

3 Resultados Numéricos

Neste capítulo serão apresentados os resultados numéricos dos experimentos realizados com a técnica proposta. Os cinco primeiros foram feitos com dados simulados e os outros dois com bancos de dados disponíveis no repositório de dados da UCI*.

O algoritmo de LVQ utilizado para atualizar os protótipos foi o LVQ 1, descrito no apêndice B. A taxa de aprendizado foi fixada em 0.01 para todos os experimentos realizados.

As classificações finais são divididas em três etapas:

- 1) Classificação dentro-da-amostra e fora-da-amostra sem eliminar nenhum dado para o treinamento;
- 2) Classificação dentro-da-amostra e fora-da-amostra marcando o corte do vetor de correlação ρ em zero;
- 3) Classificação dentro-da-amostra e fora-da-amostra marcando o corte do vetor de correlação ρ que pode ser no mínimo encontrado pela função, em zero ou em -0.2.

Note que, como afirmado no capítulo 2, quando o mínimo da função é maior que zero, o corte no vetor de correlação ρ pode ser feito no próprio mínimo da função, em zero ou em -0.2, de modo que, se em algum caso se decida cortar novamente em zero, as etapas 2 e 3 serão equivalentes, pois ambas atualizarão os protótipos com o mesmo conjunto de dados.

* <http://www.ics.uci.edu/~mlearn/MLRepository.html>

Nos experimentos com dados simulados, a maior parte das distribuições teve ruído gerado em diferentes proporções para constatar a eficiência do método, como será explicitado em cada exemplo. Nesses casos, a análise será feita não só em torno da classificação geral, mas, também, de acordo com o reconhecimento de ruído.

Para o cálculo do vetor de correlação ρ foram considerados 200 sorteios aleatórios de protótipos, pois esse número pareceu uma boa solução de compromisso entre a qualidade dos resultados e as possibilidades computacionais.

Exemplo 1 – 2 Classes, 2 Dimensões:

Classe 1:

Média: [1 2]

Matriz de Covariância: [5 4; 5 4]

Instâncias: 1000

Classe 2:

Média: [12 10]

Matriz de Covariância: [4 2; 2 4]

Instâncias: 1000

Na figura 4, vêem-se as duas distribuições e o intervalo de sorteio de protótipos de cada uma. Foram gerados 10% de ruído em cada distribuição. A figura 5 representa as distribuições após o cálculo do vetor de correlação ρ , estipulando o corte em correlação zero. Nesse caso, os pontos azuis são os mantidos e os vermelhos, os retirados. O histograma de ρ foi calculado (figura 6) e a função estimada (figura 7). O mínimo da função foi encontrado no ponto -0,2636. Finalmente, na figura 8, encontram-se as distribuições após a separação pelo mínimo da função. Novamente, os pontos azuis são aqueles que serão utilizados durante o treinamento e os vermelhos os pontos eliminados.

Note que, na figura 5, vários pontos pertencentes à fronteira de decisão estavam marcados em vermelho e após o corte no mínimo da função (figura 8) passaram a estar marcados em azul. A tabela 2 mostra a quantidade de ruído reconhecido com o corte da função em zero e após a obtenção do mínimo da função. Entre parênteses encontram-se as porcentagens de ruído reconhecido. Na tabela 3, tem-se a quantidade de pontos eliminados com o corte em zero que não eram ruídos e com o corte no mínimo da função que não eram ruídos.

Exemplo 1 – Quantidade de Ruído Detectado pelo Método

	Corte em zero	Corte no mínimo
Distribuição 1	100 (100%)	94 (94%)
Distribuição 2	100 (100%)	98 (98%)

Tabela 2

Exemplo 1 – Quantidade de Pontos Relevantes Equivocadamente Eliminados

	Corte em zero	Corte no mínimo
Quantidade de Pontos	8	0

Tabela 3

Observe que a quantidade de ruído reconhecido após o corte no mínimo da função foi ligeiramente inferior ao corte em zero. Em compensação, 8 pontos que haviam sido eliminados nessa fase foram poupados ao se estabelecer o corte da função.

Finalmente, na tabela 4, tem-se a porcentagem de classificação relativa a cada uma das etapas. Dentro-da-amostra, tivemos 2000 dados, 1000 de cada classe e fora-da-amostra, 4000 dados, 2000 de cada classe.

Exemplo 1 - Classificação

	Etapa 1 Sem seleção de dados	Etapa 2 Com seleção de dados (Corte em zero)	Etapa 3 Com seleção de dados (Corte no mínimo)
Dentro-da-amostra	89,65 %	99,89 %	99,12 %
Fora-da-amostra	97,83 %	99,53 %	99,38 %

