UNIVERSIDADE FEDERAL DE MINAS GERAIS

ELT135 - Introdução ao Reconhecimento de Padrões

Exercício 11 - 29/09/2019

KDE

Hernane Braga Pereira - 2014112627

1. Introdução

Este relatório tem como objetivo demonstrar o uso da técnica KDE, Kernel Density Estimation, aplicada à classificação de problemas não linearmente separáveis.

2. Problema espiral

Para este exercício foi utilizado o problema *Espiral* com 1000 amostras, que foi gerado usando o pacote *mlbench* do R, e foi pedido que o problema fosse classificado. Como as classes 1 (preta) e classe 2 (vermelha) não são linearmente separáveis, utilizou-se o método KDE para se obter uma gaussiana em cada ponto, e então foi utilizado o classificador bayesiano para separação de classes.

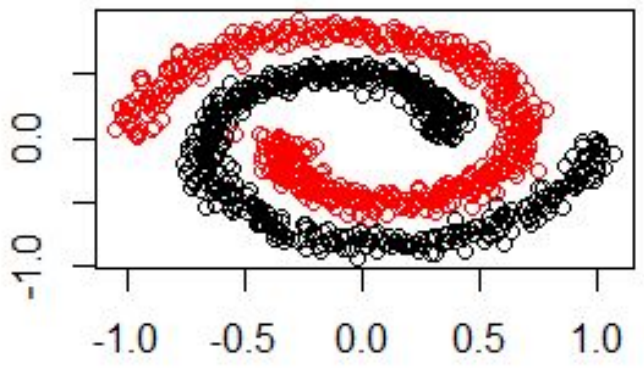

Figura 1. Problema *Espiral* utilizado

Os dados foram separados em 90% para treino e 10% para teste. Para gerar as estimativas através do KDE, foi utilizada a função de densidade normal como função de kernel. Apesar deste ser um modelo não paramétrico, é necessário definir um valor de abertura **h** na utilização do método, como visto na equação 1.

$$p(\mathbf{x}_i) = \frac{1}{N(\sqrt{2\pi}h)^n} \sum_{j=1}^{N} e^{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2h^2}}$$

Equação 1. KDE multivalorado

Onde N o número total de amostras e h a abertura da gaussiana. Após a execução do modelo para um h = 0.25, foi realizada a validação cruzada dos dados e os resultados são demonstrados nos quadros 1 e 2.

Nº Teste	Taxa de acurácia para a Classe 1	Taxa de acurácia para a Classe 2	Taxa de acurácia Total
1	100%	98%	99%
2	100%	98%	99%
3	100%	100%	100%
4	100%	94%	97%
5	100%	100%	100%
6	100%	100%	100%
7	100%	98%	99%

8	100%	100%	100%
9	98%	100%	99%
10	100%	100%	100%
Média Geral	**100%**	**99%**	**99%**

Quadro 1. Comparação de acurácia para h = 0.25

Nº Teste	Desvio padrão para a Classe 1	Desvio padrão para a Classe 2	Desvio padrão Total
1	0	0,141421	0,502418
2	0	0,141421	0,502418
3	0	0	0,502519
4	0	0,239898	0,501614
5	0	0	0,502519
6	0	0	0,502519
7	0	0,141421	0,502418
8	0	0	0,502519
9	0,141421	0	0,502418
10	0	0	0,502518
Média Geral	0,014142	0,066416	0,502388

Quadro 3. Comparação de desvio padrão para h = 0.25

Ao analisar os resultados, conclui-se que o método da mistura de gaussianas foi efetivo em separar as duas classes do problema *Espiral*, utilizando uma abertura h = 0.25. Podemos confirmar este fato através do gráfico do espaço da verossimilhança para os dados de teste na figura 2, nele é possível traçar uma reta que separa as classes 1 e 2.

