```
# -----------
# EXERCISE 2
# -----------

library(MASS)
library(e1071)

rm(list=ls())
d <- read.table('musicCountry.txt', header=TRUE)
load('mcshapiro.test.RData')
head(d)
names(d)

n <- dim(d)[1]
p <- dim(d)[2]

d$release.country <- factor(d$release.country)
levels(d$release.country)

# ------------------------ point a

pf <- 0.9
pt <- 1-0.9
prior.c <- c(pt,pf)
# WARNING!!!!! IT SHOULD BE IN THE ORDER OF THE LEVELS

us <- which(d$release.country=='US')
ge <- which(d$release.country=='Germany')

# verify assumptions 1) e 2):
# 1) normality within the groups
mcshapiro.test(d[us,1:2])$p
mcshapiro.test(d[ge,1:2])$p

# ok

# 2) equal variance (univariate)
S1 <- cov(d[us,1:2])
S2 <- cov(d[ge,1:2])

S1
S2

# ok lda

mylda <- lda(d[,1:2], d$release.country, prior=prior.c)
mylda

#predd <- predict(mylda,d[,1:2])
#table(class.true=d$release.country, class.assigned=predd)

iris <- d[,1:2]

plot(d[,1:2], col = d$release.country, pch=20)
legend("topright", legend=levels(d$release.country), fill=c('red','green'), cex=.7)

points(mylda$means, pch=4,col=c('red','green') , lwd=2, cex=1.5)
x  <- seq(min(iris[,1]), max(iris[,1]), length=200)
y  <- seq(min(iris[,2]), max(iris[,2]), length=200)
xy <- expand.grid(Sepal.Length=x, Sepal.Width=y)

z  <- predict(mylda, xy)$post  # these are P_i*f_i(x,y)
z1 <- z[,1] - pmax(z[,2])  # P_1*f_1(x,y)-max{P_j*f_j(x,y)}
z2 <- z[,2] - pmax(z[,1])  # P_2*f_2(x,y)-max{P_j*f_j(x,y)}

# Plot the contour line of level (levels=0) of z1, z2, z3:
# P_i*f_i(x,y)-max{P_j*f_j(x,y)}=0 i.e., boundary between R.i and R.j
# where j realizes the max.
contour(x, y, matrix(z1, 200), levels=0, drawlabels=F, add=T)
contour(x, y, matrix(z2, 200), levels=0, drawlabels=F, add=T)


# ------------------------ point b


ldaCV <- lda(d[,1:2], d$release.country, CV=TRUE, prior = prior.c)  # specify the argument CV
misc <- table(class.true=d$release.country, class.assignedCV=ldaCV$class)
AERCV  <- misc[1,2]*prior.c[1]/sum(misc[1,]) + misc[2,1]*prior.c[2]/sum(misc[2,])
AERCV

# ------------------------ point c





# ------------------------ point d

testdat <- data.frame(price=50,average.length=3.5)
ypred <- predict(mylda,testdat)
ypred
#$class
#[1] US
#Levels: Germany US
#
#$posterior
#Germany        US
#[1,] 0.08824657 0.9117534
#
#$x
#LD1
#[1,] -1.009987

# ------------------------ point e


set.seed (1)
tune.out <- tune(svm,release.country~.,data=d ,kernel = 'linear',
                 ranges =list(cost=c(0.001  , 0.01, 0.1, 1, 10,100) ))
summary(tune.out)
# Extract the best model from the result of tune
bestmod <- tune.out$best.model
summary(bestmod)
# cost 10
plot(bestmod , d, col =c('salmon', 'light blue'), pch=19)


predict(bestmod,testdat)
# US
```