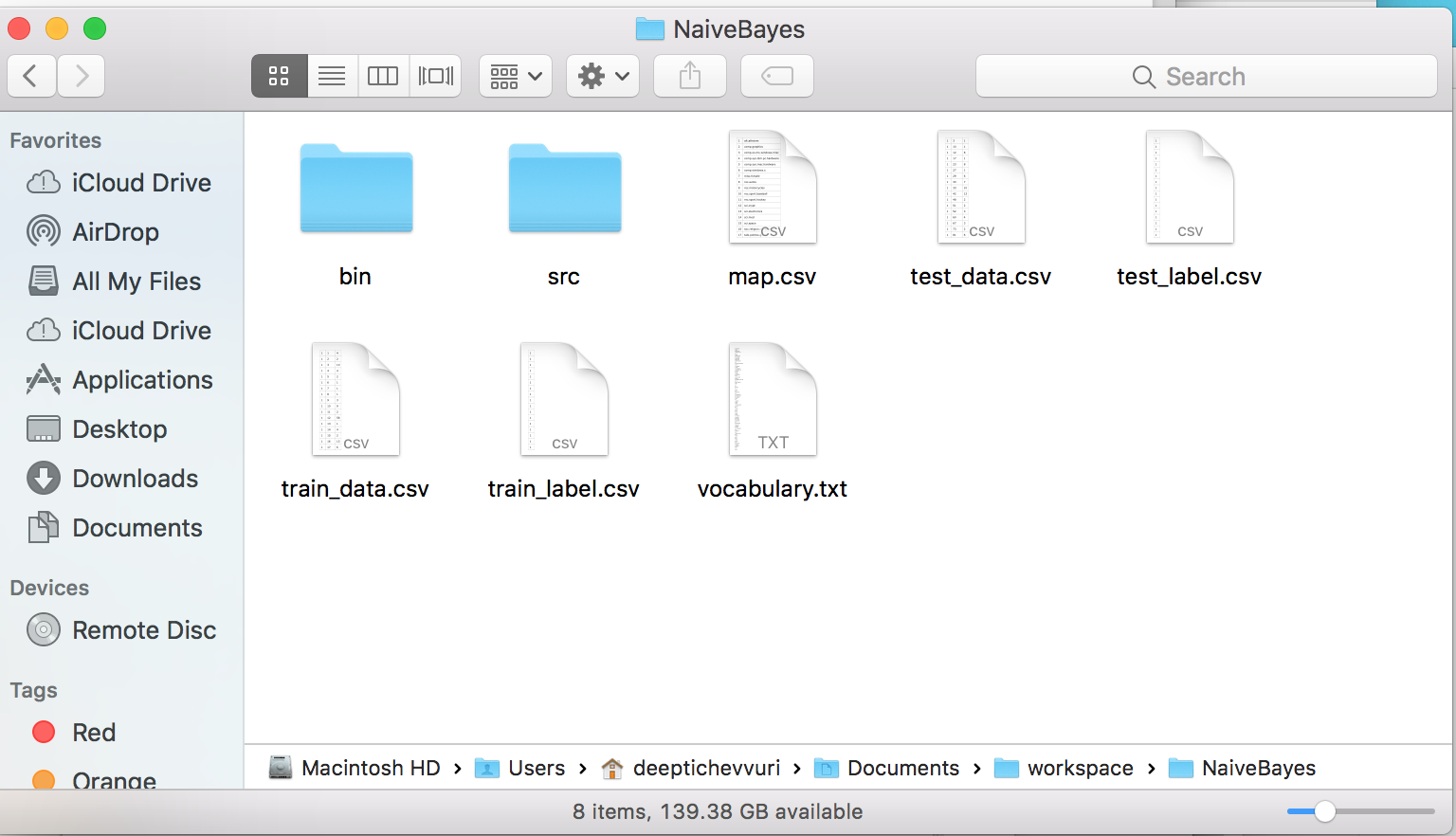
