

Tutorial 8 On-site Questions

(Naive Bayes classifier)

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.

1. Compute the probabilities  $P(Y = 1)$  (survived) and  $P(Y = 0)$  (did not survive).
2. 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  $X = \{class, sex, age\}$ .
3. Predict survival for an adult female passenger in 2<sup>nd</sup> class cabin.
4. Compare your prediction in (c) with the one performed by the `naiveBayes()` function in package ‘e1071’.