L7

Kovel

31 РѕРєС‚СЏР±СЂСЏ 2020 Рі

# Download the data

set.seed(123)  
f\_train <- read.csv2('education\_train.csv', header = TRUE, encoding = 'UNICOD')  
f\_test <- read.csv2('education\_test.csv', header = TRUE, encoding = 'UNICOD')  
f\_train <- f\_train[-1]  
f\_test <- f\_test[-1]

## Висновок: завантажано датасет,який було розподілено на навчальну та тестову вибірки.

# Fitting & predicting

library(class)  
y = knn(train = f\_train[,c('age','traveltime')],  
 test = f\_test[,c('age','traveltime')],  
 cl = f\_train[, 'school'],  
 k = 16,  
 prob = TRUE)

## Висновок: і навчання, і прогнозування за моделлю k найближчих сусідів здійснюється однією функцією. У результаті отримуємо вектор класів об’єктів.

## Confusion Matrix

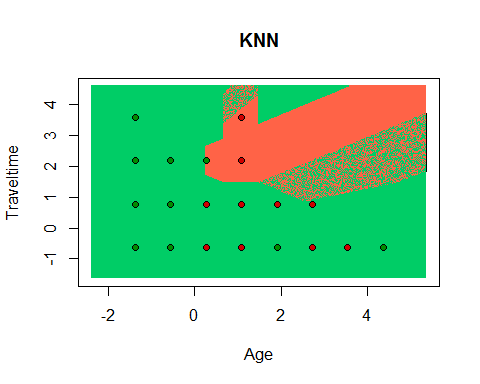
cm = table(f\_test[, 'school'], y == '1')  
print(cm)

##   
## FALSE  
## 0 116  
## 1 15

## Висновок: точність моделі – (115+1) / 131 = 88,5 %, частка невірно класифікованих випадків – (15+0) / 131 = 11,4 %. Чутливість – 1 / (15+1) = 6 %, специфічність – 115 / (115+0) = 100 %, тобто модель більш чутлива до виявлення негативних випадків.

## Visualising the Test set results

library(ggplot2)  
set = f\_test[,c('age','traveltime','school')]  
X1 = seq(min(set['age']) - 1, max(set['age']) + 1, by = 0.01)  
X2 = seq(min(set['traveltime']) - 1, max(set['traveltime']) + 1, by = 0.01)  
grid\_set = expand.grid(X1, X2)  
colnames(grid\_set) = c('age', 'traveltime')  
y\_grid = knn(train = f\_train[,c('age','traveltime')], test = grid\_set, cl = f\_train[, 'school'], k = 5)  
plot(set[, -3],  
 main = 'KNN',  
 xlab = 'Age', ylab = 'Traveltime',  
 xlim = range(X1), ylim = range(X2))  
contour(X1, X2, matrix(as.numeric(y\_grid), length(X1), length(X2)), add = TRUE)  
points(grid\_set, pch = '.', col = ifelse(y\_grid == 1, 'tomato', 'springgreen3'))  
points(set, pch = 21, bg = ifelse(set[, 3] == 1, 'red3', 'green4'))

 ##Висновок: на графіку червоним позначені випадки потрапляння до школи першого типу, зеленим – ло другого. Червоним виділена зона високої ймовірності потрапляння до першого типу. Модель описує нелінійний варіант розподіляючої кривої.