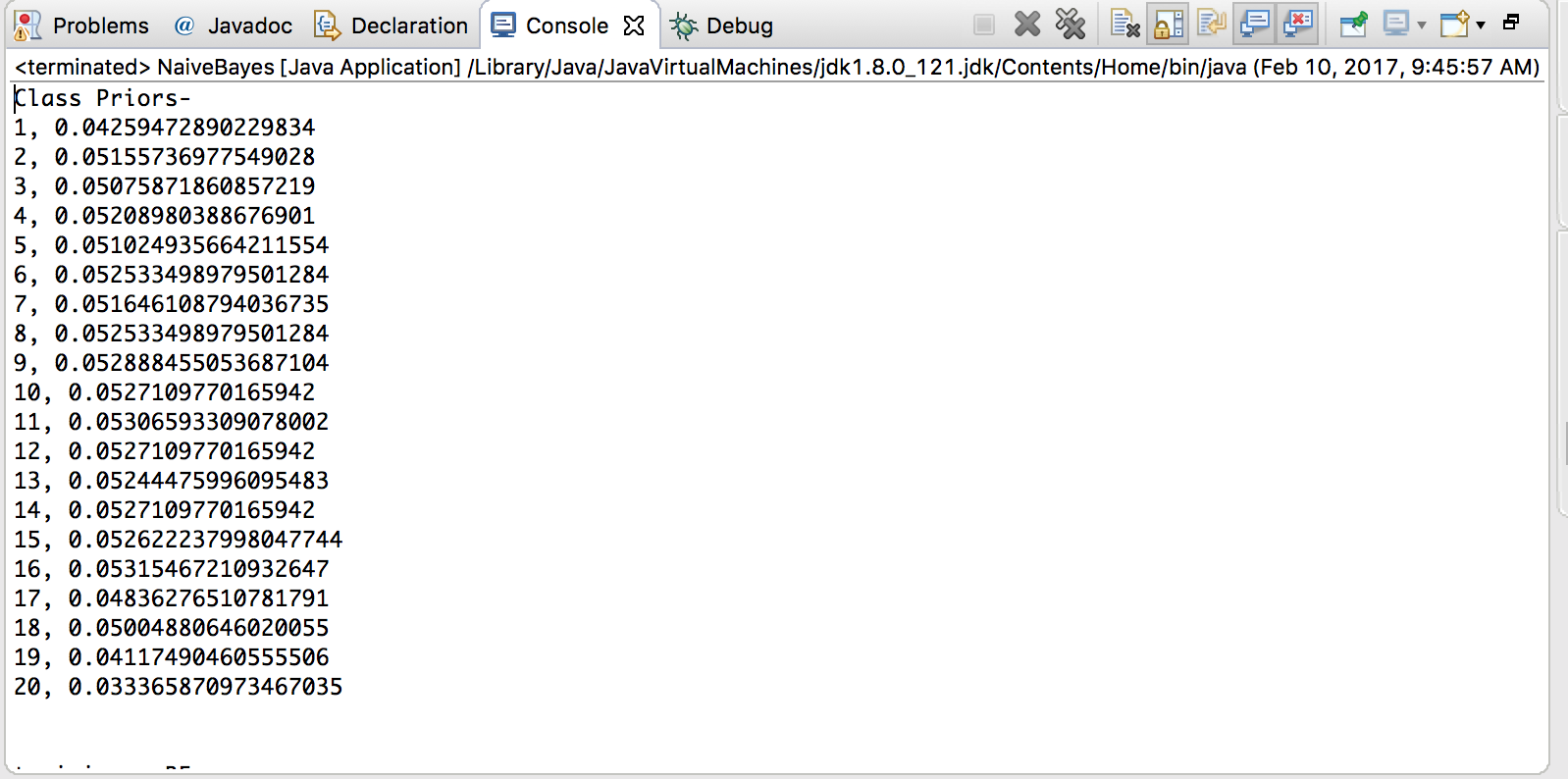
Lab 1 Report:

