Lab 14

Maria Barouti

Exercise - Support Vector Machines

For this lab we will use the `Auto.rda` dataset. In addition we should use `library(e1071)` and thus we have to install the `e1071` package.

a) Use support vector machines to classify cars into Economy and Consuming classes. For this you have to create a new variable based on the following command. `ECO = ifelse( mpg > 22.75, "Economy", "Consuming" )`. In addition, visualize the two classes on the space spanned by the variable `weight` and `horsepower`. Can we separate the two classes by a hyperplane? Is SVM applicable for this task?

b) To apply SVM we need to create a reduced dataset that will only contain `ECO`, `weight`, and `horsepower`. For this, use `d = data.frame(ECO, weight, horsepower)`. Apply SVM and summarize the results using the function `summary`. Moreover, use the `plot` function and see how does the SVM classification plot looks like. What do you observe?

c) Apply different kernels and see how do the SVM classification plots look like. What do you observe in terms of the boundaries? How do they compare with previous question?

d) The `cost` option specifies the cost of violating the margin. We can try costs 0.001, 0.01, 0.1, 1, 10, 100, 1000 and see which one provides the best performance. In addition, find the optimal kernel. The kernel options are `linear`, `polynomial`, `radial`, `sigmoid`.

e) Use the validation method to estimate the classification rate using the optimal parameters.

f) Create more categories for `ECO` and use SVM to predict the multiple classes. Use the tools from previous parts to identify the optimal kernel as well as cost.