Tabela 4

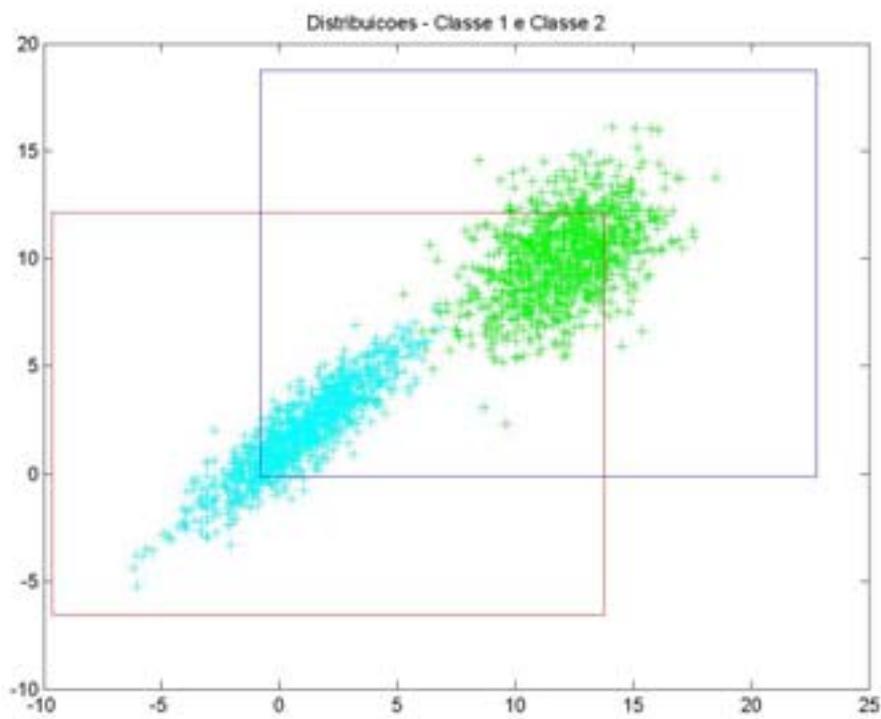


Figura 4 - Exemplo 1: Distribuições Originais

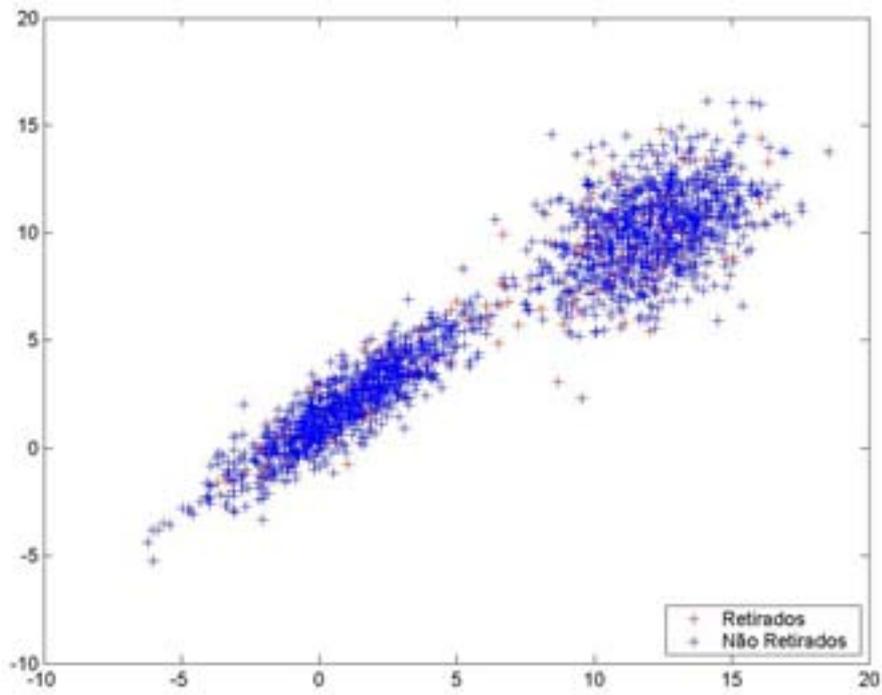


Figura 5 - Exemplo 1: Distribuições após corte em zero

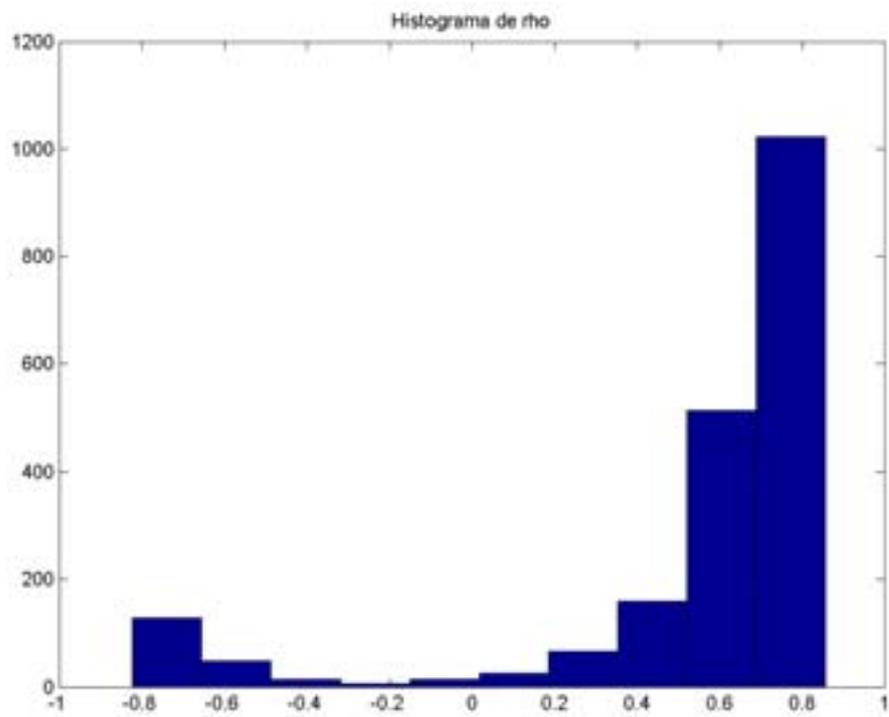


Figura 6 - Exemplo 1: Histograma de Rho

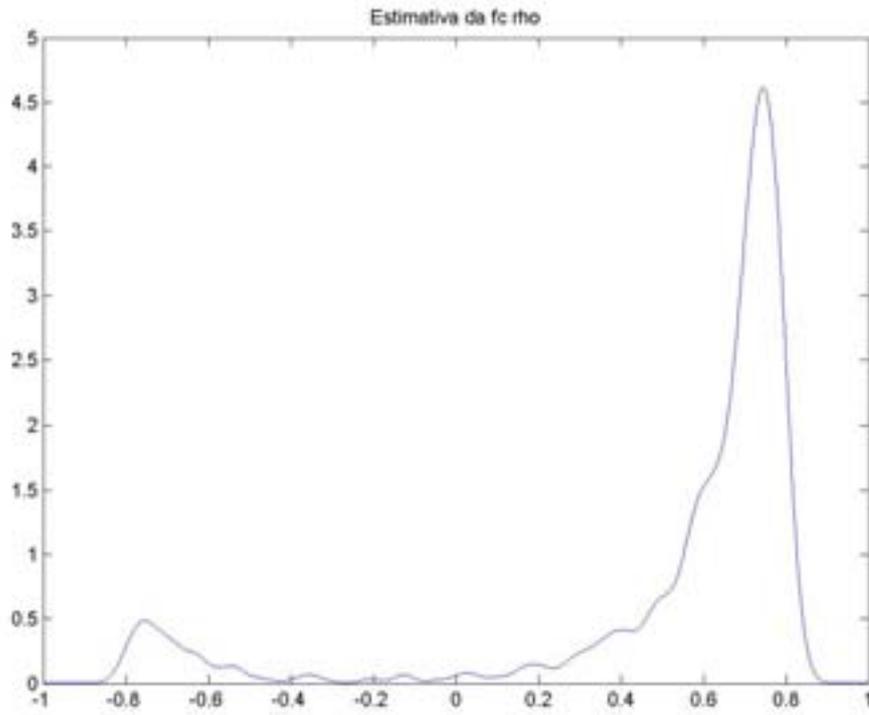


Figura 7 - Exemplo 1: Função Estimada pela Janela de Parzen

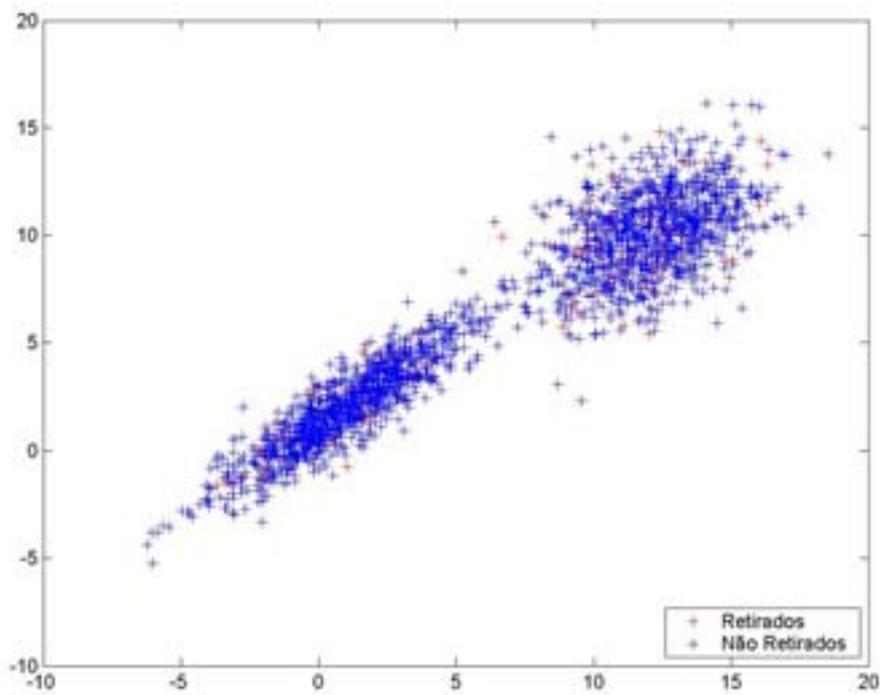


Figura 8 - Exemplo 1: Distribuições após corte no mínimo

Embora as classificações tenham sido boas de uma maneira geral, houve um ganho de 1,5% a 1,7% fora-da-amostra ao considerarmos a eliminação de ruídos. Nesse caso, em termos de classificação, não houve diferença significativa ao realizar o corte em zero ou após o mínimo da função.

Exemplo 2 – 2 Classes, 2 Dimensões:

Classe 1:

Média: [1 2]

Matriz de Covariância: [5 4; 5 4]

Instâncias: 1000

Classe 2:

Média: [9 7]

Matriz de Covariância: [1 0; 0 1]

Instâncias: 1000

Na figura 9, vêem-se as duas distribuições e o intervalo de sorteio de protótipos de cada uma. Na primeira distribuição não foi gerado ruído, enquanto na segunda, foram gerados 20% de ruído do seu total de pontos. A figura 10 representa as distribuições após o cálculo do vetor de correlação ρ , estipulando o corte em correlação zero. Nesse caso, os pontos azuis são os mantidos e os vermelhos, os retirados. O histograma de ρ foi calculado (figura 11) e a função estimada (figura 12). O mínimo da função nesse caso foi positivo, assim, a aplicação estipulou o mínimo como zero. Dessa forma, não é necessário realizar a terceira etapa de classificação já que ela seria igual à segunda.

Existe uma explicação do por que o mínimo da função foi positivo. Note que, a região de sorteio de protótipos da distribuição azul é muito grande, devido aos ruídos gerados na verde-limão. Em contrapartida, a região de sorteio de protótipos da distribuição verde-limão é pequena, pois não há ruído na azul. Dessa forma, em grande parte dos 200 sorteios aleatórios os ruídos que foram gerados na verde-limão terão sua classificação errada, pois estarão mais distantes

do protótipo azul em comparação ao verde. Assim, ao calcular a correlação, esse número se aproxima muito de zero, podendo ser positivo ou negativo.

