Assignment 4: Make-up Assignment
Due Date: 11.59 PM 12/12/2023 Maximum Points: 80

Please follow the instructions. Failure to adhere to instructions will result in a penalty.

1 Question 1 (40 pts)

We are going to design an end-to-end neural network for Sentiment Analysis: Refer to the chatbot code (in blackboard) to answer these questions. Please note that the chatbot code is many-to-many and sentiment analysis is many-to-one: We have the following sentences as dataset with the corresponding labels:

- *This restaurant is trash.* Label: **Negative**
- *Taco Bell serves better Mexican food than this.* Label: **Negative**
- *The food is fabulous.* Label: **Positive**
- *Good food, looong wait for the food.* Label: **Neutral**

Now answer the following questions:

- What is your 'X' and 'Y' here? (3 pts)
- The first step is building your vocabulary. What would the vocabulary/dictionary be? (6 pts) [You can remove common words/stopwords such as 'The','is','for','this',etc.]
- Convert all your sentences to numerical values using the indexes in the dictionary. (Do not forget to add the additional tokens you will require for the vocabulary) (6 pts)
- The next step is selecting a max length of the sentences and ensure that all sentences are of the same length. What would be your dataset after doing this? (6 pts)
- The next step is to convert each word to a one-hot vector. What would be the shape of the vector representing each word after doing this? (3 pts)
- Look at Figure 1. If you are confused, read the description to know about what the symbols represent. Consider that the hidden representation is 4 dimensions. So, *h(0)* will be [0,0,0,0] as per our assumption. What would be the dimension of W_{xh}, W_{hh}, and W_{hy}? (3 pts)
- Assume W_{xh} and W_{hh} at the beginning of the training are all '1's. i.e, if the shape of W_{xh} is 3X4 it will look like [[1,1,1,1],[1,1,1,1],[1,1,1,1]]. The same is true for W_{hh}. Calculate h(1), using the words from the first sentence. Assume that you are using ReLu activation. (6 pts)
- Use the formula for softmax in the slides. Let us say you get the values [1.3,0.8,-0.1] for Y after you multiply h(3) with W_{hy}. (For the first sentence) What are the values after you use a softmax activation? Use your calculator to get the values. (3 pts)

- The formula for cross-entropy is as follows: $-\sum_{NumClasses} y_t * log(y_p)$ where y_p are the values that you get from the softmax. y_t represents the actual classes. For the first sentence, $y_t = [1, 0, 0]$ (the first index corresponds to the Negative class, 2nd index = Positive, 3rd index = Neutral). Use the values for y_p that you get after applying softmax and the y_t described above for the first sentence to compute the loss. (4 pts)

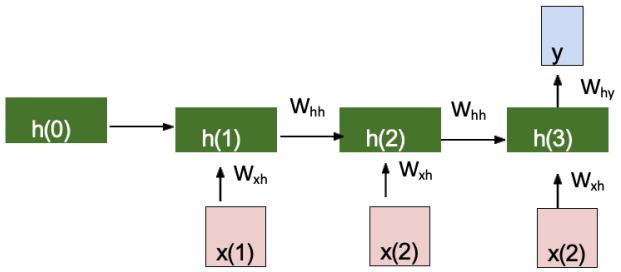
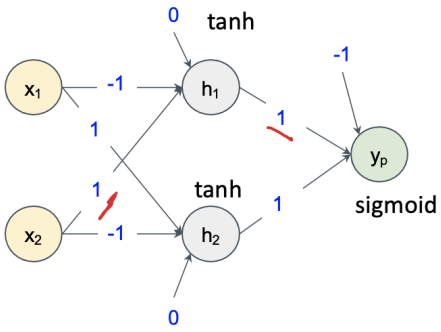


Figure 1: This is what the model would look like if your sentence has 3 words. X(i) is the ith word. h(0) is the first hidden vector, usually all 0's. Y is your predicted label.

2 Question 2 (40 pts)

:

We are going to solve a similar question that we have seen before in the mid-term. The difference here is that the hidden units use tanh activations and the final unit that gives y uses a sigmoid activation. Use this website to get the values of tanh. Use this to calculate the values of sigmoid.



x_1	x_2	h_1	h_2	yt	yp
0.5	0.5			.25	
0	1			.75	
1	0			.85	
1	1			.35	

Figure 2: The neural network and the corresponding table in Q2. Some of the weights are marked with red.

- Calculate the values of h1, h2, yp for the inputs provided. Fill the table (20 pts)
- There are two weights marked with red. Recall that by using Finite Differences we can approximate the gradient. Use this formula to calculate the loss for all the samples:

  $Loss(L) = -\sum_{allsamples} y_t * log(y_p) + (1 - y_t) * log(1_{yp})$

  $\frac{\partial L}{\partial w} \approx \frac{L[w+\delta] - L[w]}{\delta}$

  Let us call w_{21} as the weight connecting h1 and yp and w_{13} as the weight connecting x2 and h1. Calculate the gradients for these two sets of weights. (20 pts)