

5 Estudo de Casos

O principal objetivo desse estudo de casos é a avaliação do desempenho do modelo FReX_SVM em problemas envolvendo duas ou mais classes. São estudados quatro problemas, dois de classificação binária e dois de classificação em mais de duas classes. Os bancos de dados utilizados nos testes são Bupa Liver Disorders, Wisconsin Breast Cancer, classificação da subespécie da planta íris e classificação do tipo de vinho. As características dos bancos de dados acima mencionados encontram-se na Tabela 5. Esse estudo de casos também tem como objetivo a comparação do desempenho dos métodos de classificação em múltiplas classes. Para isso, os dois problemas de classificação em mais de duas classes citados acima são considerados. Assim, nos problemas de classificação da subespécie da planta íris e do tipo de vinho, além da comparação do desempenho dos métodos de classificação em múltiplas classes, também se avalia se algum desses métodos é mais apropriado para a extração de regras fuzzy.

Tabela 5 - Características dos Banco de Dados

Nome	nº de atributos	nº de classes	nº de exemplos
Bupa Liver Disorders	6	2	345
Winconsin Breast Cancer	9	2	683
Íris	4	3	150
Vinho	13	3	178

5.1. Extração de Regras para Classificação Binária

Nessa seção o desempenho do modelo FReX_SVM é avaliado em dois problemas de classificação binária: Bupa Liver Disorders e Wisconsin Breast Cancer.

Os kernels usados para o treinamento da SVM nos problemas acima são: linear e RBF com 4 valores de σ^2 (1, 5, 10 e 50). Os valores da constante de regularização C usados são 0,1, 1 e 10. Assim, são geradas 15 SVMs para cada um dos problemas.

No treinamento das SVMs para ambos os casos é empregado o software do Gavin C. Cawley³.

³ <http://theoval.sys.uea.ac.uk/%7Egccc/svm/toolbox>

5.1.1. Bupa Liver Disorders

O banco de dados Bupa Liver Disorders é bastante conhecido na literatura de classificação binária de padrões. Esse banco de dados consiste em informações médicas ligadas a doenças de fígado. Existem seis atributos de entrada e um atributo de saída que define duas classes: pacientes que tiveram doença de fígado (classe 2) e pacientes saudáveis (classe 1). O conjunto de dados contém 345 exemplos, 200 da classe 1 e 145 da classe 2. O banco de dados Bupa Liver Disorders está disponível no "UCI Machine Learning Repository" [54].

Os 345 exemplos foram divididos em 2 conjuntos disjuntos, um com 172 padrões, contendo 102 padrões da classe 1 e 70 padrões da classe 2, e o outro com 173 padrões, contendo 98 exemplos da classe 1 e 75 da classe 2. Foram feitos dois treinamentos, um com o primeiro conjunto de dados como conjunto de treinamento e o segundo como conjunto de teste, e outro usando o segundo conjunto como conjunto de treinamento e o primeiro como conjunto de teste. Isso permite que os dois conjuntos de dados sejam empregados tanto para o treinamento quanto para o teste.

Para o treinamento da SVM é necessário que todos os atributos de entrada tenham o mesmo peso, o que ocorre no caso desse banco de dados. Assim, não foi necessária a normalização dos dados, os dados originais foram utilizados.

5.1.1.1. Apresentação dos Resultados

Os resultados obtidos encontram-se na Tabela 6, onde estão listadas as taxas de pontos incorretamente classificados no treinamento ("Err Tr") e no teste ("Err Te") e o número de vetores suporte ("n° sv").

O melhor desempenho nos pontos de teste, destacado em negrito na Tabela 6, foi obtido com o kernel linear com as constantes de regularização $C = 10$ e $C = 0,1$, classificando corretamente 61,45% padrões de teste (erro de 38,55%).

Tabela 6 - Resultados das SVMs - Bupa Liver Disorders

Kernel			n° sv	Err Tr	Err Te
Linear		$C = 0,1$	151,5	40,29%	38,55%
		$C = 1$	154,5	42,03%	42,03%
		$C = 10$	151	40,29%	38,55%
RBF	$\sigma^2 = 1$	$C = 0,1$	153,5	42,03%	42,03%
		$C = 1$	153,5	31,30%	42,03%
		$C = 10$	150	17,97%	38,84%
	$\sigma^2 = 5$	$C = 0,1$	151	42,03%	42,03%
		$C = 1$	150	38,84%	42,32%
		$C = 10$	149,5	33,91%	44,06%
	$\sigma^2 = 10$	$C = 0,1$	149	42,03%	42,03%
		$C = 1$	151	40,00%	42,03%
		$C = 10$	149	37,39%	42,32%
	$\sigma^2 = 50$	$C = 0,1$	150,5	42,03%	42,03%
		$C = 1$	149	42,03%	42,03%
		$C = 10$	150	40,00%	42,32%

O kernel linear apresentou a menor média das taxas de erros no teste, exposta na Tabela 7. A média das taxas de erro de um kernel é calculada somando-se as taxas de erros desse kernel para os três valores de C e dividindo-se essa soma por 3.

Os resultados obtidos na classificação do banco de dados Bupa Liver Disorders com SVM foram comparados com os resultados obtidos por outros métodos descritos em [55]. Os resultados são apresentados na Tabela 8: HNFB⁻¹ (Inverted Hierarchical Neuro-Fuzzy BSP), HNFB (Hierarchical Neuro-

Fuzzy BSP), MFS (multiple feature subsets), HNFQ (Hierarchical Neuro-Fuzzy Quadtree), FSS (nearest neighbor with forward sequential selection of feature), BSS (nearest neighbor with backward sequential selection of feature), kNN (k nearest neighbor) e NN (nearest neighbor).

Tabela 7 - Comparação do desempenho dos Kernels - Bupa Liver Disorders

Kernel	Média da taxa de erro
Linear	39,71%
RBF com $\sigma^2 = 1$	40,97%
RBF com $\sigma^2 = 5$	42,80%
RBF com $\sigma^2 = 10$	42,13%
RBF com $\sigma^2 = 50$	42,13%

Como pode ser observado na Tabela 6, a classificação desse banco de dados com SVM apresenta uma taxa elevada de erro, tanto no treinamento quanto no teste. No entanto, como pode ser visto na Tabela 8, o Bupa Liver Disorders é um banco de dados de difícil classificação.

Tabela 8 - Comparação do desempenho da Classificação - Bupa Liver Disorders

SVM	61,45%
HNFB ⁻¹	73,33%
HNFQ	75,07%
HNFB	74,49%
MFS2	64,4%
MFS1	65,4%
BSS	60,0%
FSS	56,8%
kNN	61,3%
NN	60,4%

5.1.1.2. Extração de Regras Fuzzy

Nesta seção, avalia-se o desempenho do método de extração de regras para classificação do Bupa Liver Disorders com SVM, usando-se:

1. número de regras geradas (“nr”);
2. percentagem do conjunto de teste coberto pelas regras geradas (“cob”).

Para a extração de regras foram utilizados os resultados de classificação obtidos na seção anterior.

Para se avaliar o desempenho de FREx_SVM, variou-se o número de conjuntos fuzzy definidos para cada coordenada. Três configurações foram utilizadas, com 3, 5 e 7 conjuntos fuzzy, expostos nas Figuras 18, 19 e 20, respectivamente. No banco de dados Bupa Liver Disorders todos os atributos de entrada estão entre 0 e 1. Assim, esse é o domínio de todos os conjuntos fuzzy definidos.

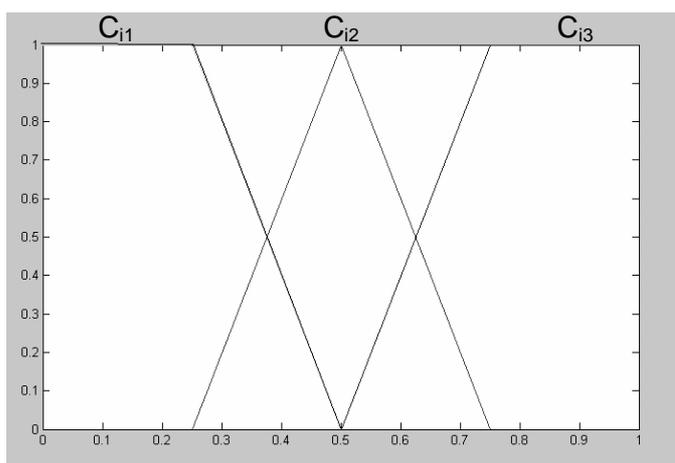


Figura 18 - 3 Conjuntos Fuzzy – Bupa Liver Disorders

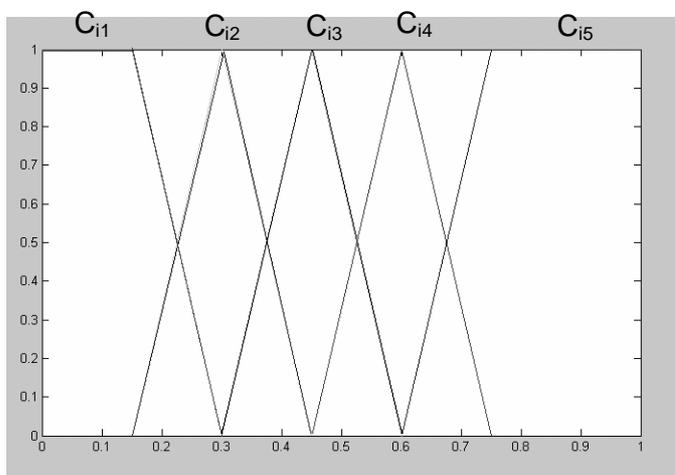


Figura 19 - 5 Conjuntos Fuzzy – Bupa Liver Disorders

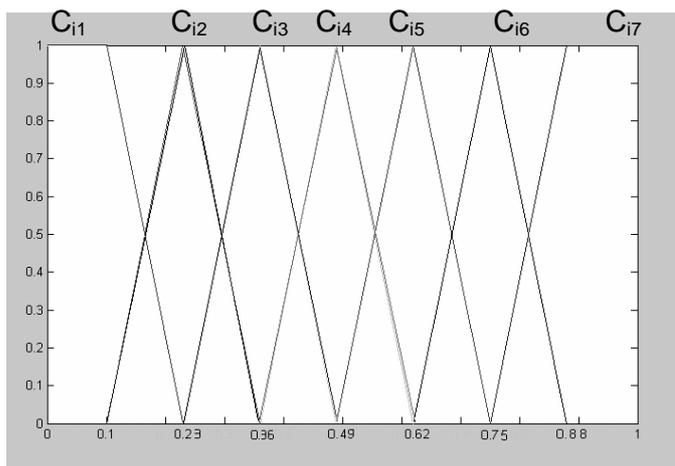


Figura 20 - 7 Conjuntos Fuzzy – Bupa Liver Disorders

O desempenho de cada configuração encontra-se exposto na Tabela 9, separado de acordo com o número de conjuntos fuzzy. Nesta Tabela são mostrados o número médio de regras (média entre o número de regra geradas em cada treinamento) e a cobertura média dos pontos de teste (média entre a cobertura das regras em cada treinamento).

