Exercises – Topic2 – Classification using Naïve Bayes

1. Apply Naïve Bayes to the iris dataset to classify to the three types of iris – setosa, versicolor, virginica. How does the accuracy compare with that of kNN (from last week)?

Suggestions: i) remember that the dataset is arranged in the three classes in sequence, you need to shuffle and randomize before creating training and test datasets; ii) consider a 80/20 split of the data to training/ test dataset.

Make sure to install (use **install.package(" ")**) and upload (use **library( )**) the necessary packages: e1071, caTools, caret.

2. Apply Naïve Bayes to this weather dataset: https://archive.ics.uci.edu/ml/datasets/SML2010 to predict weather forecast temperature (attribute 5). You may need to clean the data, remove rows with missing data or insert an average value, convert to categorical values (like warm, cold; sunny, clouded – etc.)

Note: The data from the repository is in .txt rather than .csv format. You can use:

read.table instead of **read.csv** when importing the file to RStudio

Make sure to install (use **install.package(" ")**) and upload (use **library( )**) the necessary packages: lattice, ggplot2, e1071, caTools, caret.