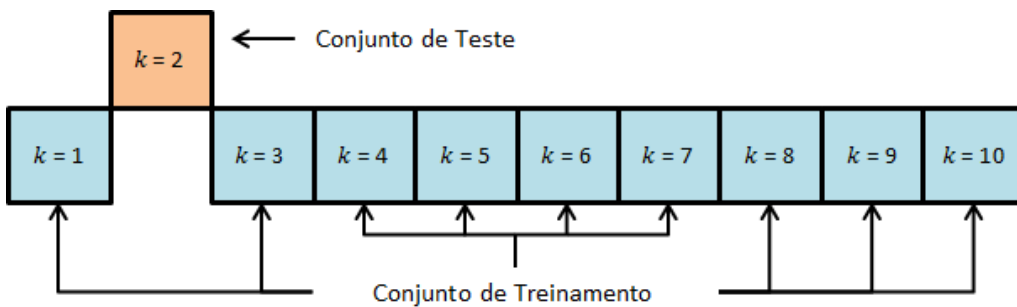
[47]. Este processo é repetido k vezes, de forma que em cada uma das repetições, o k -ésimo *fold* será reservado para o teste. O desempenho final é dado pela média dos desempenhos de cada modelo, inferida nos *fold*s de teste. A Figura 3.10 ilustra o método para $k = 10$.



(a) O conjunto de exemplos é particionado em 10 *fold*s.



(b) Na rodada #1, o primeiro *fold* é reservado para o conjunto de teste, enquanto os 9 *fold*s restantes integram o conjunto de treinamento.



(c) Na rodada #2, o segundo *fold* é reservado para o conjunto de teste, enquanto os 9 *fold*s restantes integram o conjunto de treinamento.

Figura 3.10: Ilustração do método Validação Cruzada via k -*fold* para $k = 10$ em suas duas primeiras rodadas.

A definição de um valor adequado para k é importante, visto que está associado a um compromisso viés-variância. Ao passo que k aumenta, a diferença entre o conjunto de treino e o conjunto original diminui, de forma que o viés do modelo

tende a decrescer. Uma escolha aconselhável para k é 5 ou 10, a qual ocasiona uma variância modesta e um baixo viés [48].

Caso a divisão dos exemplos nos *folds* preserve a proporção previamente existente entre as categorias existentes no conjunto original, diz-se que o processo é estratificado.

Para a definição dos melhores hiperparâmetros de cada modelo, é recomendável a utilização de um conjunto de validação, que possua exemplos diferentes daqueles presentes nos conjuntos de treinamento e teste. Desta forma, um modelo inicial é gerado através do conjunto de treinamento para uma faixa de hiperparâmetros (Figura 3.11a), e depois ajustado através do conjunto de validação (Figura 3.11b). De posse dos melhores hiperparâmetros, pode-se gerar um novo modelo considerando a união dos conjuntos de treinamento e validação — aqui chamado conjunto de desenvolvimento (Figura 3.11c). Por fim, este novo modelo terá seu desempenho final avaliado com o uso do conjunto de teste (Figura 3.11d).

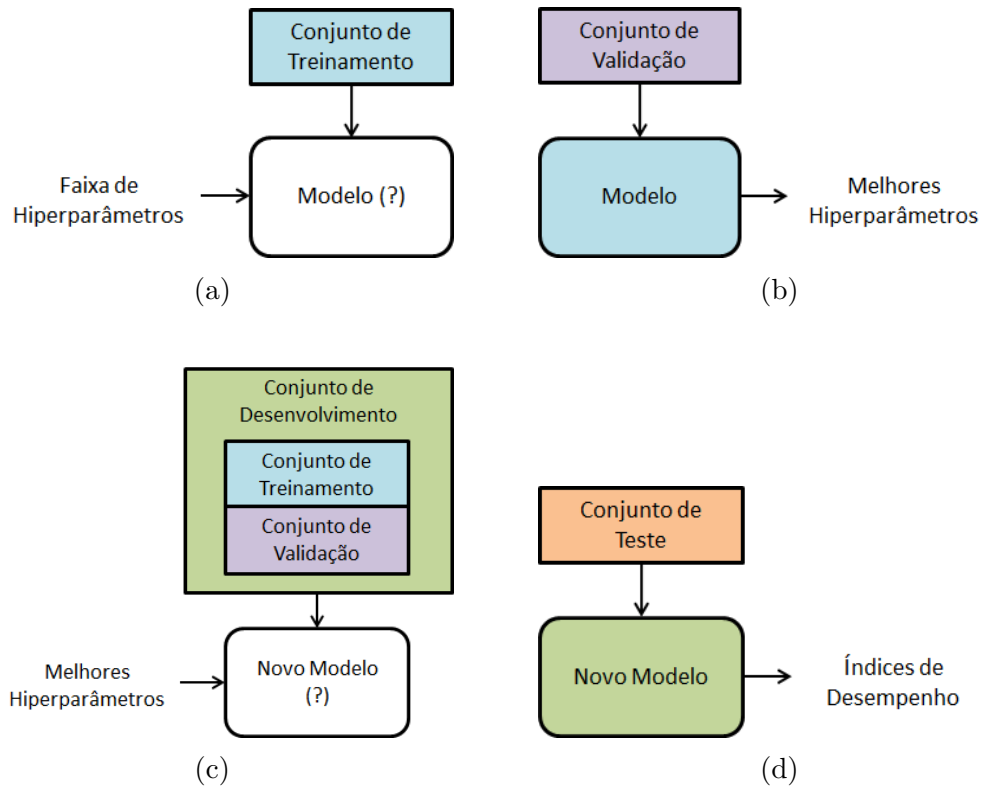
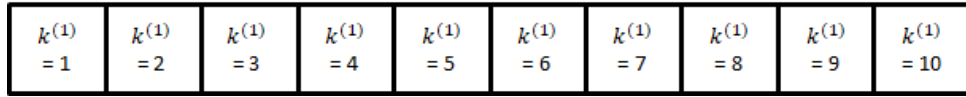
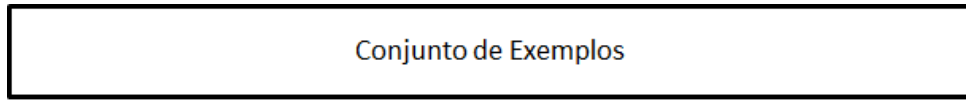
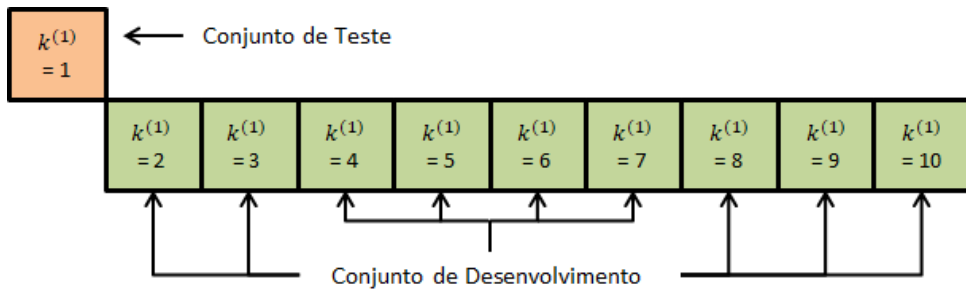


Figura 3.11: Ilustração da dinâmica dos conjuntos de Treinamento, Validação e Teste no processo de desenvolvimento e avaliação de um modelo.

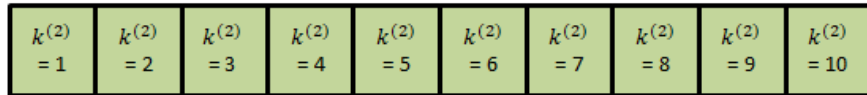
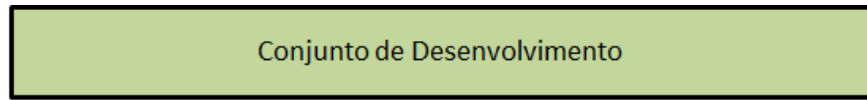
A definição de quais exemplos integrarão cada um dos três conjuntos (treinamento, validação e teste) pode ser feita através do método de reamostragem *k-fold* aninhado [49], como ilustrado na Figura 3.12. Segundo esta técnica, o conjunto de exemplos é particionado, de forma estratificada, em $k^{(1)}$ *folds* em um primeiro



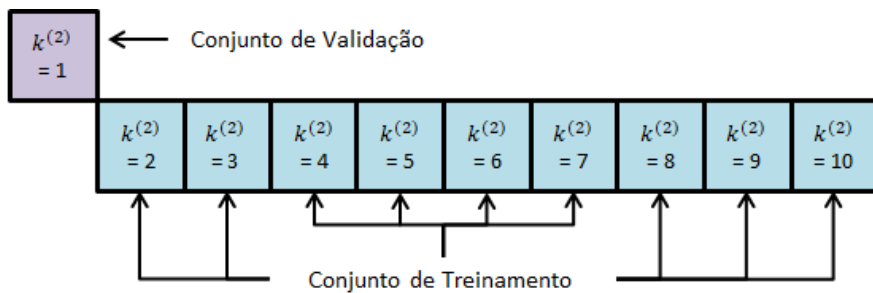
(a) O conjunto de exemplos é particionado em $k^{(1)} = 10$ *fold*s.



(b) Na rodada #1 externa, o primeiro *fold* é reservado para o conjunto de teste, enquanto os 9 *fold*s restantes integram o conjunto de desenvolvimento.



(c) O conjunto de desenvolvimento é particionado em $k^{(2)} = 10$ *fold*s.



(d) Na rodada #1 interna, o primeiro *fold* é reservado para o conjunto de validação, enquanto os 9 *fold*s restantes integram o conjunto de treinamento.

Figura 3.12: Ilustração do método *k-fold* aninhado para $k^{(1)} = k^{(2)} = 10$.

estágio (Figura 3.12a), dos quais um *fold* será reservado para o conjunto de teste, totalizando $k^{(1)}$ possíveis conjuntos de teste distintos. Na sequência, os $k^{(1)} - 1$ *folds* restantes são unidos para gerar o conjunto de desenvolvimento (Figura 3.12b).

Em um segundo estágio, o conjunto de desenvolvimento é dividido, de forma estratificada, em $k^{(2)}$ *folds* (Figura 3.12c), dos quais cada um deles será destinado para um conjunto de validação, totalizando $k^{(1)} \times k^{(2)}$ conjuntos de validação possíveis no total. Por fim, os $k^{(2)} - 1$ *folds* restantes irão definir um dos $k^{(1)} \times k^{(2)}$ conjuntos de treinamento (Figura 3.12d). Portanto, ao longo do processo, são produzidos também $k^{(1)} \times k^{(2)}$ modelos.

3.3.2 Índices de Desempenho

É imprescindível a definição de índices de desempenho consistentes para o problema, de forma que diferentes sistemas construídos por meio de um conjunto de algoritmos e variados hiperparâmetros possam ser comparados.

Para a detecção de novidades, serão consideradas: (i) a Taxa Média de Detecção de Eventos Conhecidos; (ii) a Taxa Média de Detecção de Novidades; e (iii) a Área sob a Curva de Operação do Detector. Por sua vez, para a classificação de classes conhecidas será computada a Acurácia Média dos Acertos.

Taxa Média de Detecção de Eventos Conhecidos

Seja um conjunto de navios conhecidos $\mathcal{K}_N = \{\mathcal{K}_N^1, \mathcal{K}_N^2, \dots, \mathcal{K}_N^M\}$, pelo qual \mathcal{K}_N^j é constituído por subconjuntos de espectros associados aos navios conhecidos da classe j ($1 \leq j \leq M$). Por sua vez, seja $\mathcal{K}_N^j = \{K_{N_1^j}, K_{N_2^j}, \dots, K_{N_{Q_j}^j}\}$, onde $K_{N_i^j}$ representa um conjunto de espectros associado ao i -ésimo navio conhecido da classe j . A taxa de detecção do i -ésimo navio da classe j conhecida ($TK_{N_i^j}$) é dada por:

$$TK_{N_i^j} = \frac{1}{\#K_{N_i^j}} \sum_{\mathbf{x}_l \in K_{N_i^j}} I(d(\mathbf{x}_l) = 0), \quad 1 \leq i \leq Q_j, \quad (3.1)$$

onde: $\#K_{N_i^j}$ representa a cardinalidade de $K_{N_i^j}$ e a função indicadora $I(\cdot)$ possui saída 1, caso verdadeira a condição do seu argumento; e saída 0, caso contrário. Destaca-se que esta condição está atrelada à saída d do sistema, definida na Seção 3.2.

Assumindo um conjunto de classes conhecidas \mathcal{K}_C dado por $\mathcal{K}_C = \{K_{C_1}, K_{C_2}, \dots, K_{C_M}\}$, a taxa de detecção da j -ésima classe conhecida (TK_{C_j}) corresponde a:

$$TK_{C_j} = \frac{1}{Q_j} \sum_{i=1}^{Q_j} TK_{N_i^j}, \quad 1 \leq j \leq M. \quad (3.2)$$

Uma possível medida global de desempenho é a taxa média de detecção de eventos conhecidos dada por:

$$TK = \frac{1}{M} \sum_{j=1}^M TK_{C_j}, \quad (3.3)$$

a qual simplesmente corresponde à média dos valores de TK_{C_j} .

