



KDE

Exercício

No exercício proposto, foram gerados dados em formato de espiral pertencentes às duas classes distintas, conforme mostra a figura 1, obtidos a partir da biblioteca *mlbench* do R.

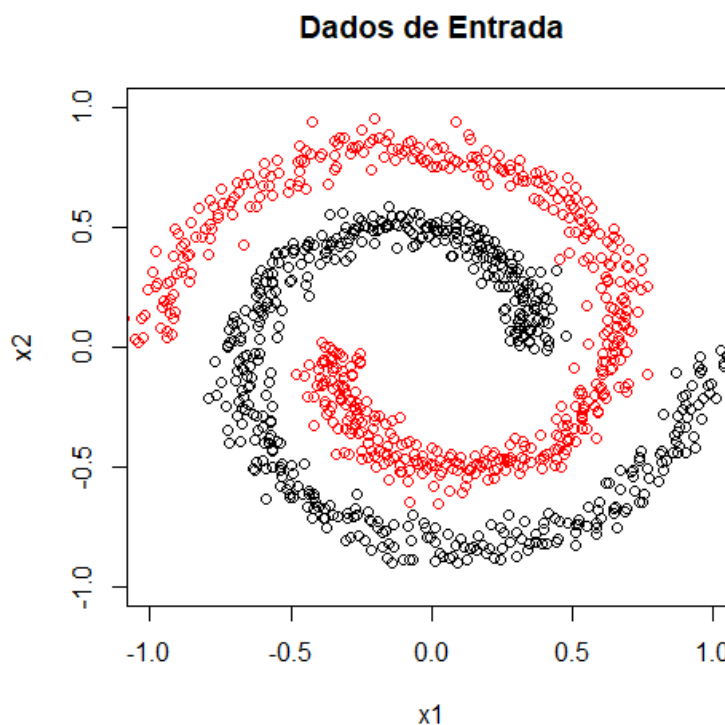


Figura 1: Dados de entrada para o problema.

Os dados foram divididos em 10 *folds* para a técnica de validação cruzada, sendo que a cada iteração, um grupo funcionava como dado de teste e os demais como dados de entrada.

Para cada *fold* criado, os dados de entrada foram utilizados para o treinamento de uma rede KDE. A superfície de separação de um dos folds é mostrada na figura 2 em 2D e na figura 3 em 3D.

O valor do parâmetro h foi definido usando a equação 1, retirada do livro "Density Estimation for Statistics and Data Analysis", de B.W. Silverman, 1ª Edição, 1986.

$$h = 0,9 \times \min\left(\frac{Q_3 - Q_1}{1,349}, \sigma\right) \times N^{-0.2} \quad (1)$$

Em que Q_3 é o terceiro quartil, Q_1 o primeiro quartil, σ é o desvio padrão e N é o número de amostras da classe utilizada.

