

MAE0217 - Estatística Descritiva - Lista 6

Natalia Hitomi Koza¹
Rafael Gonçalves Pereira da Silva²
Ricardo Geraldles Tolesano³
Rubens Kushimizo Rodrigues Xavier⁴
Rubens Gomes Neto⁵
Rubens Santos Andrade Filho⁶
Thamires dos Santos Matos⁷

Julho de 2021

Sumário

Capítulo 8	2
Exercício 6	2

¹Número USP: 10698432

²Número USP: 9009600

³Número USP: 10734557

⁴Número USP: 8626718

⁵Número USP: 9318484

⁶Número USP: 10370336

⁷Número USP: 9402940

Capítulo 8

Exercício 6

Considere o conjunto de dados `antracose`, selecionando aleatoriamente 70% das observações para treinamento e as restantes para validação.

```
antracose0 <- read_xls('data/antracose.xls')
antracose1 <- na.omit(antracose0)
```

a) Ajuste modelos de regressão linear múltipla e de regularização Ridge, Lasso e Elastic Net ao conjunto de treinamento com o objetivo de previsão da variável `antracose` com base nas variáveis preditoras `idade`, `tmunic`, `htransp`, `cargatabag`, `ses`, `densid` e `distmin`.

```
# Separação dos dados em 70% treinamento e 30% validação
set.seed(12)
dados <- data.matrix(antracose1)
train_indexes <- sample(seq_len(nrow(dados)), floor(0.7*nrow(dados)))

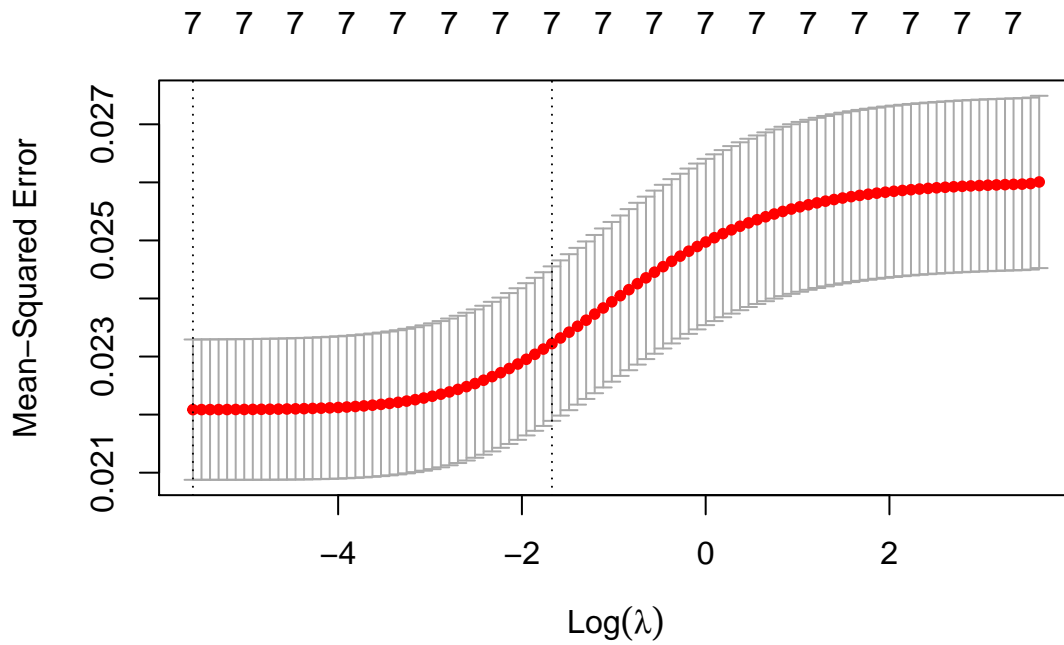
y_train <- dados[train_indexes,c(5)]
x_train <- dados[train_indexes,-c(5)]
y_val <- dados[-train_indexes,c(5)]
x_val <- dados[-train_indexes,-c(5)]
```