Taxa Média de Detecção de Novidades

Em relação ao reconhecimento dos navios desconhecidos, uma suposição prática para viabilizar a avaliação do sistema de detecção de novidades é dispor de uma base de dados \mathcal{U}_C formada por um conjunto de W classes, tal que $\mathcal{U}_C = \{\mathcal{U}_C^1, \mathcal{U}_C^2, \dots, \mathcal{U}_C^W\}$, onde \mathcal{U}_C^j responde pela j -ésima classe desconhecida ($1 \leq j \leq W$). Assuma também que cada \mathcal{U}_C^j é formado por R_j navios, conforme: $\mathcal{U}_C^j = \{U_{N_1^j}, U_{N_2^j}, \dots, U_{N_{R_j}^j}\}$, onde $U_{N_i^j}$ representa o conjunto de espectros associado ao i -ésimo navio de classe j desconhecida. De forma similar à Eq. (3.1), é possível definir uma taxa de detecção do i -ésimo navio da classe j desconhecida ($TU_{N_i^j}$) conforme:

$$TU_{N_i^j} = \frac{1}{\#U_{N_i^j}} \sum_{\mathbf{x}_i \in U_{N_i^j}} I(d(\mathbf{x}_i) = 1), \quad 1 \leq i \leq R_j, \quad (3.4)$$

onde $\#U_{N_i^j}$ representa a cardinalidade de $U_{N_i^j}$.

Por sua vez, considerando um conjunto de classes desconhecidas \mathcal{U}_C , a taxa de detecção da j -ésima classe desconhecida (TU_{C_j}) é dada por:

$$TU_{C_j} = \frac{1}{R_j} \sum_{i=1}^{R_j} TU_{N_i^j}, \quad 1 \leq j \leq W. \quad (3.5)$$

Por consequência, uma taxa média de detecção de eventos desconhecidos corresponderá a:

$$TU = \frac{1}{W} \sum_{j=1}^W TU_{C_j}, \quad (3.6)$$

a qual equivale à média aritmética dos valores de TU_{C_j} .

Área sob a Curva de Operação do Detector

Na literatura, é comum a utilização de uma curva formada por variados pares de taxas de verdadeiros positivos e falsos positivos, a chamada *Receiver Operating Characteristic curve* (curva ROC) [49], a qual é amplamente adotada para avaliar o desempenho final de modelos no campo da Medicina, por exemplo.

De forma similar, é possível definir uma curva para a taxa de reconhecimento de dados desconhecidos \times conhecidos, isto é, TU e TK , aqui referida como Curva de Operação do Detector (COD), considerando diferentes escolhas para os limiares de detecção de novidades λ_i^2 . Para tal, é possível se considerar diferentes percentis do escore de dissimilaridade obtido para os dados de validação conhecidos. Por exemplo, para o detector GL, o percentil de 20% da variável de decisão definirá um limiar λ associado a um valor de TK supostamente correspondente a 20%. Para os detectores EC, são considerados os percentis do escore de dissimilaridade computados para os dados de validação de cada classe conhecida separadamente. Assim, cada detector EC_i apresentará uma taxa de reconhecimento da classe i teoricamente igual ao percentil definido. De forma similar, para os detectores EN, consideram-se os percentis do escore de dissimilaridade gerados para os dados de validação de cada navio conhecido. Logo, cada detector EN_i apresentará um valor de reconhecimento do navio i pertencente a uma das classes conhecidas teoricamente igual ao percentil definido, para um determinado valor de limiar λ_i .

Uma medida quantitativa da qualidade da COD é a AUC (*area under the curve*) [49]. Um detector ideal apresentaria ambos TK e TU iguais a 100%, resultando em uma AUC unitária. Quanto maior for o valor da AUC, maiores serão os valores de TK associados a um mesmo TU , portanto mais eficaz será o detector. A Figura 3.13 ilustra três CODs genéricas em um problema arbitrário, cujos valores de AUC são crescentes: $A < B < C$.

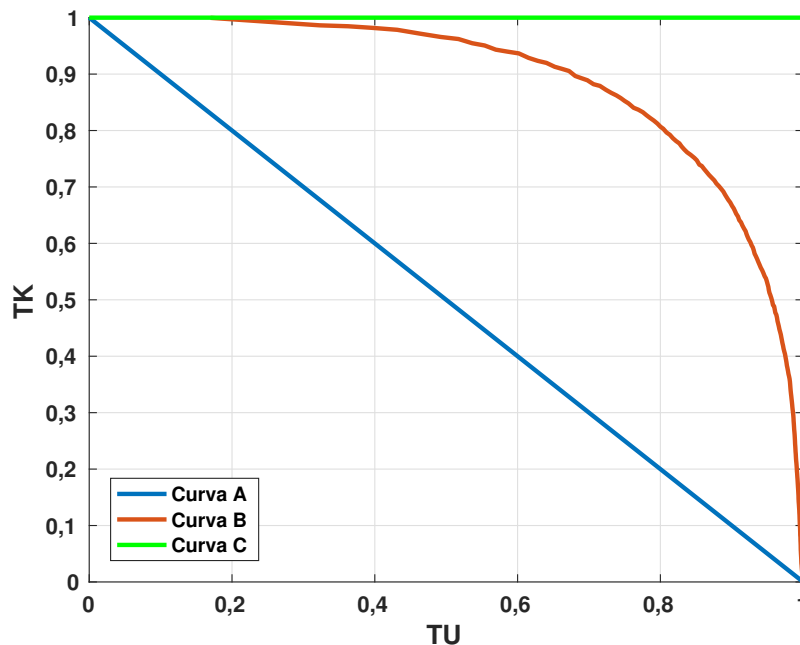


Figura 3.13: Exemplo ilustrando três CODs em um problema arbitrário.

²Definidos na Seção 3.2

Acurácia Média de Acertos

Seja o mesmo conjunto de navios conhecidos $\mathcal{K}_N = \{\mathcal{K}_N^1, \mathcal{K}_N^2, \dots, \mathcal{K}_N^M\}$ supracitado, onde $\mathcal{K}_N^j = \{K_{N_1^j}, K_{N_2^j}, \dots, K_{N_{Q_j}^j}\}$. A taxa de acertos na classificação do i -ésimo navio pertencente a j -ésima classe ($TA_{N_i^j}$) é dada por:

$$TA_{N_i^j} = \frac{1}{\#K_{N_i^j}} \sum_{\mathbf{x}_l \in K_{N_i^j}} I(r_{C_l} = y_l), \quad 1 \leq i \leq Q_j, \quad (3.7)$$

onde: $\#K_{N_i^j}$ representa a cardinalidade de $K_{N_i^j}$ e a função indicadora $I(\cdot)$ possui saída 1, caso a saída r_{C_l} produzida pelo sistema de classificação³ seja igual ao rótulo de classe real y_l ; e 0 em caso contrário.

Novamente, seja o conjunto de classes conhecidas dado por $\mathcal{K}_C = \{K_{C_1}, K_{C_2}, \dots, K_{C_M}\}$. A taxa de acertos no reconhecimento da i -ésima classe (TA_{C_i}) é dada por:

$$TA_{C_j} = \frac{1}{Q_j} \sum_{i=1}^{Q_j} TA_{N_i^j}, \quad 1 \leq j \leq M. \quad (3.8)$$

Como medida global de desempenho será considerada a acurácia média, correspondente a:

$$TA = \frac{1}{M} \sum_{j=1}^M TA_{C_j}, \quad (3.9)$$

definida pela média aritmética simples dos valores de TA_{C_j} .

3.4 Sintonia de Hiperparâmetros

No projeto dos detectores de novidades, faz-se necessária a definição dos hiperparâmetros de cada técnica considerada, os quais estão listados na Tabela 3.1.

Para a sintonia dos hiperparâmetros, neste trabalho é considerado o método de reamostragem de Validação Cruzada via k -fold aninhado, para $k = 10$. Em síntese, pelo último processo, foram produzidos um total de 100 trios de conjuntos de treinamento, validação e teste.

Para cada trio, foram gerados modelos com variados hiperparâmetros sobre o conjunto de treinamento, os quais foram avaliados considerando o conjunto de validação, tendo a AUC como figura de mérito. O melhor hiperparâmetro foi definido como aquele associado ao maior valor da AUC no conjunto de validação.

A escolha do hiperparâmetro ótimo também considerou testes estatísticos para a produção de modelos parcimoniosos. Portanto, foi aplicado o teste de *Friedman*

³Definido na Seção 3.2.

Tabela 3.1: Hiperparâmetros presentes em cada uma das técnicas.

PCA	Variância explicada (Var_e)
KPCA	Variância explicada (Var_e) Largura do <i>kernel</i> (γ)
GMM	Número de distribuições base da mistura (m)
kNN	Número de vizinhos mais próximos (k)
s-kNN	Limiar do critério de parada (ϵ) Número de vizinhos mais próximos (k)
LOF	Número de vizinhos mais próximos (k)
OCSVM	Largura do <i>kernel</i> (K_s) Penalidade associada às variáveis de folga (ν)

($p < 0,05$) para a detecção de diferenças estatisticamente significativas. Nos casos onde esta premissa se mostrou verdadeira, considerou-se o teste de *Nemenyi* ($\alpha = 0,05$) para averiguar entre quais pares de comparações tal diferença é significativa [49]. Desta forma, se dois hiperparâmetros apresentam valores quanto à figura de mérito estatisticamente similares, é escolhido o valor associado a um modelo mais simples.

Para que o processo de modelagem do detector não dependesse da utilização das classes desconhecidas, no caso dos detectores especializados, considera-se cada classe ou navio, isoladamente, no processo de sintonia, bem como apenas os dados supostamente conhecidos. Em resumo, hipoteticamente assumindo M classes conhecidas: $\mathcal{K}_C = \{K_{C_1}, K_{C_2}, \dots, K_{C_M}\}$, uma dada técnica e um valor de hiperparâmetro, foram calculados valores para a AUC considerando variados pares de K_{C_i} e K_{C_m} , onde K_{C_i} é suposta conhecida e K_{C_m} desconhecida, resultando em variados valores de AUC, aqui denotados como $AUC_{i,m}$, onde $1 \leq i \neq m \leq M$.

Para cada detector i , foi inferida a AUC_i correspondente a média das $AUC_{i,m}$ segundo:

$$AUC_i = \frac{1}{M-1} \sum_{m=1, m \neq i}^M AUC_{i,m}. \quad (3.10)$$

Para facilitar o entendimento, a Figura 3.14 ilustra as classes do conjunto de validação que participarão do cálculo da $AUC_{i,m}$. Para cada modelo construído com a classe conhecida i presente no conjunto de treinamento, este procedimento é executado $M-1$ vezes, variando a classe m que será tratada como desconhecida para esta análise.

De forma similar, o mesmo procedimento foi realizado para o enfoque especializado em navios, substituindo-se a análise de classes por uma análise de navios.

No caso do detector GL, cabe observar que não é possível a sintonia dos hiper-

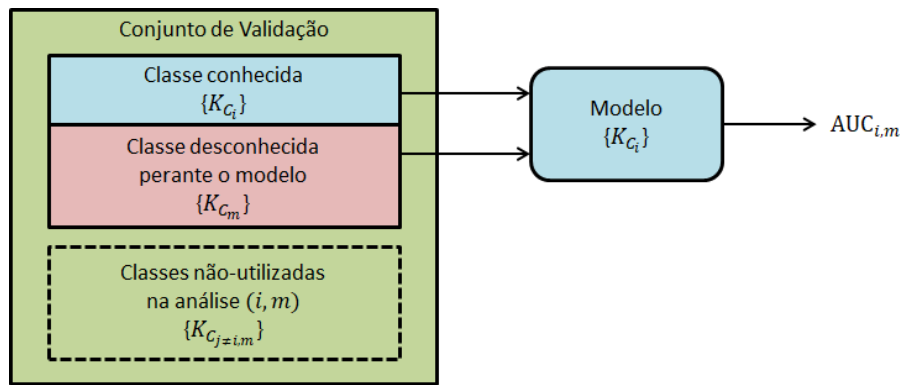


Figura 3.14: Representação dos subconjuntos do conjunto de validação utilizados para o cálculo da $AUC_{i,m}$.

parâmetros por Validação Cruzada sem o uso das classes supostamente desconhecidas. Logo o modelo será gerado diretamente sob o conjunto de desenvolvimento, sem considerar o conjunto de validação, e avaliado para o conjunto de teste considerando a faixa de hiperparâmetros.

Capítulo 4

Resultados

Este capítulo sintetiza os experimentos realizados para a avaliação das diferentes técnicas consideradas neste trabalho quanto às tarefas de detecção de novidades e de classificação de contatos. Inicialmente, a base de dados utilizada e ambas modalidades de particionamento, tanto para a definição de quais classes são conhecidas, bem como aquelas supostamente desconhecidas, assim como relativamente à formação dos conjuntos de treino, validação e teste são descritos. Em sequência, os hiperparâmetros para as simulações são determinados considerando as diferentes abordagens de construção dos detectores. Por fim, os resultados relativamente às diferentes técnicas e arquiteturas são discutidos.

4.1 Base de Dados

Os experimentos utilizaram sinais gravados em ensaios controlados, realizados na raia acústica da Marinha do Brasil, localizada em Arraial do Cabo. A base de dados, a qual foi fornecida pelo Instituto de Pesquisas da Marinha (IPqM), é constituída pelo áudio irradiado por 28 navios, pertencentes a 8 diferentes classes de embarcações, em 263 corridas de prova.