Existe uma certa quantidade de ruído que se aglomera na região central do histograma (figura 11). Ao se eliminar os 5% de pontos em cada extremo da função, os ruídos que tiveram correlação negativa acabam sendo eliminados, portanto, o mínimo acaba na parte positiva e elimina, junto com o ruído, muitos pontos considerados relevantes. Por isso pode ser indicada a política de corte considerar o corte em zero ou em -0,2 quando o mínimo é positivo.

A tabela 5 mostra a quantidade de ruído reconhecido com o corte da função em zero. Entre parênteses encontram-se as porcentagens de ruído reconhecido. Na tabela 6, tem-se a quantidade de pontos relevantes eliminados com o corte em zero.

Exemplo 2 – Quantidade de Ruído Detectado pelo Método

	Corte em zero
Distribuição 2	188 (94%)

Tabela 5

Exemplo 2 – Quantidade de Pontos Relevantes Equivocadamente Eliminados

	Corte em zero
Quantidade de Pontos	39

Tabela 6

Observe que buscar mais ruídos com o mínimo da função positivo implicaria em mais pontos relevantes eliminados.

Finalmente, na tabela 7, tem-se a porcentagem de classificação relativa a cada uma das duas etapas. Dentro-da-amostra, tivemos 2000 dados, 1000 de cada classe e fora-da-amostra, 4000 dados, 2000 de cada classe.

Exemplo 2 - Classificação

	Etapa 1 Sem seleção de dados	Etapa 2 Com seleção de dados (Corte em zero)
Dentro-da-amostra	87,6 %	97,4 %
Fora-da-amostra	92,3 %	97,8 %

Tabela 7

Nesse caso já houve uma diferença razoavelmente considerável fora-da-amostra ao eliminar ruídos.

