

Exercício k-means

Prof. Frederico Coelho

August 27, 2019

EXERCÍCIO 1

Os alunos deverão implementar o algoritmo de agrupamentos *k-means* conforme explicações dada em sala de aula. Basicamente o algoritmo do k-médias é:

1. Definir número k de *clusters*;
2. Escolhe aleatoriamente k pontos de treinamento para serem os centros dos k clusters;
3. Calcula a distância de todos os demais pontos de treinamento ao centro dos clusters e os aloca aos clusters mais próximos;
4. Calcula-se o novo centro dos clusters através das médias dos seus pontos;
5. Volta no passo 3 até que o novo cluster não difira do anterior.

EXERCÍCIO 2

Uma vez implementado os alunos deverão observar e entender o comportamento do mesmo. Para isso os alunos deverão criar um conjunto de dados com quatro distribuições Gaussianas bidimensionais e utilizar o algoritmo *k-means* para diferentes valores de k (número de agrupamentos). Em seguida os alunos deverão variar também a sobreposição das distribuições (variando o desvio padrão) e novamente utilizar o algoritmo *k-means* para diferentes valores de k .

Dica para implementação:

1. Importar a função *k-means* criada
2. Criar as quatro gaussianas bidimensionais
3. Concatenar as quatro gaussianas
4. Utilizar a função *k-means* para fazer o agrupamento
5. Plotar os grupos com cores diferentes de acordo com o resultado do agrupamento
6. Criar um loop para fazer a analisar os passos de 2-5 com $k = \{2, 4, 8\}$
7. Criar um loop para fazer os passos 2-6 variando o desvio padrão de forma que $sd = \{0.3, 0.5, 0.7\}$

FORMA DE ENTREGA

Relatório em .pdf, descrevendo o que foi feito, mostrando os gráficos e as informações pedidas e explicando os resultados obtidos.

DICAS PARA O R

1. Criando as gaussianas bidimensionais

```
s1<-0.3
s2<-0.3
s3<-0.3
s4<-0.3
nc<-100
xc1<-matrix(rnorm(nc*2),ncol=2)*s1 +
t(matrix(c(2,2),nrow=2,ncol=nc))
xc2<-matrix(rnorm(nc*2),ncol=2)*s2 +
t(matrix(c(4,4),nrow=2,ncol=nc))
xc3<-matrix(rnorm(nc*2),ncol=2)*s3 +
t(matrix(c(2,4),nrow=2,ncol=nc))
xc4<-matrix(rnorm(nc*2),ncol=2)*s4 +
t(matrix(c(4,2),nrow=2,ncol=nc))
```

2. Criar um vetor de cores

```
cores <- rainbow(k)
```

Exemplo de como ficarão os gráficos:

Exercício baseado no material do Prof. Antônio Braga

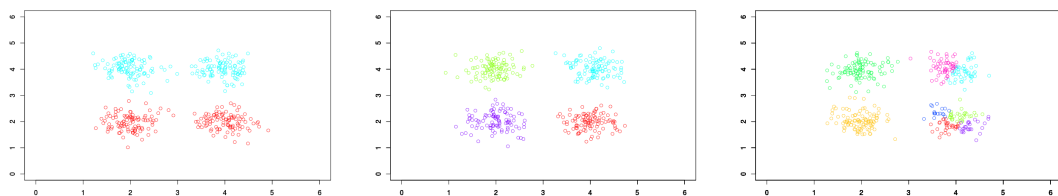


Figure 0.1: Std = 0.3

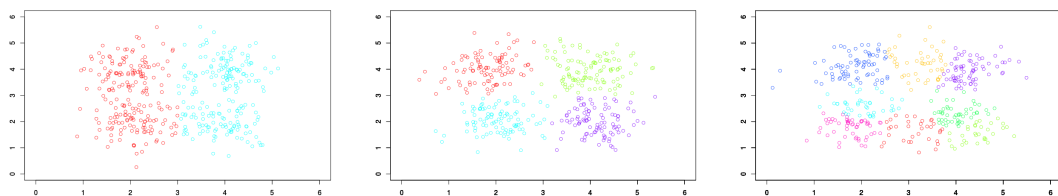


Figure 0.2: Std = 0.5

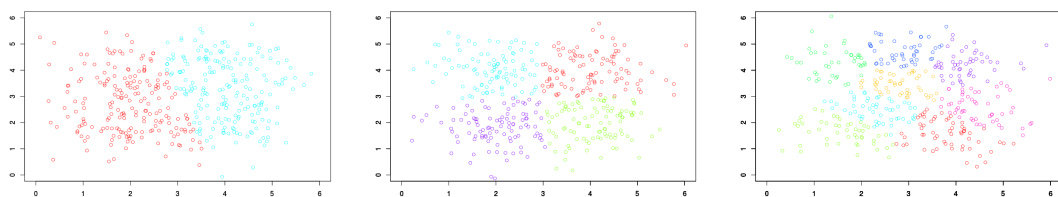


Figure 0.3: Std = 0.7