

Reconhecimento de Padrões

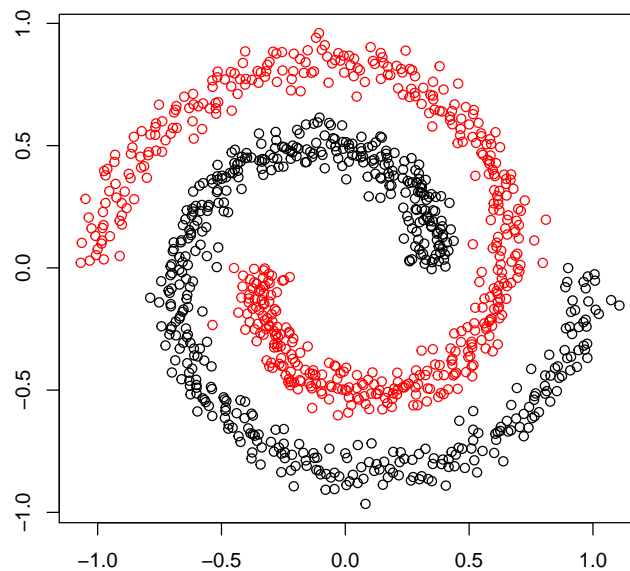
Exercício prático de Estimação de Densidades utilizando Misturas de Gaussianas

Prof. Frederico Coelho

4 de setembro de 2019

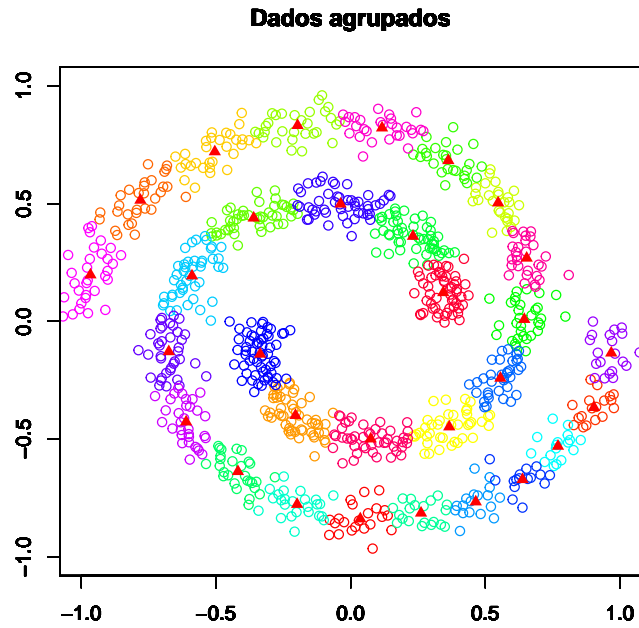
1 Exercício 1

Neste exercício o Aluno deverá usar a misturas de gaussianas para alimentar um classificador de Bayes na resolução de um problema de classificação. Primeiramente o(a) aluno(a) realizará a estimação de densidades a partir do método de misturas de Gaussianas. Neste exercício será utilizado o conjunto de dados sintéticos do problema espiral gerado a partir do pacote *mlbench* como mostrado

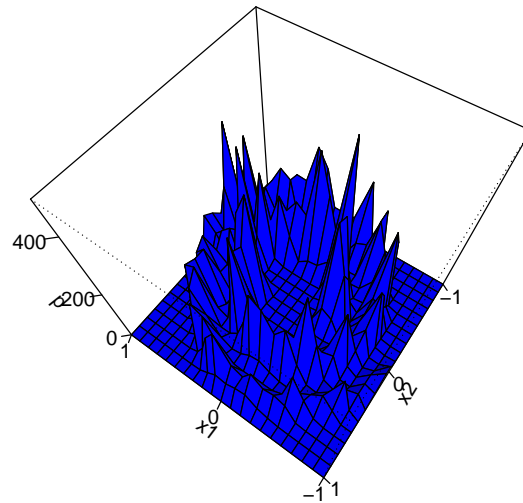


a seguir.

Para a definição dos centros de cada uma destas, o(a) aluno(a) deverá utilizar o algoritmo K-means, anteriormente implementado. A figura a seguir mostra os clusters e os centros estimados pelo K-means para $k=30$.



O gráfico a seguir mostra o resultado esperado da estimação da densidade dos dados anteriores quando se utiliza 30 gaussianas como componentes da mistura.



O aluno deverá variar o número de gaussianas utilizadas para estimar as densidades e comparar os resultados para definir o melhor número de gaussianas a serem utilizadas.

O aluno deverá seguir os seguintes passos:

1. Carregar a base de dados.
2. Separar os dados em treinamento e teste utilizando a técnica de validação cruzada com 10 *folds*.
3. Treinar o modelo de misturas de gaussianas, determinando inclusive o melhor número de gaussianas a se utilizar para o problema.
4. Resolver o problema de classificação utilizando o classificador de Bayes para cada um dos folds.
5. Gerar a tabela de acurácias de cada fold.
6. Calcular a acurácia média e o desvio padrão das soluções encontradas.
7. Plotar a superfície de separação para o treinamento com a melhor acurácia.

Apresentar no relatório a descrição de tudo o que foi feito em todos os passos acima, os resultados e suas análises e conclusões.

2 Exercício 2

Neste exercício o aluno deverá treinar um classificador de bayes aplicado a um problema real. O problema a ser considerado é o do câncer de mama (Breast Cancer) que está no pacote *mlbench*. Utilize os comando abaixo para acessar os dados:

```
> library("mlbench")  
> #pega os dados da package mlbench  
> data("BreastCancer")  
> dados <- BreastCancer
```

Execute tudo o que se pede no exercício 1, exceto plotar a superfície de separação. A descrição do banco de dados pode ser obtida na documentação do pacote.