Tutorial 8

1. (Naive Bayes)

   Data set `Titanic.csv` provides information on the fate of passengers on the fatal maiden voyage of the ocean liner Titanic,. It includes the variables: economic status (class), sex, age and survival. We will train a naive Bayes classifier using this data set, and predict survival.

   (a) Compute the probabilities $P(Y = 1)$ (survived) and $P(Y = 0)$ (did not survive).

   (b) Compute the conditional probabilities $P(X_i = x_i|Y = 1)$ and $P(X_i = x_i|Y = 0)$, where $i = 1, 2, 3, 4$ for the feature variables $X = \{class, sex, age\}$.

   (c) Predict survival for an adult female passenger in 2^{nd} class cabin.

   (d) Compare your prediction in (c) with the one performed by the `naiveBayes()` function in package 'e1071'.

2. (Naive Bayes + Logistic Regression, ROC, AUC)

   Consider the data set `Titanic.csv` again.

   (a) Perform logistic regression of 'Survived' on all the feature variables, called model M2.

   (b) Write down the fitted model (M2).

   (c) Interpret the coefficient of the variable 'Sex' in M2.

   (d) Interpret the coefficient of the variable 'Age' in M2.

   (e) Observe and compare the ROC curves and AUC for the two classifiers (naive Bayes from question 1 and logistic regression).