Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 16th, 2022

Problem n.2

The file `musicCountry.txt` contains the price (in €) and the average song length (in minutes) of 188 albums released in Germany and US.

a) Knowing that, on average, 90% of albums are released in US, build a classifier to characterize the country of release based on the price and the average song length. Report the model for the data, the estimates of its parameters (means and covariances) and verify the model assumptions. Report the plot of the classification regions.

b) Estimate the AER of the classifier through leave-one-out cross-validation.

c) Using the classifier built at point (a), what is the estimated probability that a new album is classified as US?

d) How would you classify a new album with with a price of 50€, and an average song length of 3.5 minutes?

e) Use a support vector machine with linear kernel to classify albums. Tune the cost parameter with a 10-fold cross validation choosing between the values: 0.001, 0.01, 0.1, 1, 10, 100. Report the chosen cost and a plot of the classification regions. How would you classify the album at point d) with this classifier?

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUQ1dXS1VTREw1TkVZVEpQRlRIRkFMNzZWRi4u`