Experiment 4: Naive Bayes

September 1, 2018

1   Description

In this exercise, you will use Naive Bayes to make admission recommendations. Your dataset is a series of admission decision decisions to a nursery in Ljubljana, Slovenia. We downloaded this data from http://archive.ics.uci.edu/ml/datasets/Nursery, which you can see for more details if you're interested.

The database contains one tuple for each admission decision. The features or attributes include financial status of the parents, the number of other children in the house, etc. The first three tuples in the dataset are as follows:

usual,proper,complete,1,convenient,convenient,nonprob,recommended,recommend
smallusual,proper,complete,1,convenient,convenient,nonprob,priority,priority
usual,proper,complete,1,convenient,convenient,nonprob,not_recom,not_recom

where the first 8 values are features or attributes and the 9th value is the class assigned (i.e., the admission decision recommendation).

Your job is to build a Naive Bayes classifier that will make admission recommendations. The training data have been reformatted. Details can be found in convData.m. In particular, we use digits (instead of character strings) to represent the features and classes. We also divide the data set into two subsets: one is for training the Naive Bayes model, while the other one is for purpose of test.

2   Questions

Question 1: Estimate a Naive Bayes model using Maximum Likelihood on the training data, and use it to predict the categories of the test data. Report the accuracy of your classifier.

Question 2: Try smaller data sets for training. Use data sets of different sizes to train your Naive Bayes model. Specifically, extract smaller subsets from the given training data in a randomized manner. Show how the size of the training data impact on the accuracy of the Naive Bayes model, and give corresponding analysis.