Em cada corrida, o áudio emitido pelas embarcações foi captado por um hidrofone localizado junto ao fundo da raia acústica. Em sequência, o sinal resultante foi amostrado, pré-processado e normalizado, conforme esquema detalhado em [45], visando a obter janelas espectrais de 557 pontos que cobrem a faixa de frequência de interesse (0 a 3 kHz).

O conjunto resultante é constituído por pares de janelas espectrais e seus respectivos rótulos. Tal conjunto é formado por 8 classes de embarcações: $\mathcal{W}_C = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{A, B, C, D, E, F, G, H\}, i = 1, \dots, 29277\}$ ou, de forma mais específica, por 28 navios: $\mathcal{W}_N = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{A1, A2, A4, B1, B2, B3, B5, C1, C2, C3, C4, C5, D1, D3, D4, E1, E2, E3, E4, F1, F2, F3, F4, G1, G2, H1, H2, H5\}, i = 1, \dots, 29277\}$. A Tabela 4.1 sintetiza para cada classe:

os navios existentes o número de corridas realizadas (NC) e o número de espectros (NE) disponíveis. Observa-se uma razoável variação de tais valores ao se comparar diferentes classes.

Tabela 4.1: Distribuição do quantitativo de corridas (NC) e de espectros (NE) para as diferentes classes e navios.

Classe	Navio	NC	NE	Classe	Navio	NC	NE
A	A1	11	1191	B	B1	8	914
	A2	4	504		B2	4	700
	A4	7	737		B3	9	1224
					B5	5	594
	Total	22	2432		Total	26	3432
Classe	Navio	NC	NE	Classe	Navio	NC	NE
C	C1	5	512	D	D1	13	1401
	C2	10	1081		D3	4	406
	C3	6	799		D4	12	1265
	C4	6	705				
	C5	16	1700				
	Total	43	4797		Total	29	3072
Classe	Navio	NC	NE	Classe	Navio	NC	NE
E	E1	21	2462	F	F1	6	636
	E2	7	717		F2	8	857
	E3	31	3505		F3	6	628
	E4	7	391		F4	7	813
	Total	66	7075		Total	27	2934
Classe	Navio	NC	NE	Classe	Navio	NC	NE
G	G1	8	867	H	H1	16	1680
	G2	11	1276		H2	9	975
					H5	6	737
	Total	19	2143		Total	31	3392

Cada espectro presente na base de dados consiste em uma linha do chamado espectrograma, um gráfico que traça os valores de magnitude das componentes de frequência do sinal na faixa considerada ao longo do tempo, neste problema correspondente a uma janela temporal com duração de 186 ms. A Figura 4.1a apresenta o espectrograma referente a uma corrida arbitrária do Navio A1, para a qual se observa uma concentração de raias em uma faixa de baixa frequência até 1 kHz, padrão similar ao observado para outra corrida (Figura 4.1b).

Ao se avaliar o espectrograma do Navio A2 (Figura 4.1c), embora também relativo a uma embarcação da classe A, constata-se uma concentração maior das raias em frequências inferiores a 0,3 kHz, bem como de raias nas frequências de 0,8 e 1,2 kHz. Dessa forma, conclui-se que a análise especializada em navios pode beneficiar as tarefas de detecção de novidades e classificação.

Naturalmente, classes diferentes devem apresentar outras assinaturas espectrais características, tal como exibido na Figura 4.1d.

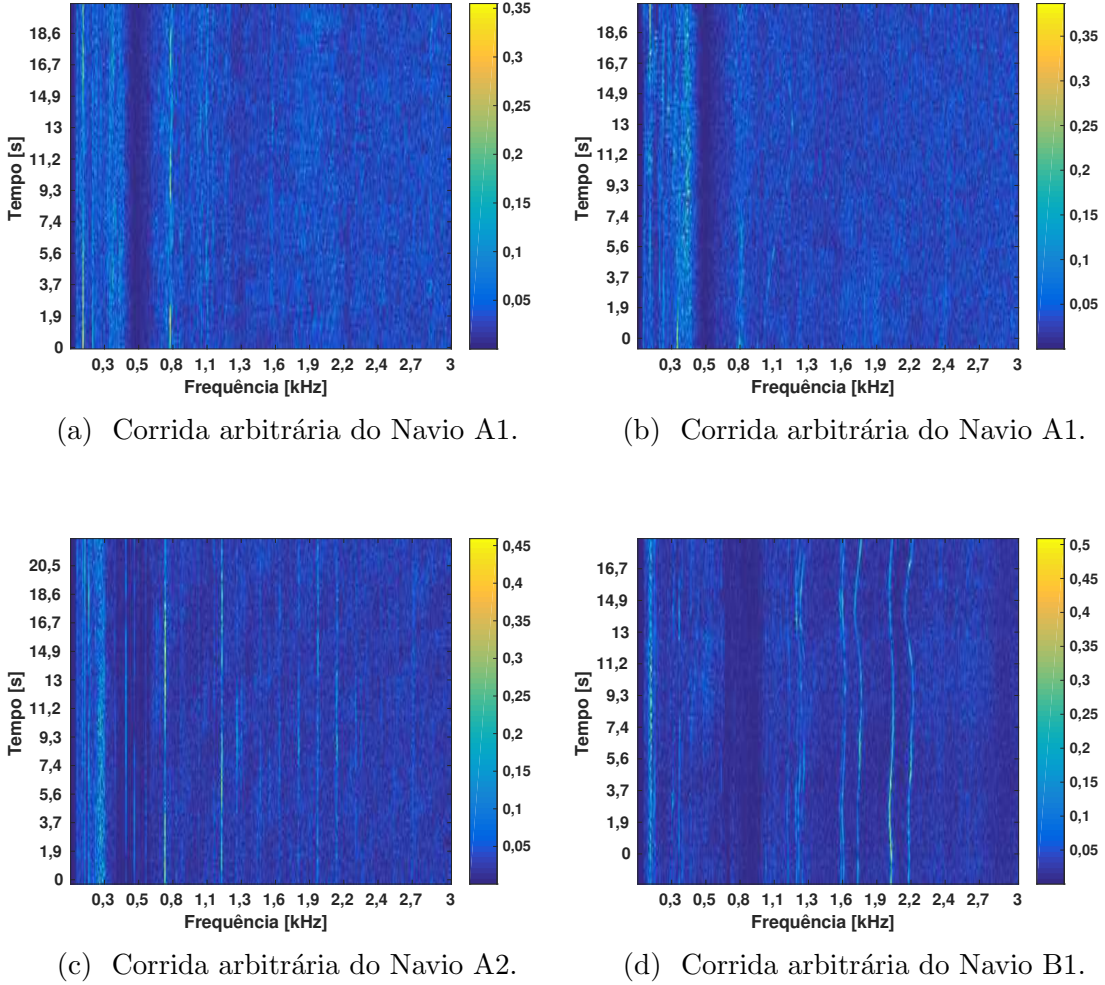


Figura 4.1: Espectrogramas referentes a quatro cenários de operação.

4.2 Particionamento dos Dados

Relativamente à divisão da base de dados adotada para os experimentos, primeiramente, definem-se quantas e quais serão as classes reservadas para compor o grupo de contatos conhecidos \mathcal{K}^1 e desconhecidos \mathcal{U}^2 . Em sequência, tem-se a aplicação do procedimento de Validação Cruzada, apenas em \mathcal{K} , para a formação dos conjuntos de treinamento, validação e teste, segundo o procedimento descrito na Seção 3.3.

A avaliação dos detectores considerou o particionamento da base de dados em: (i) um grupo de contatos de classe conhecida $\mathcal{K}_C = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{B, C,$

¹Este grupo refere-se ao conjunto de exemplos, apresentado na Seção 3.3.1.

²Este grupo sempre será utilizado, em sua totalidade, nos conjuntos de teste.

E, H}, $i = 1, \dots, 18696$ } e outro de navios conhecidos $\mathcal{K}_N = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{B1, B2, B3, B5, C1, C2, C3, C4, C5, E1, E2, E3, E4, H1, H2, H5\}, i = 1, \dots, 18696\}$; (ii) um grupo de contatos de classe desconhecida estratificados por classes $\mathcal{U}_C = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{A, D, F, G\}, i = 1, \dots, 10581\}$ ou estratificados por navios $\mathcal{U}_N = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^{557}, y_i \in \{A1, A2, A4, D1, D3, D4, F1, F2, F3, F4, G1, G2\}, i = 1, \dots, 10581\}$.

A separação da base em 4 classes conhecidas e 4 classes desconhecidas foi adotada por caráter de simetria. Ademais, a combinação escolhida para as classes que compõem cada grupo foi aquela que apresentou a detecção mais crítica, conforme resultados de [45].

4.3 Hiperparâmetros escolhidos

O sistema de classificação presente na Topologia GL e na Topologia HI foi baseado, de forma exclusiva, na técnica kNN, para uma maior simplicidade e face aos bons resultados obtidos. Sua sintonia considerou valores de $k = \{1 \leq k \leq 10 \mid k \in \mathbb{N}^*\}$, utilizando como figura de mérito: (i) a média dos acertos de classificação de cada classe conhecida, para o enfoque EC; (ii) a média dos acertos de classificação de cada navio conhecido, para o enfoque EN. O valor de $k = 3$ foi o melhor ao se considerar o sistema que contempla os modelos especializados em classes; e $k = 4$, para a abordagem especializada em navios. Por sua vez, as faixas de hiperparâmetros avaliados para cada técnica de detecção de novidades são apresentados na Tabela 4.2. Nas seções subsequentes, são apontados os valores selecionados para cada técnica, topologia e nível de especialização.

Tabela 4.2: Faixas de valores de hiperparâmetros investigados para cada técnica adotada como geradora do escore de novidade.

Técnica	Hiperparâmetro	Valores investigados
PCA	Var_e (%)	10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99
		10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99
KPCA	γ	0,01, 0,1, 1
		1, 2, 3, 4, 5
GMM	m	1, 2, 3, 4, 5
kNN	k	1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50
		0,3, 0,4, 0,5, 0,6, 0,7
s-kNN	k	1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50
		2, 3, 4, 5, 10, 15, 20, 25
LOF	K_s	10, 15, 20, 25, 30
		ν

4.4 Sistema Detector de Novidades

Ao se aplicar uma dentre as técnicas de detecção de novidades descritas na Seção 2.3 para a construção dos geradores de escore, obtiveram-se os resultados que serão apresentados a seguir.

4.4.1 PCA

A Tabela 4.3 sintetiza o percentual de variância explicada selecionado pelo procedimento de validação cruzada para a técnica PCA, considerando os três níveis de especialização: Global, Classes, e Navios. É interessante observar uma razoável variabilidade entre diferentes classes e navios. Adicionalmente, percebe-se que o enfoque Global demanda um maior percentual de representatividade dos dados, conforme esperado, visto representar um conjunto de dados mais heterogêneo.

Tabela 4.3: Percentual de variância explicada (Var_e) para os detectores baseados na técnica PCA.

Global															
85															
Especializado em Classes															
B 75			C 50			E 80			H 25						
Especializado em Navios															
B1	75	B3	75	C1	65	C3	45	C5	30	E2	75	E4	45	H4	40
B2	55	B5	60	C2	50	C4	45	E1	70	E3	70	H1	15	H5	35

Na Figura 4.2 é exibido o diagrama de caixa para valores da AUC associados a diferentes combinações de topologia (UN, CO, HI) e especialização (GL, EC, EN). Claramente, a especialização em navios produz um ganho significativo em relação à especialização em classes. Adicionalmente, para a maioria dos cenários, as topologias especializadas apresentam um resultado superior ao da topologia global.

O teste de hipóteses ($\chi^2(6) = 58,9, p < 0,001$) confirma um melhor desempenho da abordagem EN nos três cenários. A Figura 4.3 apresenta o diagrama de diferença crítica [50] que auxilia na identificação visual da melhor dentre as abordagens tratadas para o problema, neste caso, a EN-HI, sinalizada pelo valor do *rank* médio mais próximo de 1. Ademais, o diagrama indica quais abordagens são estatisticamente equivalentes³, através das ligações em vermelho. Os valores de *rank* médio estão contidos na Tabela 4.4 e o valor de DC neste ensaio⁴ corresponde a 2,85.

³Caso o módulo da diferença entre dois valores de *rank* médio obtidos para abordagens distintas não ultrapasse a diferença crítica ($DC = q_\alpha \sqrt{\frac{k(k+1)}{N}}$), estas são estatisticamente equivalentes.

⁴Para $k = 7$ (abordagens), $N = 10$ (conjuntos testados) e $q_{0,05} = 2,949$ (valor crítico para o teste de *Nemenyi* bi-caudal, conforme a Tabela 5a em [50]).

