Machine Learning and Data mining
Serie 4. Naïve Bayes

Prof. J. Savoy

October 25, 2021

1. In the folder "Exercises" you have access to the dataset Titanic.csv presenting information about travellers with their status (survived=1 (yes) or =0 (no)). In addition, you have the information about the class (Pclass), name (Name), gender (Sex), age (Age), sibling or spouse on board (1/0), parents or children aboard (1/0), and fare price (Fare).

   (a) In the correlation visualization, select the two features that have the most significant correlation to the target feature, Survived.

   (b) Using Naive Bayes classifier and the most two significant features to predict the Survival of the travellers.

   (c) Compare the performance of your model when using all the attributes of the travellers.
   You need to split the dataset into two disjoint sample, the training and the test set. For example, used 75% for the training sample, and the remaining 25% for the test set.

2. Using your select stock / market index and your decision criterion (binary or ternary) on the daily return of the next day or on the trend (daily return after 5/10 days), can you generate a correlation visualization of volume, and the moving average (with a period of 5, 10, 20 50 or 200).

   (a) select the two features that have the most significant correlation to the target feature, daily return.

   (b) Using Naive Bayes classifier and the most two significant features predict daily return.
   You can learn on all days except the last 100 (that will be used as the test set).