A Tabela 10 mostra o melhor desempenho de classificação nos pontos de teste, em termos de cobertura e taxa de acerto, para 3, 5 e 7 conjuntos fuzzy. O melhor desempenho da SVM é mostrado na mesma Tabela. Nessa Tabela, é mostrado, ainda, o kernel para o qual o FReX_SVM ou a SVM obteve melhor desempenho, as taxas de acerto e erro. Além disso, a taxa de pontos não classificados pelo FReX_SVM aparece na Tabela como “taxa de pontos nc”. O melhor desempenho obtido foi a classificação correta de 48,69% dos padrões de teste e erro de 47,25% com a utilização de 3 conjuntos com kernel RBF com $\sigma^2 = 50$ e $C = 0,1$.

As configurações com 3 conjuntos tiveram melhor desempenho do que as outras, com menos regras e maior cobertura. Assim, essa configuração se mostrou mais adequada ao problema da classificação do Bupa Liver Disorders.

Comparando-se o desempenho da classificação com FReX_SVM com o da classificação com SVM, vê-se que o primeiro é inferior ao segundo (Tabela 10). No entanto, o objetivo do modelo proposto é a extração de conhecimento de uma SVM treinada e não o aperfeiçoamento do desempenho da SVM. Assim, o mais importante é que o FReX_SVM gera um modelo interpretável (regras lingüísticas), ao contrário da SVM que gera um “modelo caixa preta”.

Tabela 9 - Desempenho do FReX_SVM - Bupa Liver Disorders

Kernel		3 conjuntos		5 conjuntos		7 conjuntos		
		fuzzy		fuzzy		fuzzy		
		nr	cob	nr	cob	nr	cob	
Linear	C = 0,1	27,5	95,94%	75,5	85,22%	107,5	79,71%	
	C = 1	28	95,94%	77	85,22%	110	79,71%	
	C = 10	28	95,94%	75,5	85,22%	107,5	79,42%	
RBF	$\sigma^2 = 1$	C = 0,1	27,5	95,94%	76	84,93%	108,5	80,58%
		C = 1	28	95,94%	77	85,22%	108,5	80,29%
		C = 10	25	95,65%	72,5	84,93%	103	80,00%
	$\sigma^2 = 5$	C = 0,1	27,5	95,94%	74,5	84,93%	106	79,71%
		C = 1	27	95,94%	75,5	85,22%	106,5	79,13%
		C = 10	27	95,94%	76	85,22%	108,5	79,71%
	$\sigma^2 = 10$	C = 0,1	27	95,94%	74	85,51%	106,5	80,00%
		C = 1	27	95,94%	76,5	85,51%	107,5	79,71%
		C = 10	27,5	95,94%	76	85,22%	105,5	79,13%
	$\sigma^2 = 50$	C = 0,1	28	95,94%	76,5	85,22%	108,5	80,58%
		C = 1	27,5	95,94%	75	85,22%	107	80,00%
		C = 10	27,5	95,94%	76,5	85,22%	108,5	79,42%

Tabela 10 - Melhor Desempenho do FReX_SVM e da SVM – Bupa Liver Disorders

	3 conjuntos fuzzy	5 conjuntos fuzzy	7 conjuntos fuzzy	SVM
Kernel	RBF com $\sigma^2 = 50$ e C = 0,1	RBF com $\sigma^2 = 10$ e C = 0,1	RBF com $\sigma^2 = 50$ e C = 0,1	Linear C = 0,1
Taxa de acerto	48,69%	44,64%	44,06%	61,45%
Taxa de erro	47,25%	40,87%	36,52%	38,55%
Taxa de pontos nc	04,06%	14,49%	19,42%	---

As melhores regras, segundo a acurácia e a abrangência fuzzy nos pontos de teste, para cada classe e configuração, são listadas a seguir:

a) Três conjuntos fuzzy

Se x_1 é C_{12} e x_2 é C_{22} e x_3 é C_{31} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} , então Classe 1.

[acurácia: 0,8156/ abrangência: 0,0113]

Se x_1 é C_{12} e x_2 é C_{22} e x_3 é C_{32} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} , então Classe 2.
[acurácia: 0,7252/ abrangência: 0,0109]

b) Cinco conjuntos fuzzy

Se x_1 é C_{15} e x_2 é C_{23} e x_3 é C_{31} e x_4 é C_{42} e x_5 é C_{51} e x_6 é C_{61} , então Classe 1.
[acurácia: 0,7394/ abrangência: 0,0140]

Se x_1 é C_{14} e x_2 é C_{21} e x_3 é C_{31} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} , então Classe 2.
[acurácia: 0,7673/ abrangência: 0,0385]

c) Sete conjuntos fuzzy

Se x_1 é C_{16} e x_2 é C_{23} e x_3 é C_{31} e x_4 é C_{42} e x_5 é C_{51} e x_6 é C_{61} , então Classe 1.
[acurácia: 0,7686/ abrangência: 0,0205]

Se x_1 é C_{15} e x_2 é C_{24} e x_3 é C_{31} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} , então Classe 2.
[acurácia: 0,7638/ abrangência: 0,0240]

Como se pode verificar, a acurácia das regras acima é bem alta (maior do que 0,7). Isso mostra que as regras geradas representam bem o conjunto de teste. Por outro lado, a abrangência das regras apresenta-se baixa, o que é de se esperar, uma vez que a abrangência fuzzy é a soma dos graus de pertinência dos padrões que satisfazem a regra dividida pelo número total de padrões.

5.1.2. Wisconsin Breast Cancer

O banco de dados Wisconsin Breast Cancer também é conhecido na literatura de classificação binária de padrões. Ele, da mesma forma que o Bupa Liver Disorders, consiste em informações médicas ligadas, nesse caso, ao câncer de mama. Há nove atributos de entrada e um atributo de saída que define duas classes: benigno (classe 1) e maligno (classe 2). Os atributos de entrada são (em inglês): *clump thickness*, *uniformity of cell size*, *uniformity of cell shape*, *marginal adhesion*, *single epithelial cell size*, *bare nuclei*, *bland chromatin*, *nucleoli normal* e *mitoses*. O conjunto de dados contém 683 exemplos, 444 da classe 1 e 239 da classe 2. O banco de dados Wisconsin Breast Cancer também está disponível no "UCI Machine Learning Repository" [54].

Os 683 padrões de entrada foram divididos em 2 conjuntos disjuntos, um com 342 exemplos, contendo 222 padrões da classe 1 e 120 padrões da classe 2, e o outro com 341 exemplos, contendo 222 padrões da classe 1 e 119 da classe 2. Da mesma maneira que no caso anterior, foram feitos dois treinamentos, um com o primeiro conjunto de dados como conjunto de treinamento e o segundo como conjunto de teste, e outro usando o segundo conjunto como conjunto de treinamento e o primeiro como conjunto de teste.

Como mencionado, para o treinamento da SVM é necessário que todos os atributos de entrada tenham o mesmo peso. Ao contrário do caso anterior em que isso já ocorre, o conjunto de treinamento tem que ser normalizado para todos os atributos estarem entre 0 e 1 e nenhum ter maior influência.

5.1.2.1. Apresentação dos Resultados

Os resultados da classificação de Wisconsin Breast Cancer com SVM estão expostos na Tabela 11 com as mesmas taxas da seção anterior.

Tabela 11 - Resultados das SVMs - Wisconsin Breast Cancer

Kernel		n ^o sv	Err Tr	Err Te	
Linear	C = 0,1	30,05	03,07%	02,78%	
	C = 1	29,00	02,63%	03,21%	
	C = 10	26,50	02,19%	01,27%	
RBF	$\sigma^2 = 1$	C = 0,1	179,50	04,84%	05,56%
		C = 1	141,00	01,17%	04,68%
		C = 10	136,00	---	04,54%
	$\sigma^2 = 5$	C = 0,1	96,50	03,37%	03,37%
		C = 1	59,50	02,34%	03,37%
		C = 10	48,00	01,32%	03,66%
	$\sigma^2 = 10$	C = 0,1	95,00	03,07%	03,07%
		C = 1	44,00	02,63%	02,78%
		C = 10	42,00	01,76%	03,07%
	$\sigma^2 = 50$	C = 0,1	156,50	03,66%	03,96%
		C = 1	57,00	02,93%	02,93%
		C = 10	38,00	03,08%	02,93%

A classificação com SVM apresentou excelentes desempenhos para todos os kernels e valores de C testados. O kernel linear com constante de regularização C = 10 proporcionou o melhor desempenho nos pontos de teste com taxa de erro de 1,27%.

Como no caso do Bupa Liver Disorders, o kernel linear apresentou a menor média de taxa de erro nos pontos de teste, calculadas de forma análoga ao caso anterior. A média de taxa de erro de todos os kernels está exposta na Tabela 12.

Tabela 12 - Comparação do desempenho dos Kernels - Wisconsin Breast Cancer

Kernel	Média da taxa de erro
Linear	02,42%
RBF com $\sigma^2 = 1$	04,93%
RBF com $\sigma^2 = 5$	03,47%
RBF com $\sigma^2 = 10$	02,97%
RBF com $\sigma^2 = 50$	03,27%

5.1.2.2. Extração de Regras Fuzzy

Nesta seção, o desempenho do modelo FReX_SVM é avaliado para classificação do Wisconsin Breast Cancer, usando-se, como no caso anterior, o número de regras geradas (“nr”) e a percentagem dos pontos de teste cobertos pelas regras geradas (“cob”).

