

Solução Lista 01

Nome: Luana Andrade Oliveira
E-mail: luana.a@aluno.ufabc.edu.br
Nome: Larissa Santos Costa
E-mail: costa.larissa@aluno.ufabc.edu.br
(Não é preciso informar os RAs)

25 February, 2025

Exercício 01

Problema de Classificação

- Aplicação: Classificação de produtos em categorias ou identificadores únicos.
- Vetores de Características: Informações sobre o produto, como tamanho, peso, nome, cor, material e modo de uso.
- Rótulos ou Respostas: Identificador único (como um código SKU).

Problema de Regressão

- Aplicação: Previsão do tempo de vida útil de um produto.
- Vetores de Características: Características do produto e de produtos semelhantes, como material, condições de uso, tempo médio de vida, marca e frequência de manutenção.
- Rótulos ou Respostas: Tempo estimado de vida útil do produto (em anos, meses, etc.).

Problema de Agrupamento

- Aplicação: Segmentação avançada de clientes para estratégias de recomendação e marketing personalizado.
- Vetores de Características: Histórico de compras, produtos visitados, produtos adquiridos, frequência de visitas ao site, tempo médio de navegação e outras interações com a loja.
- Rótulos ou Respostas: Agrupamento automático de clientes com perfis de comportamento semelhantes, sem necessidade de categorias pré-definidas.

Exercício 02

A maldição da dimensionalidade acontece quando trabalhamos com dados que têm muitas variáveis ou características. Conforme o número de dimensões aumenta, a quantidade de dados cresce exponencialmente, tornando a análise mais complexa. Isso pode não significar que estamos obtendo mais informações úteis, mas sim que há mais ruído e redundância.

Exercício 03

```
library(tidyverse)

classificacao <- function(k, x, dados) {
  dados %>%
    mutate(distancia = (x[1] - x_1)^2 + (x[2] - x_2)^2) %>%
```

```

    arrange(distancia) %>%
    slice_head(n = k) %>%
    count(y, sort = T) %>%
    slice(1) %>%
    pull(y)
}

D <- tibble(
  x_1 = rnorm(100, 1, 1),
  x_2 = rnorm(100, -1, 2),
  y = factor(sample(c("one", "two", "three"), 100, replace = T)))

num_k <- 10
vetor_x <- c(1, 2)

classe <- classificacao(num_k, vetor_x, D)
print(paste("A classe mais frequente é:", classe))

## [1] "A classe mais frequente é: two"

```

Exercício 04

```

library(class)
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)

caracteristicas <- iris %>% select(x_1, x_2)
rotulos <- iris$y

l_iris <- as.list(iris)

acuracia <- function(k) {
  predicao <- knn(train = caracteristicas, test = caracteristicas, cl = rotulos, k = k)

  acertos <- pmap_lgl(l_iris, function(x_1, x_2, y) {
    i <- which(iris$x_1 == x_1 & iris$x_2 == x_2)[1]
    predicao[i] == y
  })

  sum(acertos)
}

cat("Acertos com k = 1:", acuracia(1), "\n")

## Acertos com k = 1: 149

cat("Acertos com k = 10:", acuracia(10), "\n")

## Acertos com k = 10: 143

```