

221004 Classification

Jongrak

2022-12-31

KSS model exercise

```
library(ISLR)
library(class)
library(MASS)

set.seed(42)

#randomly sampling
Default$student = as.numeric(Default$student) - 1
default_index <- sample(nrow(Default), 5000)

training_data <- Default[default_index, ]
test_data <- Default[-default_index, ]

head(training_data)

##      default student  balance  income
## 2369      No        0  943.1324 30178.734
## 5273      No        0 1035.4859 41714.374
## 9290      No        1 1009.1093  8868.657
## 1252      No        0 1195.4840 38452.641
## 8826      No        1  728.9858 17360.523
## 356       No        0  717.3419 44266.377

nrow(training_data)

## [1] 5000

head(test_data)

##      default student  balance  income
## 5         No        0  785.6559 38463.496
## 6         No        1  919.5885  7491.559
## 7         No        0  825.5133 24905.227
## 10        No        0   0.0000 29275.268
## 12        No        1 1220.5838 13268.562
## 13        No        0  237.0451 28251.695

nrow(test_data)

## [1] 5000

sum(training_data$default == "Yes")
```

```

## [1] 170

# training data
X_train <- training_data[, -1]
Y_train <- training_data$default

# test data
X_test <- test_data[, -1]
Y_test <- test_data$default

Y_predict <- knn(train = X_train, test = X_test, cl = Y_train, k = 3)

# accuracy
sum(Y_test == Y_predict)

## [1] 4828

sum(Y_predict == "Yes")

## [1] 57

y_pred_yes <- which(Y_test == "Yes")
y_pred_test_yes <- which(Y_predict == "Yes")

score = 0

for (yes in y_pred_yes) {
  score <- score + ifelse (yes %in% y_pred_test_yes, 1, 0)
}

acc_k <- rep(x = 0, times = 100) # container
for(i in 1:100) {
  pred <- knn(train = X_train, test = X_test, cl = Y_train, k = i)
  acc_k[i] <- sum(Y_test == pred)
}
plot(acc_k, type = "b")

```

