

CS114B (Spring 2022) Lab 3 Exercise

Logistic Regression

February 10, 2022

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

Given the following short movie reviews (from the Jurafsky and Martin book, Exercise 4.2), 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

we will train a logistic regression classifier. 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 i in the mini-batch.

$$(\nabla L)^{(i)} = \begin{bmatrix} \left(\frac{\partial L}{\partial w_{\text{fast}}} \right)^{(i)} \\ \left(\frac{\partial L}{\partial w_{\text{couple}}} \right)^{(i)} \\ \left(\frac{\partial L}{\partial w_{\text{shoot}}} \right)^{(i)} \\ \left(\frac{\partial L}{\partial w_{\text{fly}}} \right)^{(i)} \\ \left(\frac{\partial L}{\partial b} \right)^{(i)} \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

Then, compute the overall gradient. 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 a new document D:

fast couple shoot fly

Compute the probability that D has the class “comedy”.