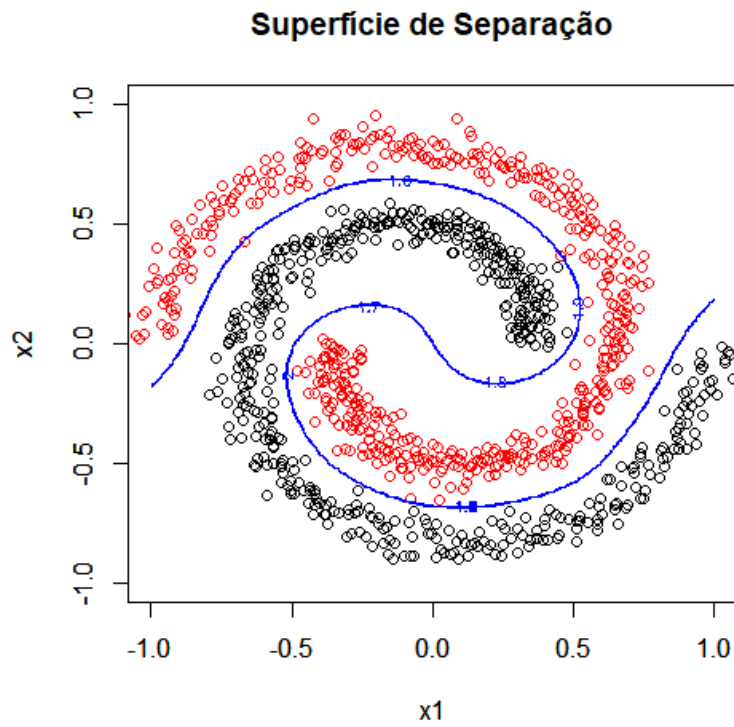


Figura 2: Superfície de Separação em 2D.

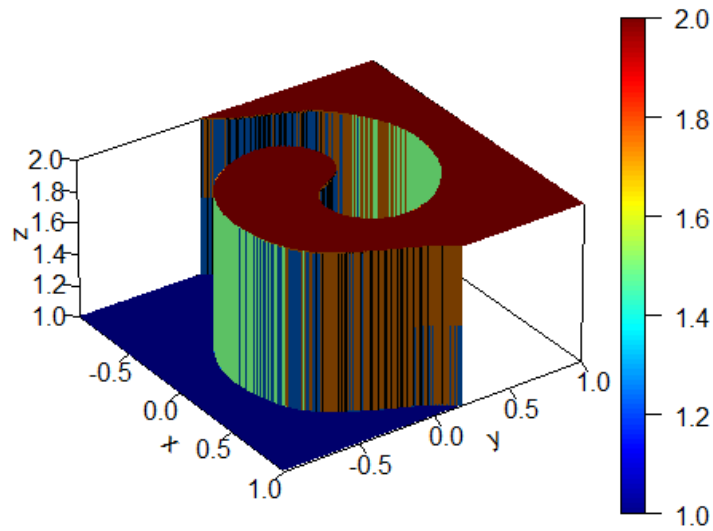


Figura 3: Superfície de Separação em 3D.

A acurácia média dos 10 *folds* foi de 99,8%, enquanto o desvio padrão foi de $\sigma = 0,42\%$, mostrando a eficiência do algoritmo do *KDE* em separar os dados de forma linearmente classificável. Essa separação

para um dos *folds* cuja acurácia foi de 100% é mostrada na figura 4, além da reta de separação.

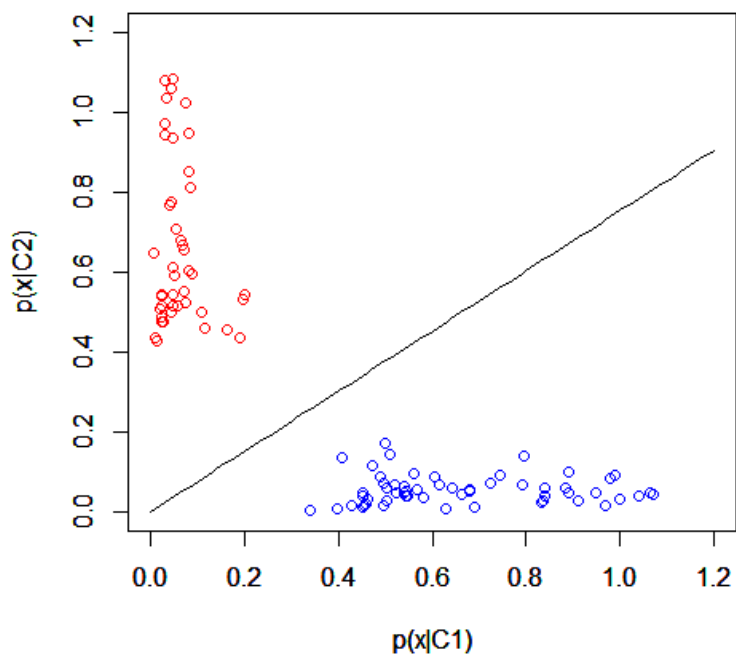


Figura 4: Superfície de separação do espaço de verossimilhanças pra um dos *folds*.