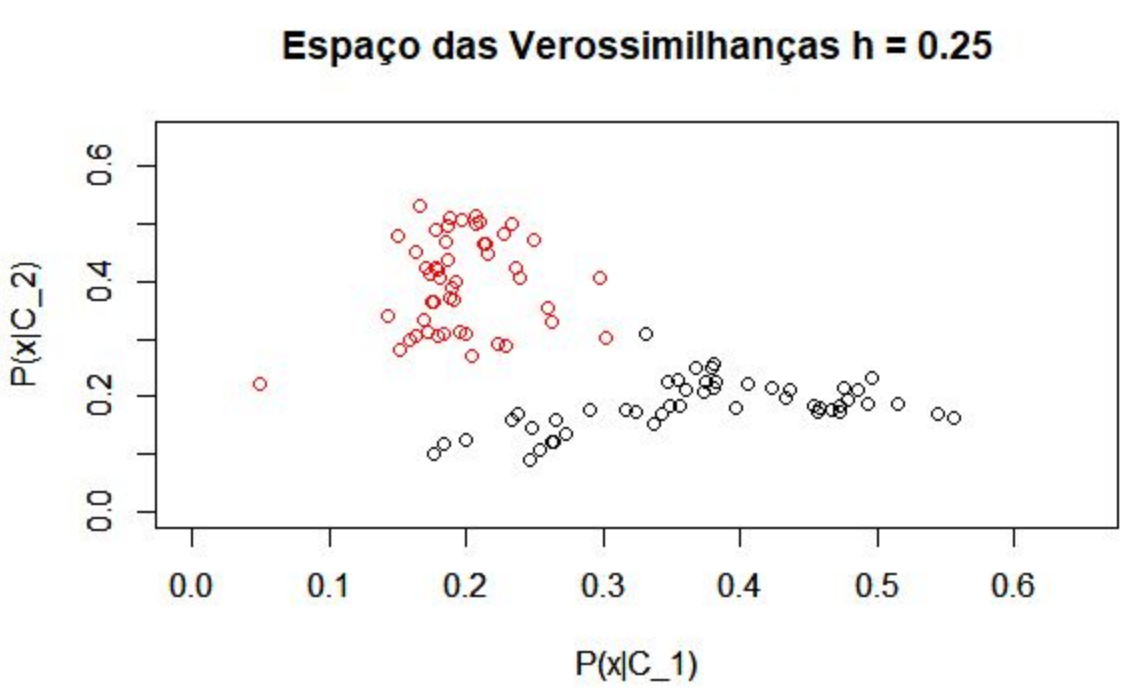

Figura 3. Espaço de Verossimilhanças dos dados de teste para h = 0.25

A superfície de contorno do problema, superfície de separação e superfície de densidade de probabilidade podem ser vistas nas figuras 4, 5 e 6 respectivamente.