Os vetores suporte obtidos pelas SVMs da seção anterior são empregados para a extração de regras fuzzy.

Para se verificar a importância do número de conjuntos fuzzy construídos no modelo, duas configurações são utilizadas com 3 e 5 conjuntos fuzzy, expostos nas Figuras 21 e 22. Como o domínio dos atributos de entrada é [0,10], esse intervalo é usado como intervalo para definição dos conjuntos fuzzy.

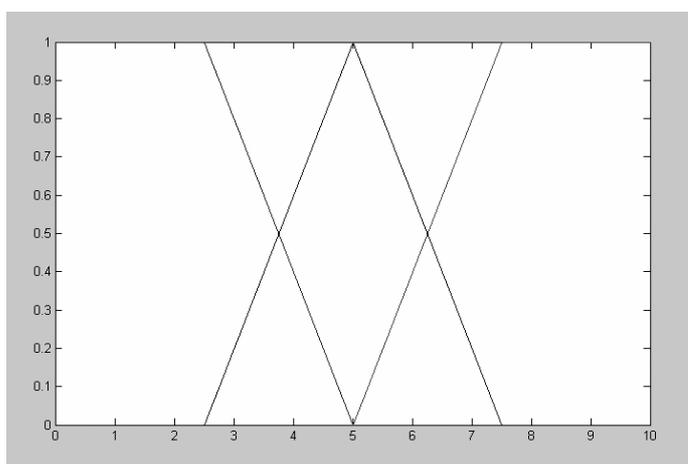


Figura 21 - 3 Conjuntos Fuzzy - Wisconsin Breast Cancer

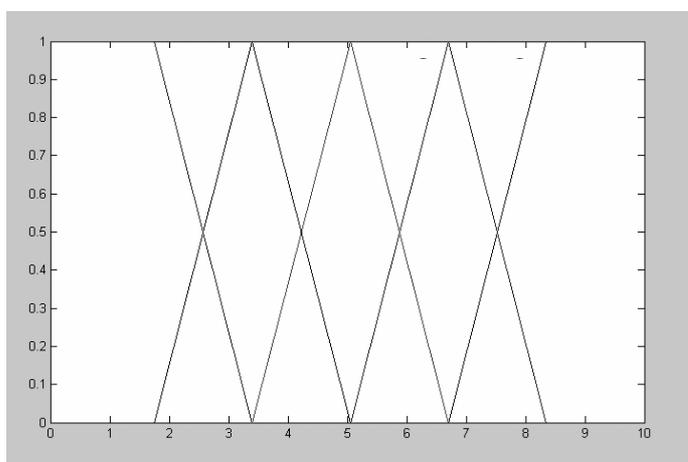


Figura 22 - 5 Conjuntos Fuzzy - Wisconsin Breast Cancer

O desempenho de ambas configurações está exposto na Tabela 13, de acordo com o número de conjuntos fuzzy. Nesta Tabela mostra-se o número médio de regras (média entre os dois treinamentos) e a cobertura média dos pontos de teste.

A Tabela 14 mostra o melhor desempenho de classificação nos pontos de teste, em termos de cobertura e taxa de acerto, para 3 e 5 conjuntos fuzzy. Além disso, também é mostrado na Tabela, o melhor desempenho da SVM. Na Tabela, como no caso do exemplo anterior, é mostrado o kernel para o qual o FReX_SVM ou a SVM obteve melhor desempenho, a taxa de acerto, de erro e de pontos não classificados, que aparece na Tabela como “taxa de pontos nc”. A maior cobertura, 77,90%, foi obtida com a utilização de 3 conjuntos fuzzy para a SVM com kernel RBF, $\sigma^2 = 1$ e $C = 0,1$, em que se obteve a classificação correta de 75,41% dos padrões de teste e erro de 02,49%.

As configurações com 3 conjuntos fuzzy proporcionaram uma maior cobertura, mostrando que essa configuração é mais adequada ao problema da classificação do Wisconsin Breast Cancer.

Tabela 13 - Desempenho do FReX_SVM - Wisconsin Breast Cancer

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		
		nr	cob	nr	cob	
Linear	C = 0,1	31	39,10%	34,50	11,27%	
	C = 1	27	38,23%	28,50	09,36%	
	C = 10	24,5	38,81%	26	08,78%	
RBF	$\sigma^2 = 1$	C = 0,1	131,5	77,90%	157	63,98%
		C = 1	127,5	77,02%	137,5	50,50%
		C = 10	74	76,58%	133	50,06%
	$\sigma^2 = 5$	C = 0,1	75,5	70,57%	92	37,61%
		C = 1	56	44,52%	58,5	10,97%
		C = 10	46	50,98%	47	10,83%
	$\sigma^2 = 10$	C = 0,1	71	70,72%	87,5	35,58%
		C = 1	41	32,33%	43,5	11,41%
		C = 10	37	32,03%	42,5	09,07%
	$\sigma^2 = 50$	C = 0,1	99,5	72,04%	123	48,91%
		C = 1	51	67,02%	56	25,31%
		C = 10	32	40,28%	32	09,07%

Tabela 14 - Melhor Desempenho do FReX_SVM e da SVM – Wisconsin Breast Cancer

	3 conjuntos fuzzy	5 conjuntos fuzzy	SVM
Kernel	RBF com $\sigma^2 = 1$ e C = 0,1	RBF com $\sigma^2 = 1$ e C = 0,1	Linear C = 10
Taxa de acerto	75,41%	63,25%	98,73%
Taxa de erro	02,49%	00,73%	01,27%
Taxa de pontos nc	22,10%	36,02%	---

As melhores regras, segundo a acurácia e a abrangência fuzzy nos pontos de teste, para cada classe e configuração, são listadas a seguir:

a) Três conjuntos fuzzy

Se x_1 é C_{11} e x_2 é C_{21} e x_3 é C_{31} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} e x_7 é C_{71} e x_8 é C_{81} e x_9 é C_{91} , então Classe 1. [acurácia: 1,0000/ abrangência: 0,3618]

Se x_1 é C_{12} e x_2 é C_{23} e x_3 é C_{33} e x_4 é C_{43} e x_5 é C_{52} e x_6 é C_{63} e x_7 é C_{73} e x_8 é C_{83} e x_9 é C_{91} , então Classe 2. [acurácia: 1,0000/ abrangência: 0,0079]

b) Cinco conjuntos fuzzy

Se x_1 é C_{11} e x_2 é C_{21} e x_3 é C_{31} e x_4 é C_{41} e x_5 é C_{51} e x_6 é C_{61} e x_7 é C_{71} e x_8 é C_{81} e x_9 é C_{91} , então Classe 1. [acurácia: 1,0000/ abrangência: 0,1371]

Se x_1 é C_{15} e x_2 é C_{25} e x_3 é C_{35} e x_4 é C_{45} e x_5 é C_{55} e x_6 é C_{65} e x_7 é C_{75} e x_8 é C_{85} e x_9 é C_{91} , então Classe 2. [acurácia: 1,0000/ abrangência: 0,0030]

Como pode ser observado também nesse caso, a acurácia das regras é muito alta, igual a 1 em todos os casos expostos, mostrando que as regras fornecem uma boa representação para o banco de dados. Para ambas configurações, a abrangência fuzzy das regras da classe 1 é boa, mas a da classe 2 é baixa.

5.1.3. Discussão dos Resultados

Os resultados dos testes feitos nos problemas de classificação do Bupa Liver Disorders e do Wisconsin Breast Cancer mostram que é possível construir um conjunto de regras que classifique os dados de teste de maneira confiável. A cobertura das regras se mostrou muito boa. Para o problema da classificação do Bupa Liver Disorders, a configuração com 3 conjuntos fuzzy obteve uma excelente cobertura de 95,94% dos padrões de teste; a configuração com 5 conjuntos cobriu 85,51% dos padrões de teste e com sete conjuntos fuzzy a cobertura é de 80,58% dos padrões de teste. Para o problema da classificação do Wisconsin Breast Cancer, a configuração com 3 conjuntos fuzzy obteve uma cobertura de 77,90% dos padrões de teste e a configuração com 5 conjuntos cobriu 63,98% dos padrões de teste. Além disso, as regras obtidas são fáceis de se interpretar com a utilização de conjuntos fuzzy nos antecedentes, ao contrário da obtida pelo método SVM + Prototypes, apresentado no capítulo 3.

5.2. Extração de Regras para Classificação em Múltiplas Classes

Nesta seção, o desempenho dos métodos de classificação em múltiplas classes, discutidos na seção 2.3, é avaliado em dois problemas: classificação da subespécie da planta íris e do tipo de vinho.

Esse estudo teve dois objetivos principais: analisar o desempenho dos métodos de classificação em múltiplas classes em diferentes bancos de dados e avaliar a adequação dos métodos em termos de capacidade de extração de regras e de desempenho.

Para os dois problemas foram utilizados os mesmos kernels e parâmetros empregados no caso de classificação binária, ou seja, kernels linear e RBF e constantes de regularização C iguais a 0,1, 1 e 10.

Como já mencionado no capítulo 2, SVMs foram originalmente construídas para a classificação binária. Há duas maneiras básicas de se lidar com um problema de classificação em mais de duas classes: reduzir esse problema a um conjunto de problemas binários (métodos de decomposição um por classe e de separação das classes duas a duas) ou estender a definição de SVM binária (método de Crammer e Singer).

O treinamento das SVMs, usando o método de decomposição um por classe, foi feito com a utilização do software de Gavin C. Cawley. O software LIBSVM⁴ foi usado para o treinamento do método de separação das classes duas a duas com decisão pela estratégia do voto. A implementação do método de Crammer e Singer foi feita com a utilização do software BSVM⁵.

Em um vetor de entrada $x \in \mathbb{R}^n$, cada uma das n componentes representa um atributo. Para que todos os atributos tenham o mesmo peso, deve-se realizar previamente a normalização de cada um deles, assegurando que todos estejam entre 0 e 1. Desse modo, nenhum atributo terá maior influência na resolução do problema.

