10-605 Homework #1 Report

Chen Sun
chens1@andrew.cmu.edu

Did you receive any help whatsoever from anyone in solving this assignment? No.
Did you give any help whatsoever to anyone in solving this assignment? No.

1. What changes could you make to reduce the amount of RAM required for the dictionary?
Sort the output from NRTrain, and constructs dictionary(the hash tables one by one). Also, instead of hash table, it could be stored in a sorted array list (,which will increase the searching time).

2.Right now we're basically ignoring the fact that there are multi-labeled instances in the train/test sets. How would you extend your algorithm to enable it to predict multiple labels?
In addition to the training instance with single label, treat the multiple class as a stand-alone label, such as “Y=ACGT,ECGT” and also constructs dictionaries based on them and test on them.
Or we could record the testing result of each test documents for the all four labels. And for each label, we set a threshold value (such as average, etc) to decide whether it belongs to this label.

3.Why should we use Laplace smoothing? What will happen if we don't use any smoothing?
Laplace smoothing will give at least some probability for the instances that did not occur in the training data set. Since our data are categorical, Laplace smoothing is easy and effective to reduce the bias between observed data and assumed normal distribution. If we do not use smoothing, some instances will have probability of 0, which may lead to bias in the prediction stage.

4. What is the relationship between Laplace smoothing and Dirichlet prior?
From a Bayesian point of view, **Laplace smoothing** corresponds to the expected value of the posterior distribution, using a symmetric Dirichlet distribution with parameter α as a prior, [1] where α is the parameter in Laplace smoothing.A symmetric Dirichlet distribution is equivalent to a uniform distribution, where all the points in the space have uniform distribution.

[1] http://en.wikipedia.org/wiki/Additive_smoothing