

## Clustering de problemas não linearmente separáveis utilizando o método *K-means*

Hernane Braga Pereira - 2014112627

### 1. Introdução

Este relatório tem como objetivo exemplificar o uso da técnica de identificação de grupos (*clustering*) utilizando o algoritmo K-means em um conjunto de dados que possui classes não linearmente separáveis.

### 2. Formação de grupos usando K-means

Para este exercício foi utilizado o problema *Espiral* com 1000 amostras, que foi gerado usando o pacote *mlbench* do R, onde foi pedido que fosse encontrado um número mínimo de agrupamentos destes dados, para que em cada agrupamento houvesse apenas dados da classe 1 (vermelha) ou classe 2 (preta).

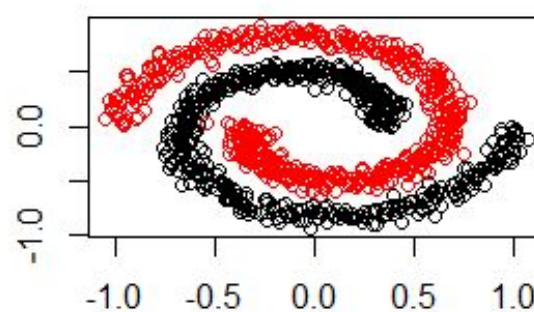


Figura 1. Problema *Espiral* utilizado

Para resolver o problema, foi utilizado o algoritmo *K-means* com 5, 10 e 20 agrupamentos, onde houve sucesso apenas no último caso. Após encontrado o número de grupos, foi plotado um gráfico demonstrando a função de densidade (PDF) de cada um dos 20 clusters definidos.

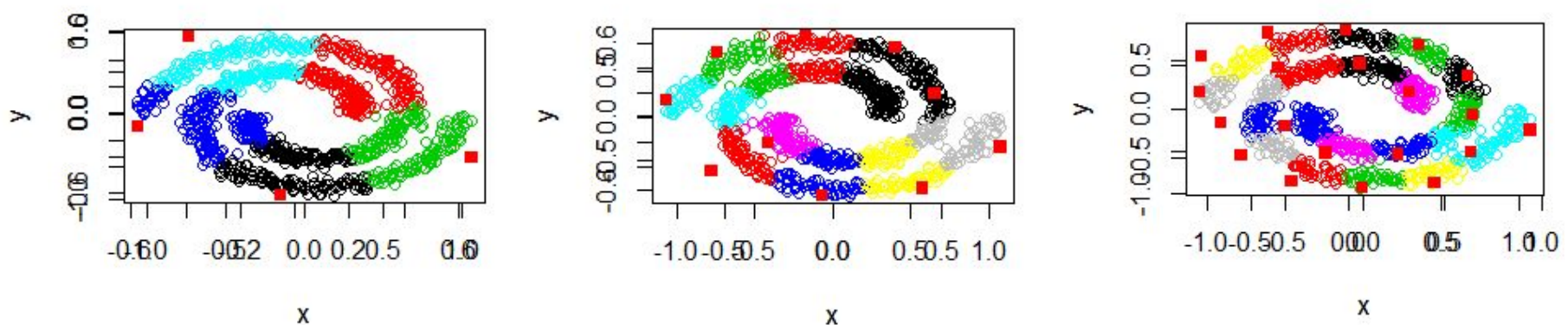


Figura 2. Clustering usando 5, 10 e 20 agrupamentos. Da esquerda para à direita

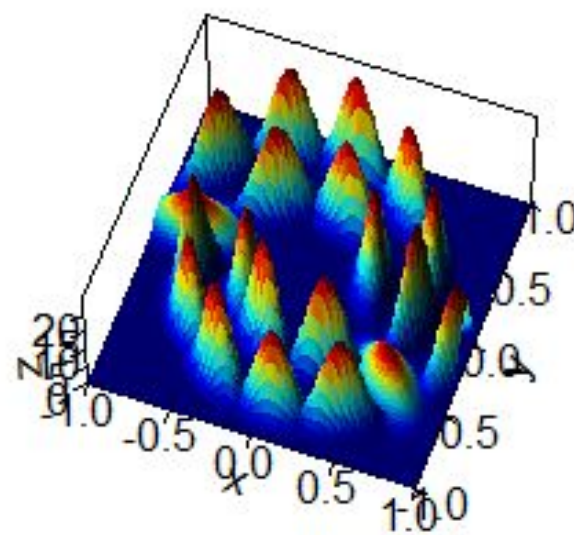


Figura 3. PDF de cada um dos 20 clusters formados

Ao analisar os resultados, conclui-se que o método k-means foi efetivo em formar os grupos para separar as duas classes do problema *Espiral*, porém nota-se que há uma dificuldade em se determinar qual é o número mínimo de agrupamentos necessários para a solução do problema. Atrelado à esta dificuldade, existe o problema da apuração do resultado encontrado. Neste exemplo foi utilizada a inspeção visual, para decidir se a solução encontrada era válida, porém em problemas multivariáveis não é possível utilizar este mesmo método para a validação.

### 3. Referências

[1] Clustering, Notas de aula, agosto de 2019.