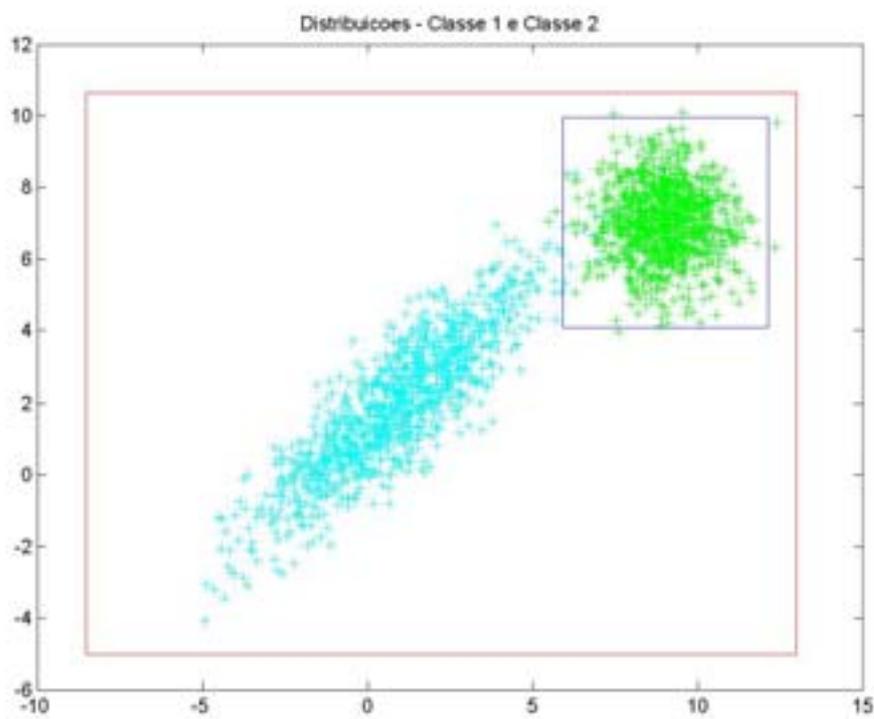


Figura 9 - Exemplo 2: Distribuições Originais

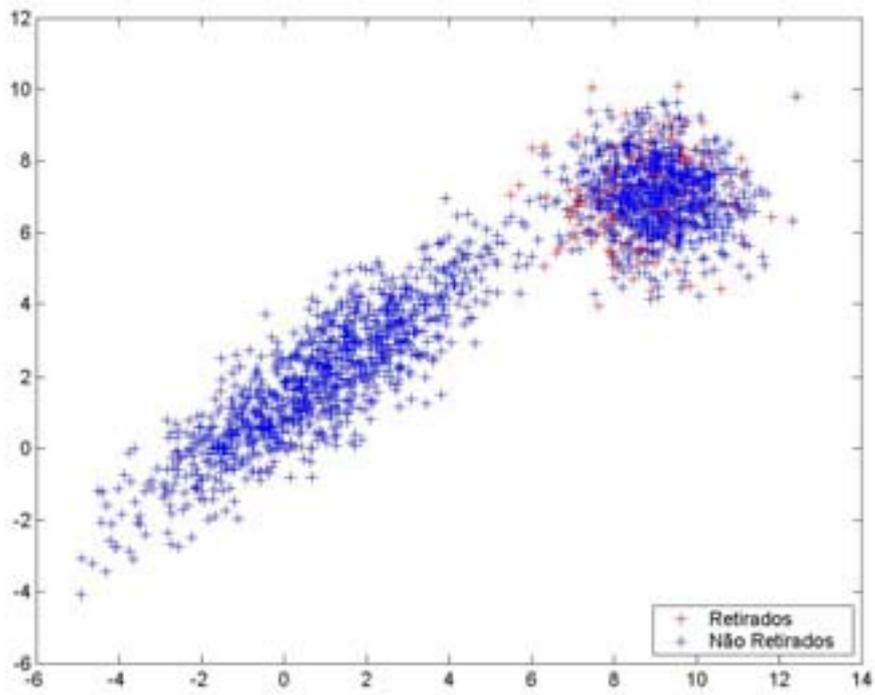


Figura 10 - Exemplo 2: Distribuições após corte em zero

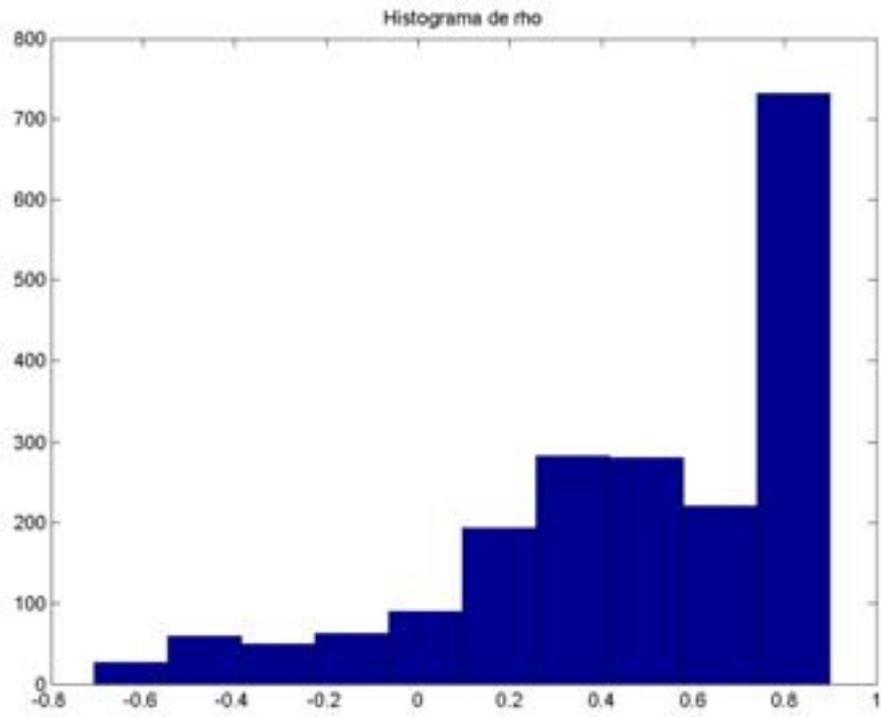


Figura 11 - Exemplo 2: Histograma de Rho

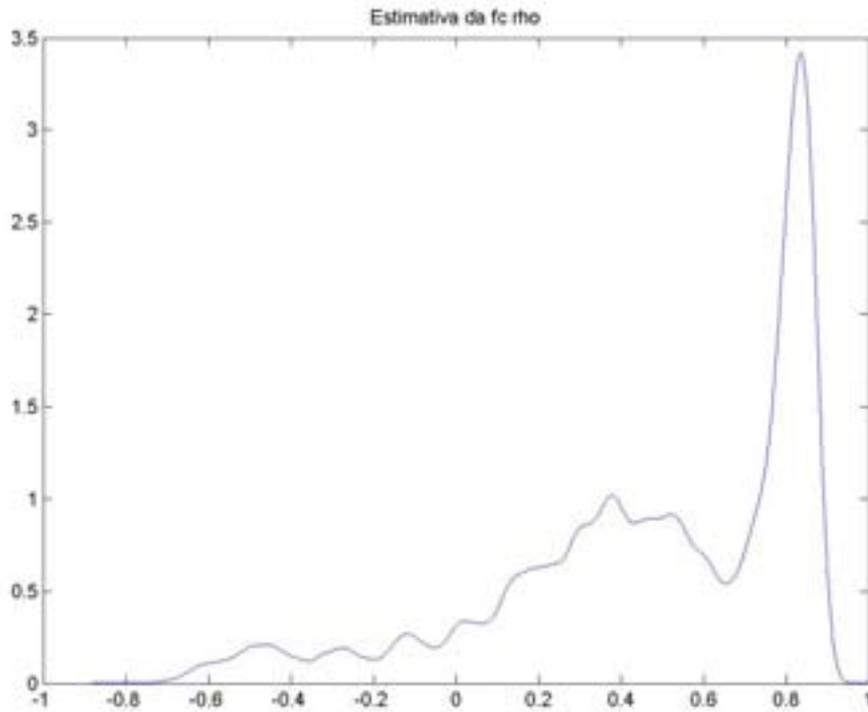


Figura 12 - Exemplo 2: Função Estimada pela Janela de Parzen

Exemplo 3 – 2 Classes, 2 Dimensões:

Classe 1:

Média: [1 2]

Matriz de Covariância: [10 -4; -4 10]

Instâncias: 1000

Classe 2:

Média: [10 8]

Matriz de Covariância: [6 2; 2 6]

Instâncias: 1000

Na figura 13, vêem-se as duas distribuições e o intervalo de sorteio de protótipos de cada uma. Foram gerados 30% de ruído em cada distribuição. A figura 14 representa as distribuições após o cálculo do vetor de correlação ρ , estipulando o corte em correlação zero. Nesse caso, os pontos azuis são os

mantidos e os vermelhos, os retirados. O histograma de ρ foi então calculado (figura 15) e a função estimada (figura 16). O mínimo da função foi encontrado no ponto -0,0945. Finalmente, na figura 17, encontram-se as distribuições após a separação pelo mínimo da função. Novamente, os pontos azuis são aqueles que serão utilizados durante o treinamento e os vermelhos os pontos eliminados.

A tabela 8 mostra a quantidade de ruído reconhecido com o corte da função em zero e após a obtenção do mínimo da função. Entre parênteses encontram-se as porcentagens de ruído reconhecido. Na tabela 9, tem-se a quantidade de pontos eliminados com o corte em zero que não eram ruídos e com o corte no mínimo da função que não eram ruídos.

Exemplo 3 – Quantidade de Ruído Detectado pelo Método

	Corte em zero	Corte no mínimo
Distribuição 1	293 (97,68 %)	289 (96,33 %)
Distribuição 2	291 (97 %)	284 (94,67 %)

Tabela 8

Exemplo 3 – Quantidade de Pontos Relevantes Equivocadamente Eliminados

	Corte em zero	Corte no mínimo
Quantidade de Pontos	27	17

Tabela 9

Novamente a quantidade de ruído reconhecido é menor após o corte no mínimo, entretanto, houve um ganho de pontos relevantes.

Finalmente, na tabela 10, tem-se a porcentagem de classificação relativa a cada uma das etapas. Dentro-da-amostra, tivemos 2000 dados, 1000 de cada classe e fora-da-amostra, 4000 dados, 2000 de cada classe.

Exemplo 3 - Classificação

	Etapa 1 Sem seleção de dados	Etapa 2 Com seleção de dados (Corte em zero)	Etapa 3 Com seleção de dados (Corte no mínimo)
Dentro-da-amostra	65,2 %	99,50 %	98,51 %
Fora-da-amostra	62,68 %	97,95 %	97,85 %

Tabela 10

Em termos de classificação, os resultados das etapas 2 e 3 foram muito próximos, até porque, nesse caso, o mínimo foi muito próximo de zero. Porém, como se tratavam de distribuições muito ruidosas (30% dos pontos de cada uma eram ruídos), houve uma grande diferença em relação a classificar usando todos os dados.

