CU6051NI - Artificial Intelligence

Lab work – 5

For this lab work, you will be working with the **Naïve Bayes Classifier** for spam detection**.** Given below are 2 tables containing training data and test data respectively.

Training examples consist of text (sms) labeled as spam or not spam. Use the examples to build the vocabulary for the classifier. Then using the ***bag of words*** approach, transform the texts into feature vectors.

Then following the algorithm for the Naïve Bayes Classifier, classify the 2 texts in the test data table as spam or not spam.

The algorithm for the Naïve Bayes Classifier is as follows (also refer to the lecture slides):

Learning: Based on the frequency counts in the dataset:

1. Estimate all $p(y)$, $\forall y \in Y$
2. Estimate all $p\left(a_j \middle| y\right)$, $\forall y \in Y$, $\forall a_j$

Classification: For a new example, use:

$$y_{new} = argmax_{y\epsilon Y}\ p(y) \prod_{j=1}^{d} p\left(a_j \middle| y\right)$$

You are required to submit your work in form of a simple report showing all the calculations that you have done and your final result. As the process will require lots of redundant work, you can write code to speed up the process.

Training Data:

Text	**Label**
Congrats, You have won!! reply to our sms for a free nokia mobile + free camcorder.	spam
Congrats! 1 year special cinema pass for 2 is yours. reply to this sms to claim your prize.	spam
I am pleased to tell you that you are awarded with a 1500 Bonus Prize, reply to this sms to claim your prize.	spam
Dont worry. I guess he is busy.	not spam
Going for dinner. msg you later.	not spam
Ok, I will call you up when I get some cash.	not spam

Test Data:

Text	**Label**
I am busy. I will msg you later.	?
Congrats! You are awarded a free mobile.	?