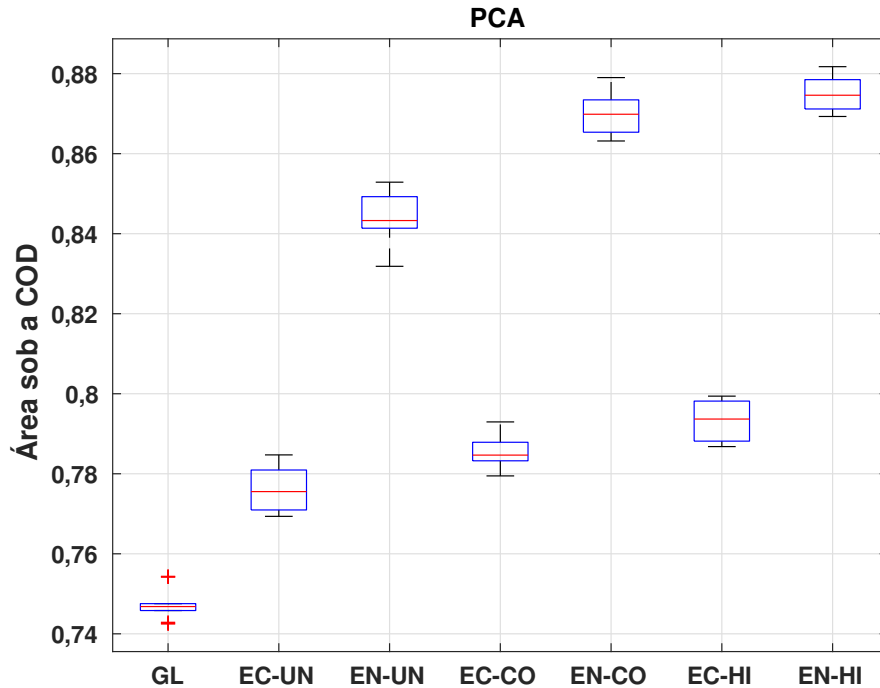


Figura 4.2: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica PCA.

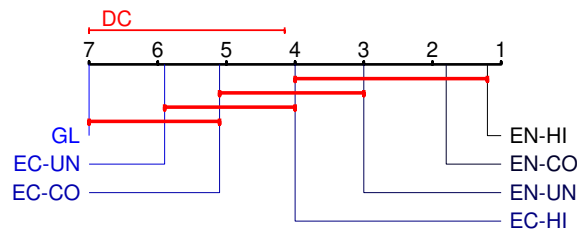


Figura 4.3: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica PCA.

Tabela 4.4: *Rank* médio para as diferentes topologias e especializações considerando a técnica PCA.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
7	5,9	3	5,1	1,8	4	1,2

Analisando em termos de p -valores, ao se considerar $\alpha = 0,05$, a combinação EN-HI é similar à EN-CO ($p = 0,996$), à EN-UN ($p = 0,505$), e à EC-HI ($p = 0,058$). Dado que a EN-HI também apresenta a maior mediana da AUC (0,8746 — $IIQ^5 = 0,0073$), esta será a escolhida.

⁵Intervalo Interquartil: calculado pela diferença entre o 3º quartil e o 1º quartil. É um indicador da dispersão de dados.

4.4.2 KPCA

Dada a necessidade da extração de autovetores de matrizes de dimensões elevadas, da ordem do quantitativo de dados envolvidos, as quais impactam fortemente no custo computacional, optou-se por realizar os experimentos por meio de uma subamostragem dos conjuntos de treinamento, isto é, utilizando uma parcela menor dos seus dados disponíveis. Assim, para o enfoque Global foram selecionados aleatoriamente 10% das amostras, enquanto para o enfoque especializado por classes considerou-se 20% das amostras⁶.

As Tabelas 4.5 e 4.6 indicam os valores de percentual de variância explicada (Var_e) e de largura do *kernel* (γ) escolhidos pelo procedimento de validação cruzada para a técnica KPCA. De forma similar à PCA, observa-se que o enfoque Global exige um maior percentual de representatividade dos dados. Quanto à largura do *kernel*, o valor $\gamma = 1$ foi eleito o melhor para a maioria dos detectores especializados.

Tabela 4.5: Percentual de variância explicada (Var_e) para os detectores baseados na técnica KPCA.

Global															
90															
Especializado em Classes															
B 80		C 60			E 70			H 70							
Especializado em Navios															
B1	70	B3	70	C1	60	C3	75	C5	65	E2	80	E4	70	H4	75
B2	60	B5	65	C2	85	C4	70	E1	75	E3	65	H1	55	H5	30

Tabela 4.6: Largura do *kernel* (γ) para os detectores baseados na técnica KPCA.

Global															
0,01															
Especializado em Classes															
B 0,01		C 1			E 1			H 1							
Especializado em Navios															
B1	1	B3	1	C1	1	C3	1	C5	1	E2	0,1	E4	1	H4	1
B2	0,1	B5	1	C2	1	C4	1	E1	1	E3	1	H1	1	H5	1

O diagrama de caixa para valores da AUC associados a diferentes combinações de topologia e nível de especialização é apresentado na Figura 4.4. Observa-se um

⁶Tais valores foram definidos por uma estratégia heurística, na qual, a partir da totalidade de amostras contidas no conjunto de treinamento, buscou-se um percentual quantitativo com o qual o algoritmo fosse capaz de lidar em um tempo viável (da ordem de 1h por simulação), para o desenvolvimento deste trabalho. Esta questão pode ser aprofundada, mas foge do escopo atual.

contraste entre os resultados obtidos ao se considerar a topologia CO com outras abordagens. Tal fato é explicado pelos diferentes valores de γ utilizados nos detectores, uma vez que este hiperparâmetro está diretamente associado ao escore de dissimilaridade. Quanto menor o γ adotado, menor o escore resultante, logo o detector especializado na classe ou navio modelado sob o menor γ tende a ser sempre escolhido na topologia CO, acarretando uma ineficiência desta abordagem. Por sua vez, as demais abordagens especializadas foram mais eficazes quando comparadas ao caso Global.

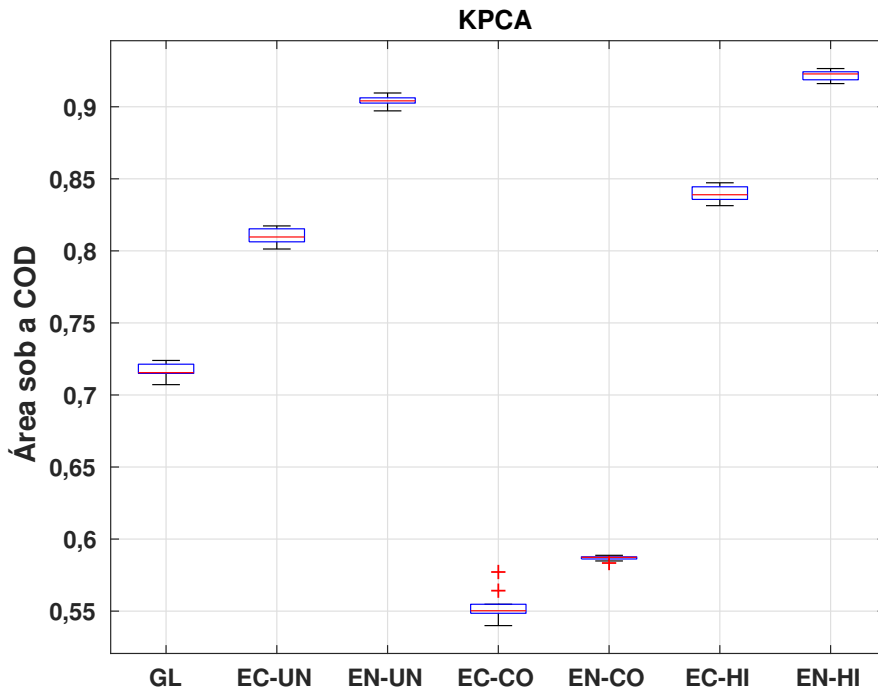


Figura 4.4: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica KPCA.

O teste de hipóteses ($\chi^2(6) = 60, p < 0,001$) resulta no diagrama de diferença crítica exibido na Figura 4.5, o qual indica que a combinação EN-HI é a melhor e estatisticamente similar às EN-UN e EC-HI. Neste ensaio, $DC = 2,85$, e os valores de *rank* médio são apresentados na Tabela 4.7.

Tabela 4.7: *Rank* médio para as diferentes topologias e especializações considerando a técnica KPCA.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
5	4	2	7	6	3	1

Tratando-se de p -valores ($\alpha = 0,05$), a combinação EN-HI é estatisticamente

similar à EN-UN ($p = 0,946$), e à EC-HI ($p = 0,370$); porém, como apresenta a maior mediana da AUC ($0,9228 - \text{IIQ} = 0,0056$), a EN-HI será a escolhida.

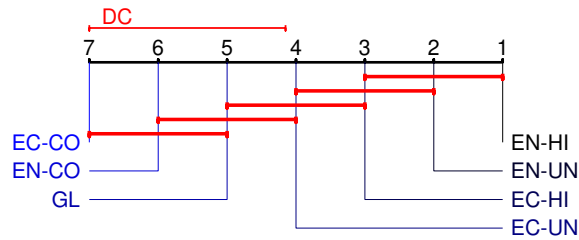


Figura 4.5: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica KPCA.

4.4.3 OCSVM

Os experimentos para o enfoque Global demandaram uma subamostragem dos dados, considerando apenas 50% das amostras⁷ disponíveis no conjunto de treinamento.

As Tabelas 4.8 e 4.9 apresentam o par largura do *kernel* (K_s) e penalidade associada às variáveis de folga (ν) selecionado por validação cruzada para a técnica OCSVM. Os modelos especializados em navios exigiram maiores valores de K_s em relação aos outros enfoques. Entretanto, quanto ao valor da penalidade, todos os modelos elegeram $\nu = 0,3$ como a melhor escolha, exceto em um dos casos (classe C - modelo especializado em classes).

Tabela 4.8: Largura do *kernel* (K_s) para os detectores baseados na técnica OCSVM.

Global															
15															
Especializado em Classes															
B	15	C	15	E	15	H	20								
Especializado em Navios															
B1	25	B3	20	C1	30	C3	30	C5	25	E2	30	E4	30	H4	30
B2	30	B5	30	C2	30	C4	30	E1	30	E3	25	H1	30	H5	30

Na Figura 4.6 é exposto o diagrama de caixa para valores de AUC associados às diferentes combinações investigadas. Nitidamente, quanto maior é o nível de especialização, melhor o desempenho dos modelos. Adicionalmente, observa-se que as topologias UN e CO apresentaram os mesmos resultados ao se considerar um determinado enfoque especializado (EC ou EN).

⁷Valor definido por uma estratégia heurística, assim como no caso da KPCA.

Tabela 4.9: Penalidade associada às variáveis de folga (ν) para os detectores baseados na técnica OCSVM.

Global															
0,3															
Especializado em Classes															
B 0,3			C 0,5			E 0,3		H 0,3							
Especializado em Navios															
B1	0,3	B3	0,3	C1	0,3	C3	0,3	C5	0,3	E2	0,3	E4	0,3	H4	0,3
B2	0,3	B5	0,3	C2	0,3	C4	0,3	E1	0,3	E3	0,3	H1	0,3	H5	0,3

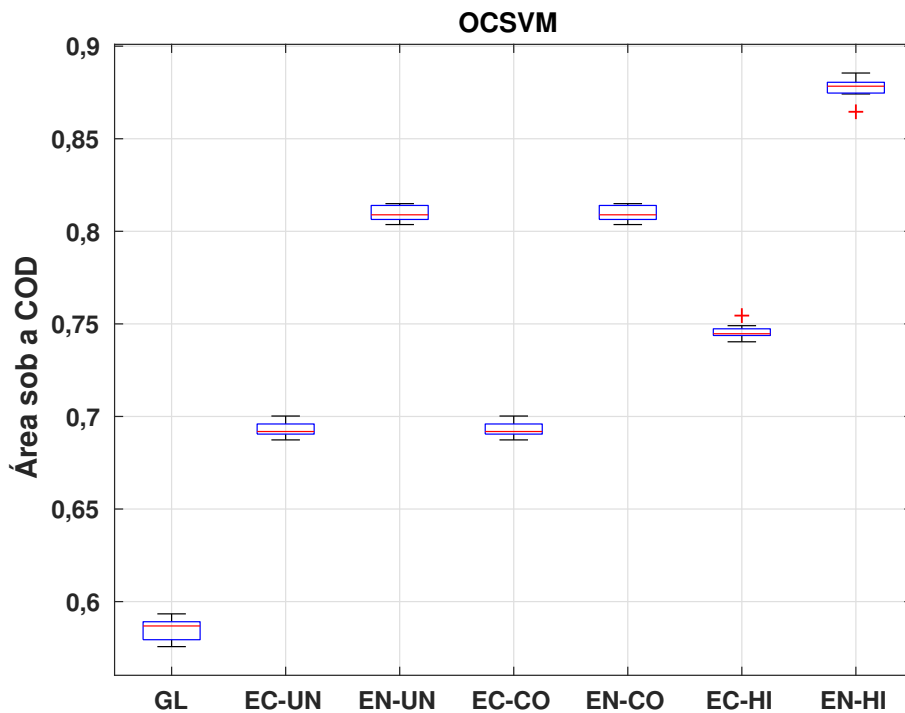


Figura 4.6: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica OCSVM.

As diferenças de desempenho dos métodos são confirmadas pelo teste de hipóteses ($\chi^2(6) = 60, p < 0,001$). A Figura 4.7 ilustra o diagrama de diferença crítica correspondente, que sinaliza a combinação EN-HI como a melhor, porém estatisticamente similar às EN-UN e EN-CO, ambas com o mesmo *rank* médio. A Tabela 4.10 exhibe os valores de *rank* médio, onde $DC = 2,85$.

Quanto aos p -valores ($\alpha = 0,05$), a combinação EN-HI é semelhante às EN-UN e EN-CO ($p = 0,695$), porém resulta no maior valor de mediana da AUC (0,8784 — IIQ = 0,0058).