Please place the lab1 input files beside the src file in the project folder as shown below before executing



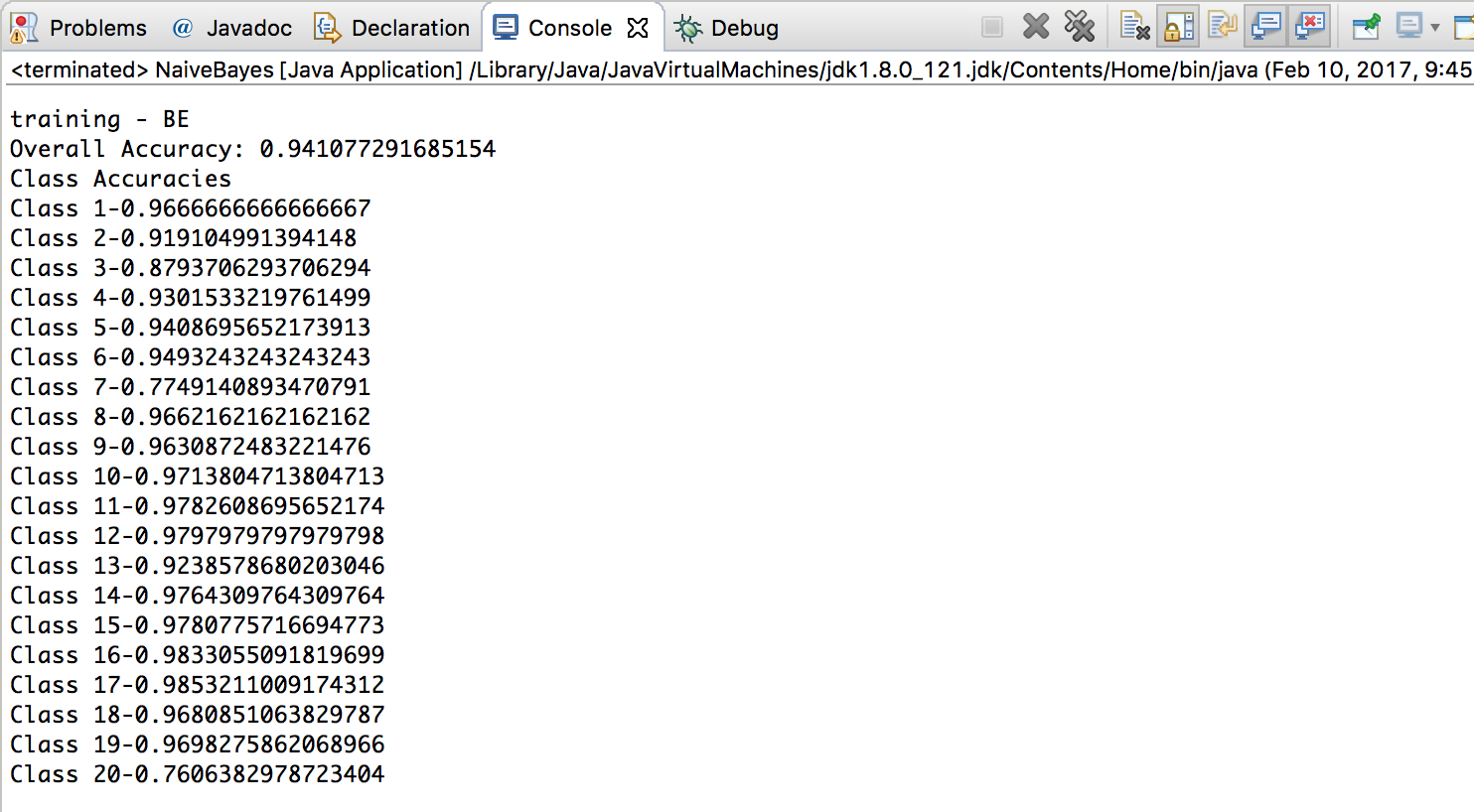
On Executing the java file, you will see the following output.