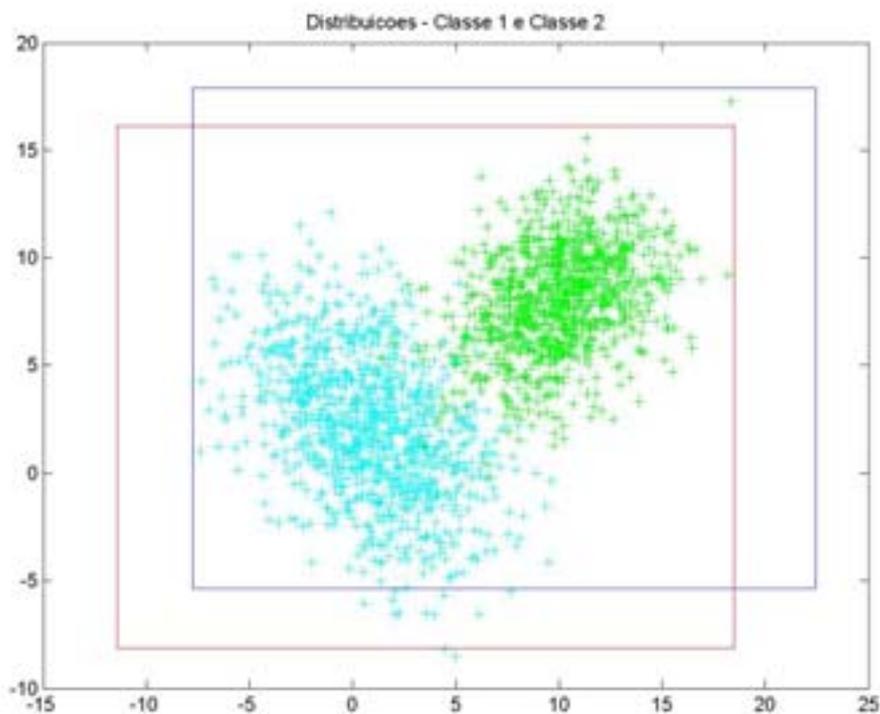


Figura 13 - Exemplo 3: Distribuições Originais

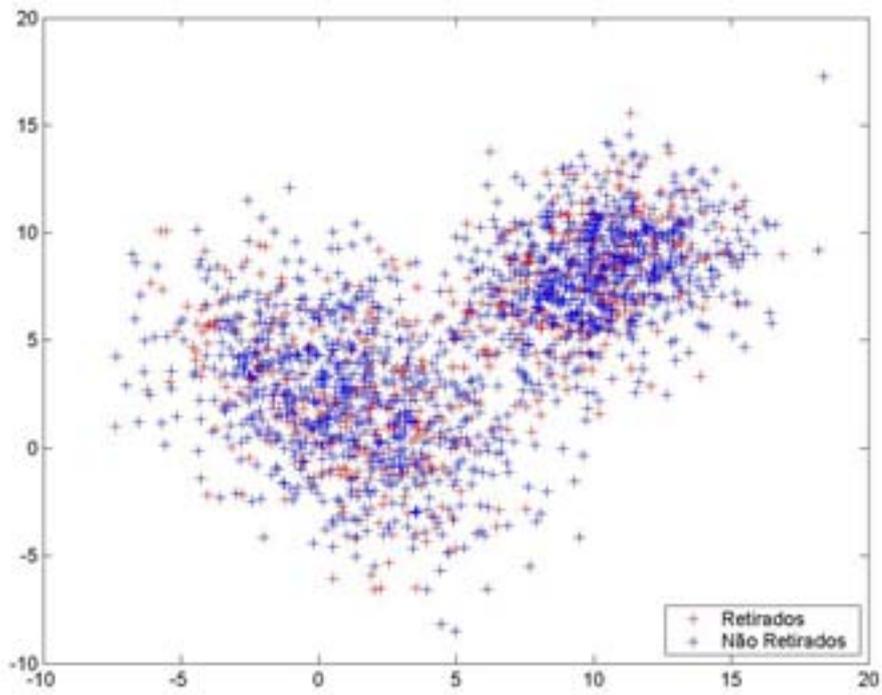


Figura 14 - Exemplo 3: Distribuições após corte em zero

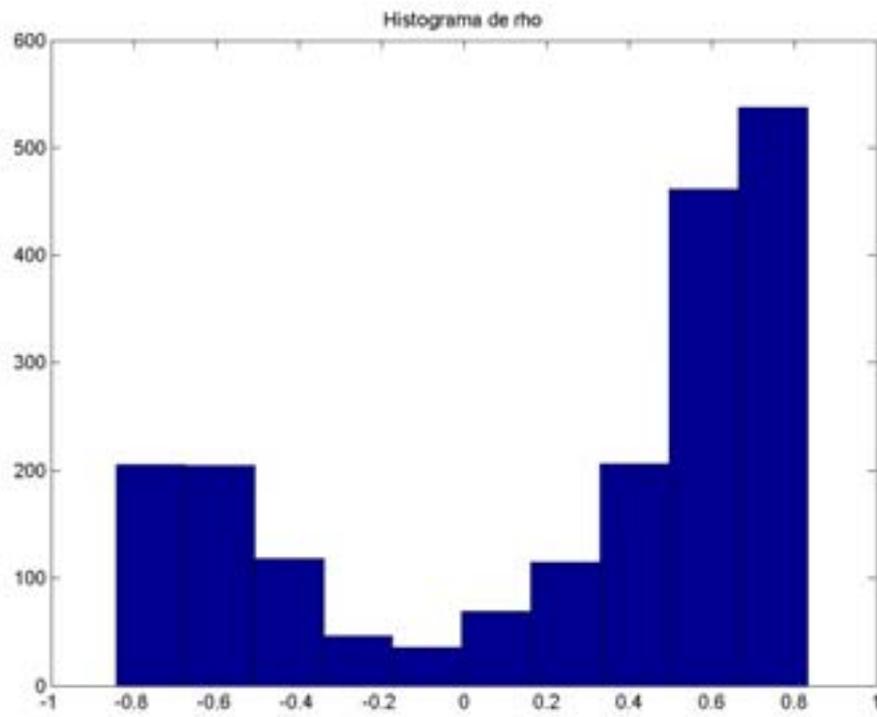


Figura 15 - Exemplo 3: Histograma de Rho

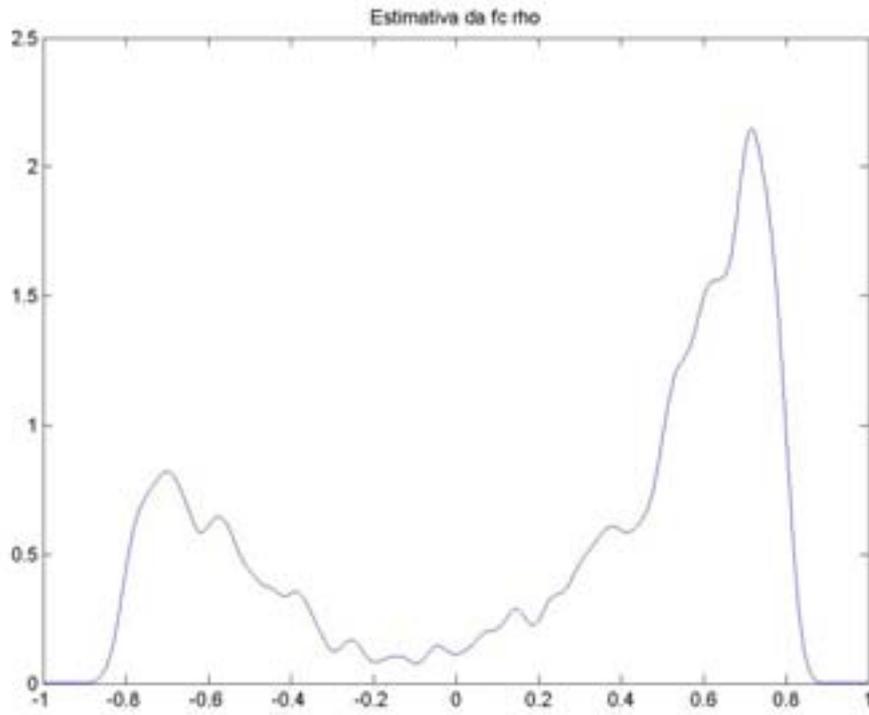


Figura 16 - Exemplo 3: Função Estimada pela Janela de Parzen

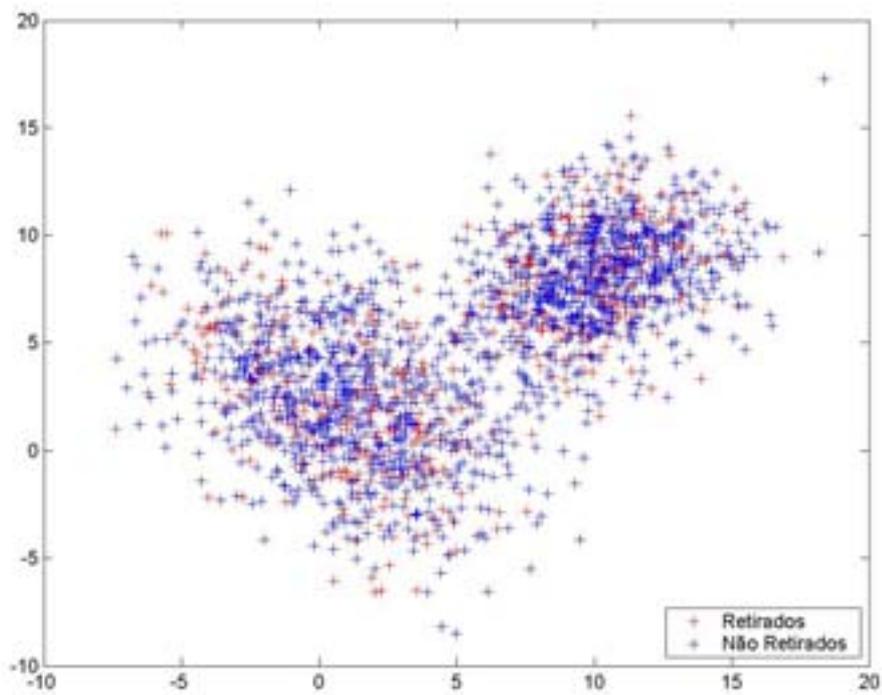


Figura 17 - Exemplo 3: Distribuições após corte no mínimo

Exemplo 4 – 2 Classes, 5 Dimensões:**Classe 1:**

Média: [8 3 6 10 9]

Matriz de Covariância: [1 0 0 0 0; 0 1 0 0 0; 0 0 1 0 0; 0 0 0 1 0; 0 0 0 0 1]

Instâncias: 1000

Classe 2:

Média: [15 9 12 19 16]

Matriz de Covariância: [9 7 3 2 1; -2 10 5 4 2; 3 2 8 7 6; 1 1 2 10 4; 1 1 2 3 10]

Instâncias: 1000

Foram gerados 30% de ruído em cada distribuição. As figuras 18 e 19 representam o histograma de ρ e a função estimada. O mínimo da função estimada foi encontrado no ponto -0,2366.

A tabela 11 mostra a quantidade de ruído reconhecido com o corte da função em zero e após a obtenção do mínimo da função. Entre parênteses encontram-se as porcentagens de ruído reconhecido. Na tabela 12, tem-se a quantidade de pontos eliminados com o corte em zero que não eram ruídos e com o corte no mínimo da função que não eram ruídos.

Exemplo 4 – Quantidade de Ruído Detectado pelo Método

	Corte em zero	Corte no mínimo
Distribuição 1	300 (100 %)	300 (100 %)
Distribuição 2	287 (95,67 %)	266 (88,67 %)

Tabela 11

Exemplo 4 – Quantidade de Pontos Relevantes Equivocadamente Eliminados

	Corte em zero	Corte no mínimo
Quantidade de Pontos	34	12

Tabela 12

Novamente a quantidade de ruído reconhecido é menor após o corte no mínimo, entretanto, houve um ganho de pontos relevantes.

Finalmente, na tabela 13, tem-se a porcentagem de classificação relativa a cada uma das etapas. Dentro-da-amostra, tivemos 2000 dados, 1000 de cada classe e fora-da-amostra, 4000 dados, 2000 de cada classe.