⁴ <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

⁵ <http://www.csie.ntu.edu.tw/~cjlin/bsvm>

5.2.1. Classificação das Subespécies da planta Íris

Esse banco de dados é, provavelmente, o mais conhecido na literatura de reconhecimento de padrões. A espécie de planta íris é dividida em três subespécies: Íris Setosa (classe 1), Íris Versicolor (classe 2) e Íris Virginica (classe 3). O conjunto de dados contém 150 exemplos, 50 de cada uma das três classes. O banco de dados usado para a classificação da subespécie da planta íris foi obtido do "UCI Machine Learning Repository" [54].

Para a classificação são usados quatro atributos numéricos: o comprimento e a largura da sépala em centímetros e o comprimento e a largura da pétala também em centímetros. Assim, a entrada do sistema é uma matriz de 150 linhas (um padrão por linha) com 4 colunas.

Os 150 exemplos disponíveis foram divididos em 2 conjuntos disjuntos com 75 padrões, 25 de cada classe, em cada conjunto. Foram feitos dois treinamentos, um usando o primeiro conjunto como conjunto de treinamento e o segundo como conjunto de teste, e outro usando o segundo conjunto como conjunto de treinamento e o primeiro como conjunto de teste.

Como o problema de classificação do tipo de íris aqui abordado é um problema de classificação em 3 classes, devem-se usar os métodos para classificação em múltiplas classes descritos na seção 2.3: decomposição um por classe, separação das classes duas a duas e método de Crammer e Singer.

Como anteriormente mencionado, foram usados os kernels linear e RBF (com valores de σ^2 iguais a 1, 5, 10 e 50) e 3 valores da constante de regularização C (0,1, 1 e 10). Logo, foram geradas 15 SVMs para cada um dos três métodos para cada treinamento.

5.2.1.1. Apresentação dos Resultados

A média dos resultados obtidos nos pontos de teste em cada treinamento encontram-se nas Tabelas 15, 16 e 17, respectivamente, para os métodos de decomposição um por classe, separação das classes duas a duas com decisão por voto e método de Crammer e Singer. Na Tabela 15, encontram-se listados o número de vetores suporte (“nº sv”), as taxas de pontos incorretamente classificados no treinamento (“Err Tr”) e no teste (“Err Te”) e de pontos não classificados tanto para o treinamento (Nc Tr) quanto para o teste (Nc Te). Nas Tabelas 16 e 17, em virtude de todos os pontos terem sido classificados, estão listados apenas o número de vetores suporte e as taxas de erro no treinamento e no teste, com a mesma notação da Tabela anterior.

Tabela 15 - Resultado do método de decomposição um por classe - Íris

Kernel		nº sv	Err Tr	Nc Tr	Err Te	Nc Te	
Linear	C = 0,1	55	02,67%	26,67%	03,33%	28,67%	
	C = 1	48,5	03,33%	18,67%	03,33%	19,33%	
	C = 10	46,5	04,00%	14,00%	05,33%	16,67%	
RBF	$\sigma^2 = 1$	C = 0,1	72,5	00,67%	34,67%	--	39,33%
		C = 1	42	02,00%	--	03,33%	01,33%
		C = 10	31,5	00,67%	--	05,33%	00,67%
	$\sigma^2 = 5$	C = 0,1	73	--	52,00%	--	55,33%
		C = 1	44	04,00%		05,33%	--
		C = 10	24,5	01,33%	00,67%	02,67%	01,33%
	$\sigma^2 = 10$	C = 0,1	75	--	66,67%	--	65,33%
		C = 1	51,5	06,00%	03,33%	06,67%	06,00%
		C = 10	28,5	01,33%	00,67%	02,00%	00,67%
	$\sigma^2 = 50$	C = 0,1	75	--	100%	--	100%
		C = 1	65,5	04,00%	35,53%	05,33%	33,33%
		C = 10	50	02,67%	10,67%	03,33%	12,67%

O melhor resultado obtido nos padrões de teste foi a classificação correta de 98% com o kernel RBF para largura $\sigma^2 = 10$ e constante de regularização $C = 10$ usando o método de Crammer e Singer.

O kernel RBF com $\sigma^2 = 1$ apresentou a menor média da taxa de erros nos pontos de teste, exposto na Tabela 18. A média da taxa de erro para um dado

kernel é a soma dos resultados das três configurações (uma para cada valor de C) para os três métodos dividida por nove.

Tabela 16 - Resultado do método de separação das classes duas a duas - Íris

Kernel		nº sv	Err Tr	Err Te	
Linear	C = 0,1	47,5	03,33%	05,33%	
	C = 1	19,5	02,00%	02,67%	
	C = 10	12	01,33%	04,00%	
RBF	$\sigma^2 = 1$	C = 0,1	70	05,33%	12,00%
		C = 1	38	02,00%	04,00%
		C = 10	30,5	00,67%	04,00%
	$\sigma^2 = 5$	C = 0,1	72,5	11,33%	12,67%
		C = 1	38,5	02,67%	05,33%
		C = 10	23	02,00%	04,00%
	$\sigma^2 = 10$	C = 0,1	75	11,33%	12,67%
		C = 1	45,5	04,00%	06,00%
		C = 10	23	01,33%	02,67%
	$\sigma^2 = 50$	C = 0,1	70	56,00%	56,00%
		C = 1	65,5	10,67%	10,00%
		C = 10	34,5	01,33%	03,33%

Como já foi dito, a constante de regularização C controla o equilíbrio entre o número de erros e a margem do classificador, sendo que valores maiores de C tendem a permitir menos erros. Assim, como era de se esperar, as configurações com C = 10, o maior valor entre os utilizados, proporcionaram, de uma maneira geral, um melhor resultado.

A comparação do desempenho dos métodos de classificação em múltiplas classes está exposta na Tabela 19. O método de classificação com maior média de taxa de acertos nos pontos de teste foi o de separação das classes duas a duas com decisão por voto com 90,36% de acertos. A média da taxa de acertos é calculada somando-se todos os resultados das 15 SVMs treinadas para cada método dividido por 15.

Apesar de ser o método mais utilizado para a resolução de problemas de classificação em mais de duas classes, o método de decomposição um por classe apresentou o pior desempenho, com 71,55% de média da taxa de acertos.

Tabela 17 - Resultado do método de Crammer e Singer - Íris

Kernel		nº sv	Err Tr	Err Te	
Linear	C = 0,1	40,5	16,00%	18,00%	
	C = 1	30,5	14,67%	17,33%	
	C = 10	28	14,00%	17,33%	
RBF	$\sigma^2 = 1$	C = 0,1	60	04,00%	08,00%
		C = 1	33	02,00%	04,67%
		C = 10	25,5	00,67%	05,33%
	$\sigma^2 = 5$	C = 0,1	63,5	12,00%	14,00%
		C = 1	32	01,33%	03,33%
		C = 10	17	02,00%	03,33%
	$\sigma^2 = 10$	C = 0,1	65,5	28,67%	31,33%
		C = 1	39	04,00%	05,33%
		C = 10	19	01,33%	02,00%
	$\sigma^2 = 50$	C = 0,1	74,5	33,33%	33,33%
		C = 1	53,5	23,33%	27,33%
		C = 10	32	02,67%	03,33%

Tabela 18 - Comparação do desempenho dos Kernels - Íris

Kernel	Média da taxa de erro
Linear	08,52%
RBF com $\sigma^2 = 1$	05,19%
RBF com $\sigma^2 = 5$	05,63%
RBF com $\sigma^2 = 10$	07,63%
RBF com $\sigma^2 = 50$	15,78%

Tabela 19 - Comparação do desempenho dos métodos de classificação -Íris

Método de classificação	Média da taxa de acerto
Decomposição um por classe	71,55%
Separação das classes duas a duas com decisão por voto	90,36%
Crammer e Singer	87,07%

5.2.1.2. Extração de Regras Fuzzy

Nesta seção, avalia-se o desempenho do método de extração de regras para a classificação da subespécie da planta íris de maneira análoga ao caso binário, ou seja, considerando o número de regras geradas (“nr”) e a porcentagem do conjunto de teste coberto pelas regras geradas (“cob”). Como foram feitos dois treinamentos, calculou-se a média desses treinamentos para cada uma das medidas acima citadas.

Na extração de regras foram utilizadas as informações obtidas pelas SVMs da seção anterior. Para a avaliação do desempenho de FReX_SVM, construiu-se um número diferente de conjuntos fuzzy para cada coordenada. Três configurações foram utilizadas com 3, 5 e 7 conjuntos fuzzy, expostos nas Figuras 23, 24 e 25, respectivamente. Como pode ser observado nestas Figuras, o domínio dos conjuntos fuzzy construídos não está normalizado. Os dados de entrada utilizados foram obtidos diretamente do “UCI Machine Learning Repository” [54], já que é mais interessante as regras serem aplicadas diretamente ao conjunto de entrada original.

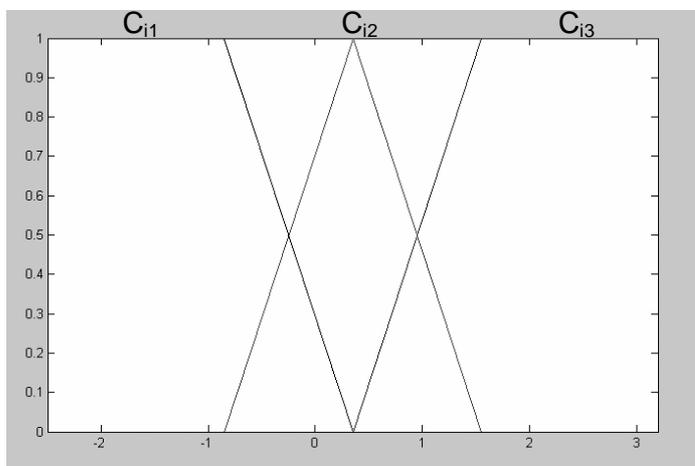


Figura 23 - 3 Conjuntos Fuzzy – Íris

Os desempenhos das três configurações encontram-se expostos abaixo, separados de acordo com os métodos de classificação em mais de duas classes. As Tabelas 20, 21 e 22 mostram os resultados para os métodos de decomposição um por classe, separação da classe duas a duas com decisão pela estratégia do voto e Crammer e Singer, respectivamente.

