Question

Given the following tweet and frequencies , what is the value of the sum of negative frequencies?

$$\sum_{w} freqs(w, 0)]$$

Vocabulary	PosFreq (1)	NegFreq (0)
I	3	3
am	3	3
happy	2	0
because	1	0
learning	1	1
NLP	1	1
sad	0	2
not	0	1

I am sad, I am not learning NLP

- (selected) 11
- 17
- 8
- 7

Correct

That's right.

Question

Given the following list of tweets, how many parameters would a logistic regression classifier have to learn with inputs represented as discussed in this lecture? (*Remember to add in 1 as well because of the bias*)

[“I am happy because I am learning NLP”, “I hated that movie”, “I love working at DL”]

- (●) 14
- ( ) 18
- ( ) 42
- ( ) 4

✓ **Correct**

That's right.

1. When performing logistic regression on sentiment analysis, you represented each tweet as a vector of ones and zeros. However your model did not work well. Your training cost was reasonable, but your testing cost was just not acceptable. What could be a possible reason?

1 / 1 point

- (●) The vector representations are sparse and therefore it is much harder for your model to learn anything that could generalize well to the test set.
- ( ) You probably need to increase your vocabulary size because it seems like you have very little features.
- ( ) Logistic regression does not work for sentiment analysis, and therefore you should be looking at other models.
- ( ) Sparse representations require a good amount of training time so you should train your model for longer

Correct
This is correct.

2. Which of the following are examples of text preprocessing?

1 / 1 point

- [x] Stemming, or the process of reducing a word to its word stem.

Correct
This is correct.

- [x] Lowercasing, which is the process of removing changing all capital letter to lower case.

Correct
This is correct.

- [x] Removing stopwords, punctuation, handles and URLs

Correct
This is correct.

- [ ] Adding new words to make sure all the sentences make sense

3. The sigmoid function is defined as $h(x^{(i)}, \theta) = \frac{1}{1+e^{-\theta^T x^{(i)}}}$. Which of the following is true.

1 / 1 point

- ○ Large positive values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ closer to 1 and large negative values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ close to -1.
- ◉ Large positive values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ closer to 1 and large negative values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ close to 0.
- ○ Small positive values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ closer to 1 and large positive values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ close to 0.
- ○ Small positive values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ closer to 0 and large negative values of $\theta^T x^{(i)}$ will make $h(x^{(i)}, \theta)$ close to -1.

Correct

This is correct.

4. The cost function for logistic regression is defined as $J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log h\left(x^{(i)},\theta\right) + \left(1-y^{(i)}\right)\log\left(1-h\left(x^{(i)},\theta\right)\right)\right]$. Which of the following is true about the cost function above. Mark all the correct ones.

1 / 1 point

- [x] When $y^{(i)} = 1$, as $h(x^{(i)}, \theta)$ goes close to 0, the cost function approaches ∞.

Correct
This is correct.

- [ ] When $y^{(i)} = 1$, as $h(x^{(i)}, \theta)$ goes close to 0, the cost function approaches 0.
- [x] When $y^{(i)} = 0$, as $h(x^{(i)}, \theta)$ goes close to 0, the cost function approaches 0.

Correct
This is correct.

- [ ] When $y^{(i)} = 0$, as $h(x^{(i)}, \theta)$ goes close to 0, the cost function approaches ∞.

5. For what value of $\theta^T x$ in the sigmoid function does $h(x^{(i)}, \theta) = 0.5$.

1 / 1 point

0

Correct

6. Select all that apply. When performing logistic regression for sentiment analysis using the method taught in this week's lecture, you have to:

1 / 1 point

- [x] Performing data processing.

Correct
This is correct.

- [x] Create a dictionary that maps the word and the class that word is found in to the number of times that word is found in the class.

Correct
This is correct.

- [ ] Create a dictionary that maps the word and the class that word is found in to see if that word shows up in the class.

- [x] For each tweet, you have to create a **positive feature** with the sum of positive counts of each word in that tweet. You also have to create a **negative feature** with the sum of negative counts of each word in that tweet.

Correct
This is correct.

7. When training logistic regression, you have to perform the following operations in the desired order.

1 / 1 point

- ○ Initialize parameters, get gradient, classify/predict, update, get loss, repeat
- ◉ Initialize parameters, classify/predict, get gradient, update, get loss, repeat
- ○ Initialize parameters, get gradient, update, classify/predict, get loss, repeat
- ○ Initialize parameters, get gradient, update, get loss, classify/predict, repeat

Correct

This is correct.

9. What is the purpose of gradient descent? Select all that apply.

1 / 1 point

- [x] Gradient descent allows us to learn the parameters θ in logistic regression as to minimize the loss function J.

Correct
This is correct.

- [ ] Gradient descent allows us to learn the parameters θ in logistic regression as to maximize the loss function J.

- [x] Gradient descent, *grad_theta* allows us to update the parameters θ by computing $\theta = \theta - \alpha * grad\_theta$

Correct
This is correct.

- [ ] Gradient descent, *grad_theta* allows us to update the parameters θ by computing $\theta = \theta + \alpha * grad\_theta$

10. What is a good metric that allows you to decide when to stop training/trying to get a good model? Select all that apply.

1 / 1 point

- [x] When your accuracy is good enough on the test set.

Correct
This is correct.

- [ ] When your accuracy is good enough on the train set.

- [x] When you plot the cost versus (# of iterations) and you see that your the loss is converging (i.e. no longer changes as much).

Correct
This is correct.

- [ ] When α, your step size is neither too small nor too large.

8. Assuming we got the classification correct, where $y^{(i)} = 1$ for some specific example i. This means that $h(x^{(i)}, \theta) > 0.5$. Which of the following has to hold:

0 / 1 point

- Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is exactly equal to its corresponding label $y^{(i)}$.
- Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is less than $(1 - y^{(i)})$.
- Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is less than $(1 - h(x^{(i)}, \theta))$.
- Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is greater than $(1 - h(x^{(i)}, \theta))$.

Incorrect

This is not correct.

8. Assuming we got the classification correct, where $y^{(i)} = 1$ for some specific example i. This means that $h(x^{(i)}, \theta) > 0.5$. Which of the following has to hold:

1 / 1 point

- ○ Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is exactly equal to its corresponding label $y^{(i)}$.
- ○ Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is less than $(1 - y^{(i)})$.
- ○ Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is less than $(1 - h(x^{(i)}, \theta))$.
- ◉ Our prediction, $h(x^{(i)}, \theta)$ for this specific training example is greater than $(1 - h(x^{(i)}, \theta))$.

Correct

This is correct.