

Zadanie KNN

Magdalena Opielińska

14 02 2020

```
chooseCRANmirror(graphics=FALSE, ind=1)
knitr::opts_chunk$set(echo = TRUE)
library(asbio)

## Warning: package 'asbio' was built under R version 3.6.2
## Loading required package: tcltk
library(tidyverse)

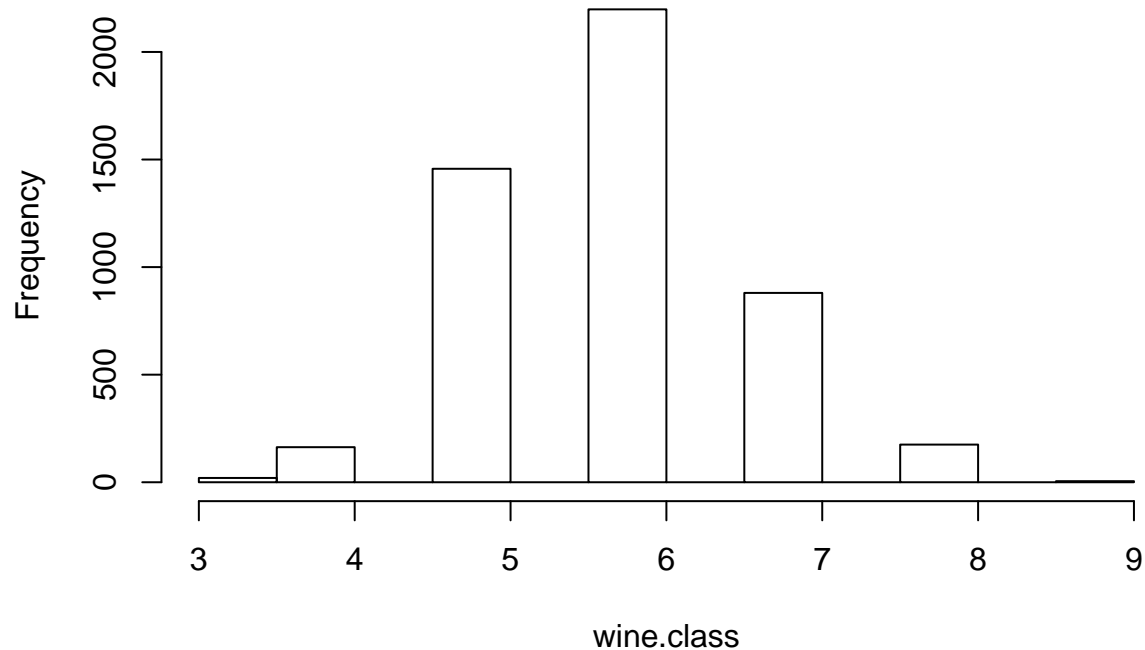
## Warning: package 'tidyverse' was built under R version 3.6.2
## -- Attaching packages -----
## v ggplot2 3.2.1      v purrr  0.3.3
## v tibble  2.1.3      v dplyr  0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
## Warning: package 'ggplot2' was built under R version 3.6.2
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
dane <- read.csv('http://mlr.cs.umass.edu/ml/machine-learning-databases/cpu-performance/machine.data')

#1
data("wine")

#2
wine.class <- wine[,1]
wine.data <- wine[,-1]

#3
hist(wine.class)
```

Histogram of wine.class



#4

```
wine.data <- scale(wine.data, center = TRUE, scale = TRUE)
```

#5

```
set.seed(123)
inx <- sample(nrow(wine), 4000)
wine.train.data <- wine.data[inx, ]
```

#6

```
wine.test.data <- wine.data[-inx, ]
```

#7

```
wine.train.class <- wine.class[inx]
wine.test.class <- wine.class[-inx]
```