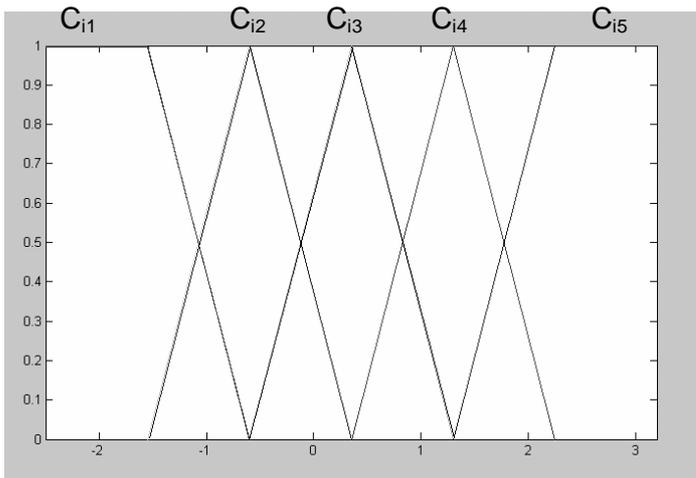


Figura 24 - 5 Conjuntos Fuzzy - Íris

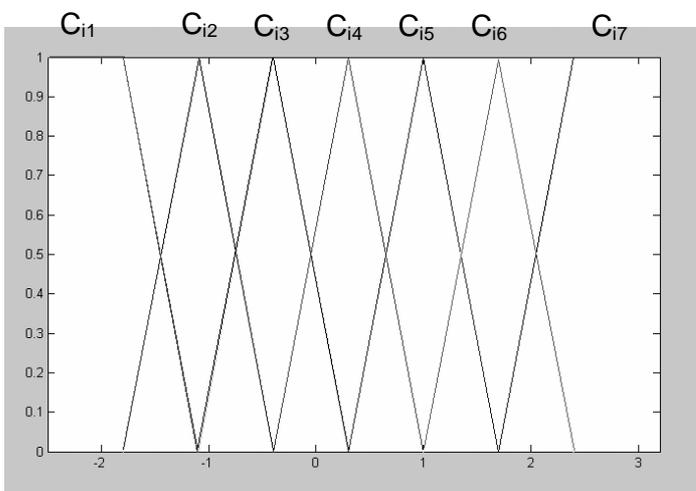


Figura 25 - 7 Conjuntos Fuzzy - Íris

A Tabela 23 mostra o melhor desempenho de classificação nos pontos de teste, em termos de cobertura e taxa de acerto, para 3, 5 e 7 conjuntos fuzzy. O melhor desempenho da SVM também é mostrado na Tabela. Na Tabela, são mostrados os kernels e os métodos de classificação em múltiplas classes para os quais o FREx_SVM ou a SVM obteve melhor desempenho, as taxas de acerto e erro. A taxa de pontos não classificados pelo FREx_SVM aparece na Tabela como “taxa de pontos nc”. No caso de várias configurações obterem as mesmas taxas de acertos e cobertura, foi considerado o melhor resultado a configuração que gerou menos regras. O melhor desempenho obtido pelo FREx_SVM foi a classificação correta de 94% dos padrões de teste e erro de 6% com a utilização de 5 conjuntos. Tanto a classificação com o método de decomposição um por classe, quanto a com o método de separação da classe duas a duas com decisão pela estratégia do voto proporcionaram esse resultado. No entanto, o desempenho do método de Crammer e Singer foi

apenas um pouco pior, com a classificação correta de 93,33% dos padrões de teste e erro de 6,67%.

Tabela 20 - Desempenho do FReX_SVM - Decomposição um por classe - Íris

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		7 conjuntos fuzzy		
		nr	cob	nr	cob	nr	cob	
Linear	C = 0,1	17,5	96,00%	27	94,00%	34,5	90,00%	
	C = 1	17,5	96,00%	25	94,00%	29,5	88,67%	
	C = 10	16,5	96,00%	23,5	92,67%	29	86,00%	
RBF	$\sigma^2 = 1$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	18,5	100%	28	100%	34	96,67%
		C = 10	17	100%	24,5	98,67%	27,5	94,00%
	$\sigma^2 = 5$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	17,5	98,00%	24,5	97,33%	29,5	95,33%
		C = 10	13,5	96,00%	18	91,33%	22	86,67%
	$\sigma^2 = 10$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	19,5	98,67%	28	98,67%	32,5	92,00%
		C = 10	12,5	95,33%	17,5	90,00%	23,5	86,67%
	$\sigma^2 = 50$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	20	99,33%	30,5	98,00%	38,5	96,00%
		C = 10	29	97,33%	26	94,00%	31	89,33%

As configurações com 5 conjuntos tiveram melhor desempenho do que as outras, quando se considera a taxa de acerto. Assim, essa configuração se mostrou mais adequada ao problema da classificação da subespécie da planta íris. A Tabela 24 mostra o desempenho médio de cada método para cada configuração, ou seja, a média das taxas de acerto de cada método para 3, 5 e 7 conjuntos fuzzy. Para um dado método e um número de conjuntos fuzzy, essa média é calculada somando-se todas as taxas de acertos das regras para cada uma das 15 SVMs e dividindo-se por 15.

Quanto aos métodos de classificação em mais de duas classes, o método de decomposição um por classe teve o melhor desempenho nas três configurações. O principal motivo desse bom desempenho é que nesse método são obtidos mais vetores suporte, gerando mais regras. O método de Crammer e Singer, por outro lado, teve o pior desempenho dentre os três métodos. Isso

pode ser explicado novamente em relação ao número de vetores suporte, pois esse método define menos vetores suporte e, por isso, menos regras são geradas.

Tabela 21 - Desempenho do FReX_SVM - Separação das classes duas a duas - Íris

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		7 conjuntos fuzzy		
		nr	cob	nr	cob	nr	cob	
Linear	C = 0,1	16,5	95,33%	23,5	92,67%	27,5	88,67%	
	C = 1	10	90,00%	12,5	82,00%	18	80,00%	
	C = 10	8	88,67%	9	79,33%	11	67,33%	
RBF	$\sigma^2 = 1$	C = 0,1	20,5	100%	34	100%	42	96,67%
		C = 1	18	100%	26	100%	32,5	96,67%
		C = 10	16	100%	24	98,00%	26,5	94,00%
	$\sigma^2 = 5$	C = 0,1	20,5	100%	33,5	100%	42	96,67%
		C = 1	18	98,00%	25	98,67%	27,5	90,00%
		C = 10	13	96,00%	16	92,00%	20	84,00%
	$\sigma^2 = 10$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	18,5	98,00%	26,5	98,67%	30,5	95,33%
		C = 10	12	95,33%	15	88,67%	20	86,67%
	$\sigma^2 = 50$	C = 0,1	20	99,33%	34	100%	39,5	94,00%
		C = 1	20	99,33%	32,5	99,33%	36,5	93,33%
		C = 10	14,5	89,33%	22	94,00%	25	84,00%

Novamente o desempenho da classificação com FReX_SVM é inferior ao da classificação com SVM. No entanto, como já mencionado, o objetivo do modelo proposto não é o aperfeiçoamento do desempenho de uma SVM treinada e sim a extração de seu conhecimento. A características mais importante do FReX_SVM é a interpretabilidade lingüística.

As melhores regras, segundo a acurácia fuzzy, para cada classe e configuração, são listadas a seguir:

a) Três conjuntos fuzzy

Se x_1 é C_{11} e x_2 é C_{22} e x_3 é C_{31} e x_4 é C_{41} , então Classe 1.

[acurácia: 0,9990/ abrangência: 0,1623]

Se x_1 é C_{11} e x_2 é C_{21} e x_3 é C_{32} e x_4 é C_{42} , então Classe 2.

[acurácia: 0,9629/ abrangência: 0,0525]

Se x_1 é C_{12} e x_2 é C_{21} e x_3 é C_{32} e x_4 é C_{43} , então Classe 3.

[acurácia: 0,9645/ abrangência: 0,0235]

b) Cinco conjuntos fuzzy

Se x_1 é C_{11} e x_2 é C_{23} e x_3 é C_{31} e x_4 é C_{41} , então Classe 1.

[acurácia: 1,0000/ abrangência: 0,0477]

Se x_1 é C_{13} e x_2 é C_{22} e x_3 é C_{33} e x_4 é C_{43} , então Classe 2.

[acurácia: 0,8245/ abrangência: 0,0630]

Se x_1 é C_{14} e x_2 é C_{23} e x_3 é C_{34} e x_4 é C_{44} , então Classe 3.

[acurácia: 0,9767/ abrangência: 0,0385]

c) Sete conjuntos fuzzy

Se x_1 é C_{12} e x_2 é C_{25} e x_3 é C_{32} e x_4 é C_{42} , então Classe 1.

[acurácia: 1,0000/ abrangência: 0,0316]

Se x_1 é C_{14} e x_2 é C_{23} e x_3 é C_{34} e x_4 é C_{44} , então Classe 2.

[acurácia: 0,9629/ abrangência: 0,0351]

Se x_1 é C_{14} e x_2 é C_{23} e x_3 é C_{35} e x_4 é C_{45} , então Classe 3.