Exemplo 4 - Classificação

	Etapa 1 Sem seleção de dados	Etapa 2 Com seleção de dados (Corte em zero)	Etapa 3 Com seleção de dados (Corte no mínimo)
Dentro-da-amostra	59,9 %	98,91 %	95,99 %
Fora-da-amostra	58,73 %	97,1 %	95,5 %

Tabela 13

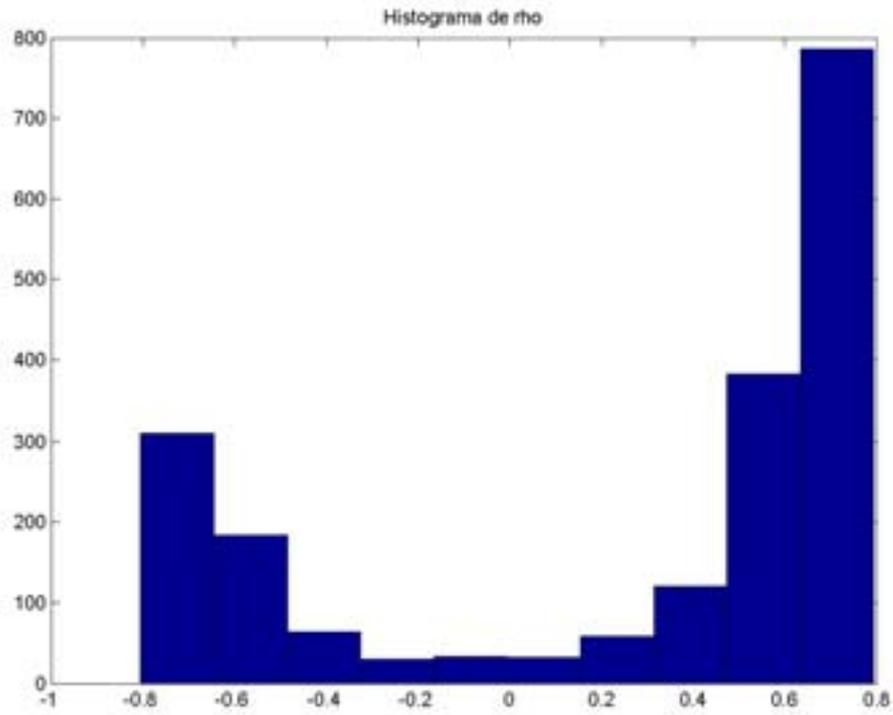


Figura 18 - Exemplo 4: Histograma de Rho

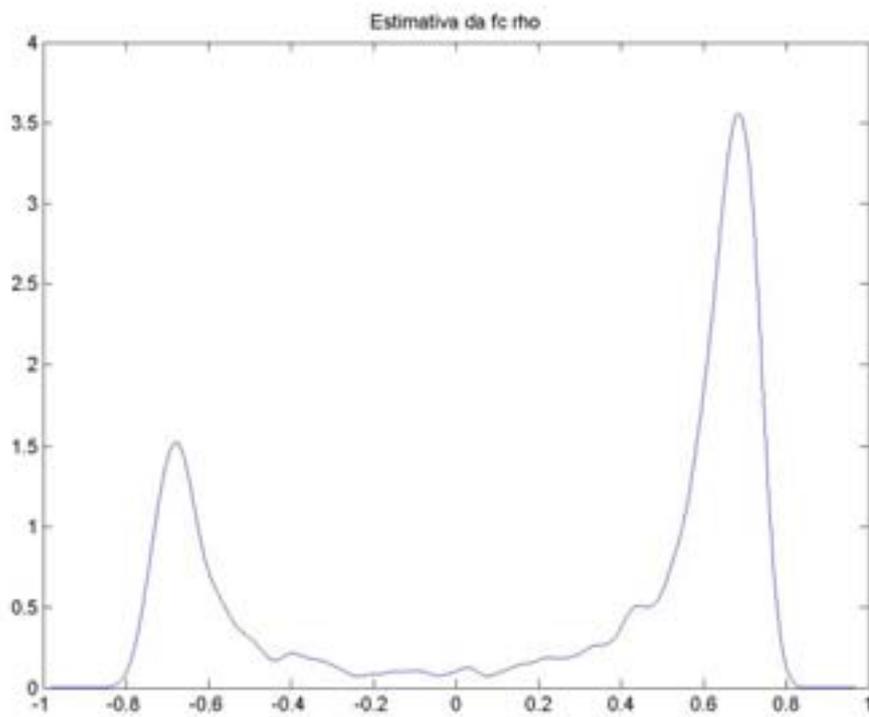


Figura 19 - Exemplo 4: Função Estimada pela Janela de Parzen

Por se tratarem novamente de distribuições muito ruidosas, a diferença de classificação para a etapa que não teve seleção de dados é muito grande. Em relação às duas propostas de seleção, nesse caso, a classificação foi melhor com corte em zero.

Exemplo 5 – 2 Classes, 10 Dimensões:

Classe 1:

Média: [1 1 1 1 1 1 1 1 1 1]

Matriz de Covariância: matriz identidade

Instâncias: 1000

Classe 2:

Média: [3 3 3 3 3 3 3 3 3 3]

Matriz de Covariância: matriz identidade

Instâncias: 1000

Foram gerados 30% de ruído em cada distribuição. As figuras 20 e 21 representam o histograma de ρ e a função estimada. O mínimo da função estimada foi encontrado no ponto -0,0315.

A tabela 14 mostra a quantidade de ruído reconhecido com o corte da função em zero e após a obtenção do mínimo da função. Entre parênteses encontram-se as porcentagens de ruído reconhecido. Na tabela 15, tem-se a quantidade de pontos eliminados com o corte em zero que não eram ruídos e com o corte no mínimo da função que não eram ruídos.

Exemplo 5 – Quantidade de Ruído Detectado pelo Método

	Corte em zero	Corte no mínimo
Distribuição 1	300 (100 %)	299 (99,67 %)
Distribuição 2	300 (100 %)	300 (100 %)

Tabela 14

Exemplo 5 – Quantidade de Pontos Relevantes Equivocadamente Eliminados

	Corte em zero	Corte no mínimo
Quantidade de Pontos	1	0

Tabela 15

Finalmente, na tabela 16, tem-se a porcentagem de classificação relativa a cada uma das etapas. Dentro-da-amostra, tivemos 2000 dados, 1000 de cada classe e fora-da-amostra, 4000 dados, 2000 de cada classe.

Exemplo 5 - Classificação

	Etapa 1 Sem seleção de dados	Etapa 2 Com seleção de dados (Corte em zero)	Etapa 3 Com seleção de dados (Corte no mínimo)
Dentro-da-amostra	60,4 %	99,93 %	99,86 %
Fora-da-amostra	57,63 %	99,95 %	99,95 %

Tabela 16

Por se tratarem novamente de distribuições muito ruidosas, a diferença de classificação para a etapa que não teve seleção de dados é muito grande. Houve um equilíbrio muito grande entre as duas etapas de seleção.