* Class priors

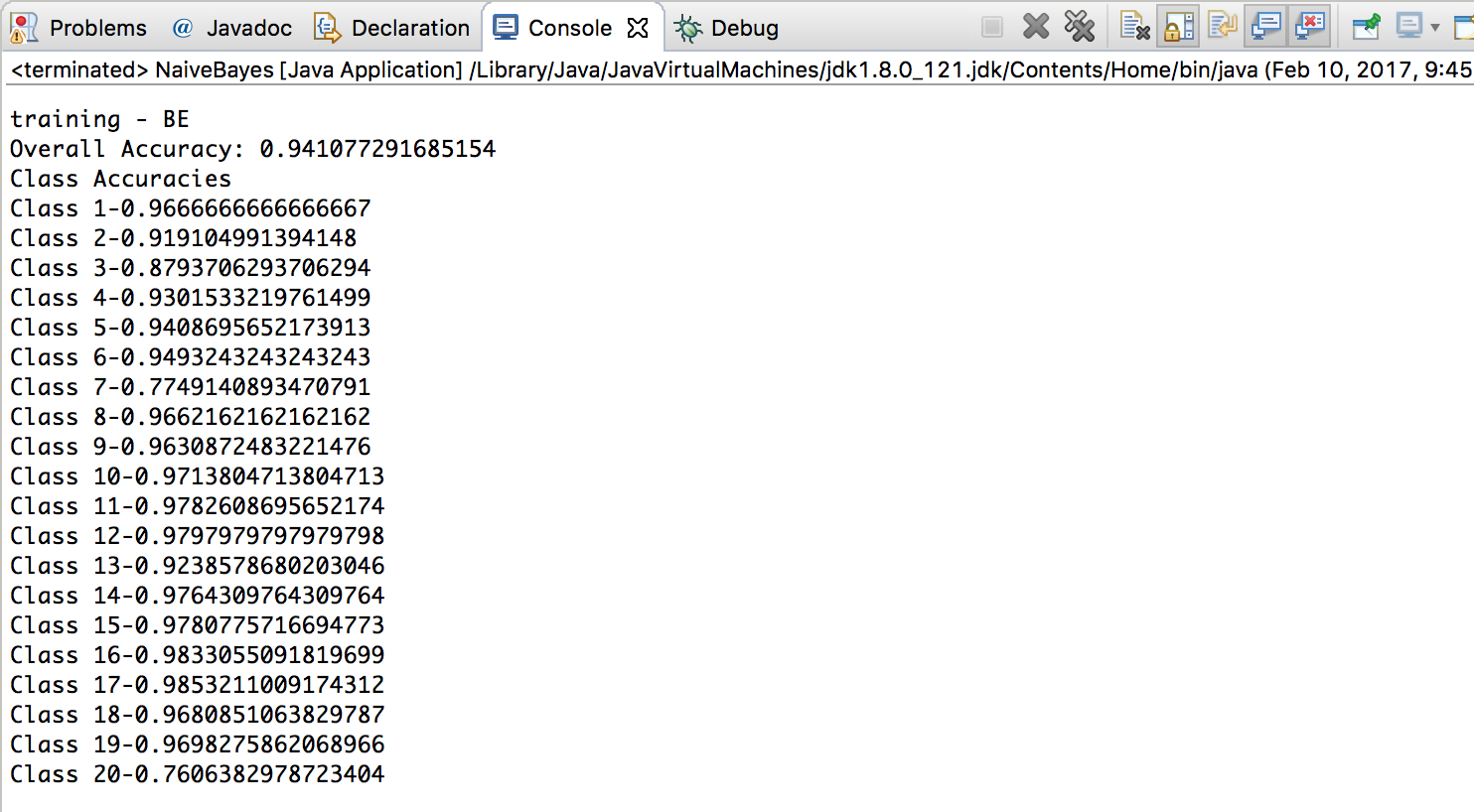
All class priors are obtained and printed, please find the screen shot below

* Performance on training data (using Bayesian estimators): overall accuracy, class accuracy, confusion matrix.