[acurácia: 0,9300/ abrangência: 0,0202]

Tabela 22 - Desempenho do FREx_SVM - Método de Crammer e Singer - Íris

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		7 conjuntos fuzzy		
		nr	cob	nr	cob	nr	cob	
Linear	C = 0,1	14,5	89,33%	22,5	88,67%	25,5	79,33%	
	C = 1	11	80,67%	15	68,00%	20	60,00%	
	C = 10	10	79,33%	15	66,67%	18,5	64,00%	
RBF	$\sigma^2 = 1$	C = 0,1	20,5	100%	32,5	100%	40	96,67%
		C = 1	17	99,33%	23,5	98,67%	29,5	96,67%
		C = 10	13,5	98,00%	20	97,33%	22,5	90,00%
	$\sigma^2 = 5$	C = 0,1	20,5	100%	32,5	100%	39,5	96,67%
		C = 1	15,5	96,67%	21,5	97,33%	25,5	91,33%
		C = 10	10,5	94,67%	12	86,67%	15,5	75,33%
	$\sigma^2 = 10$	C = 0,1	20,5	100%	33	100%	40	96,67%
		C = 1	16,5	98,00%	22,5	95,33%	26	90,00%
		C = 10	10,5	95,33%	13,5	85,33%	17,5	77,33%
	$\sigma^2 = 50$	C = 0,1	20,5	100%	34	100%	42,5	96,67%
		C = 1	18	96,67%	26,5	96,00%	33,5	86,00%
		C = 10	14	95,33%	19	94,00%	23	83,33%

Tabela 23 - Melhor Desempenho do FReX_SVM e da SVM - Íris

	3 conjuntos fuzzy	5 conjuntos fuzzy	7 conjuntos fuzzy	SVM
Método	Crammer e Singer	Separação das classes	Crammer e Singer	Crammer e Singer
Kernel	RBF com $\sigma^2 = 50$ e $C = 0,1$	RBF com $\sigma^2 = 5$ e $C = 0,1$	RBF com $\sigma^2 = 5$ e $C = 0,1$	RBF com $\sigma^2 = 10$ e $C = 10$
Taxa de acerto	89,33%	94,00%	92,00%	98,00%
Taxa de erro	10,67%	06,00%	04,67%	02,00%
Taxa de pontos nc	---	---	03,33%	---

Verifica-se que a acurácia das regras acima é bem alta (maior do que 0,8) mostrando que as regras geradas são muito precisas quanto à classificação do conjunto de teste. No entanto, a abrangência das regras é baixa, já que a abrangência fuzzy calcula a soma dos graus de pertinência dos padrões que satisfazem a regra e divide pelo número total de padrões.

Tabela 24 - Desempenho Médio do FReX_SVM – Íris

n° de conj. / Método	3 conjuntos fuzzy	5 conjuntos fuzzy	7 conjuntos fuzzy
Decomposição um por classe	86,02% acertos	89,20% acertos	86,67% acertos
Separação das classes duas a duas	84,00% acertos	86,31% acertos	83,24% acertos
Crammer e Singer	81,33% acertos	82,04% acertos	79,60% acertos

5.2.2. Classificação dos tipos de Vinhos

O banco de dados de classificação do tipo de vinho é também conhecido da literatura de reconhecimento de padrões. Os dados foram obtidos da análise química de vinhos cultivados na mesma região da Itália, mas por três diferentes videiras. A análise determinou 13 constituintes achados em cada um dos três tipos de vinho. Os atributos são: álcool, ácido málico, cinzas, alcalinidade das cinzas, magnésio, fenóis totais, flavonóides, fenóis não flavonóides, proantocianinas, intensidade da cor, matiz, OD280/OD315 do vinho diluído e prolina. O conjunto de dados contém 178 exemplos, 59 da classe 1, 71 da classe 2 e 48 da classe 3.

O banco de dados usado para a classificação do tipo de vinho foi obtido do "UCI Machine Learning Repository" [54]. Assim, a entrada do sistema é uma matriz de 178 linhas (um padrão por linha) com 13 colunas.

Os 178 exemplos disponíveis foram divididos em 2 conjuntos disjuntos com 89 padrões cada, sendo que um conjunto contém 30 padrões da classe 1, 35 padrões da classe 2 e 24 padrões da classe 3, e o outro contém 29 exemplos da classe 1, 36 da classe 2 e 24 da classe 3. Foram feitos dois treinamentos, um usando o primeiro conjunto como conjunto de treinamento e o segundo como conjunto de teste e outro usando o segundo conjunto como conjunto de treinamento e o primeiro como conjunto de teste.

Como no problema anteriores, os métodos para classificação em múltiplas classes descritos no seção 2.3 foram utilizados neste caso. Os kernels e valores da constante de regularização C também foram mantidos. Assim, como nos casos anteriores, foram geradas 15 SVMs para cada um dos três métodos.

5.2.2.1. Apresentação dos Resultado

Os resultados obtidos encontram-se nas Tabelas 25, 26, e 27. Como no caso da classificação das subespécies da planta íris, encontram-se listadas nas Tabelas as percentagens de pontos incorretamente classificados e de pontos não classificados, tanto no teste (“Err Te” e “Nc Te”, respectivamente) quanto no treinamento (“Err Tr” e “Nc Tr”, respectivamente). Os casos em que o erro é igual a 0 são indicados com o sinal “--” para serem facilmente destacados.

O melhor desempenho foi obtido pelo método de separação das classes duas a duas com decisão por voto para o kernel RBF com largura $\sigma^2 = 10$ e constante de regularização $C = 10$, apresentando erro de teste igual a 01,12% dos padrões de teste.

O kernel RBF com $\sigma^2 = 10$ apresentou a menor média de taxa de erros, exposta na Tabela 28. O cálculo dessa média para um dado kernel é a soma dos resultados das três configurações (uma para cada valor de C) para os três métodos dividida por nove.

Tabela 25 - Resultado do método de decomposição um por classe - Vinho

Kernel		nº sv	Err Tr	Nc Tr	Err Te	Nc Te	
Linear	C = 0,1	35,5	00,56%	01,12%	02,25%	01,69%	
	C = 1	21	--	--	01,12%	02,81%	
	C = 10	19	--	--	01,12%	01,69%	
RBF	$\sigma^2 = 1$	C = 0,1	89	--	100%	--	100%
		C = 1	88	--	--	--	46,07%
		C = 10	87,5	--	--	--	43,26%
	$\sigma^2 = 5$	C = 0,1	87	--	43,26%	--	56,18%
		C = 1	53	--	--	01,12%	02,81%
		C = 10	50	--	--	00,56%	03,93%
	$\sigma^2 = 10$	C = 0,1	89	--	26,97%	--	31,46%
		C = 1	50	00,56%	--	01,69%	01,12%
		C = 10	39	--	100%	01,12%	01,12%
	$\sigma^2 = 50$	C = 0,1	89	--	--	--	100%
		C = 1	64	00,56%	02,25%	02,25%	05,62%
		C = 10	29,5	--	--	01,12%	02,25%

Tabela 26 - Resultado do método de separação das classes duas a duas - Vinho

Kernel			nº sv	Err Tr	Err Te
Linear		C = 0,1	30,5	00,56%	02,81%
		C = 1	17	--	02,81%
		C = 10	17	--	02,81%
RBF	$\sigma^2 = 1$	C = 0,1	88,5	60,11%	60,11%
		C = 1	87,5	--	17,42%
		C = 10	87,5	--	16,29%
	$\sigma^2 = 5$	C = 0,1	86,5	10,67%	14,04%
		C = 1	49	00,56%	02,25%
		C = 10	46,5	--	01,69%
	$\sigma^2 = 10$	C = 0,1	86	01,69%	03,93%
		C = 1	44,5	00,56%	01,69%
		C = 10	36,5	--	01,12%
	$\sigma^2 = 50$	C = 0,1	89	58,43%	60,11%
		C = 1	55,5	00,56%	03,93%
		C = 10	26	00,56%	02,25%

Tabela 27 - Resultado do método de Crammer e Singer - Vinho

Kernel			nº sv	Err Tr	Err Te
Linear		C = 0,1	23,5	01,12%	05,62%
		C = 1	16	--	06,74%
		C = 10	16	--	06,74%
RBF	$\sigma^2 = 1$	C = 0,1	89	--	06,18%
		C = 1	88	--	07,30%
		C = 10	88	--	07,30%
	$\sigma^2 = 5$	C = 0,1	79,5	01,12%	05,06%
		C = 1	49,5	--	04,49%
		C = 10	48	--	05,06%
	$\sigma^2 = 10$	C = 0,1	79,5	01,69%	05,62%
		C = 1	39,5	00,56%	04,49%
		C = 10	36	--	05,06%
	$\sigma^2 = 50$	C = 0,1	89	03,37%	06,74%
		C = 1	44,5	00,56%	05,06%
		C = 10	22	--	04,49%

Como já foi mencionado, a constante de regularização C controla o peso que o número de erros e a margem tem no problema quadrático da SVM: quanto maior C , menos erros são permitidos. Assim, era esperado que as configurações com $C = 10$ obtivessem um melhor resultado, o que de fato, de uma maneira geral, ocorreu.

A comparação do desempenho dos métodos de classificação em múltiplas classes aparece na Tabela 29, onde são mostradas as médias das taxas de acerto nos pontos de teste obtidas de maneira análoga ao caso anterior. O método de classificação com melhor desempenho foi o de Crammer e Singer. Novamente o método mais utilizado para a resolução de problemas de classificação em mais de duas classes, o método de decomposição um por classe, apresentou o pior desempenho.

Tabela 28 - Comparação do desempenho dos Kernels - Vinho

Kernel	Média da taxa de erro
Linear	03,56%
RBF com $\sigma^2 = 1$	12,73%
RBF com $\sigma^2 = 5$	03,80%
RBF com $\sigma^2 = 10$	02,75%
RBF com $\sigma^2 = 50$	09,55%

Tabela 29 - Comparação do desempenho dos métodos de classificação - Vinho

Método de classificação	Média da taxa de acerto
Decomposição um por classe	72,51%
Separação das classes duas a duas com decisão por voto	87,08%
Crammer e Singer	94,31%

5.2.2.2. Extração de Regras Fuzzy

Nesta seção, o desempenho do método de extração de regras para classificação dos tipos de vinho é avaliado usando-se as mesmas medidas que nos casos anteriores.

Para a avaliação do desempenho de FReX_SVM, varia-se o número de conjuntos fuzzy construídos para cada coordenada. Como a dimensão do conjunto de entrada é alta, são utilizadas apenas duas configurações com 3 e 5 conjuntos, expostos nas Figuras 26 e 27, respectivamente. Como o domínio de cada atributo é muito diferente e o método FReX_SVM utiliza o mesmo conjunto fuzzy para todos os atributos, os dados de entrada foram previamente modificados para todos estarem no mesmo domínio, entre -3 e 3.