Regressão linear múltipla:

```
antracose_lm <- lm(y_train~x_train)
```

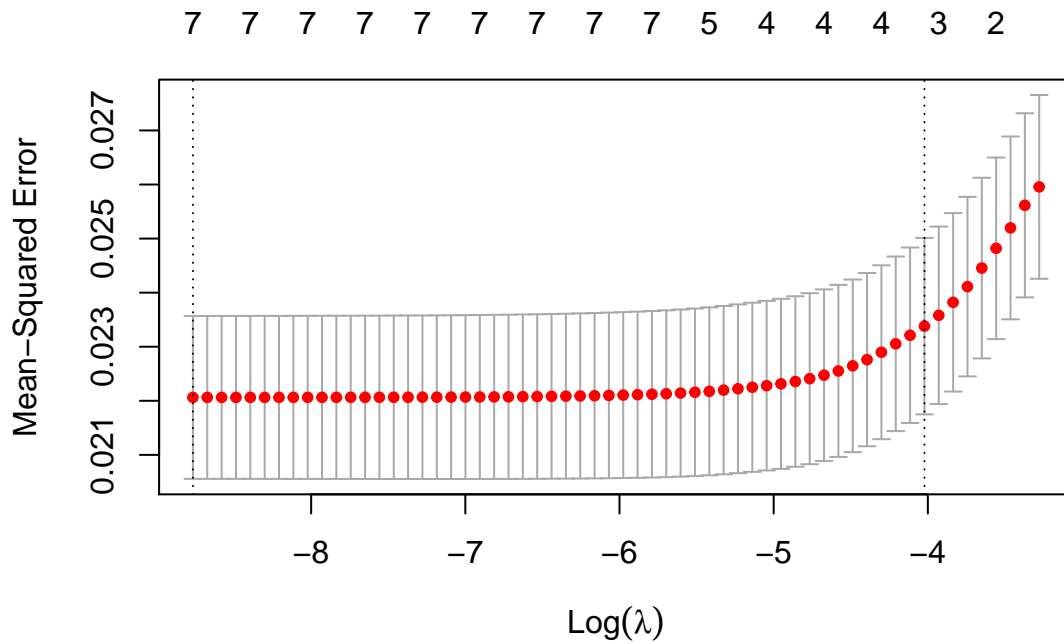
Regularização Ridge:

```
antracose_ridgecv = cv.glmnet(x_train, y_train, alpha = 0)
plot(antracose_ridgecv)
```



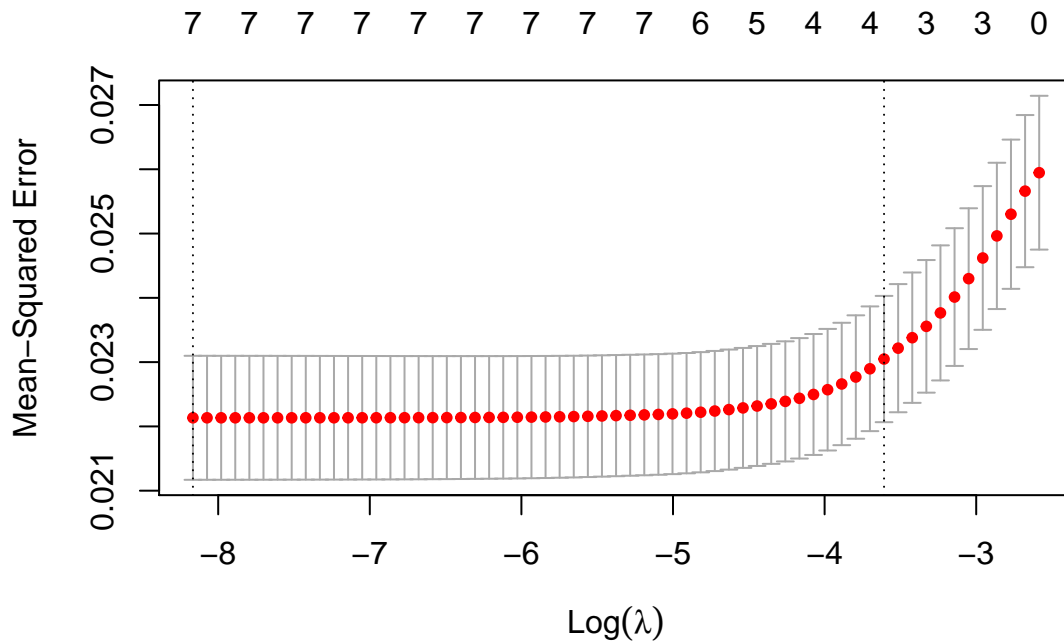
Regularização Lasso:

```
antracose_lassocv = cv.glmnet(x_train, y_train, alpha = 1)
plot(antracose_lassocv)
```