Over all accuracy and class accuracy:

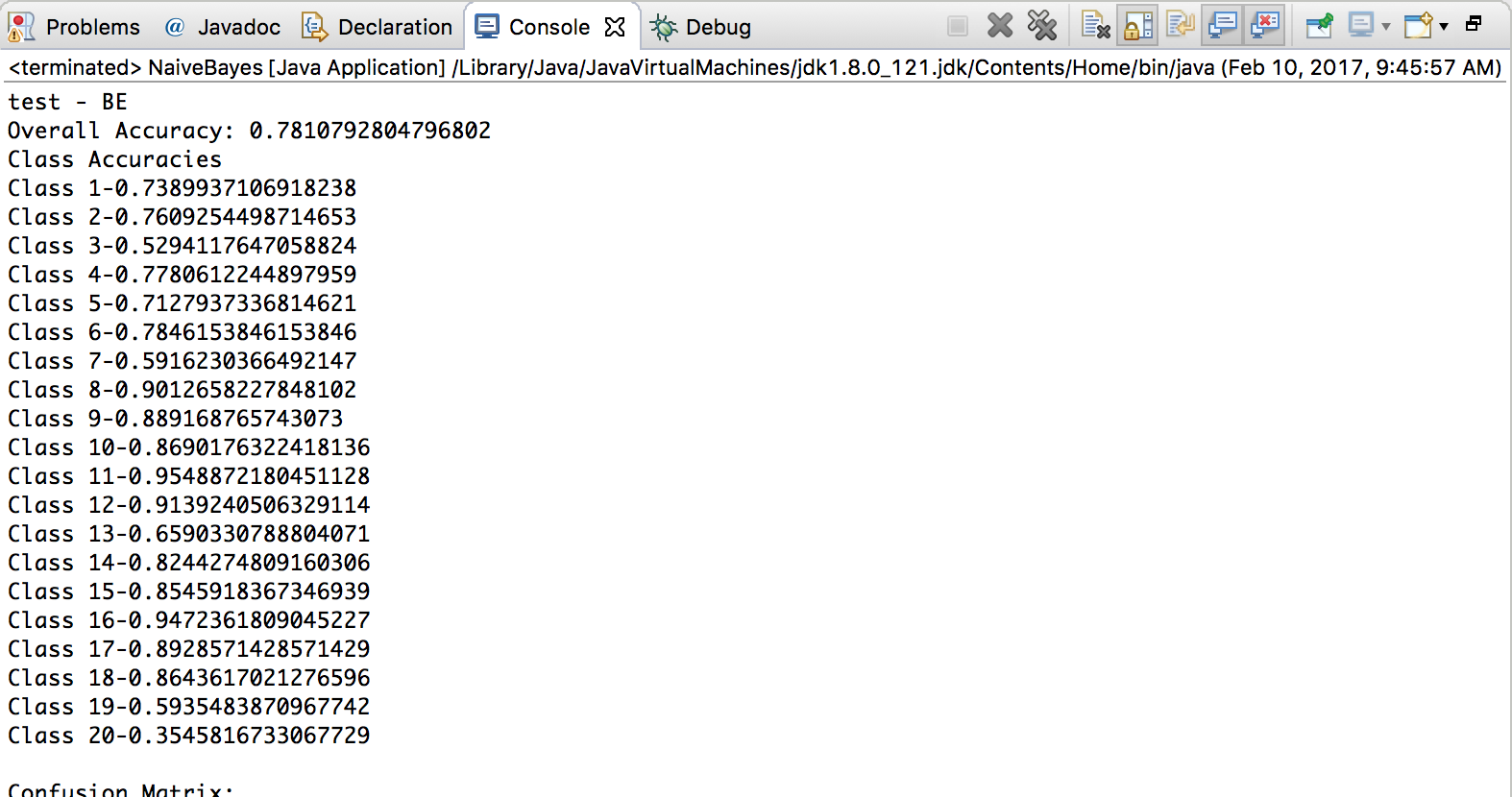


Confusion Matrix:

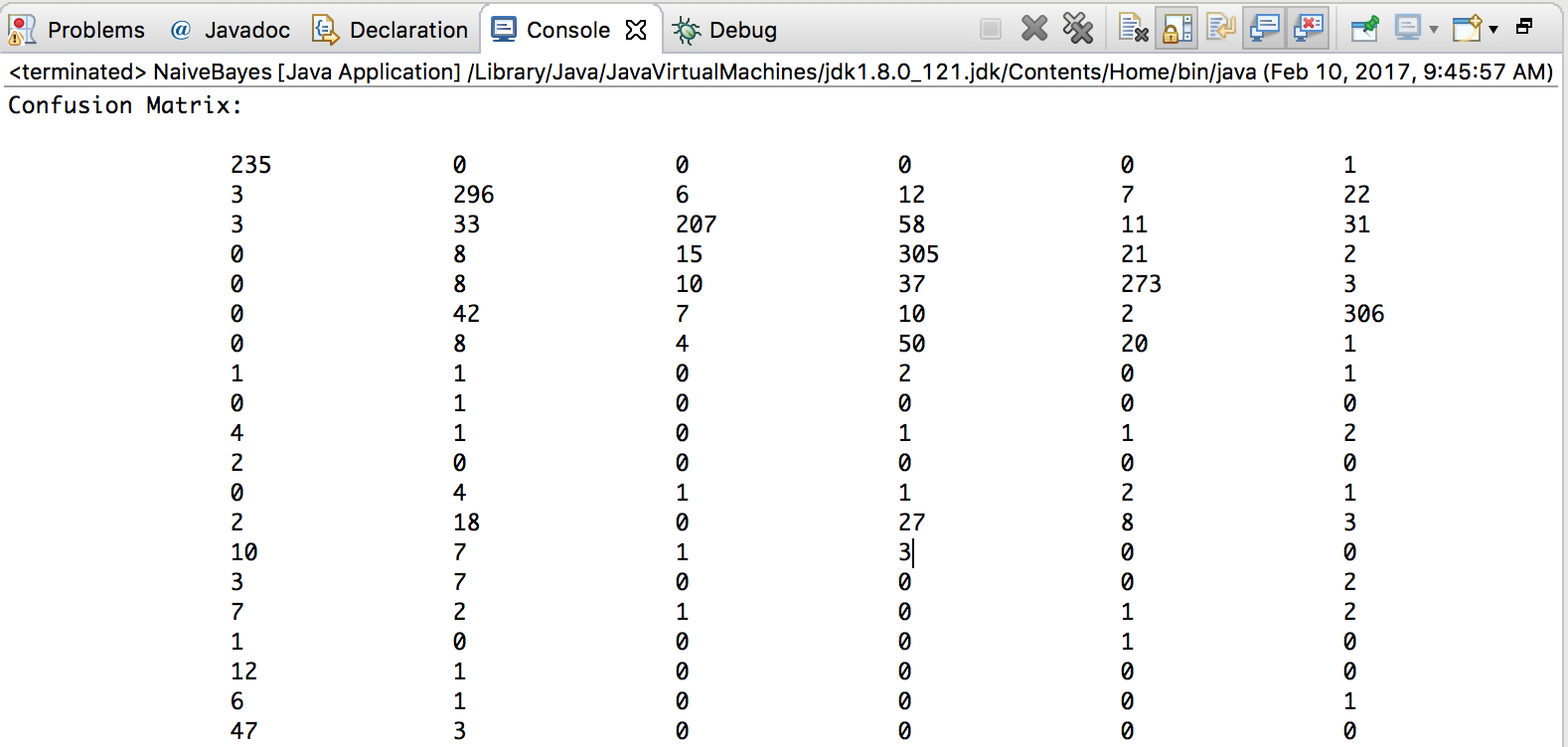


* Performance on testing data (using both MLE and BE estimators): overall accuracy, class accuracy, confusion matrix for MLE, overall accuracy, class accuracy, confusion matrix for BE.

BE: Over all accuracy and class accuracy:

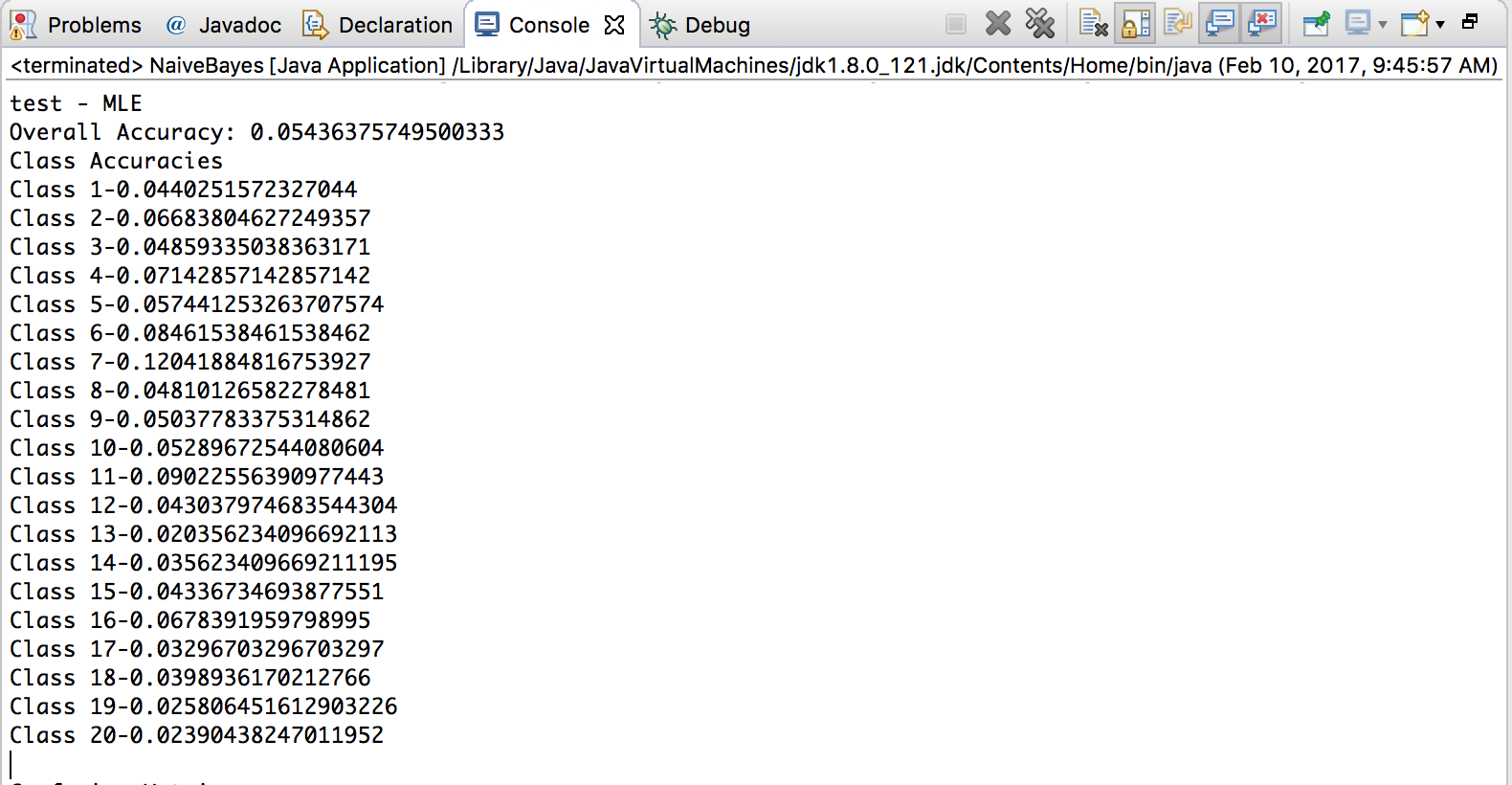


Confusion Matrix:

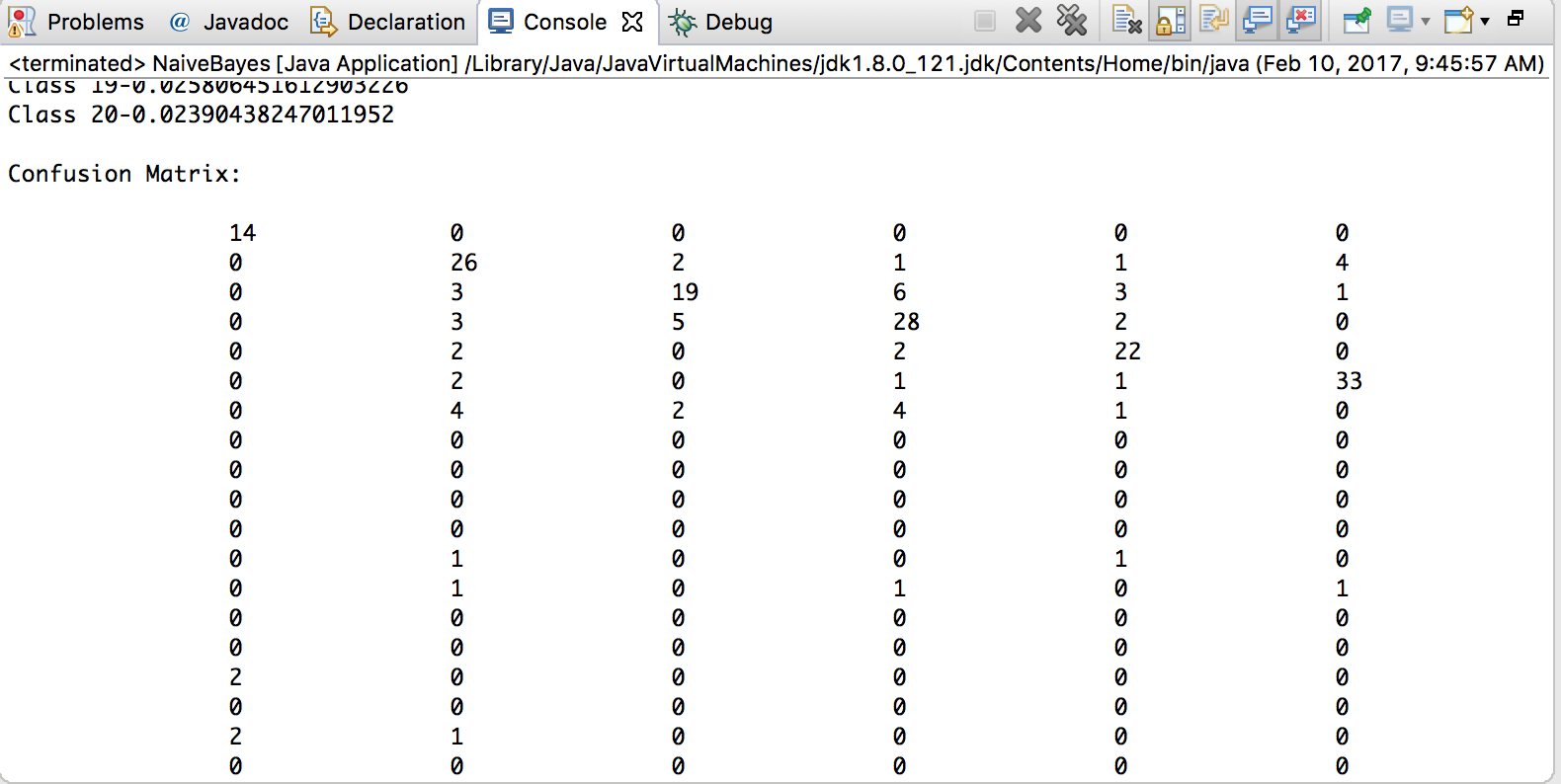


MLE: Over all accuracy and class accuracy:

Over all accuracy and class accuracy:



Confusion Matrix:



Observations:

1. The classifier’s performance is 94% with the training data sets and seems to decrease 78% (BE) and 5% with MLE with the test data set. This happens due to various reasons:

* The test data might contain documents which might fall into different classes at once.
* Or the test data might contain words which were not present in the training data set.

1. Comparing the performances of BE and MLE with the test data.

* The above point might also be a reason for the low performance of the MLE when compared to BE. BE gives better results when the test data sets contain words which are not present in training data set.
* If any words does not occur in the training data set, MLE for the corresponding probability will be zero
* But even with uniform prior, Bayesian estimate for this same probability will be non-zero

(probability estimates of zero can have b=very bad effect on the performance of any learning algorithm)

When you might expect the test data to contain words which are not present in the training data set (which happens in most of the cases)it is better to use BE over MLE.