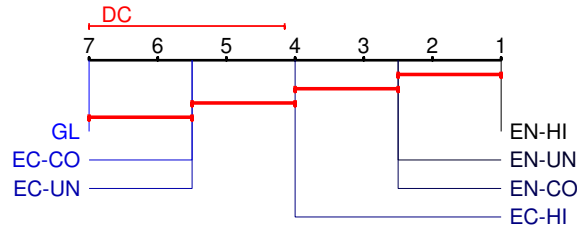


Figura 4.7: Diagrama de diferença crítica obtido na comparação de ranks médios das medidas AUC considerando a técnica OCSVM.

Tabela 4.10: *Rank* médio para as diferentes topologias e especializações considerando a técnica OCSVM.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
7	5,5	2,5	5,5	2,5	4	1

4.4.4 GMM

A quantidade de distribuições base (m) escolhida para os detectores GMM está listada na Tabela 4.11, cujo maior valor aparece no enfoque Global. A especialização em navios resultou em cenários com um número insuficiente de espectros (em número menor que a dimensão dos dados), inviabilizando a derivação dos modelos pela forma de GMM usual.

Tabela 4.11: Número de distribuições base da mistura Gaussiana (m) para os detectores baseados na técnica GMM.

Global							
5							
Especializado em Classes							
B	4	C	4	E	4	H	3

A Figura 4.8 mostra o diagrama de caixa de valores da AUC para este cenário de avaliação. Novamente, a topologia HI se destaca, haja vista a maior mediana da AUC.

A confirmação do melhor resultado é obtida pelo teste de hipóteses ($\chi^2(3) = 30, p < 0,001$). O diagrama de diferença crítica é apresentado na Figura 4.9, constatando a topologia HI como a melhor, porém estatisticamente similar à topologia CO. Nestes ensaios⁸, DC = 1,48 e os valores de *rank* médio são resumidos na Tabela 4.12.

⁸Para $k = 4$ (abordagens), $N = 10$ (conjuntos testados) e $q_{0,05} = 2,569$ (valor crítico para o teste de *Nemenyi* bi-caudal conforme a Tabela 5 em [50]).

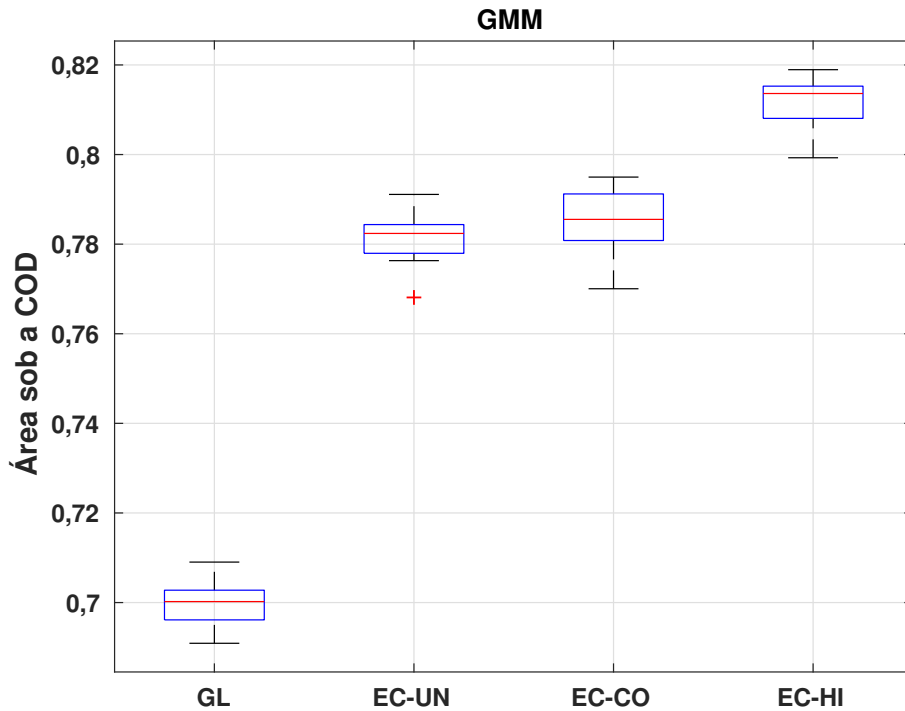


Figura 4.8: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica GMM.

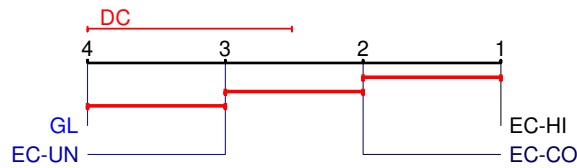


Figura 4.9: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica GMM.

Tabela 4.12: *Rank* médio para as diferentes topologias e especializações considerando a técnica GMM.

GL	EC-UN	EC-CO	EC-HI
4	3	2	1

Assumindo $\alpha = 0,05$, em termos de p -valores, as combinações EC-HI e EC-CO ($p = 0,3069$) apresentaram resultados similares, porém com uma mediana da AUC para EC-HI (0,8136 — IIQ = 0,0072) superior.

4.4.5 kNN

A Tabela 4.13 sumariza o número de vizinhos (k) selecionados para a técnica kNN, considerando três níveis de especialização. Observa-se que a variabilidade dos hi-

perparâmetros não é tão expressiva, exceto para alguns navios e classes.

Tabela 4.13: Número de vizinhos (k) para os detectores baseados na técnica kNN.

Global									
2									
Especializado em Classes									
B 1		C 2			E 2		H 10		
Especializado em Navios									
B1 1	B3 3	C1 1	C3 2	C5 15	E2 1	E4 4	H4 10		
B2 1	B5 2	C2 2	C4 2	E1 1	E3 2	H1 4	H5 4		

O diagrama de caixa para valores da AUC associados a diferentes simulações é apresentado na Figura 4.10. Conforme esperado, a abordagem EN se destaca em relação à GL e à EC, para a qual as topologias CO e HI se sobressaem com resultados similares, porém com uma maior mediana associada à CO.

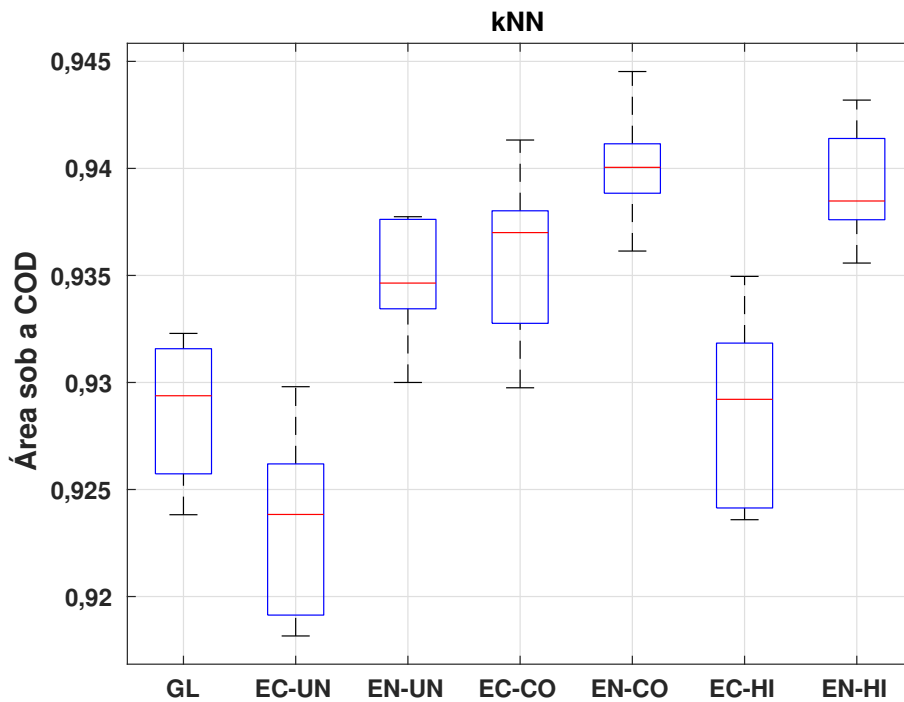


Figura 4.10: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica kNN.

Testes estatísticos confirmam as diferenças de desempenho dos métodos ($\chi^2(6) = 56,6, p < 0,001$). O diagrama de diferença crítica é apresentado na Figura 4.11, sinalizando que a combinação EN-CO é a de melhor *rank* médio e estatisticamente similar às EN-HI, EC-CO e EN-UN. Neste caso, DC = 2,85 e os valores de *rank* médio estão listados na Tabela 4.14.

Em termos de p -valores, assumindo $\alpha = 0,05$, a combinação EN-CO é semelhante à EN-HI ($p = 0,999$), à EC-CO ($p = 0,575$), bem como à EN-UN ($p = 0,165$), porém apresenta maior mediana da AUC (0,9401 — IIQ = 0,0023).

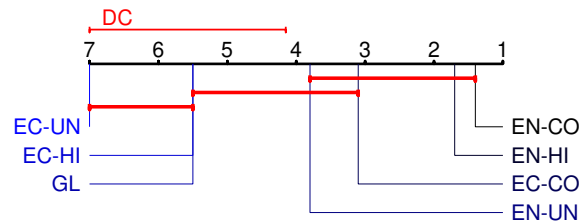


Figura 4.11: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica kNN.

Tabela 4.14: *Rank* médio para as diferentes topologias e especializações considerando a técnica kNN.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
5,5	7	3,8	3,1	1,4	5,5	1,7

4.4.6 s-kNN

Primeiramente, cada amostra do conjunto de validação foi expandida individualmente considerando as amostras do conjunto de treinamento para gerar Φ , explorando uma faixa de valores para o hiperparâmetro ϵ .

Um indicador relevante consiste no percentual médio de amostras do conjunto de treinamento original presentes nos Φ utilizados, conforme o valor de ϵ adotado. Tal informação é exibida na Figura 4.12. Observa-se que o intervalo para os valores de ϵ considerados é bem representativo, visto que os conjuntos obtidos abrangem cerca de 10% a 90% do conjunto de treinamento original.

Por sua vez, para cada um dos conjuntos Φ foi efetuado um kNN, elegendo-se o melhor número k de vizinhos através do conjunto de validação. Contudo, a avaliação dos conjuntos de teste neste caso é diferenciada. Como o conjunto de desenvolvimento corresponde ao conjunto de treinamento, portanto é constituído apenas pelo Φ obtido, há um conjunto de teste para 10 diferentes conjuntos de treinamento, totalizando 100 avaliações possíveis. Para preservar a coerência com as análises das outras técnicas, foi considerado, no entanto, o valor médio dos diferentes modelos associados a uma mesma partição de teste. A Tabela 4.15 exibe os valores da mediana da AUC para as combinações estudadas, de acordo com o valor de ϵ pré-determinado. É interessante notar que mesmo com uma parcela significativamente menor de amostras que o conjunto original, a técnica atinge resultados consideráveis.

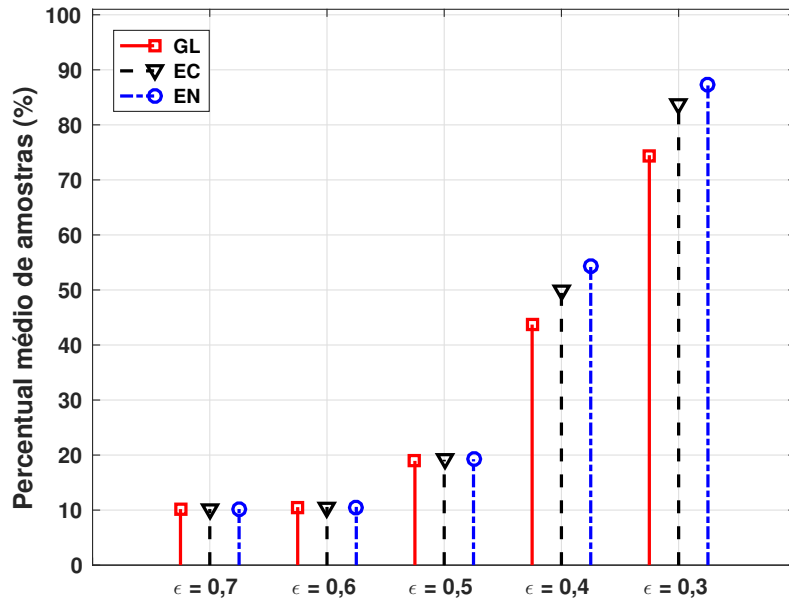


Figura 4.12: Percentual médio de amostras para os diferentes níveis de especialização considerando a faixa de ϵ investigada para a técnica s-kNN.

Tabela 4.15: Mediana da AUC para as diferentes topologias e especializações considerando a faixa de ϵ investigada para a técnica s-kNN.

	$\epsilon = 0,7$	$\epsilon = 0,6$	$\epsilon = 0,5$	$\epsilon = 0,4$	$\epsilon = 0,3$
GL	0,8490	0,8507	0,8820	0,9122	0,9272
EC-UN	0,8571	0,8589	0,8856	0,9149	0,9231
EN-UN	0,8520	0,8558	0,8927	0,9310	0,9337
EC-CO	0,8644	0,8661	0,8905	0,9249	0,9350
EN-CO	0,8729	0,8747	0,9030	0,9358	0,9395
EC-HI	0,8746	0,8766	0,8975	0,9229	0,9290
EN-HI	0,8916	0,8932	0,9121	0,9366	0,9386

Apesar de o valor $\epsilon = 0,3$ gerar os melhores resultados, seleciona-se o $\epsilon = 0,4$ para a conseguinte análise, visto que os seus resultados são estatisticamente equivalentes ($p = 0,6184$) pelo teste de hipóteses ($\chi^2(4) = 40, p < 0,001$). Ademais, este é mais interessante por operar com menos amostras, no caso, cerca de 45% a 55%.

