

## Algoritmo de agrupamento (*clustering*) 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.

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

Para este exercício foram geradas 100 amostras de cada uma das quatro classes de distribuições normais e depois estes dados foram separados em 2, 4 e 8 grupos.

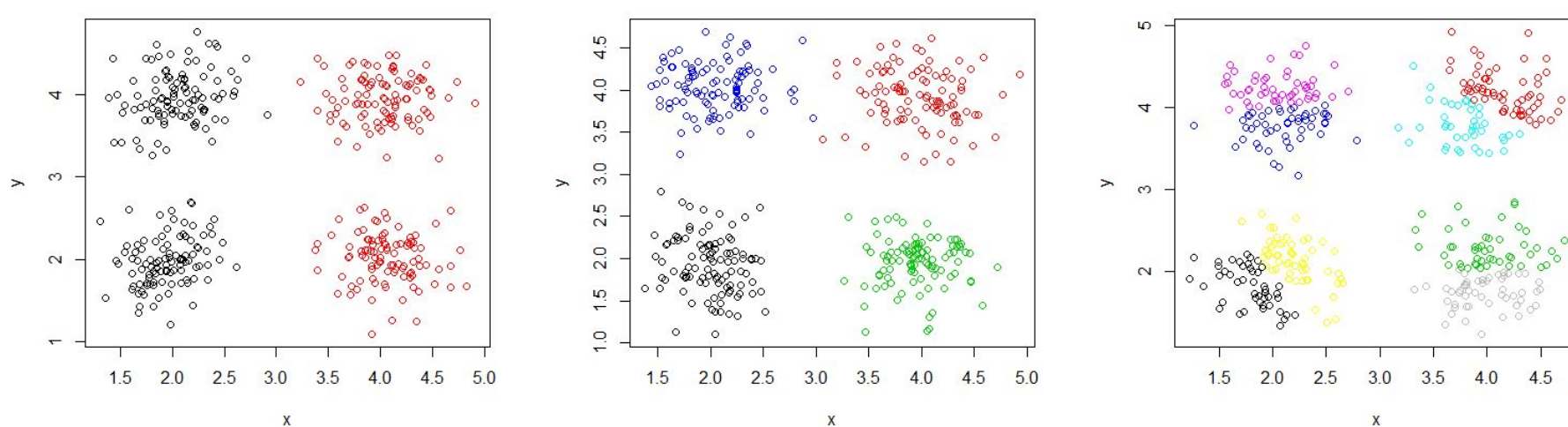


Figura 1. Clusterização em 2, 4 e 8 grupos para dados com desvio padrão de 0.3

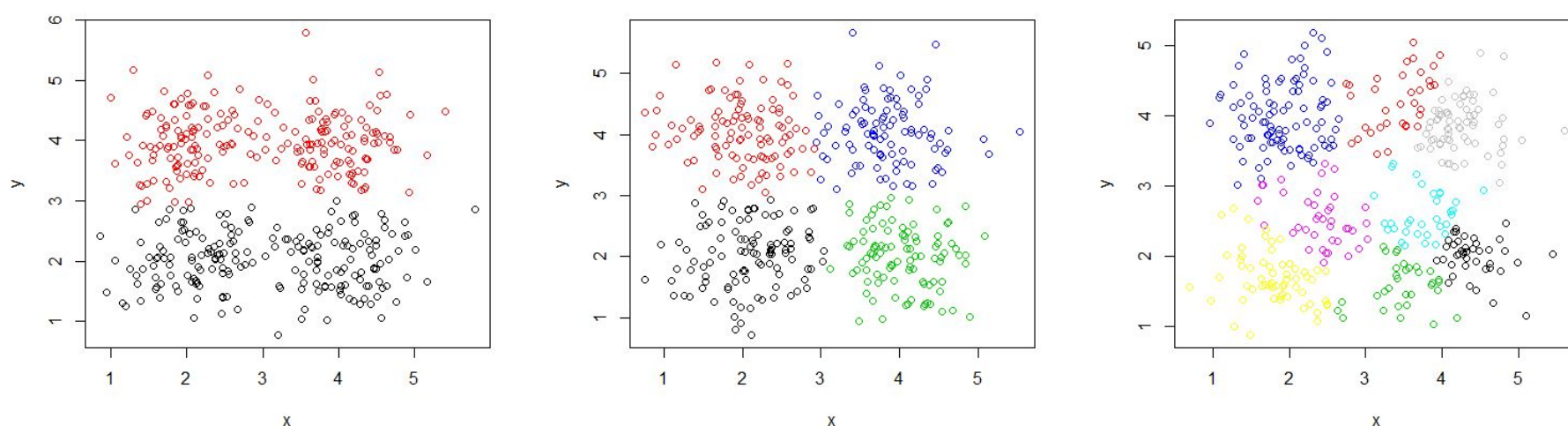
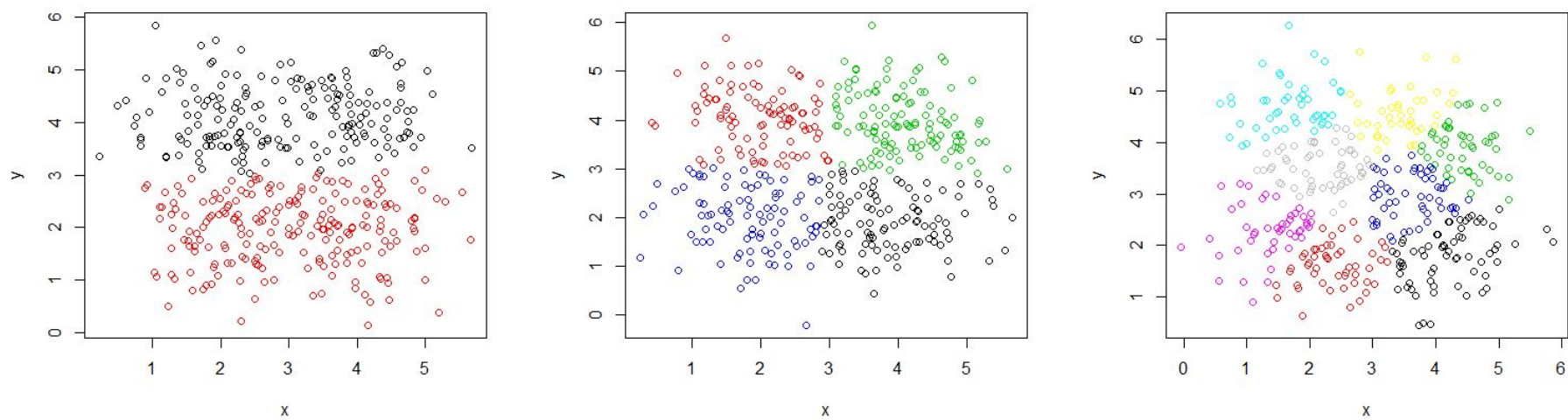


Figura 2. Clusterização em 2, 4 e 8 grupos para dados com desvio padrão de 0.5



**Figura 3. Clusterização em 2, 4 e 8 grupos para dados com desvio padrão de 0.7**

Ao analisar os resultados, conclui-se que o método k-means foi efetivo em formar os grupos para dados com distribuição normal, entretanto verifica-se que um dos desafios dos métodos de *clustering* é estimar qual o número adequado de grupos a serem criados. Esta dificuldade é exemplificada na Figura 1 quando tenta-se criar 8 grupos, porém os dados são melhor agrupados em 4 cluster, ao invés de 8.

### 3. Referências

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