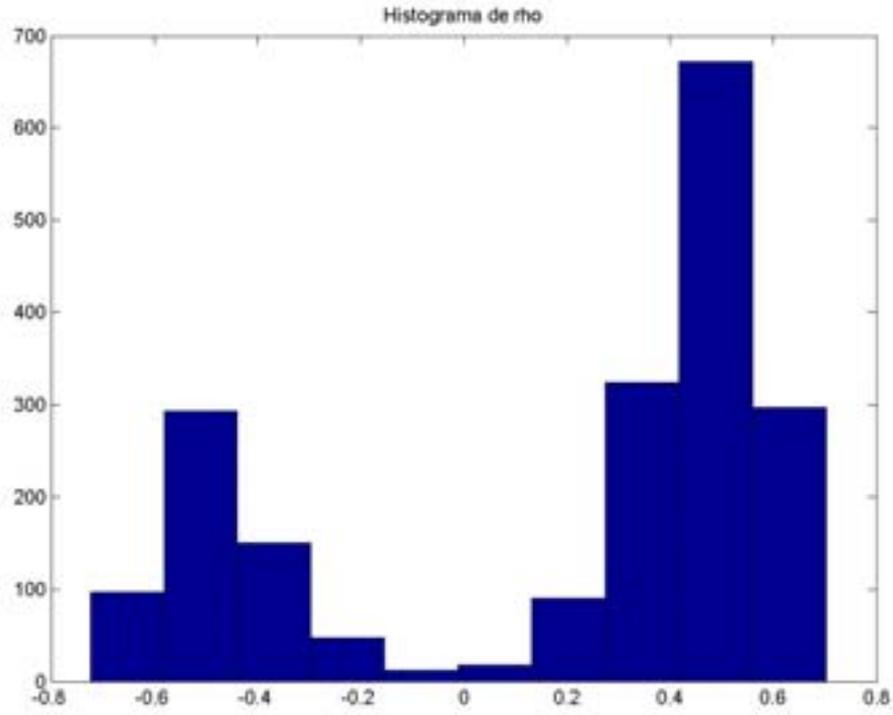


Figura 20 - Exemplo 5: Histograma de Rho

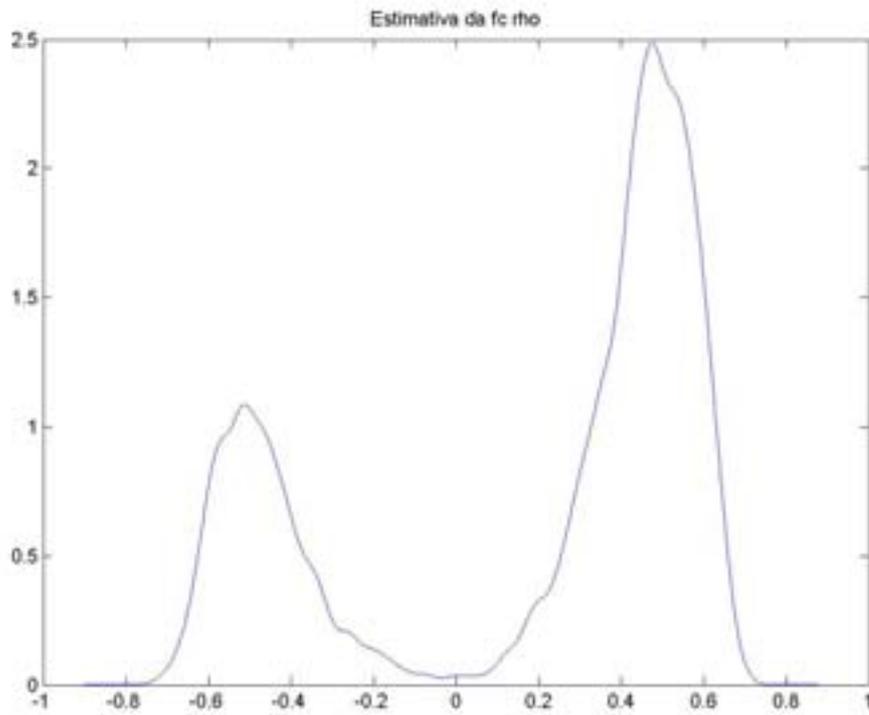


Figura 21 - Exemplo 5: Função Estimada pela Janela de Parzen

Exemplo 6 – Íris

Íris é um banco de dados clássico, um dos mais conhecidos na literatura de reconhecimento de padrões [6], sempre usado como benchmark. Criado por Fisher [24], o banco* contém três classes (Íris Setosa, Íris Versicolour e Iris Virginica) com cinquenta instâncias cada, e quatro atributos de entrada.

Esse banco foi escolhido por ser amplamente utilizado na literatura, permitindo comparações com outros métodos. Além disso, possui uma particularidade interessante: com poucos dados disponíveis para treinamento, são apenas 150 instâncias, permite ilustrar que a seleção de dados não é só interessante para bancos com grande quantidade de dados disponíveis para treinamento. Mesmo em bancos com poucos dados, deseja-se excluir padrões que possam ser ruídos, pois eles prejudicarão a classificação.

Por causa do problema da pequena quantidade de dados, durante o processo sem seleção de dados, as 150 instâncias foram utilizadas e o erro foi testado usando a técnica de leave-one-out [13]. As três possibilidades de corte foram testadas nesse experimento: com o corte no mínimo da função, que foi em 0,3020, foram selecionadas 112 amostras, com o corte em zero, foram selecionadas 134 amostras e, finalmente, com o corte em -0,2 foram selecionadas 147 amostras. Em todas elas, a classificação foi realizada também por leave-one-out.

Na tabela 17, estão as diferentes classificações para o problema. Nas figuras 22 e 23 vemos o histograma do vetor de correlação e a função estimada, respectivamente.

Note que as classificações em 98,51% e em 100% podem ser consideradas excelentes em comparação com a literatura, ainda mais considerando que apenas um protótipo por classe foi utilizado. A seguir, serão feitas algumas especulações a respeito desses resultados.

* <http://www.ics.uci.edu/~mlearn/MLSummary.html>

Exemplo 6 – Classificação do banco Íris (leave-one-out)

	Etapa 1 Sem seleção de dados (150 instâncias)	Etapa 2 Com seleção de dados (Corte no mínimo) (112 instâncias)	Etapa 3 Com seleção de dados (Corte em zero) (134 instâncias)	Etapa 4 Com seleção de dados (Corte em -0.2) (147 instâncias)
Acertos	92,67 %	100 %	98,51 %	97,32 %

Tabela 17

Ao eliminar grande porcentagem de dados como nas etapas 2 e 3, é muito provável que a classificação esteja muito boa em virtude da eliminação de pontos pertencentes às zonas de decisão. Se o banco fosse mais extenso, permitindo uma melhor análise de classificação, o sucesso das etapas 2 e 3 poderia não se repetir dado que, se parte dos pontos excluídos por elas pertencerem as fronteiras de decisão ou zonas de risco eles não estariam sendo levados em conta nas atualizações dos protótipos. Assim, possivelmente, as fronteiras de decisão não estariam bem delineadas, podendo levar a aplicação a ter um percentual de erro maior.

Desta forma, a separação efetiva de ruídos e pontos relevantes em problemas reais mostra-se, como era esperado, um problema extremamente difícil, pois obter uma classificação muito boa, como nesse caso, pode ser reflexo da escolha dos pontos. Além disso, uma dificuldade importante está na capacidade de efetivamente avaliar a separação entre pontos que sejam ruídos ou não, já que os dados são reais e não há como saber isso a priori, como era o caso nos problemas com dados sintéticos.

Por outro lado, surge uma interpretação muito interessante que abre a porta para futuros desenvolvimentos. Com a aplicação do método proposto, podemos assumir que os 112 pontos presentes na etapa 2 fornecem um subconjunto dos dados em que a classificação para eles é absolutamente certa. Assim sendo, futuramente podem-se investigar métodos que considerem múltiplos protótipos,

resolvendo primeiramente de forma certa a maioria dos pontos (112 neste caso) e utilizando posteriormente novos protótipos para 'dar conta' dos restantes 38 pontos.