A Tabela 4.16 apresenta o número de vizinhos (k) utilizados para a técnica s-kNN ($\epsilon = 0,4$) em seus três níveis de especialização. Observa-se que os valores obtidos são menores ou iguais aos encontrados para a técnica kNN original, conforme a Tabela 4.13.

O diagrama de caixa para valores da AUC associados às diferentes simulações é apresentado na Figura 4.13. Conforme esperado, a abordagem EN se destaca em

relação à GL e à EC. Neste cenário, a HI se sobressai com uma maior mediana.

Tabela 4.16: Número de vizinhos (k) para os detectores baseados na técnica s-kNN ($\epsilon = 0,4$).

Global															
2															
Especializado em Classes															
B	1	C	1	E	1	H	10								
Especializado em Navios															
B1	1	B3	2	C1	1	C3	2	C5	10	E2	1	E4	4	H4	10
B2	2	B5	2	C2	2	C4	2	E1	1	E3	2	H1	4	H5	4

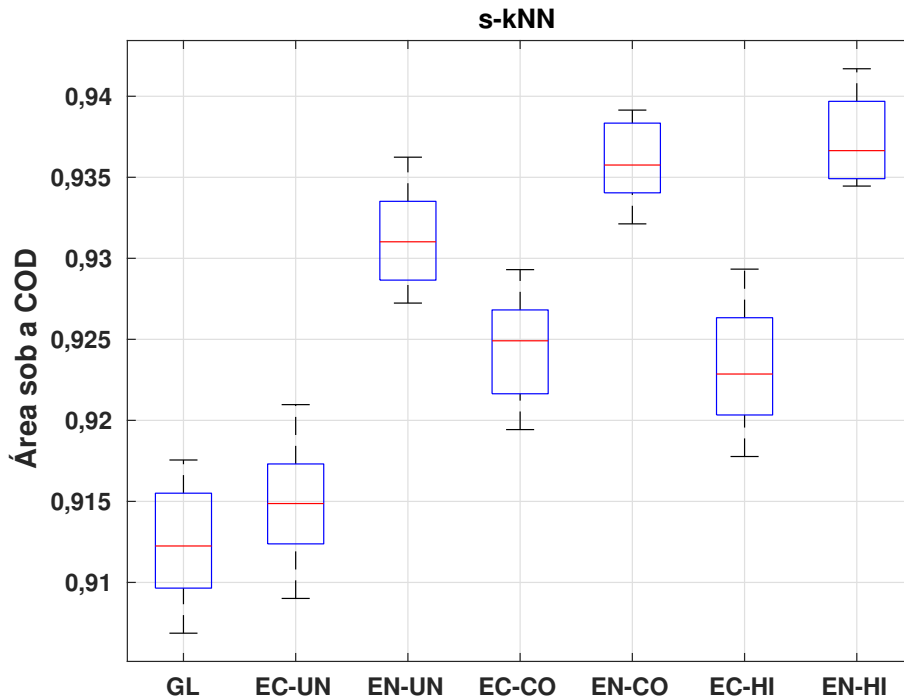


Figura 4.13: Diagrama de caixa da área sob a COD para diferentes topologias e especializações considerando a técnica s-kNN ($\epsilon = 0,4$).

Os testes estatísticos relatam as diferenças de desempenho dos métodos ($\chi^2(6) = 58,0, p < 0,001$). O diagrama de diferença crítica é apresentado na Figura 4.14, indicando que a combinação EN-HI é a de melhor *rank* médio e estatisticamente similar às EN-CO e EN-UN. Aqui, $DC = 2,85$ e os valores de *rank* médio estão contidos na Tabela 4.17.

Quanto aos p -valores ($\alpha = 0,05$), a combinação EN-HI é semelhante à EN-CO ($p = 0,982$) e à EN-UN ($p = 0,436$); porém, como apresenta maior mediana da AUC (0,9367 — IIQ = 0,0048), a EN-HI será a escolhida.

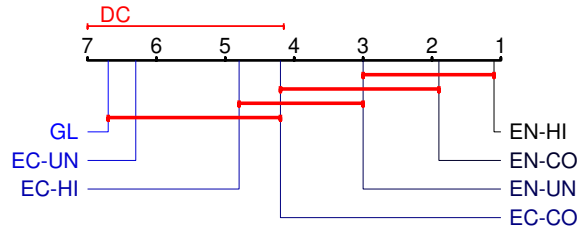


Figura 4.14: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica s-kNN.

Tabela 4.17: *Rank* médio para as diferentes topologias e especializações considerando a técnica s-kNN.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
6,7	6,3	3	4,2	1,9	4,8	1,1

4.4.7 LOF

A Tabela 4.18 reúne o número de vizinhos (k) utilizados para a técnica LOF em seus três níveis de especialização, na qual há uma razoável variabilidade dos hiperparâmetros.

Tabela 4.18: Número de vizinhos (k) para os detectores baseados na técnica LOF.

Global															
5															
Especializado em Classes															
B	10	C	3	E	4	H	20								
Especializado em Navios															
B1	3	B3	10	C1	2	C3	10	C5	5	E2	4	E4	4	H4	4
B2	5	B5	4	C2	20	C4	10	E1	10	E3	3	H1	5	H5	20

A Figura 4.15 exibe o diagrama de caixa para valores da AUC relativos aos diferentes experimentos. Observa-se que os enfoques que utilizam a Topologia HI se sobressaem, em especial o EN.

As diferenças de desempenho dos métodos são confirmadas pelo teste de hipóteses ($\chi^2(6) = 58,3, p < 0,001$). O diagrama de diferença crítica é mostrado na Figura 4.16, relatando que a combinação EN-HI é estatisticamente igual às EC-HI, EC-CO e EN-CO e possui o melhor *rank* médio. A Tabela 4.19 resume os valores de *rank* médio, onde $DC = 2,85$.

Ao se considerar $\alpha = 0,05$, quanto aos p -valores, a combinação EN-HI é semelhante à EC-HI ($p = 0,999$), à EC-CO ($p = 0,575$) e à EN-CO ($p = 0,058$), contudo

apresenta maior mediana da AUC (0,8879 — IIQ = 0,0092).

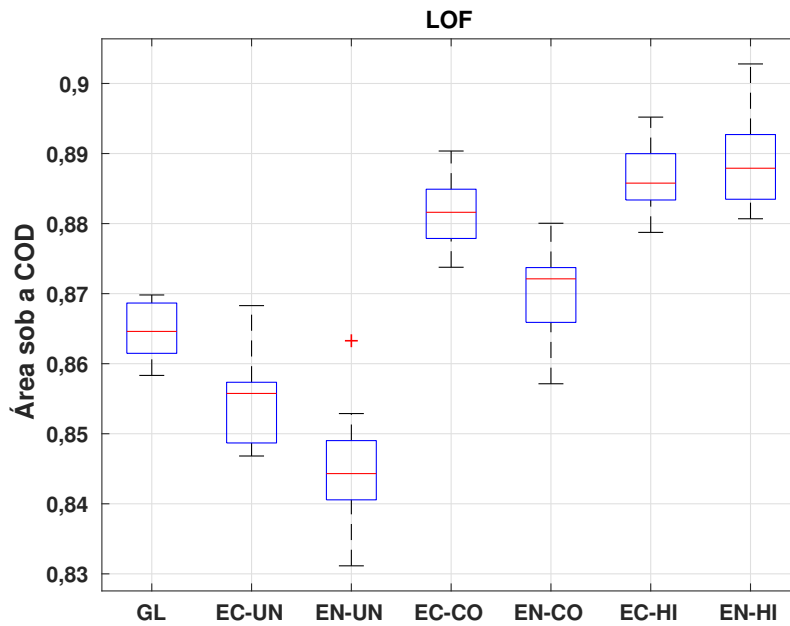


Figura 4.15: Diagrama de caixas da área sob a COD para diferentes topologias e especializações considerando a técnica LOF.

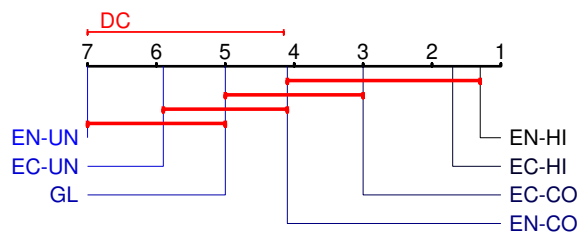


Figura 4.16: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando a técnica LOF.

Tabela 4.19: *Rank* médio para as diferentes topologias e especializações considerando a técnica LOF.

GL	EC-UN	EN-UN	EC-CO	EN-CO	EC-HI	EN-HI
5	5,9	7	3	4,1	1,7	1,3

4.4.8 Comparativo das Técnicas

A Figura 4.17 exibe o diagrama de caixa para valores da AUC associados às melhores combinações de topologia e nível de especialização associadas às técnicas anteriores: EN-HI (PCA), EN-HI (KPCA), EN-HI (OCSVM), EC-HI (GMM),

EN-CO (kNN), EN-HI (s-kNN) e EN-HI (LOF). O modelo de maior mediana da AUC é o kNN (0,9401 — IIQ = 0,0023), com desempenho estatisticamente similar ($\chi^2(6) = 58,7, p < 0,001$) ao s-kNN ($p = 0,982$) e à KPCA ($p = 0,436$).

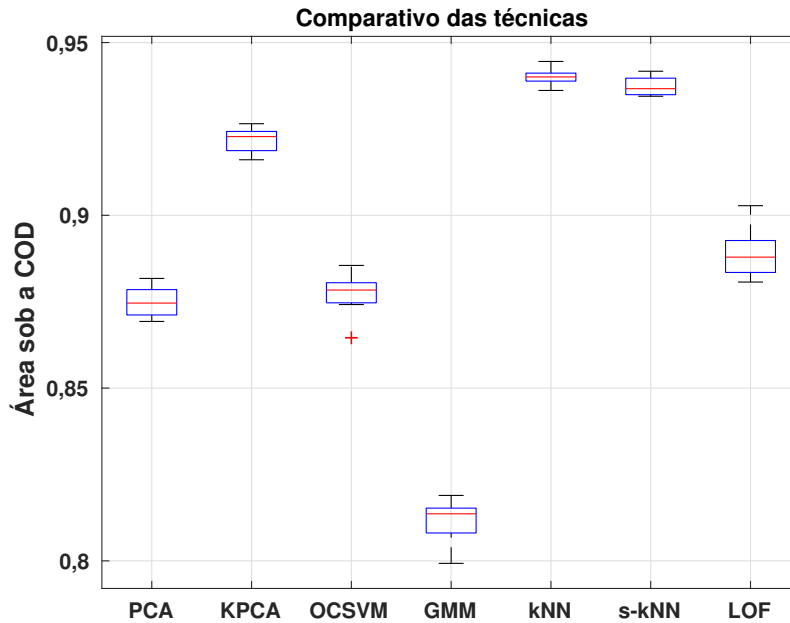


Figura 4.17: Diagrama de caixa da área sob a COD para a melhor combinação de topologia e especialização identificada para cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).

O diagrama de diferença crítica é apresentado na Figura 4.18, indicando que a técnica kNN possui melhor *rank* médio, porém é estatisticamente similar às s-kNN e KPCA. A Tabela 4.20 sintetiza os valores de *rank* médio, onde $DC = 2,85$.

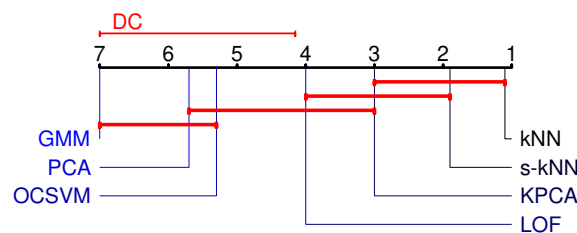


Figura 4.18: Diagrama de diferença crítica obtido na comparação de *ranks* médios das medidas AUC considerando cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).

A seguir, na Figura 4.19, são exibidas CODs médias para cada técnica, constituídas pela média das 10 curvas obtidas com os conjuntos de teste. Constata-se que o s-kNN apresenta um desempenho tão bom quanto o kNN tradicional e superior ao das demais técnicas, com o benefício de utilizar menos amostras, logo com uma redução do custo computacional da ordem de 50%, uma vez que é necessário

um menor quantitativo de cálculos de distâncias Euclidianas no sistema detector. Em contrapartida, o GMM apresenta o pior desempenho dentre as técnicas.

Tabela 4.20: *Rank* médio para a melhor topologia e especialização de cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).

