Natural Language Processing
Lab 3

January 29, 2024

This lab sheet is to practice the concepts taught this week so far: Naive Bayes classification and Sentiment Analysis.

1. Assume the following likelihoods for each word being part of a positive or negative movie review, and equal prior probabilities for each class.

	POS	NEG
I	0.09	0.16
usually	0.07	0.06
enjoy	0.29	0.06
foreign	0.04	0.15
film	0.08	0.11

   What class will a Naive Bayes classifier assign to the sentence "I usually enjoy foreign film"? Show your workings.

2. Given the following short, feature-selected, movie reviews, each labelled with a genre, either comedy or action:

   1. fun, couple, love, love **comedy**
   2. fast, furious, shoot **action**
   3. couple, fly, fast, fun, fun **comedy**
   4. furious, shoot, shoot, fun **action**
   5. fly, fast, shoot, love **action**

   and a new document D:

   - fast, couple, shoot, fly

   compute the most likely class for D. Assume a Naive Bayes classifier and use add-1 smoothing for the likelihoods.

3. Why might a Naive Bayes classifier be used as a baseline classifier in sentiment analysis?

4. Why is the 'Naive' assumption made in Naive Bayes, and how does it affect sentiment classification?