Regularização Elastic Net

```
antracose_elncv = cv.glmnet(x_train, y_train, alpha = 0.5)
plot(antracose_elncv)
```



b) Compare o desempenho dos modelos no conjunto de validação.

c) Repita os itens a) e b) com outra seleção aleatória dos conjuntos de treinamento e de validação.

```
# Separação dos dados em 70% treinamento e 30% validação
set.seed(13)
dados2 <- data.matrix(antracose1)
train_indexes2 <- sample(seq_len(nrow(dados2)), floor(0.7*nrow(dados2)))

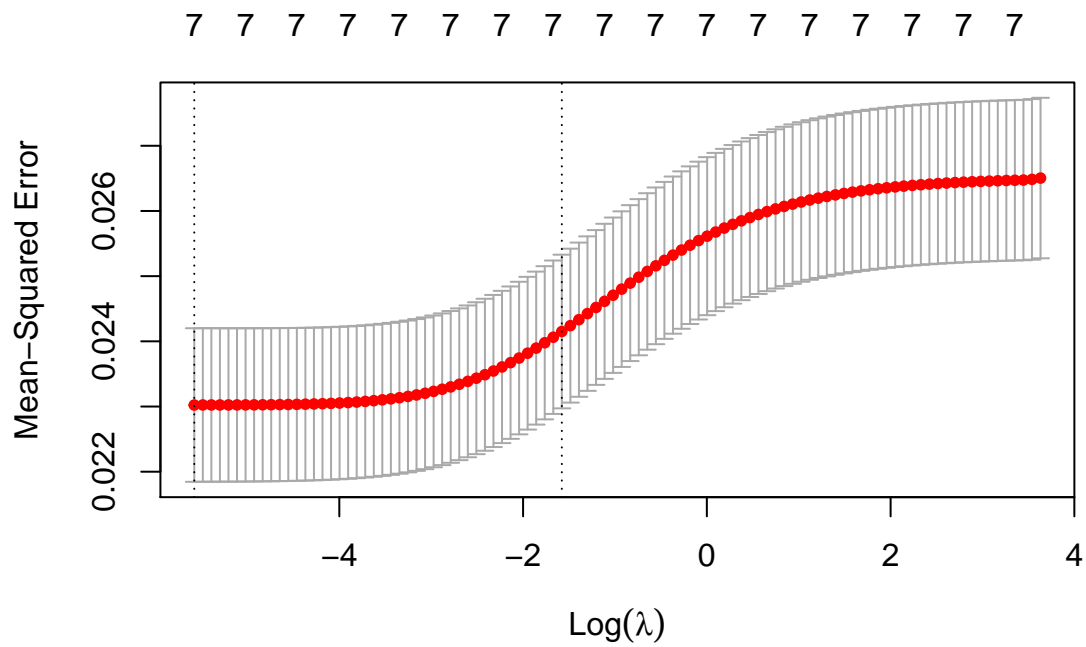
y_train2 <- dados2[train_indexes2,c(5)]
x_train2 <- dados2[train_indexes2,-c(5)]
y_val2 <- dados2[-train_indexes2,c(5)]
x_val2 <- dados2[-train_indexes2,-c(5)]
```

