

Solução Lista 01

Nome: Pedro Henrique Fonseca Teixeira da Silva
E-mail: pedro.fonseca@aluno.ufabc.edu.br

25 fevereiro, 2025

Exercício 01

- a) É quando temos uma variável resposta (rótulo) qualitativa. Exemplos: Locais mais viáveis para reflorestamento de uma determinada flora. Variáveis possíveis: temperaturas médias, duração das chuvas, umidade do ar.
- b) É quando temos uma variável resposta quantitativa, por exemplo: Estimativa de remuneração. Variáveis possíveis: idade, anos de estudo etc
- c) O agrupamento não utiliza rótulos ou respostas, pois não há uma variável dependente a ser predita. Por exemplo, descobrir grupos de clientes a partir da frequência de compras, categoria de produtos comprados e localização geográfica.

Exercício 02

A “maldição da dimensionalidade” refere-se à dificuldade em analisar e organizar dados em espaços de alta dimensão. À medida que o número de dimensões cresce, os pontos de dados se tornam mais dispersos, tornando as técnicas tradicionais de aprendizado menos eficazes. Além disso, aumenta o custo computacional e a necessidade de mais dados para generalizar corretamente os modelos.

Exercício 03

```
knn_classify <- function(k, x, D) {  
  library(dplyr)  
  
  D2 <- D %>%  
    mutate(dist = (x[1] - x_1)^2 + (x[2] - x_2)^2) %>%  
    arrange(dist) %>% head(k) %>% count(y)  
  
  return(D2$y[which.max(D2$n)])  
}
```

Exercício 04

```

library(tidyverse)
data("iris")
iris <- as_tibble(iris) %>%
  select(Petal.Length, Sepal.Length, Species) %>%
  rename(x_1 = Petal.Length, x_2 = Sepal.Length, y = Species)

l_iris <- as.list(iris)

correct_classifications_k10 <- sum(pmap_lgl(l_iris, function(x_1, x_2, y) {
  knn_classify(10, c(x_1, x_2), iris) == y
})))

correct_classifications_k1 <- sum(pmap_lgl(l_iris, function(x_1, x_2, y) {
  knn_classify(1, c(x_1, x_2), iris) == y
})))

```