

Solução Lista 01

Nome: Pedro Araujo
E-mail: araujo.souza@aluno.ufabc.edu.br
Nome: Rafael Coelho
E-mail: rafael.coelho@aluno.ufabc.edu.br
(Não é preciso informar os RAs)

26 February, 2025

Exercício 01

a) Problema de Classificação

Reconhecimento de imagens com visão computacional (ex.: Identificar se na câmera de segurança o movimento é de um invasor ou um animal)

- Vetor de características: valores de pixel, bordas, texturas.
- Rótulos: humano, cachorro, carro, animal.

b) Problema de Regressão

Previsão de demanda energética para antecipar valor cobrado em contas de luz

- Vetor de características: temperatura, dia da semana, feriados, dados antigos de consumo.
- Resposta: Consumo em reais

c) Problema de Agrupamento

Identificar transações financeiras fraudulentas

- Vetor de características: valor da transação, localização, horário, histórico.
- Saída: Grupos normais vs anomalias.

Exercício 02

A “maldição da dimensionalidade” acontece quando a dimensão dos dados é tão grande que o conceito de vizinho mais próximo deixa de fazer sentido por muitos pontos terem distâncias bem similares, diminuindo a eficiência de algoritmos como o kNN. Isso acontece, pois em grandes dimensões acaba que os dados ficam mais esparsos um do outro, por haver um número limitado de dados num espaço geométrico muito amplo.

Exercício 03

```
library(dplyr)
library(tibble)

k_nn_classification = function(k, x, D) return(D %>%
  mutate( dist = sqrt((x[1] - x_1)^2 + (x[2] - x_2)^2) ) %>% # Distância Euclidiana
  arrange( dist ) %>% head(k) %>% count(y, sort = T) %>% head(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))

x = c(1,2)
k = 10

head(k_nn_classification(k, x, D))

## [1] three
## Levels: one three two

```

Exercício 04

```

library(tidyverse)

data("iris") # Carrega o banco no ambiente global
iris <- as_tibble(iris) %>% # Converte para a dataframe tibble
  select(Petal.Length,Sepal.Length,Species) %>% # Seleciona colunas da dataframe
  rename( x_1 = Petal.Length, x_2 = Sepal.Length, y = Species) # Renomeia as colunas

l_iris <- as.list(iris)

v_bool_10 <- pmap_lgl(l_iris, function(x_1,x_2,y){
  x = c(x_1, x_2)

  class = k_nn_classification(10, x, iris)

  return( y == class )
})

v_bool_1 <- pmap_lgl(l_iris, function(x_1,x_2,y){
  x = c(x_1, x_2)

  class = k_nn_classification(1, x, iris)

  return( y == class )
})

resultados <- c(
  "k = 10" = sum(v_bool_10),
  "k = 1" = sum(v_bool_1),
  "total dados" = count(iris)
)

print(resultados)

## $`k = 10`
## [1] 143
##
## $`k = 1`
## [1] 149
##
## $`total dados.n`
## [1] 150

```