Regressão linear múltipla:

```
antracose_lm2 <- lm(y_train2~x_train2)
```

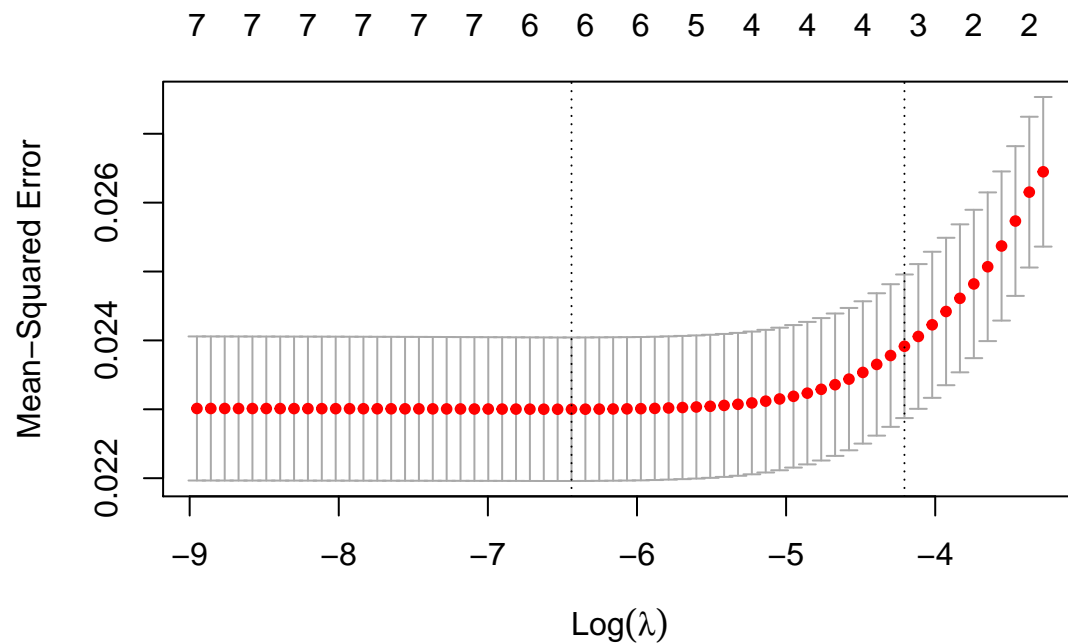
Regularização Ridge:

```
antracose_ridgecv2 = cv.glmnet(x_train2, y_train2, alpha = 0)
plot(antracose_ridgecv2)
```



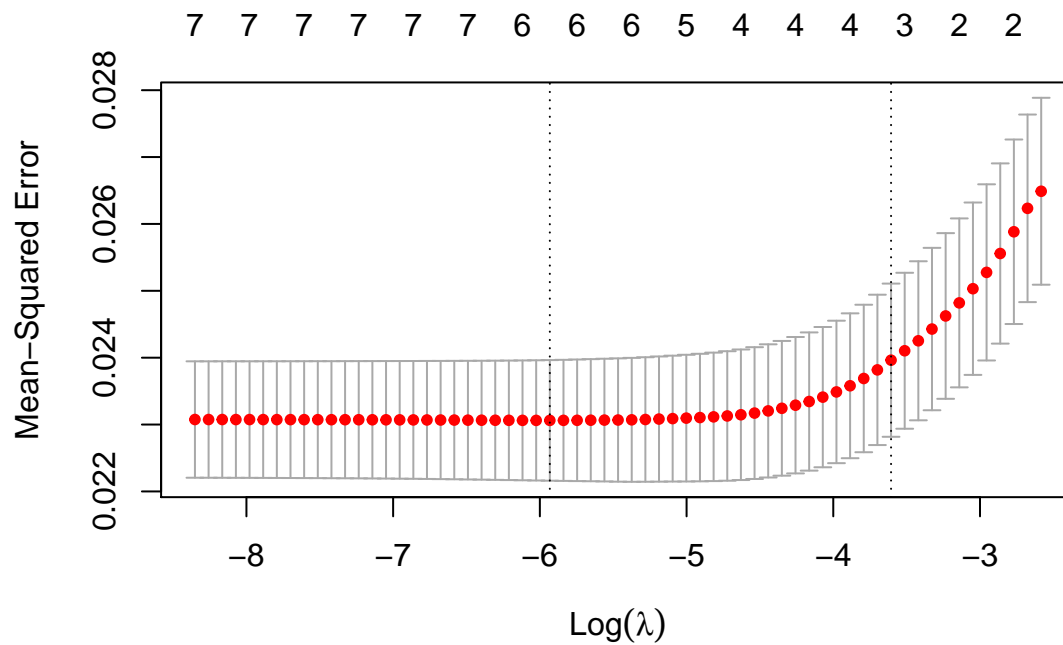
Regularização Lasso:

```
antracose_lassocv2 = cv.glmnet(x_train2, y_train2, alpha = 1)
plot(antracose_lassocv2)
```



Regularização Elastic Net

```
antracose_elncv2 = cv.glmnet(x_train2, y_train2, alpha = 0.5)
plot(antracose_elncv2)
```



d) Construa uma tabela com as medidas de desempenho de todos os modelos ajustados e comente os resultados.

e) Utilize modelos GAM para avaliar se as variáveis preditoras estão linearmente associadas com a variável resposta e compare o resultado do modelo adotado sob esse enfoque com aqueles dos demais modelos.