

Lista 3 - Mineração

Victor Alves Dogo Martins, RA: 744878 Ana Beatriz Alves Monteiro, RA: 727838
Larissa Torres, RA: 631914

14-08-2022

Itens 1 e 2

Para a transformação das variáveis categóricas (**US**, **Urban** e **ShelveLoc**) em dummies, utilizamos a função `model.matrix()`, que as transforma automaticamente e são apresentadas da seguinte maneira:

- **US:** transformada na variável X_{USYes} :

$$X_{USYes} = \begin{cases} 1, \text{ caso a loja estiver localizada nos EUA;} \\ 0, \text{ caso contrário.} \end{cases}$$

- **Urban:** transformada na variável $X_{UrbanYes}$:

$$X_{UrbanYes} = \begin{cases} 1, \text{ caso a loja estiver localizada na zona urbana;} \\ 0, \text{ caso contrário.} \end{cases}$$

- **ShelveLoc:** transformada nas variáveis $X_{ShelveLocGood}$ $X_{ShelveLocMedium}$:

$$X_{ShelveLocGood} = \begin{cases} 1, \text{ caso o produto tiver localização boa na prateleira;} \\ 0, \text{ caso contrário.} \end{cases}$$

$$X_{ShelveLocMedium} = \begin{cases} 1, \text{ caso o produto tiver localização média (mas não boa) na prateleira;} \\ 0, \text{ caso contrário.} \end{cases}$$

Por outro lado, como feito na Lista 2, a divisão do banco de dados entre treino e teste foi feita com o auxílio da função `initial_split()` do pacote `{rsample}`:

```
## Lendo pacotes

set.seed(1)

library(tidyverse)
library(rsample)
library(knitr)
library(kableExtra)
```

```

## Lendo Dados

df <- ISLR::Carseats |>
  mutate(US=as.factor(US),
         Urban=as.factor(Urban),
         ShelveLoc=as.factor(ShelveLoc))

## Divisão entre treino e teste

split <- initial_split(df, prop=0.6)

tre <- training(split)
tes <- testing(split)

x_tre <- model.matrix(Sales~., tre)
y_tre <- tre[,1]

x_tes <- model.matrix(Sales~., tes)
y_tes <- tes[,1]

```

Item 3

Item 4

Item 5

Item 6