CS114 (Spring 2020) Written Assignment 1
Text Classification

Due February 14, 2020

1 Naïve Bayes

The following problem is from the Jurafsky and Martin book, Exercise 4.2, reproduced below.

Given the following short movie reviews, each labeled with a genre, either comedy or action:

document	class
fly fast shoot love	action
fun couple love love	comedy
fast furious shoot	action
couple fly fast fun fun	comedy
furious shoot shoot fun	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.

Note: **Show your work!** In particular, show all of the probability distributions involved in the model (namely, P(class) and P(feature|class)) and all of the steps used to calculate them. It is recommended that you create (conditional) probability tables such as those shown below.

class	$P(\text{class})$
comedy	
action	

$P(\text{feature}\|\text{class})$		class	
		comedy	action
feature	fly		
	fast		
	shoot		
	love		

2 Logistic Regression

(You may find the discussion in Chapter 5 of the Jurafsky and Martin book helpful.)

Next, we will train a logistic regression classifier, using the same five training examples as before. We will use the word counts of "fast", "couple", "shoot", and "fly" as our features x_{fast}, x_{couple}, x_{shoot}, and x_{fly}, respectively, but we will not do any smoothing. Suppose also that we map the class "comedy" to $y = 1$ and the class "action" to $y = 0$. Finally, we will initialize our feature weights and bias term to 0: $w_{\text{fast}} = w_{\text{couple}} = w_{\text{shoot}} = w_{\text{fly}} = b = 0$.

We will train our classifier using gradient descent. We divide our training data into two mini-batches: the first three examples in the first batch, and the last two in the second batch.

1. Process the first mini-batch.

   (a) First, compute the gradient for each example in the mini-batch.

$$\nabla = \begin{bmatrix} \frac{\partial L}{\partial w_{\text{fast}}} \\ \frac{\partial L}{\partial w_{\text{couple}}} \\ \frac{\partial L}{\partial w_{\text{shoot}}} \\ \frac{\partial L}{\partial w_{\text{fly}}} \\ \frac{\partial L}{\partial b} \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

Then, compute the overall gradient. We will follow Jurafsky and Martin, and say that the mini-batch gradient is the average of the individual gradients.

(b) Update the weight vector. Use a constant learning rate $\eta = 0.1$.

$$\theta = \begin{bmatrix} w_{\text{fast}} \\ w_{\text{couple}} \\ w_{\text{shoot}} \\ w_{\text{fly}} \\ b \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

2. Next, process the second mini-batch. Compute the gradient and update the weight vector as before, using the weights and bias you learned in step 1.

3. Suppose we stop training here (i.e., after one epoch). You are given the same document D as before:

   fast couple shoot fly

   Compute the probability that D has the class "comedy".

Reminder: **Show your work!**

Submission Instructions

Please submit your solutions (in PDF format) to LATTE.