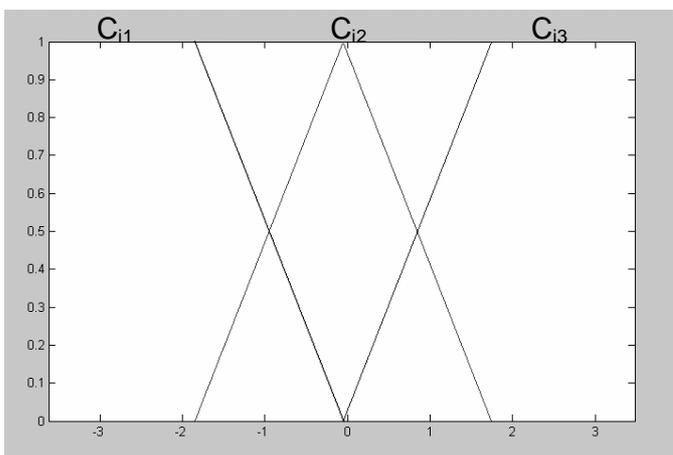


Figura 26 - 3 Conjuntos Fuzzy - Vinho

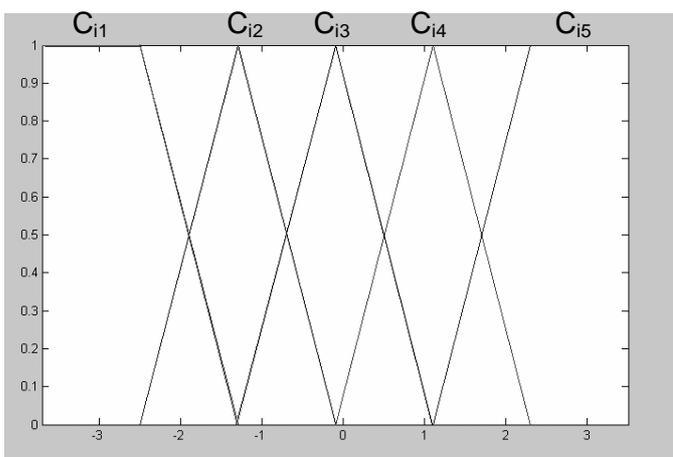


Figura 27 - 5 Conjuntos Fuzzy – Vinho

As Tabelas 30, 31 e 32 mostram o desempenho das configurações acima citadas, separados de acordo com os métodos de classificação em mais de duas classes. Uma vez que foram feitos dois treinamentos, as Tabelas apresentam a média desses treinamentos (soma de cada treinamento, dividido por 2).

Tabela 30 - Desempenho do FReX_SVM – Decomposição um por classe - Vinho

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		
		nr	cob	nr	cob	
Linear	C = 0,1	35	79,22%	35,5	32,58%	
	C = 1	20,5	65,17%	21	15,73%	
	C = 10	18,5	64,04%	19	14,61%	
RBF	$\sigma^2 = 1$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	86	92,13%	87,5	51,69%
		C = 10	85,5	91,57%	87	51,69%
	$\sigma^2 = 5$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	53	88,20%	53	32,02%
		C = 10	49,5	88,20%	50	30,90%
	$\sigma^2 = 10$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	54	86,52%	50	37,08%
		C = 10	36	83,71%	39	28,09%
	$\sigma^2 = 50$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	63,5	91,01%	64	40,45%
		C = 10	29	76,97%	29,5	25,28%

A Tabela 33 mostra o melhor desempenho de classificação nos pontos de teste, em termos de cobertura e taxa de acerto, para 3 e 5 conjuntos fuzzy. O melhor desempenho da SVM também é mostrado na Tabela. Na Tabela, são mostrados os kernels e os métodos de classificação em múltiplas classes para os quais o FReX_SVM ou a SVM obteve melhor desempenho, as taxas de acerto e erro. A taxa de pontos não classificados pelo FReX_SVM aparece na Tabela como “taxa de pontos nc”. No caso de várias configurações obterem as mesmas taxas de acertos e cobertura, foi considerado o melhor resultado a configuração que gerou menos regras. O melhor desempenho do FReX_SVM foi obtido para 3 conjuntos fuzzy, com uma classificação correta de 84,26% dos padrões de teste e erro de 7,87%. Todos os três métodos proporcionaram esse resultado.

Tabela 31 - Desempenho do FReX_SVM – Separação das classes duas a duas - Vinho

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		
		nr	cob	nr	cob	
Linear	C = 0,1	16,5	62,92%	30,5	26,97%	
	C = 1	16,5	62,92%	17	12,92%	
	C = 10	30	76,96%	17	12,92%	
RBF	$\sigma^2 = 1$	C = 0,1	86,5	92,13%	88	51,69%
		C = 1	85,5	91,57%	87	51,69%
		C = 10	85,5	91,57%	87	51,69%
	$\sigma^2 = 5$	C = 0,1	84,5	92,13%	86	51,69%
		C = 1	49	87,08%	49	30,90%
		C = 10	46,5	87,64%	48	29,21%
	$\sigma^2 = 10$	C = 0,1	84	92,13%	85,5	50,56%
		C = 1	44	85,96%	44,5	32,02%
		C = 10	36,5	83,71%	37	25,84%
	$\sigma^2 = 50$	C = 0,1	86	92,13%	87,5	51,69%
		C = 1	54,5	88,20%	55,5	38,76%
		C = 10	25,5	76,96%	22	23,03%

A Tabela 34 mostra o desempenho médio de cada método para cada uma das configurações. Dados um método e um número de conjuntos fuzzy, o desempenho médio é a soma das taxas de acertos de todas as SVMs para esse método e número de conjuntos fuzzy.

Da mesma forma que no problema anterior, o método de decomposição um por classe teve o melhor desempenho médio nas duas configurações. Novamente isso pode ser atribuído ao grande número de vetores suporte que o método obtém. O método de separação das classes duas a duas, por outro lado, teve o pior desempenho dentre os três métodos para a configuração com 3 conjuntos fuzzy. Já o método de Crammer e Singer apresentou o pior desempenho para 5 conjuntos fuzzy. Como o número de vetores suporte desses dois métodos é praticamente igual, não houve um método que tivesse o pior desempenho geral.

Como nos casos anteriores, o desempenho da classificação com FReX_SVM é pior do que o da classificação com SVM. No entanto, o modelo de extração de regras fuzzy proposto gerou regras que auxiliam no entendimento de como se chegou a essa classificação.

Tabela 32 - Desempenho do FReX_SVM – Método de Crammer e Singer - Vinho

Kernel		3 conjuntos fuzzy		5 conjuntos fuzzy		
		nr	cob	nr	cob	
Linear	C = 0,1	23,5	73,60%	23,5	17,98%	
	C = 1	16	65,17%	16	10,67%	
	C = 10	16	65,17%	16	10,67%	
RBF	$\sigma^2 = 1$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	86	92,13%	86,5	51,69%
		C = 10	86	92,13%	86,5	51,69%
	$\sigma^2 = 5$	C = 0,1	78	91,01%	79,5	47,75%
		C = 1	49	87,64%	49,5	25,28%
		C = 10	47,5	87,64%	48	24,16%
	$\sigma^2 = 10$	C = 0,1	77	91,57%	79,5	47,19%
		C = 1	38	86,52%	39,5	25,28%
		C = 10	35,5	84,27%	36	23,03%
	$\sigma^2 = 50$	C = 0,1	87	92,13%	88,5	51,69%
		C = 1	44	83,71%	44	33,15%
		C = 10	22	73,60%	22	17,42%

Tabela 33 - Melhor Desempenho do FReX_SVM e da SVM - Vinho

	3 conjuntos fuzzy	5 conjuntos fuzzy	SVM
Método	Separação das classes	Separação das classes	Separação das classes
Kernel	RBF com $\sigma^2 = 10$ e C = 0,1	RBF com $\sigma^2 = 5$ e C = 0,1	RBF com $\sigma^2 = 10$ e C = 10
Taxa de acerto	84,26%	51,69%	98,88%
Taxa de erro	07,87%	---	01,12%
Taxa de pontos nc	07,87%	48,31%	---

As melhores regras, segundo a acurácia fuzzy, para cada classe e configuração, são listadas a seguir:

a) Três conjuntos fuzzy

Se x_1 é C_{13} e x_2 é C_{22} e x_3 é C_{32} e x_4 é C_{41} e x_5 é C_{52} e x_6 é C_{63} e x_7 é C_{73} e x_8 é

C_{82} e x_9 é C_{93} e x_{10} é C_{102} e x_{11} é C_{113} e x_{12} é C_{122} e x_{13} é C_{133} , então Classe 1.
[acurácia: 1,0000/ abrangência: 0,0014]

Se x_1 é C_{11} e x_2 é C_{22} e x_3 é C_{32} e x_4 é C_{43} e x_5 é C_{51} e x_6 é C_{62} e x_7 é C_{72} e x_8 é C_{82} e x_9 é C_{92} e x_{10} é C_{101} e x_{11} é C_{112} e x_{12} é C_{122} e x_{13} é C_{132} , então Classe 2.
[acurácia: 1,0000/ abrangência: 0,0008]

Se x_1 é C_{12} e x_2 é C_{23} e x_3 é C_{32} e x_4 é C_{42} e x_5 é C_{52} e x_6 é C_{62} e x_7 é C_{71} e x_8 é C_{83} e x_9 é C_{92} e x_{10} é C_{103} e x_{11} é C_{111} e x_{12} é C_{121} e x_{13} é C_{132} , então Classe 3.
[acurácia: 0,9999/ abrangência: 0,0011]

b) Cinco conjuntos fuzzy

Se x_1 é C_{14} e x_2 é C_{22} e x_3 é C_{32} e x_4 é C_{42} e x_5 é C_{53} e x_6 é C_{64} e x_7 é C_{74} e x_8 é C_{82} e x_9 é C_{94} e x_{10} é C_{103} e x_{11} é C_{113} e x_{12} é C_{124} e x_{13} é C_{134} , então Classe 1.
[acurácia: 1,0000/ abrangência: 0,0009]

Se x_1 é C_{13} e x_2 é C_{22} e x_3 é C_{31} e x_4 é C_{42} e x_5 é C_{52} e x_6 é C_{63} e x_7 é C_{73} e x_8 é C_{82} e x_9 é C_{93} e x_{10} é C_{103} e x_{11} é C_{114} e x_{12} é C_{123} e x_{13} é C_{132} , então Classe 2.
[acurácia: 1,0000/ abrangência: 0,0010]

Se x_1 é C_{14} e x_2 é C_{25} e x_3 é C_{33} e x_4 é C_{43} e x_5 é C_{53} e x_6 é C_{62} e x_7 é C_{72} e x_8 é C_{84} e x_9 é C_{92} e x_{10} é C_{104} e x_{11} é C_{112} e x_{12} é C_{122} e x_{13} é C_{133} , então Classe 3.
[acurácia: 1,0000/ abrangência: 0,0014]

Por apresentar muitos atributos de entrada e poucos padrões para treinamento, esse banco de dados apresentou baixa abrangência para todas as regras. As somas das abrangências de todas regras fuzzy para a configuração com 3 conjuntos fuzzy são 0,0344 e 0,0299, para o primeiro e segundo treinamentos, respectivamente. Para 5 conjuntos fuzzy, as somas são 0,0269 e 0,0283, para o primeiro e segundo treinamentos, respectivamente. Já a acurácia da regras é bem alta, maior do que 0,99, mostrando que as regras são muito precisas em relação à separação das classes.

