L8

Kovel

31 .10.2020

# 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

# install.packages('e1071')  
library(e1071)  
f\_train$school <- as.factor(f\_train$school)  
f\_test$school <- as.factor(f\_test$school)  
class\_nb = naiveBayes(school ~ age, traveltime, data = f\_train)

## Висновок: для навчання моделі використано функцію naiveBayes.

# Predicting

y <- predict(class\_nb, f\_test[, c('age','traveltime')])

## Висновок: визначено класи об’єктів (вектор у).

## Confusion Matrix

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

## y  
## 0 1  
## 0 116 0  
## 1 15 0

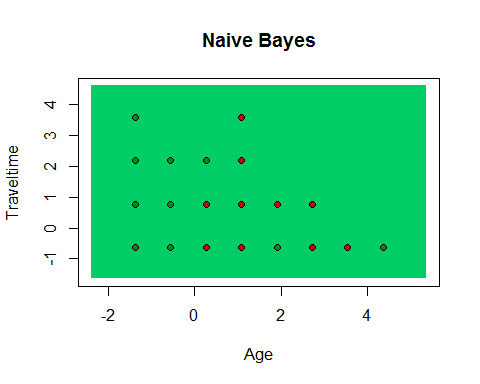
## Висновок: точність моделі – (115+0) / 131 = 87,7 , частка неправильно класифікованих випадків – (16+0) / 131 = 12 %. Чутливість – 0 / (16+0) = 0 %, специфічність – 115 / (115+0) = 100 %, тобто модель чутлива тільки до виявлення негативних випадків. У цьому разі – тих,хто потрапляє до шкои першого типу.

# 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 = predict(class\_nb, grid\_set)  
plot(set[, -3],  
 main = 'Naive Bayes',  
 xlab = 'Age', ylab = 'Traveltime',  
 xlim = range(X1), ylim = range(X2))  
contour(X1, X2, matrix(as.numeric(y\_grid), length(X1), length(X2)), add = TRUE)

## Warning in contour.default(X1, X2, matrix(as.numeric(y\_grid), length(X1), : все  
## значения z равны

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. Зеленим виділена зона високої ймовірності потрапляння до другого типу. Модель описує нелінійний варіант розподіляючої кривої.