Homework 2

Due 11/13/2017 11:59 pm

1 THE NAÏVE BAYES CLASSIFIER

1.1 MOVIE REVIEW CLASSIFICATION USING NAÏVE BAYES

Assume that you have trained a Naïve Bayes classifier for the task of sentiment classification (please refer to Chapter 6, pp. 1-9 in the J&M book). The classifier uses only bag-of-word features. Assume the following parameters for each word being part of a positive or negative movie review, and the prior probabilities are 0.4 for the positive class and 0.6 for the negative class.

	pos	neg
I	0.09	0.16
always	0.07	0.06
like	0.29	0.06
foreign	0.04	0.15
films	0.08	0.11

Question: What class will Naïve Bayes assign to the sentence "I always like foreign films"? **Show your work.**

1.2 TRAINING THE NAÏVE BAYES CLASSIFIER FOR MOVIE REVIEW CLASSIFICATION

1. Implement in Python a Naïve Bayes classifier with bag-of-word features and add-1 smoothing. Note: Smoothing should be used for the context features (bag-of-word features) only. Do not use smoothing for the prior parameters.

2. Use the following small corpus of movie reviews to train your classifier. Save the parameters of your model in a file called movie-review.NB

   a) fun, couple, love, love **comedy**

   b) fast, furious, shoot **action**

   c) couple, fly, fast, fun, fun **comedy**

   d) furious, shoot, shoot, fun **action**

   e) fly, fast, shoot, love **action**

3. Test you classifier on the new document below: *{fast, couple, shoot, fly}*. Compute the most likely class. Report the probabilities for each class.

2 PART-OF-SPEECH TAGGING

Do exercise 10.1 at the end of Chapter 10 of the book:
`https://web.stanford.edu/j̃urafsky/slp3/10.pdf`

3 PARSING

Do exercise 11.2 at the end of Chapter 11 of the book:
`https://web.stanford.edu/j̃urafsky/slp3/11.pdf`

3.1 SUBMISSION

Please place the following on the server venus.cs.qc.edu and email me the path to the directory:

1. The Python code along with a README file that has instructions on how to run it in order to obtain the answers to questions in Section 1.2

2. The writeup that should be named **Homework2** that includes the answers to the questions.

Your grade will be based on the *correctness* of your answers, the *clarity* and completeness of your responses, and the *quality* of the code that you submitted.
Please refer to the course webpage on late submission policy.