

Exercícios sobre Clustering com R

1 Questões sobre cluster plano

1. Utilizando o dataset **ruspini** do pacote **cluster**, execute uma análise de cluster utilizando o algoritmo **k-means**.
2. Faça uma análise de cluster utilizando o dataset sobre **abalos sísmicos** do exercício anterior. Considere apenas as variáveis de **longitude** e **latitude**.
3. Imprima a informação dos clusters em um mapa georeferenciado - o mesmo utilizado no exercício anterior.
4. Faça uma análise de cluster utilizando o dataset **survey**, atributos **Age** e **Height**, do pacote **UsingR**. Implemente três cenários diferentes:
 - (a) Com os valores originais;
 - (b) Com os valores da altura (**Height**) em metros, e;
 - (c) Com os valores da altura (**Height**) e idade (**Age**) devidamente normalizados.

Comente o que acontece em cada um dos casos.

5. Faça uma análise de cluster utilizando o dataset **survey**, atributos **Wr.Hnd**, **NW.Hnd**, **Age** e **Height**, do pacote **UsingR**. Faça o pré-processamento adequado dos dados e utilize a função **heatmap** para auxiliá-lo na sua análise.
6. Utilizando o dataset **survey**, atributos **Exer** (sobre o hábito de fazer exercícios) e **Smoke** (sobre o hábito de fumar), faça uma análise de cluster deste dataset.
7. Levando-se em consideração o dataset do item anterior, inclua o atributo de sexo (**Sex**) e faça a mesma análise.
8. Faça a carga do dataset **Animals** do pacote **MASS** e faça uma análise de cluster considerando os dados que estão no dataset **Animals**. Responda as seguintes perguntas:
 - Que método você utilizou para determinar o número de clusters ideal?
 - Houve algum tipo de pré-processamento? Em caso positivo, qual foi a razão para este pré-processamento?
 - Qual é o número de clusters ideal que você encontrou?
 - Faça uma descrição mais alto nível para cada cluster encontrado. Experimente rotular os agrupamentos.

2 Questões sobre cluster hierárquico

1. Faça uma análise de cluster hierárquico utilizando o dataset **survey**, atributos **Wr.Hnd**, **NW.Hnd**, **Age** e **Height**, do pacote **UsingR**.
2. Em determinadas situações é difícil identificar o número adequado de clusters utilizando o método *elbow*. Outro método que pode ser utilizado para identificar um número adequado de clusters planos é o próprio cluster hierárquico. Neste caso, faça uma análise com cluster hierárquico utilizando o dataset sobre **abalos sísmicos**. Considere apenas as variáveis de **longitude** e **latitude**.

3 Códigos que podem ser úteis

```
elbow <- function(dataset){
  wss <- numeric(15)
  for (i in 1:15)
    wss[i] <- sum(kmeans(dataset,centers=i,
                        nstart=100)$withinss)
  plot(1:15, wss, type="b", main="Elbow method",
       xlab="Number of Clusters",
       ylab="Within groups sum of squares",
       pch=8, col="red")
}
```