PCA	KPCA	OCSVM	GMM	kNN	s-kNN	LOF
5,7	3	5,3	7	1,1	1,9	4

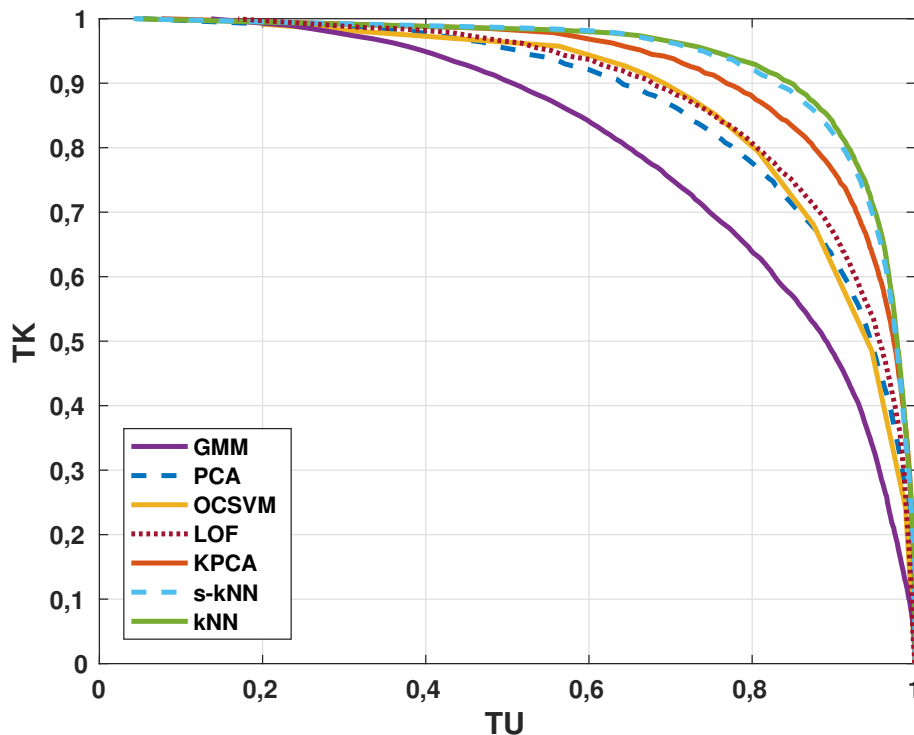


Figura 4.19: Considerando uma ordem crescente de AUC: $GMM < PCA < OCSVM < LOF < KPCA < s-kNN < kNN$.

A Tabela 4.21 exhibe alguns pares de TU e TK extraídos da Figura 4.19, que sintetizam pares de valores de TU e TK para os diferentes detectores.

4.5 Classificação de Classes de Contatos

Ao se operar com as topologias UN e CO, a classificação dos contatos é realizada de forma idêntica, com base nos escores de dissimilaridade computados através das técnicas de detecção de novidades. Assim, a Tabela 4.22 resume o valor médio para TA obtida em cada técnica e nível de especialização (EC ou EN) associado. Destaca-se que a técnica KPCA apresenta um resultado insatisfatório, devido ao fato

Tabela 4.21: TU e TK para a melhor topologia e especialização de cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF).

	$[TU — TK]$	
PCA	[90,1% — 62,7%]	[95,1% — 47,8%]
KPCA	[90,0% — 76,2%]	[95,0% — 62,7%]
OCSVM	[87,5% — 68,1%]	[94,7% — 48,9%]
GMM	[90,0% — 48,0%]	[95,0% — 32,6%]
kNN	[90,1% — 83,4%]	[95,0% — 70,5%]
s-kNN	[90,2% — 81,8%]	[95,0% — 68,7%]
LOF	[90,0% — 66,6%]	[95,0% — 52,7%]

de os valores de γ serem diferentes entre os modelos. Para contornar essa situação, considerou-se a KPCA*, na qual foi fixado $\gamma = 1$ para todos os modelos.

Percebe-se que o maior nível de especialização eleva os valores de TA , evidenciando o fato de que os melhores resultados ficaram sob o domínio das técnicas baseadas em distância: kNN, s-kNN e LOF.

Tabela 4.22: TA médio para cada técnica (PCA, KPCA, OCSVM, GMM, kNN, s-kNN e LOF) de acordo com o nível de especialização empregado (EC ou EN) para as topologias UN/CO.

	PCA	KPCA	KPCA*	OCSVM	GMM	kNN	s-kNN	LOF
EC	58,7%	25,0%	76,9%	82,9%	89,0%	97,0%	95,6%	97,7%
EN	77,3%	33,7%	87,6%	90,1%	-	97,7%	97,6%	97,8%

A classificação na Topologia HI ocorre de forma independente das técnicas de detecção de novidades utilizadas nos geradores de escore. Assim, como no caso Tradicional, foi considerado apenas o classificador que trata o problema como multiclasse usual, o qual atinge $TA = 98,74\%$ para modelos EC, e $TA = 98,69\%$ para modelos EN.

4.6 Discussão

Uma vez que a classificação de contatos alcança os melhores resultados ao se utilizar a topologia HI e considerando que a EN-CO no kNN é estatisticamente similar à EN-HI, pode-se considerar a técnica kNN-EN-HI como a melhor, cuja COD é exibida na Figura 4.20. Desta curva, alguns pares de valores de TU e TK podem ser extraídos: [95% — 69,6%]; [89,8% — 82,7%]; [83,4% — 90,5%], representando diferentes compromissos de TU e TK .

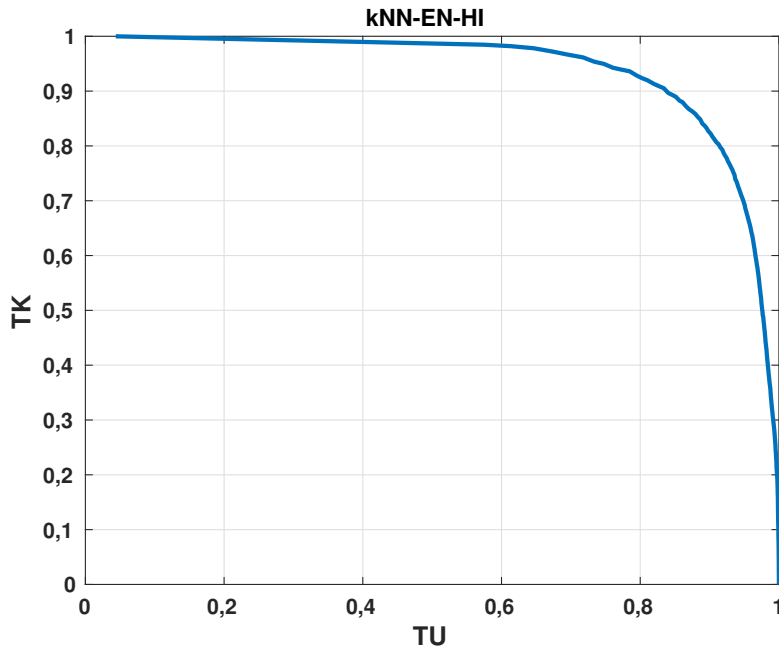


Figura 4.20: COD para a técnica kNN sob o enfoque EN-HI.

Ainda que uma comparação mais direta entre os resultados não seja possível, tendo em vista o uso de diferentes metodologias de avaliação, observa-se que a especialização em navios atingiu melhores resultados que os descritos na literatura para modelos especializados em classes, sintetizados na Tabela 4.23. Quanto à tarefa de classificação de contatos conhecidos em rótulos de classe, o sistema final alcançou uma acurácia média de acertos de 98,7%, também superior à descrita na literatura, conforme apresentado na Tabela 4.24.

Tabela 4.23: Resultados descritos na literatura quanto à detecção de novidades.

Sistema proposto	$\#\mathcal{K}_C$	$\#\mathcal{U}_C$	$[TU - TK]$
Componentes Principais [16]	4	4	[62,9% — 70,0%]
kNN [2]	4	4	[80,0% — 90,0%]
OCSVM [9]	3	1	[75,0% — 62,4%]

Tabela 4.24: Resultados descritos na literatura quanto à classificação de contatos conhecidos em rótulos de classe.

Sistema proposto	$\#\text{Classes}$	$\#\text{Navios}$	TA
CM-MLP [11]	8	34	84,4%
Redes Neurais convolucionais [12]	8	25	95,5%
Curvas Principais hierárquicas [14]	8	28	96,2%

Capítulo 5

Conclusões e Trabalhos Futuros

No âmbito militar, o sonar passivo é uma tecnologia primordial no sistema de defesa dos submarinos, uma vez que atua na identificação de ameaças durante sua operação. Através da captação e análise de ruídos provenientes do meio externo, na sua forma espectral, operadores de sonar treinados efetuam a complexa tarefa de tomada de decisão a respeito da natureza de um dado contato. Prover estes profissionais com sistemas automáticos de classificação pode reduzir o esforço físico e mental empregado na tarefa, além de agilizar o processo, o qual, naturalmente, exige decisões rápidas.

Usualmente, os sistemas supracitados são desenvolvidos a partir do conhecimento de um número restrito de classes de navios, permitindo que atuem na classificação de contatos. Ademais, destaca-se a necessidade de mecanismos para a identificação daquelas classes não utilizadas durante o seu desenvolvimento, isto é, detectar novidades. Este trabalho se baseou num banco de dados fornecido pelo IPqM, contemplando o ruído captado em 263 corridas de 28 navios pertencentes a 8 classes adquiridas em raia acústica. Destas classes, optou-se pela investigação de um cenário equilibrado com 4 classes tratadas como conhecidas, e outras 4 restantes como novidades.

Com o intuito de obter ganhos na eficiência quanto à detecção de novidades, o sistema de apoio à decisão se baseou na premissa “dividir para conquistar”, ao gerar modelos especializados em navios (EN), cujo desempenho foi comparado a modelos sem nenhuma especialização (GL) e com especialização por classes (EC), já investigados na literatura.

Os enfoques especializados exigiram um mecanismo integrador das saídas de cada modelo-especialista, consistindo na utilização de variadas topologias para a tarefa, elencadas a seguir: (i) na Unânime (UN), o contato é uma novidade, se esta decisão coincidir para todos os modelos; (ii) na Competitiva (CO), a decisão final sobre o contato é tomada com base no modelo de maior verossimilhança ao dado; (iii) pela Hierárquica (HI), é implantado um sistema prévio de classificação que levará

o contato sob avaliação à análise por apenas um dos modelos de identificação de novidades. Destaca-se que nestas três abordagens, (i) e (ii) classificam o contato quanto a seu rótulo de classe com base nos modelos de detecção de novidades. Em contrapartida, segundo a (iii), a classificação é independente do sistema detector.

Diferentes técnicas clássicas de Aprendizagem de Máquina foram empregadas para a construção dos modelos, cujos princípios se baseiam em: reconstrução (PCA e KPCA), descrição do domínio (OCSVM), probabilidade (GMM) e distância (kNN e LOF). Em especial, para o kNN foi proposta uma adaptação em seu algoritmo que introduz mecanismos de esparsidade (s-kNN), visando a reduzir o quantitativo de amostras utilizadas em sua operação.

A avaliação de desempenho dos modelos individuais e o consequente sistema completo, bem como a seleção dos hiperparâmetros, foi executada através da técnica de reamostragem Validação Cruzada por 10-*fold* aninhado. Cada modelo foi construído a partir de um conjunto de treinamento, e seus hiperparâmetros foram definidos pelo conjunto de validação. Na maioria dos casos, reintegraram-se os conjuntos de treinamento e validação em um único conjunto de desenvolvimento, avaliando-se o desempenho final sob o conjunto de teste. Os índices de desempenho considerados incluíram a Taxa Média de Detecção de Eventos Conhecidos, a Taxa Média de Detecção de Novidades, a Área sob a Curva de Operação do Detector (construída a partir das duas Taxas anteriores) e a Acurácia Média de Acertos.

A especialização em navios se mostrou mais eficaz na maioria dos cenários de avaliação, bem como a técnica de k -vizinhos mais próximos. O sistema alcançou uma taxa de detecção de novidades de 83,4%, associada ao reconhecimento de 90,5% dos sinais conhecidos. Quanto à tarefa de classificação das classes conhecidas, 98,7% são corretamente rotuladas. Tais resultados foram obtidos para a técnica kNN sob o enfoque EN, considerando a topologia HI.

Destaca-se ainda o desempenho da técnica s-kNN, que, utilizando cerca de 54,2% das amostras originais para a construção dos modelos do sistema detector, atingiu um resultado significativamente superior às outras técnicas, com exceção do kNN tradicional, porém estatisticamente equivalente a este.

Como trabalhos futuros, pretende-se explorar novas técnicas para a construção do sistema detector, bem como proceder à fusão de modelos utilizando técnicas variadas. Ademais, convém aprimorar o s-kNN de forma que este opere em batelada, e assim produza dicionários menos redundantes — logo, com um quantitativo menor de amostras. Outra adição possível é efetuar um pré-processamento nos dados via PCA ou técnica similar, visando a diminuir o número de componentes de cada sinal, o que reduziria o custo computacional associado à técnica kNN.

Referências Bibliográficas

- [1] MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- [2] ROCHA, G. G. M. *Detecção de Novidades em Sistemas de Sonar Passivo*. Dissertação (Mestrado em Engenharia Elétrica), CEFET/RJ, 2016.
- [3] TAX, D. M. J., DUIN, R. P. W. “Outlier detection using classifier instability”. In: Amin, A., Dori, D., Pudil, P., et al. (Eds.), *Advances in Pattern Recognition*, pp. 593–601, Berlin, Germany, 1998. Springer Berlin Heidelberg.
- [4] BREUNIG, M. M., KRIEGEL, H.-P., NG, R. T., et al. “LOF: identifying density-based local outliers”. In: *2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104, New York, NY, USA, 2000. ACM.
- [5] BREIMAN, L. “Random Forests”, *Machine Learning*, v. 45, n. 1, pp. 5–32, 10 2001. ISSN: 1573-0565. doi: 10.1023/A:1010933404324. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>.
- [6] URICK, R. J. *Principles of Underwater Sound*. New York, McGraw-Hill Book Company, 1983.
- [7] NOAA. “What is sonar?” Disponível em: <<https://oceanservice.noaa.gov/facts/sonar.html>>. Atualizado em: 25/06/2018. Acessado em: 20/04/2019.
- [8] YAKUBOVSKIY, A. “Bistatic sonar, explained”. Disponível em: <http://www.farsounder.com/files/Bistatic%20Sonar_explained.pdf>. Acessado em: 04/07/2019.
- [9] MOURA JUNIOR, N. N. *Detecção de Novidade em Sistemas de Sonar Passivo*. Tese (Doutorado em Engenharia Elétrica), Universidade Federal do Rio de Janeiro, 2018.