Tabela 34 - Desempenho Médio do FREx_SVM - Vinho

nº de conjuntos Métodos	3 conjuntos fuzzy	5 conjuntos fuzzy
Decomposição um por classe	75,50% acertos	37,64% acertos
Separação das classes duas a duas	73,67% acertos	36,19% acertos
Crammer e Singer	74,30% acertos	32,40% acertos

5.2.3. Discussão dos Resultados

5.2.3.1. Métodos de Classificação

A Tabela 35 apresenta a taxa de acerto do melhor resultado de cada um dos três métodos de classificação em múltiplas classes, para efeito de comparação, usando kernel linear. Note que a coluna "1-vs-todos" designa o método de decomposição um por classe, a coluna "1-vs-1", o método de separação das classes duas a duas, e a coluna "C&S", o método de Crammer e Singer. O método de separação das classes duas a duas apresentou o melhor desempenho nos dois problemas. Para a classificação da subespécie da planta íris, o método de separação das classes duas a duas apresentou um desempenho muito bom, com taxa de acerto de 97,33%. Para a classificação do tipo de vinho, os métodos de decomposição um por classe e separação das classes duas a duas apresentaram um desempenho muito bom, com taxa de acerto de 97,19%.

Tabela 35 - Comparação do desempenho dos métodos de classificação para Kernel Linear

Problema	1-vs-todos	1-vs-1	C&S
íris	78%	97,33%	82,67%
vinho	97,19%	97,19%	94,38%

Por fim, a Tabela 36 apresenta o melhor resultado da comparação dos três métodos usando kernel RBF. Para o problema de classificação da subespécie da planta íris, em todos os três métodos, a largura da RBF usada foi igual a 10 e o valor da constante de regularização C igual a 10. Para o problema de classificação do tipo de vinho, em todos os três métodos, o valor da constante de regularização C foi igual a 10. Já a largura da RBF usada foi igual a 10 nos métodos de decomposição das classes e de separação das classes duas a duas, e no método de Crammer e Singer a largura da RBF foi 50. Para o problema de classificação da subespécie da planta íris, o método de Crammer e Singer apresentou o melhor desempenho, classificando corretamente 98% dos padrões de teste. Para o problema de classificação do tipo de vinho, o método de separação das classes duas a duas apresentou o melhor desempenho, classificando corretamente 98,88% dos os padrões de teste.

Considerando os dois tipos de kernel para os dois problemas avaliados, o método de separação das classes duas a duas obteve o melhor desempenho.

Tabela 36 - Comparação do desempenho dos métodos de classificação para Kernel RBF

Problema	1-vs-todos	1-vs-1	C&S
íris	97,33%	97,33%	98%
vinho	97,76%	98,88%	95,51%

5.2.3.2. Extração de Regras

A aplicação do modelo FReX_SVM nos problemas de classificação da subespécie da planta Íris e do tipo de vinho, quando são empregados os três métodos de classificação em múltiplas classes, permite observar que é possível a construção de um conjunto de regras que classifique os dados de teste de maneira confiável (Tabelas 24 e 34). A cobertura das regras mostra-se ótima. Para o problema da íris, apenas a configuração com sete conjuntos fuzzy não teve cobertura de 100% dos padrões de teste, mas ainda assim com sete conjuntos fuzzy consegue-se classificar 96,67% dos padrões de teste. No caso da classificação do tipo de vinho, a configuração com 3 conjuntos fuzzy obteve cobertura de 92,13% dos padrões de teste e a com 5 cobriu 51,69% dos padrões de teste.

De forma análoga ao caso binário, a estratégia de extração de regras com a utilização de conjuntos fuzzy mostrou-se bastante promissora. Além disso, as regras obtidas utilizam conjuntos fuzzy nos antecedentes, o que as torna mais flexíveis.

Quanto aos métodos de classificação em múltiplas classes, o que apresentou melhor desempenho na extração de regras foi o método de decomposição um por classe. Para a classificação da subespécie da planta íris, a taxa média de acerto desse método foi de 86,02% no caso de 3 conjuntos fuzzy, 89,20% no caso de 5 conjuntos fuzzy e 86,67% no caso de 7 conjuntos fuzzy (Tabela 24). Para o problema do vinho, as taxas médias de acerto desse método foram de 75,50% e 37,64%, nos casos de 3 e 5 conjuntos fuzzy, respectivamente (Tabela 34).

O método que apresentou o pior desempenho para o problema da íris foi o método de Crammer e Singer, com taxa média de acerto de 81,33% no caso de 3 conjuntos fuzzy, 82,04% no caso de 5 conjuntos fuzzy e 79,60% no caso de 7 conjuntos fuzzy (Tabela 24). Já para o problema do vinho, o método de separação das classes duas a duas teve o pior desempenho, com taxa média de acerto de 73,67% no caso de 3 conjuntos fuzzy; o método de Crammer e Singer apresentou o pior desempenho no caso de 5 conjuntos fuzzy com taxa média de acerto de 32,40% (Tabela 34).

5.3.

Comparação do Desempenho do FREx_SVM e do Método de Mendel

Nessa seção, compara-se o melhor resultado obtido pelo FREx_SVM com o desempenho do método de Mendel, implementado usando o gerador automático de regras *Fuzzy Rules 2001 – Sistema de Inferência Fuzzy com Geração Automática de Regras* [56], para cada banco de dados e cada número de conjuntos fuzzy. As Tabelas 37, 38, 39 e 40 mostram essa comparação. Nas Tabelas estão expostos o número de regras geradas (“nr”), a cobertura (“cob”) e o erro nos pontos de teste (“erro”) para cada um dos dois métodos. O melhor resultado obtido para cada banco de dados está em negrito.

Tabela 37 - Comparação entre FREx_SVM e Método de Mendel - Bupa Liver Disorders

		FREx_SVM	Mendel
3 conjuntos fuzzy	nr	28	36,5
	cob	95,94%	96,52%
	erro	47,25%	48,97%
5 conjuntos fuzzy	nr	74	84,5
	cob	85,51%	87,25%
	erro	40,87%	39,71%
7 conjuntos fuzzy	nr	108,5	120
	cob	80,58%	100%
	erro	36,52%	39,71%

Tabela 38 - Comparação entre FREx_SVM e Método de Mendel - Wisconsin Breast Cancer

		FREx_SVM	Mendel
3 conjuntos fuzzy	nr	131,5	136
	cob	77,90%	98,09%
	erro	02,49%	03,07%
5 conjuntos fuzzy	nr	157	173
	cob	63,98%	100%
	erro	00,73%	05,41%

No caso do banco de dados Bupa Liver Disorders, os desempenhos do FREx_SVM e do método de Mendel são praticamente iguais para as

configurações com 3 e 5 conjuntos fuzzy, sendo que o FREx_SVM gera menos regras. Para a configuração com 7 conjuntos fuzzy, o método de Mendel obtém o melhor desempenho, com cobertura de 100%.

As coberturas obtidas pelo método de Mendel com 3 e 5 conjuntos fuzzy para o banco de dados Wisconsin Breast Cancer são maiores do que as coberturas obtidas pelo FREx_SVM. No entanto, FREx_SVM apresenta menor taxa de erro nas duas configurações.

Para o banco de dados Iris, tanto o FREx_SVM quanto o método de Mendel obtém ótima cobertura, sendo 100% para 3 e 5 conjuntos fuzzy. Os desempenhos dos dois métodos estão bem próximos, sendo que o FREx_SVM apresenta menor taxa de erro para duas das três configurações.

Tabela 39 - Comparação entre FREx_SVM e Método de Mendel - Iris

		FREx_SVM	Mendel
3 conjuntos fuzzy	nr	20,5	16,5
	cob	100%	100%
	erro	10,67%	07,33%
5 conjuntos fuzzy	nr	33,5	34,5
	cob	100%	100%
	erro	06,00%	10,67%
7 conjuntos fuzzy	nr	39,5	47
	cob	96,67%	99,33%
	erro	04,67%	09,33%

Tabela 40 - Comparação entre FREx_SVM e Método de Mendel - Vinho

		FREx_SVM	Mendel
3 conjuntos fuzzy	nr	84	86
	cob	92,13%	100%
	erro	07,87%	21,91%
5 conjuntos fuzzy	nr	86	89
	cob	51,69%	100%
	erro	--	43,26%

Da mesma forma que no caso do Wisconsin Breast Cancer, as coberturas do método de Mendel para o banco de dados Vinho são maiores do que as do FREx_SVM e as taxas de erro do FREx_SVM são menores do que as do método de

Mendel, sendo que a configuração de 5 conjuntos fuzzy do FREx_SVM apresenta taxa de erro igual a 0.

Pode-se verificar que FREx_SVM apresenta pequenas taxas de erro para os bancos de dados Wisconsin Breast Cancer, Íris e Vinho, em muitos casos menor do que a do método de Mendel. Para o banco de dados Bupa Liver Disorders, as taxas de erro do FREx_SVM são grandes, mas as taxas de erro do método de Mendel também são. Além disso, exceto em um caso, FREx_SVM gera menos regras do que o método de Mendel. Dessa forma, FREx_SVM se mostra um excelente extrator de regras.

A maior desvantagem de FREx_SVM em relação ao método de Mendel é a cobertura das regras. Em todos os casos, a cobertura do método de Mendel superou a do FREx_SVM.

Vale a pena mencionar que não foi possível realizar a comparação do desempenho de FREx_SVM com o dos métodos RuExSVM [13] e SVM+Prototypes [14], [15]. Em ambos os casos, os respectivos artigos não apresentam dados suficientes para a implementação dos algoritmos.