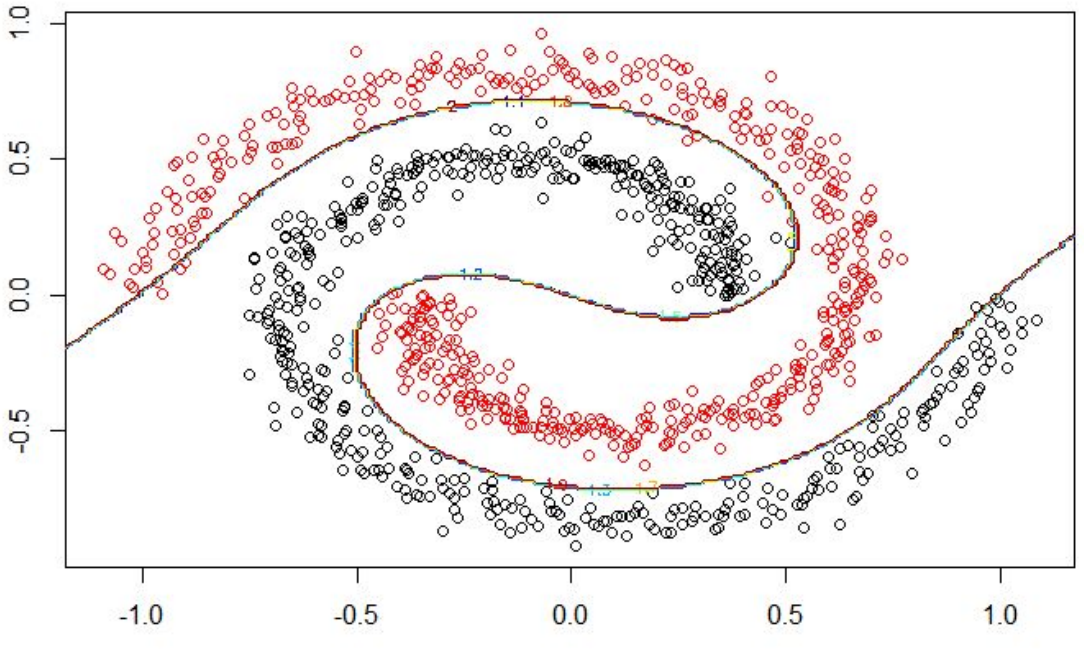

Figura 4. Superfície de contorno do problema

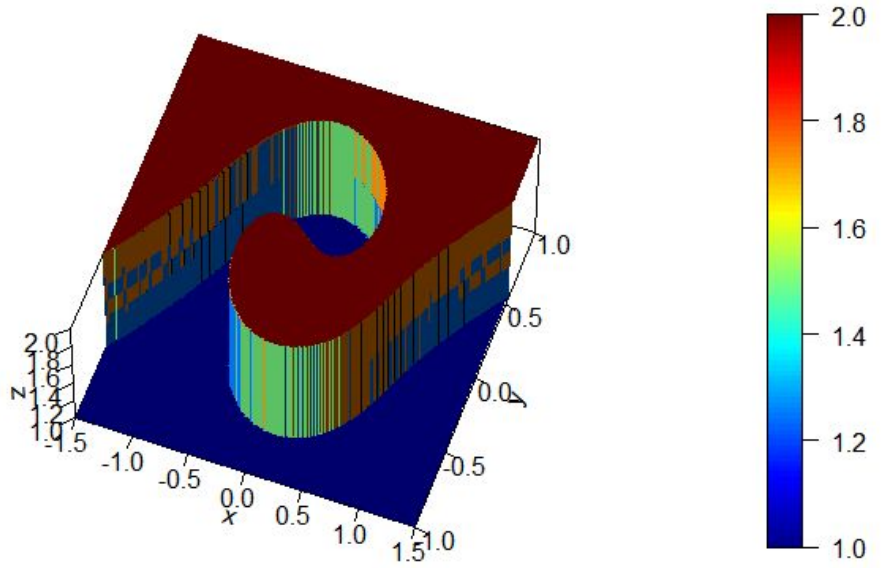

Figura 5. Superfície de separação do problema

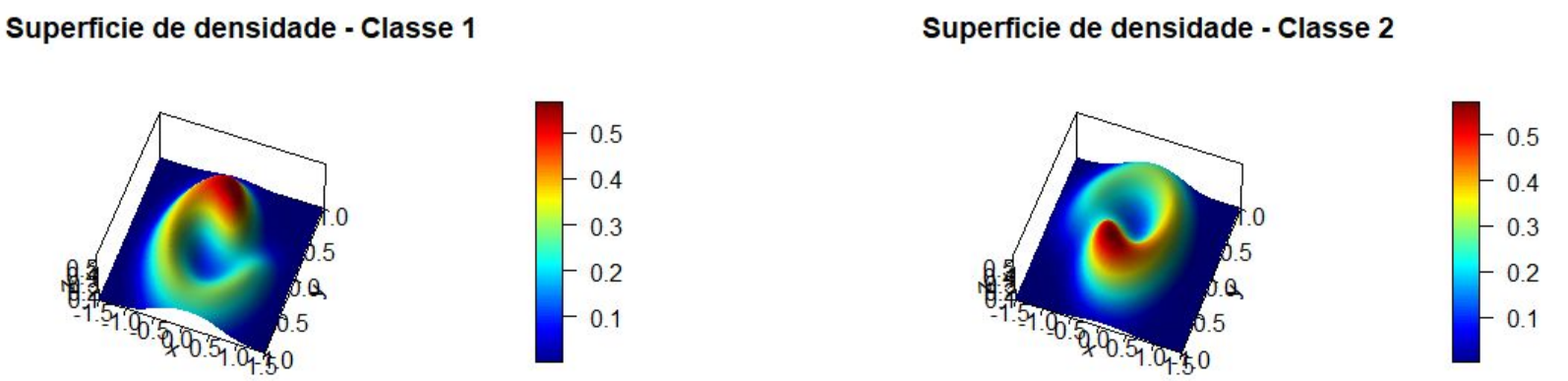

Figura 6. Superfície de densidade de probabilidade encontrada para cada classe

Para demonstrar a importância de uma bola escolha do parâmetro h, foi realizada a reclassificação para h = 0.8. O resultado de seu gráfico do espaço de verossimilhanças está presente na figura 7.

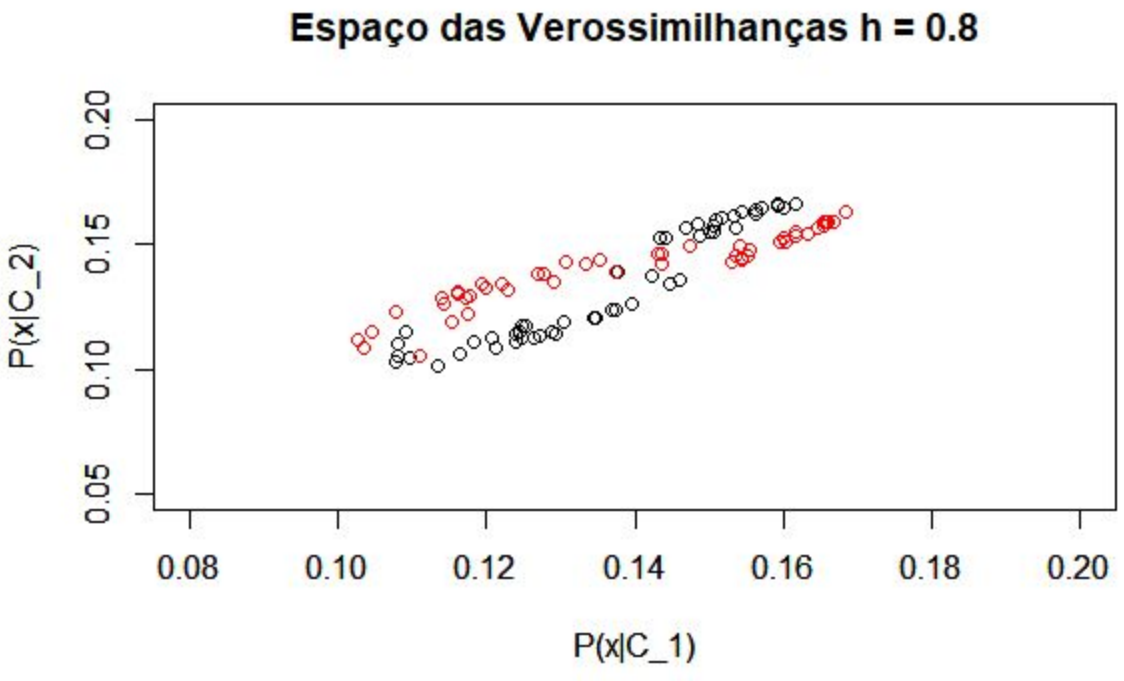

Figura 7. Espaço de Verossimilhanças dos dados de teste para h = 0.8

Analisando o gráfico da figura 7, nota-se que não é possível separar as duas classes utilizando apenas de uma reta, e que portanto, este valor de h não foi uma boa escolha na classificação do problema.

Ao realizar esta prática conclui-se que o método não paramétrico KDE, possui uma boa acurácia na classificação de problemas não lineares, porém apresenta o problema na definição do parâmetro da função de kernel, que neste caso é a função de densidade normal, que possui parâmetro h.

3. Referências

[1] KDE, Notas de aula, setembro de 2019.