- [10] RODRIGUES, C. V. S., SOARES FILHO, W. “Classificação de sinais acústicos submarinos utilizando tons característicos”. In: *VII Simpósio de Pesquisa Operacional e Logística da Marinha*, pp. 103–111, Rio de Janeiro, RJ, Brasil, 2004.
- [11] SOUZA FILHO, J. B. O., SEIXAS, J. M. “Class-modular Multi-layer Perceptron Networks for Supporting Passive Sonar Signal Classification”, *IET Radar Sonar and Navigation*, v. 10, n. 2, pp. 311–317, 02 2015.
- [12] ROCHA, G. M. R., SOUZA FILHO, J. B. O. “Redes convolucionais para a classificação de sinais de sonar passivo”. In: *XI Encontro de Tecnologia em Acústica Submarina*, pp. 205–212, Rio de Janeiro, RJ, Brasil, 2014.
- [13] HASTIE, T., STUETZLE, W. “Principal Curves”, *Journal of the American Statistical Association*, v. 84, n. 406, pp. 502–516, 06 1989.
- [14] OLIVEIRA, T. C. C., SOUZA FILHO, J. B. O. “Otimização de classificadores de sinais de sonar passivo baseados em curvas principais hierárquicas”. In: *Simpósio Brasileiro de Automação Inteligente*, pp. 569–574, Porto Alegre, RS, Brasil, 2017.
- [15] FERNANDES, T. S. P., SOUZA FILHO, J. B. O., ALVES, E. S. “Identificação de navios em sinais de sonar passivo por curvas principais”. In: *XII Congresso Brasileiro de Inteligência Computacional*, pp. 1–6, Curitiba, PR, Brasil, 2015.
- [16] GUEDES, R. M., , SOUZA FILHO, J. B. O. “Identificação de classes desconhecidas em sinais de sonar passivo através de componentes principais”. In: *XX Congresso Brasileiro de Automática*, pp. 121–127, Belo Horizonte, MG, Brasil, 2014.
- [17] CHIH-WEI HSU, CHIH-JEN LIN. “A Comparison of Methods for Multiclass Support Vector Machines”, *IEEE Transactions on Neural Networks*, v. 13, n. 2, pp. 415–425, 03 2002. ISSN: 1045-9227. doi: 10.1109/72.991427.
- [18] TAX, D. M. *One-class Classification*. Tese (Doutorado em Física), Delft University of Technology, 2001.
- [19] M. MOYA, M. KOCH, L. H. “One-class classifier networks for target recognition applications”. In: *World Congress on Neural Networks, International Neural Network Society*, pp. 797–801, Portland, OR, USA, 1993.
- [20] PIMENTEL, M. A., CLIFTON, D. A., CLIFTON, L., et al. “A Review of Novelty Detection”, *Signal Processing*, v. 99, pp. 215–249, 06 2014.

- [21] WANG, W., ZHANG, M., WANG, D., et al. “Anomaly detection based on kernel principal component and principal component analysis”. In: *International Conference in Communications, Signal Processing, and Systems*, pp. 2222–2228, Venice, Italy, 2019. ISBN: 978-981-10-6570-5. doi: 10.1007/978-981-10-6571-2_271.
- [22] HOFFMANN, H. “Kernel PCA for novelty detection”, *Pattern Recognition*, v. 40, n. 3, pp. 863–874, 03 2007. doi: 10.1016/j.patcog.2006.07.009.
- [23] CHEN, X.-F., XING, H.-J., WANG, X.-Z. “A modified AdaBoost method for one-class SVM and its application to novelty detection”. In: *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3506–3511, Anchorage, AK, USA, 2011. doi: 10.1109/ICSMC.2011.6084212.
- [24] FLEXER, A., PAMPALK, E., WIDMER, G. “Novelty detection based on spectral similarity of songs”. In: *6th International Conference on Music Information Retrieval (ISMIR)*, pp. 260–263, London, UK, 2005.
- [25] KIM, T., CHO, N.-W. “Outlier detection approach based on local outlier factor for datasets with mixed attributes”. In: *11th International Conference on Innovative Computing, Information and Control*, v. 7, pp. 2155–2160, Harbin, China, 2016.
- [26] HAUTAMÄKI, V., KARKKAINEN, I., FRANTI, P. “Outlier detection using k-nearest neighbour graph”. In: *17th International Conference on Pattern Recognition (ICPR)*, v. 3, pp. 430–433, Cambridge, UK, 2004. ISBN: 0-7695-2128-2. doi: 10.1109/ICPR.2004.1334558.
- [27] OKFALISA, GAZALBA, I., MUSTAKIM, et al. “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification”. In: *2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 294–298, Yogyakarta, Indonesia, 2017. doi: 10.1109/ICITISEE.2017.8285514.
- [28] SANDAG, G. A., TEDRY, N. E., LOLONG, S. “Classification of lower back pain using k-nearest neighbor algorithm”. In: *6th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–5, Danau Toba, Indonesia, 2018.
- [29] VIDAL, R., MA, Y., SASTRY, S. S. *Generalized Principal Component Analysis*. Springer, 2016.
- [30] BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York, Springer, 2006.

- [31] SCHOLKOPF, B., WILLIAMSON, R., SMOLA, A., et al. “Support vector method for novelty detection”. In: *12th International Conference on Neural Information Processing Systems (NIPS-99)*, pp. 582–588, Cambridge, MA, USA, 1999. MIT Press.
- [32] SCHÖLKOPF, B., PLATT, J., SHAWE-TAYLOR, J., et al. “Estimating Support of a High-Dimensional Distribution”, *Neural Computation*, v. 13, pp. 1443–1471, 07 2001. doi: 10.1162/089976601750264965.
- [33] WU, Z.-L., LI, C.-H. “On improving sequential minimal optimization”. In: *2004 International Conference on Machine Learning and Cybernetics*, v. 7, pp. 4308–4313, 09 2004. ISBN: 0-7803-8403-2. doi: 10.1109/ICMLC.2004.1384594.
- [34] MURPHY, K. *Fitting a Conditional Linear Gaussian Distribution*. [S.l.], 1998. Disponível em: <<https://www.cs.ubc.ca/~murphyk/Papers/learnCG.pdf>>. Revisado em: 29/01/2003. Acessado em: 16/04/2019.
- [35] PATI, Y. C., REZAIIFAR, R., KRISHNAPRASAD, P. S. “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition”. In: *27th Annual Asilomar Conference on Signals, Systems, and Computers*, pp. 40–44, Pacific Grove, CA, USA, 1993.
- [36] LAY, D. C., LAY, S. R., MCDONALD, J. J. *Linear Algebra and Its Applications*. Pearson, 2015.
- [37] PENROSE, R. “A Generalized Inverse for Matrices”, *Mathematical Proceedings of the Cambridge Philosophical Society*, v. 51, n. 3, pp. 406—413, 1955. doi: 10.1017/S0305004100030401.
- [38] RUBINSTEIN, R., ZIBULEVSKY, M., ELAD, M. “Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit”, *Computer Science Dept., Technical Report CS-2008-08*, 01 2008.
- [39] DUAN, L., XIE, M., BAI, T., et al. “A New Support Vector Data Description Method for Machinery Fault Diagnosis with Unbalanced Datasets”, *Expert Systems with Applications*, v. 64, pp. 239–246, 12 2016. doi: 10.1016/j.eswa.2016.07.039.
- [40] BAI, M., WANG, X., XIN, J., et al. “An Efficient Algorithm for Distributed Density-based Outlier Detection on Big Data”, *Neurocomputing*, v. 181, pp. 19–28, 03 2016. doi: 10.1016/j.neucom.2015.05.135.

- [41] LIU, B., XIAO, Y., CAO, L., et al. “SVDD-based Outlier Detection on Uncertain Data”, *Knowledge and Information Systems*, v. 34, n. 3, 03 2012. doi: 10.1007/s10115-012-0484-y.
- [42] CABRAL, G., OLIVEIRA, A. “One-Class Classification Based on Searching for the Problem Features Limits”, *Expert Systems with Applications*, v. 41, n. 16, pp. 7182–7199, 11 2014. doi: 10.1016/j.eswa.2014.05.037.
- [43] SOUZA FILHO, J. B. O., SEIXAS, J. M. D. “Classificação neural classe escalável de sinais de sonar passivo”. In: *XIX Congresso Brasileiro de Automática*, v. 1, pp. 1–6, Campina Grande, PB, Brasil, 2012.
- [44] OLIVEIRA, T. C. C., SOUZA FILHO, J. B. O. “Uso de curvas principais hierárquicas na classificação de sinais de sonar passivo”. In: *XXI Congresso Brasileiro de Automática*, pp. 2796–2801, Vitória, ES, Brasil, 2016.
- [45] ROCHA, G. G. M., SOUZA FILHO, J. B. O. “Sistemas de classificação de contatos de sonar passivo com rejeição à classes desconhecidas”. In: *XXI Congresso Brasileiro de Automática*, pp. 467–472, Vitória, ES, Brasil, 2016.
- [46] SOUZA FILHO, J. B. O., GUEDES, R. M., GESUALDI, A. R. “Classificadores multi-classe de contatos por curvas principais”. In: *Anais do X Simpósio Brasileiro de Automação Inteligente*, pp. 1013–1018, São João del-Rei, MG, Brasil, 2011.
- [47] JAMES, G., WITTEN, D., HASTIE, T., et al. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.
- [48] KUHN, M., JOHNSON, K. *Applied Predictive Modeling*. Springer, 2013.
- [49] JAPKOWICZ, N., SHAH, M. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 2011.
- [50] DEMSAR, J. “Statistical Comparisons of Classifiers over Multiple Data Sets”, *Journal of Machine Learning Research*, v. 7, pp. 1–30, 01 2006.

Apêndice A

Produção Científica

Este apêndice lista os artigos científicos submetidos em Simpósios e Congressos Nacionais ao longo do período de desenvolvimento desta Dissertação de Mestrado.

1. MUNIZ, V. H. S.; Souza Filho, J. B. O. Detecção Especializada de Novidades em Sonar Passivo. 14º Simpósio Brasileiro de Automação Inteligente, Ouro Preto, MG, Brasil, 2019.

Resumo: Em submarinos, possíveis ameaças (referidas como contatos) são identificadas e classificadas por operadores humanos através do sistema de sonar passivo. A automatização deste processo é de extrema relevância, porém demanda o desenvolvimento de um sistema capaz de detectar classes desconhecidas (novidades). Este trabalho propõe um comitê de modelos altamente especializados para a construção de tal sistema. Três níveis de especialização foram considerados: não-especializado e especializados por classes e navios, assumindo diferentes técnicas para a detecção de novidades: Análise de Componentes Principais, Modelos de Mistura de Gaussianas, e k -vizinhos mais próximos. Experimentos conduzidos com dados reais adquiridos em uma raia acústica mostraram um melhor desempenho dos modelos especializados em navios, que atingiram uma taxa de detecção de novidades de 86,1%, conjugada com uma taxa média de reconhecimento de casos conhecidos de 90,1%.

2. MUNIZ, V. H. S.; Honorato, E. S.; Souza Filho, J. B. O. Aprendizado por Instância para a Identificação de Classes Desconhecidas em Sonares Passivos. XIV Congresso Brasileiro de Inteligência Computacional, Belém, PA, Brasil, 2019.

Resumo: Em submarinos, a tarefa dos operadores de sonar consiste na identificação de possíveis ameaças (contatos), utilizando, principalmente, o sistema de sonar passivo. Sistemas de classificação automática de contatos requerem a

identificação de embarcações de classes desconhecidas durante a sua operação. Este trabalho apresenta um sistema hierárquico baseado em técnicas de aprendizado por instância para o reconhecimento de tais ocorrências. Os experimentos, explorando dados coletados em raia acústica de 28 navios pertencentes a 8 classes em diferentes condições operacionais, apontaram um melhor desempenho da técnica *k-Nearest Neighbors*.

3. MUNIZ, V. H. S.; Oliveira, M. B.; Souza Filho, J. B. O. Classificação Hierárquica de Navios em Sinais de Sonar Passivo. XIV Congresso Brasileiro de Inteligência Computacional, Belém, PA, Brasil, 2019.

Resumo: A classificação de possíveis ameaças a submarinos é realizada por operadores de sonar, que analisam as características tonais e espectrais dos ruídos adquiridos no ambiente aquático. Automatizar esta tarefa através da construção de sistemas automáticos de reconhecimento de classes e navios reduziria o desgaste físico e psíquico deste profissional. Este artigo discute a classificação de navios, propondo um modelo de classificação hierárquica. Duas técnicas são avaliadas para a construção deste classificador: *Random Forest* e *k-Nearest Neighbors*. Resultados envolvendo sinais reais adquiridos de 28 navios apontam um melhor desempenho da técnica *k-Nearest Neighbors*.