#8

```
library(class)
```

```
## Warning: package 'class' was built under R version 3.6.2
```

```
wine_knn <- function(k) {
  wine_pred <- knn(wine.train.data, wine.test.data, wine.train.class, k=7)
  accuracy <- mean(wine_pred == wine.test.class)
}
```

```

    return(accuracy)
}

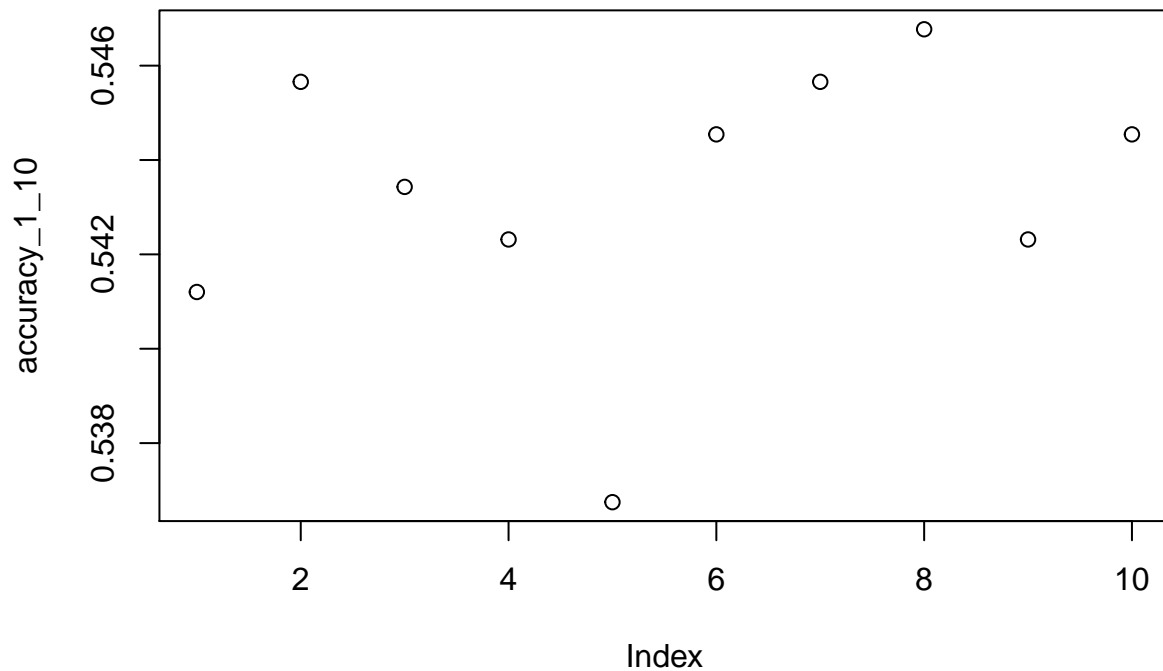
```

#9

```

k.values <- 1:10
accuracy_1_10 <- sapply(1:10, wine_knn, simplify = TRUE, USE.NAMES = TRUE)
plot(accuracy_1_10)

```



```

max_val <- max(accuracy_1_10)
sprintf('Najlepszą klasyfikację otrzymujemy dla k :%f value: %f', max_val, which(accuracy_1_10 == max_val))

```

```
## [1] "Najlepszą klasyfikację otrzymujemy dla k :0.546771 value: 8.000000"
```

#Implementacja klasyfikatora KNN

```

my_knn <- function (x, x_klas, K) {
  odl_euk <- apply(wine.train.data,1,
    function(v){sqrt(sum(x-v)^2)})
  top_k <- which(rank(odl_euk) <= K)
  getmode <- function(v) {
    uniqv <- unique(v)
    uniqv[which.max(tabulate(match(v, uniqv)))]
  }
  ccc <- wine.train.class[top_k]
  getmode(ccc)
}

```

```
K = 7
```

```
x <- wine.test.data[1, ]  
my_knn(x, wine.test.class, K)
```

```
## [1] 6
```