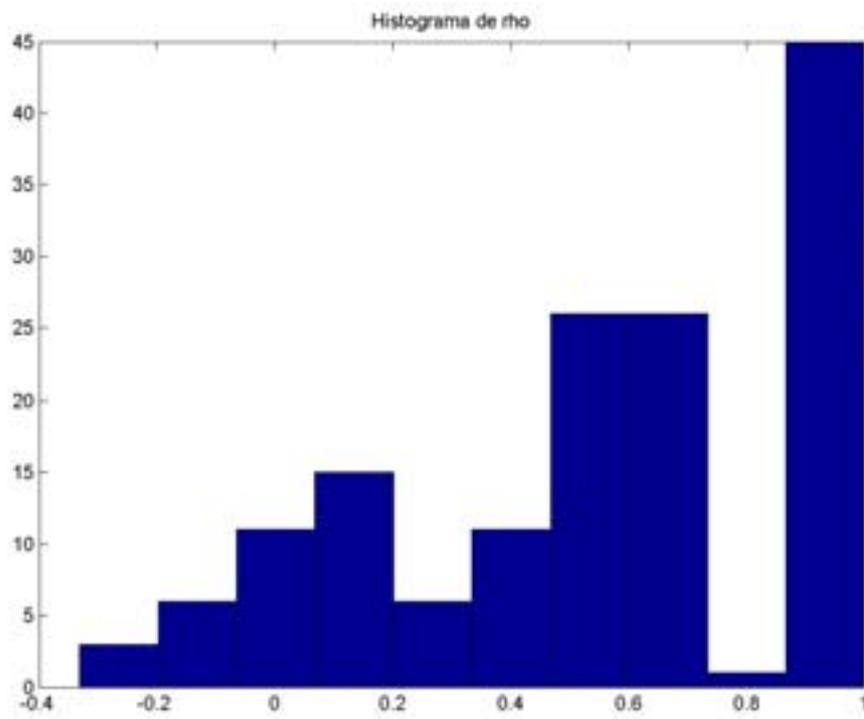


Figura 22 - Exemplo 6: Histograma de Rho

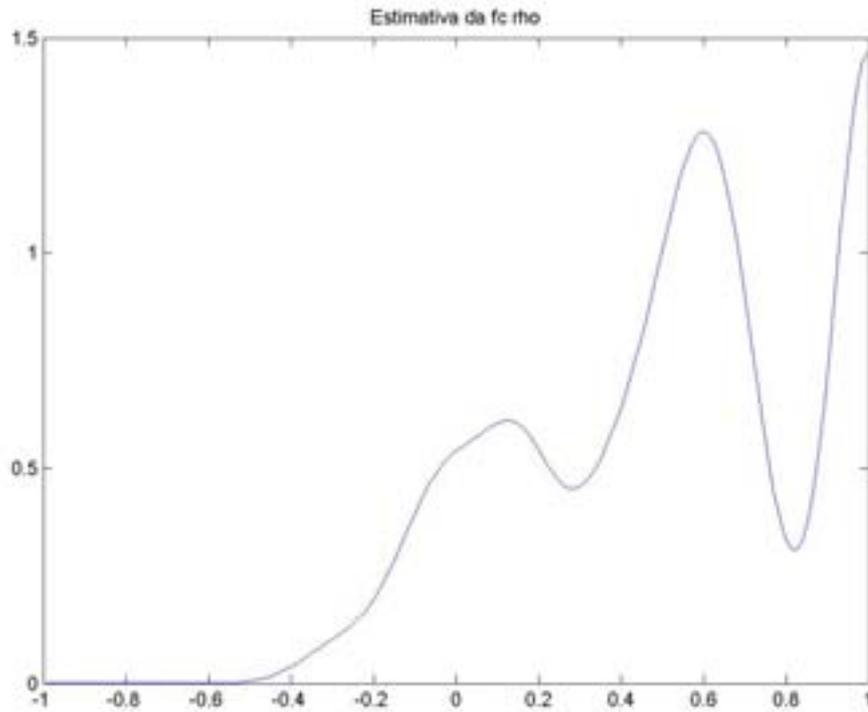


Figura 23 - Exemplo 6: Função Estimada pela Janela de Parzen

Exemplo 7 – Glass

O banco Glass^{*}, pertencente ao Serviço de Ciência Forense norte-americano contém seis classes diferentes e 214 instâncias, cada uma com nove atributos de entrada. Segundo [9], este banco é muito difícil de classificar, o que explica a baixa taxa de acertos. O treinamento foi realizado com todos os dados possíveis. Assim, durante o processo sem seleção de dados, as 214 instâncias foram utilizadas e foi aplicada a técnica leave-one-out [13]. Com o corte em zero, foram selecionadas 164 amostras para treinar o modelo, e com o mínimo em -0,2, o treinamento ficou com 197 amostras. Em ambos os casos, foram aplicados leave-one-out [13]. Os resultados estão na tabela 18.

* <http://www.ics.uci.edu/~mllearn/MLSummary.html>

Exemplo 7 – Classificação do banco Glass (leave-one-out)

	Etapa 1 Sem seleção de dados (214 instâncias)	Etapa 2 Com seleção de dados (Corte em zero) (164 instâncias)	Etapa 2 Com seleção de dados (Corte em -0.2) (197 instâncias)
Acertos	47,20 %	56,10 %	52,28 %

Tabela 18

Nas figuras 24 e 25 vemos o histograma do vetor de correlação e a função estimada, respectivamente.

Note que, apesar da dificuldade de classificação, o ganho em relação a utilizar 214 instâncias é grande, chegando a quase 9 % ao utilizar 164 dados. Da mesma forma que no banco Íris, não é provável que esses 50 pontos sejam ruído.

Assim, a mesma análise feita no exemplo anterior cabe neste problema. O fato é que, em problemas com dados reais, fica a dúvida se o ponto é um ruído ou relevante. Por isso a opção por um corte mais conservador pode ser, em um primeiro momento, interessante, mesmo que a classificação não seja a melhor entre os testes realizados.

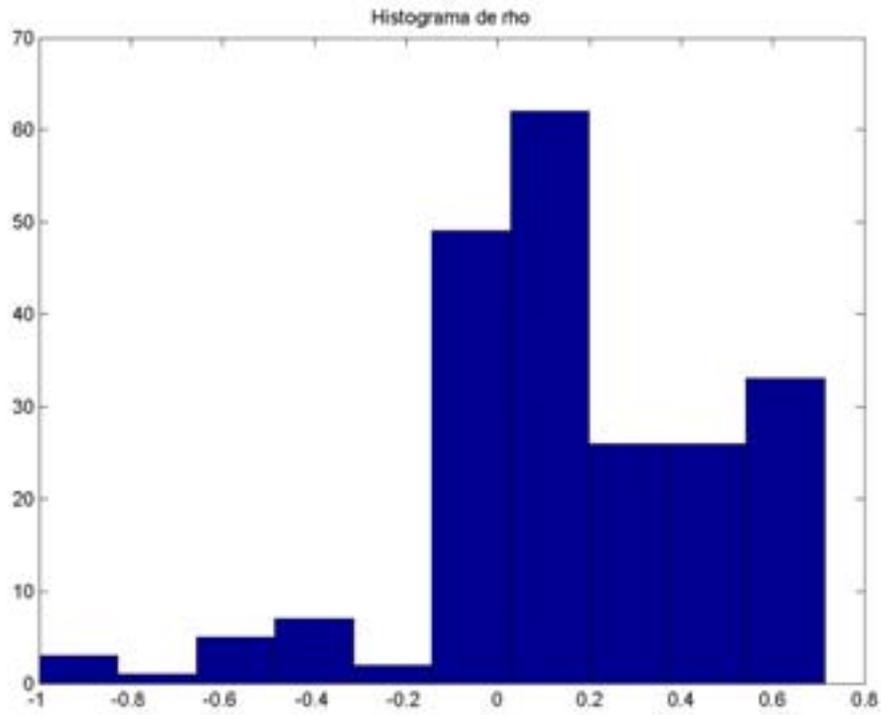


Figura 24 - Exemplo 7: Histograma de Rho

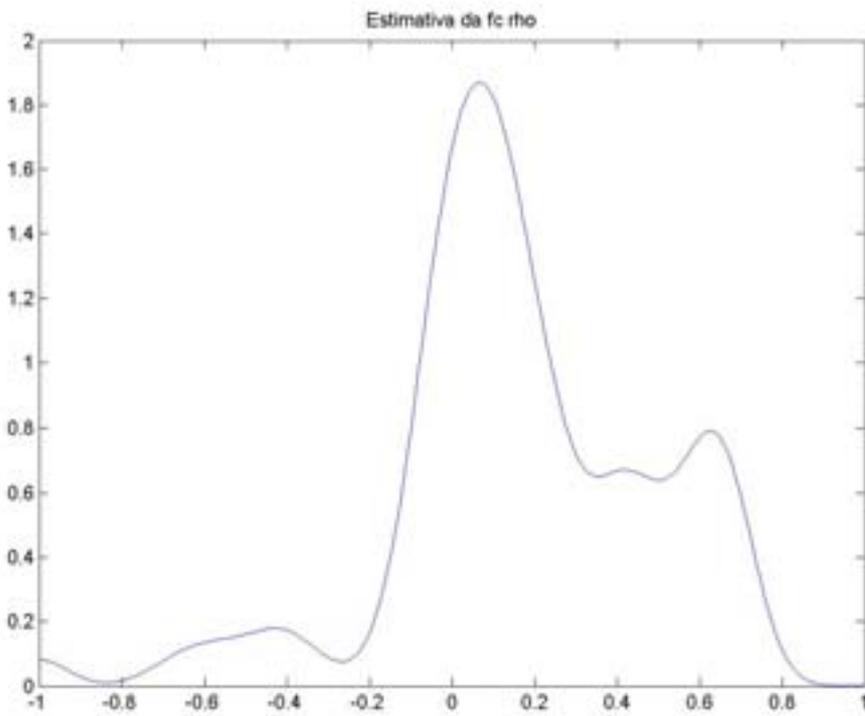


Figura 25 - Exemplo 7: Função Estimada pela Janela de Parzen