Bucci Teo
10621873

# Exercise 2

I verify the assumption of Gaussianity for each group by performing a Shapiro test (H0: “data is Gaussian”, H1: $H0^C$), which gives p-values of 0.9872, 0.1272.
These are high enough to not reject H0, thus we have evidence to say that H0 is indeed true.

At this point, I could perform Quadratic Discriminant Analysis (QDA), but I also test if groups have the same covariance structure to see if I can perform a Linear Discriminant Analysis (LDA) as well. I do qualitatively

```
> S1
                   price average.length
price          173.53448      -1.524040
average.length  -1.52404       1.543192
> S2
                    price average.length
price          108.936681      -7.621866
average.length  -7.621866       8.505943
```

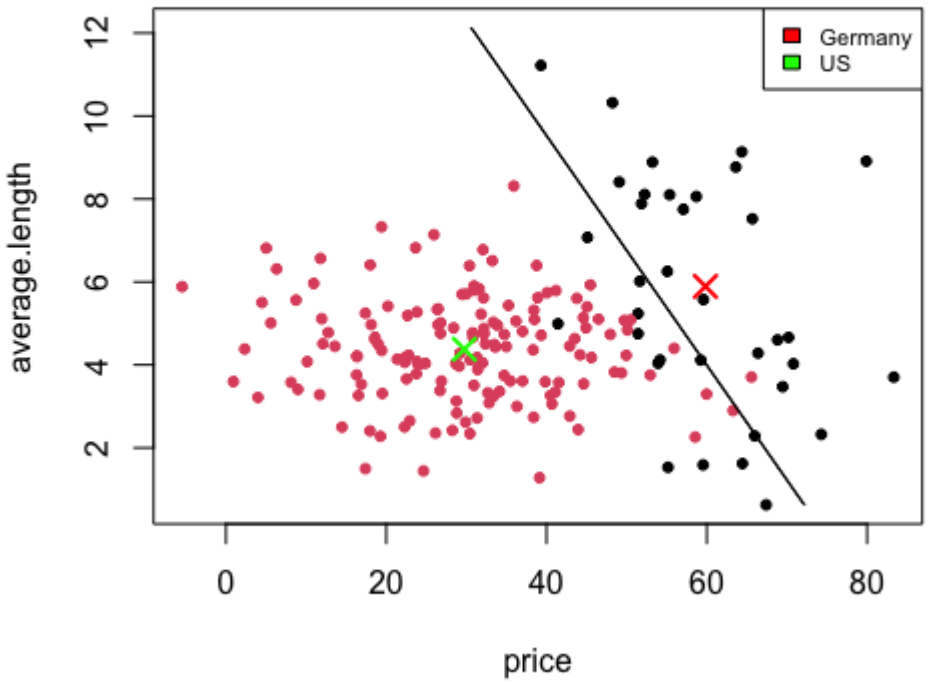
These are very similar, so I continue with LDA.

```
# 1) if L=i, X.i ~ N(mu.i, sigma.i^2), i=A,B
# 2) sigma.A=sigma.B
# 3) c(A|B)=c(B|A) (equal misclassification costs)
```

The means of the groups identified by the classifier are:

```
Group means:
           price average.length
Germany 59.82982       5.891951
US      29.74888       4.377556
```

The regions identified in the space of the variables are the following




The AER obtained by leave-one-out cross-validation is 0.04795322 and is good.

The estimated probability that a new album is classified as US is 0.9.

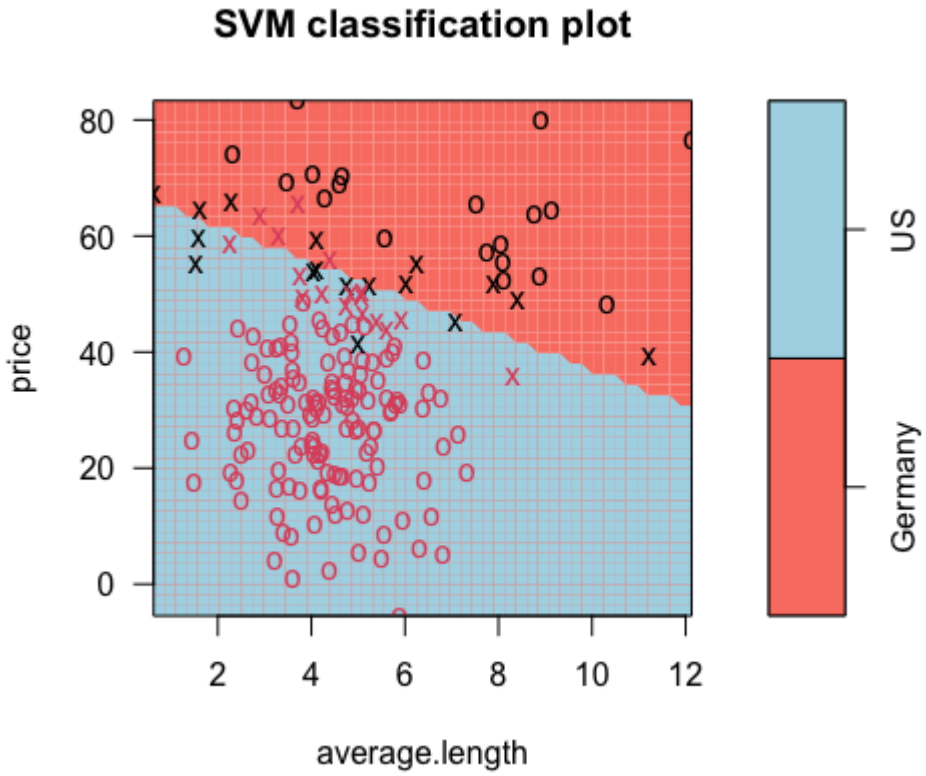
I predict the new observation using the classifier, the results are:

```
         Germany        US
[1,] 0.08824657 0.9117534
```

with 0.91 posterior probability it belongs to US.

Finally, I fit a Support Vector Machine using the required parameters, the best cost is 1.

The classification region obtained using the SVM is:




Using the SVM to predict the class of the new observation I obtain US as before.