Question

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



Vocabulary	PosFreq (1)	NegFreq (0)
1	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,	l am not	learning	NLP
-----------	----------	----------	-----

1117

0.

0 7

✓ Correct That's right.



zer	en performing logistic regression on sentiment analysis, you represented each tweet as a vector of ones and os. However your model did not work well. Your training cost was reasonable, but your testing cost was just not eptable. What could be a possible reason?	
•	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.	
0	You probably need to increase your vocabulary size because it seems like you have very little features.	
0	Logistic regression does not work for sentiment analysis, and therefore you should be looking at other models.	
0	Sparse representations require a good amount of training time so you should train your model for longer	
(~	Correct This is correct.	

1.

1 / 1 point

Which of the following are examples of text preprocessing?
Stemming, or the process of reducing a word to its word stem.
✓ Correct This is correct.
✓ Lowercasing, which is the process of removing changing all capital letter to lower case.
✓ Correct This is correct.
Removing stopwords, punctuation, handles and URLs
Adding new words to make sure all the sentences make sense

2.

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

This is correct.

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

This is correct.

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

0

⊘ Correct

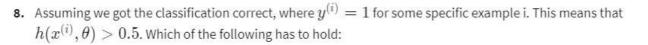
7.	When training logistic regression, you have to perform the following operations in the desired order.	1 / 1 point
	O Initialize parameters, get gradient, classify/predict, update, get loss, repeat	
	Initialize parameters, classify/predict, get gradient, update, get loss, repeat	
	O Initialize parameters, get gradient, update, classify/predict, get loss, repeat	
	O Initialize parameters, get gradient, update, get loss, classify/predict, repeat	
	✓ Correct This is correct.	

What is the purpose of gradient descent? Select all that apply.
lacksquare Gradient descent allows us to learn the parameters $ heta$ in logistic regression as to minimize the loss function J.
Correct This is correct.
\square Gradient descent allows us to learn the parameters $ heta$ in logistic regression as to maximize the loss function J.
Gradient descent, $\mathit{grad_theta}$ allows us to update the parameters θ by computing $\theta = \theta - \alpha * \mathit{grad_theta}$
○ Correct This is correct.
\Box Gradient descent, $\mathit{grad_theta}$ allows us to update the parameters θ by computing $\theta = \theta + \alpha * \mathit{grad_theta}$

9.

1/1 point

10. What is a apply.	a good metric that allows you to decide when to stop training/trying to get a good model? Select all that
✓ Whe	en your accuracy is good enough on the test set.
\sim	orrect nis is correct.
☐ Whe	en your accuracy is good enough on the train set.
	en you plot the cost versus (# of iterations) and you see that your the loss is converging (i.e. no longer nges as much).
\sim	orrect nis is correct.
☐ Whe	en $lpha$, your step size is neither too small nor too large.



Our prediction, $h(x^{(i)}, heta)$ for this specific training example is exactly equal to its corresponding label $y^{(i)}$.

Our prediction, $h(x^{(i)}, heta)$ 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 $(1 h(x^{(i)}, \theta))$.

(X) Incorrect

This is not correct.

