```
# Week 1 Homework (My comments are included with #)

#############################################################################################################
rm(list = ls())

library(kernlab)
set.seed(1)


#Question 2.1
#Classification model could be used in predicting customer purchasing habbits. I recently had to purchase an electric shaving machine as a gift for my brother on Amazon. I did have a budget of not spending more than $30 to $50  but did not want to compromise on the quality of the product. Some of the predictors that I would use to evaluate the product would be the following: cost, quality, Amazon user reviews, number of reviews, and warranty.


#Question 2.1 (1) Vanilladot model
data <- read.table("2.2credit_card_data-headersSummer2018.txt", header = TRUE, stringsAsFactors = FALSE)
head(data)
tail(data)

model1 <- ksvm(as.matrix(data[,1:10]), as.factor(data[,11]), type = "C-svc", kernel = "vanilladot", C= 100, scaled = TRUE)
# of support vectors used = 189, Training error = .1360856
model2 <- ksvm(as.matrix(data[,1:10]), as.factor(data[,11]), type = "C-svc", kernel = "vanilladot", C= 10, scaled = TRUE)
# of support vectors used = 190, Training error = .1360856
model3 <- ksvm(as.matrix(data[,1:10]), as.factor(data[,11]), type = "C-svc", kernel = "vanilladot", C= .1, scaled = TRUE)
# of support vectors used = 197, Training error = .1360856
# The three models at the top predicts that the classifer is about 87% accurate.

a <- colSums(model1@xmatrix[[1]] * model1@coef[[1]])
a0 <- -model1@b
# a0 = 0.081584921659538

#Model1 is around 86% accurate
pred <- predict(model1, data[,1:10])
sum(pred == data$R1)/nrow(data)


#Question 2.1 (2) In addition to vanilla dot, I will choose anovadot.
data <- read.table("2.2credit_card_data-headersSummer2018.txt", header = TRUE, stringsAsFactors = FALSE)
model1 <- ksvm(as.matrix(data[,1:10]), as.factor(data[,11]), type = "C-svc", kernel = "anovadot", C= 100, scaled = TRUE)
# of support vectors used = 205, Training error = 0.093272

a <- colSums(model1@xmatrix[[1]] * model1@coef[[1]])
a0 <- -model1@b
# a0 = 1.174074

#Model1 is around 90.6% accurate
pred <- predict(model1, data[,1:10])
sum(pred == data$R1)/nrow(data)

#Question 3
library(kknn)
set.seed(1)

pred_knn <- rep(0, (nrow(data)))
for (i in 1:nrow(data)){
  model= kknn(R1~., data[-i,], data[i,], k=3, scale = TRUE)
  pred_knn[i] <- as.integer(fitted(model)+0.5)
}
sum(pred_knn == data$R1)/nrow(data)


#Question 3.1
#2/3rds of data will be allocated for training and the rest for testing

data2 <- read.table("2.2credit_card_data-headersSummer2018.txt", header = TRUE, stringsAsFactors = FALSE)

#number of rows in my data
o.rows <- nrow(data2)
o.sample <- sample(1: o.rows, size = round(o.rows/3), replace = FALSE)

#2/3rds of the original data
o.train <- data2[-o.sample,]

#1/3rds of the original data
o.test <- data2[o.sample,]

cval <- train.kknn(R1~., o.train, kmax = 100, kernel = c("triangular", "rectangular", "gaussian", "inv", "optimal","cos"), scale = TRUE)
#Type of response variable: continuous
```

```
#minimal mean absolute error: 0.1826325
#Minimal mean squared error: 0.09998105
#Best kernel: inv
#Best k: 19

cval2 <- train.kknn(R1~., o.test, kmax = 100, kernel = c("triangular", "rectangular", "gaussian", "inv",
"optimal","cos"), scale = TRUE)
#Type of response variable: continuous
#minimal mean absolute error: 0.2041284
#Minimal mean squared error: 0.1082999
#Best kernel: rectangular
#Best k: 8

pred <- predict(cval, o.test)
prediction <- round(pred)
pred_accuracy <- table(prediction, o.test$R1)
pred_accuracy
sum(prediction == o.test$R1)/ length(o.test$R1)
#Prediction accuracy of cross validation method is 81.65%
```