Compulsory Exercise 2: Title (give your project an informative title)

Karianne Strand Bergem Marte Ragnhild Hotvedt Erlend Winje

27 March, 2025

Abstract

This is the place for your abstract (max 350 words)

Abstract

Introduction: Scope and purpose of your project

Problemstillingen vår kan være om vi kan predikere alder basert på de variablene i Heart Failure-datasettet? Er det vi skal finne ut av liksom

Descriptive data analysis/statistics

```
data <- read.csv("heart.csv")
```

Methods

```
n <- nrow(data) # Number of observations

# Indexes for the training set (70% of the data)
train_idx <- sample(1:n, size = round(0.7 * n), replace = FALSE)

# Split the data
train_data <- data[train_idx, ]
test_data <- data[-train_idx, ]
```

Multiple linear regression

Ridge/Lasso

Apply Ridge regression to the Heart dataset.

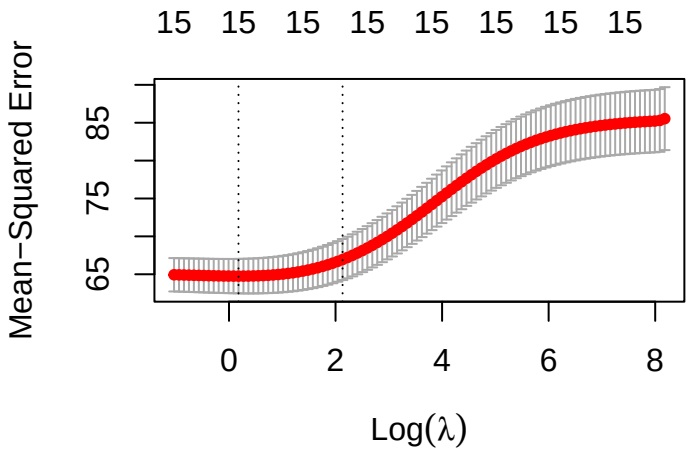
```r
library(glmnet)

# Create design matrices
x_train <- model.matrix(Age ~ ., data = train_data)[, -1]
y_train <- train_data$Age
x_test <- model.matrix(Age ~ ., data = test_data)[, -1]
y_test <- test_data$Age

# `alpha=0` is the ridge penalty
ridge_mod <- glmnet(x_train, y_train, alpha = 0)

# Cross-validation to find the best lambda
set.seed(123)
cv_ridge <- cv.glmnet(x_train, y_train, alpha = 0)

plot(cv_ridge)
```



Now, we want to find the best λ.

```r
best_lambda <- cv_ridge$lambda.min
best_lambda
```

```
## [1] 1.195248
```

Evaluate the method

```r
ridge_pred <- predict(cv_ridge, s = best_lambda, newx = x_test)
mse <- mean((y_test - ridge_pred)^2)
r2 <- 1 - sum((y_test - ridge_pred)^2) / sum((y_test - mean(y_test))^2)

cat("MSE:", mse, "\n")
```

```
## MSE: 70.9204
```

```r
cat("R²:", r2, "\n")
```

```
## R²: 0.2677013
```

Results and interpretation

Evaluere modellene på testsettet. Sammenligne metodene – hvilken ga best resultater? Diskutere hvilke variabler som har størst betydning for prediksjon av alder.

Summary