1 point

1. In the function **feature_derivative_with_L2**, was the intercept term regularized?

- ( ) Yes
- (•) No

---

1 point

2. Does the term with L2 regularization increase or decrease the log likelihood $\ell\ell(\mathbf{w})$?

- ( ) Increases
- (•) Decreases

---

1 point

3. Which of the following words is **not** listed in either **positive_words** or **negative_words**?

- ( ) love
- ( ) disappointed
- ( ) great
- ( ) money
- (•) quality

---

1 point

4. Questions 5 and 6 use the coefficient plot of the words in **positive_words** and **negative_words**.

(True/False) All coefficients consistently get smaller in size as the L2 penalty is increased.

- (•) True
- ( ) False

---

1 point

5. Questions 5 and 6 use the coefficient plot of the words in **positive_words** and **negative_words**.

(True/False) The relative order of coefficients is preserved as the L2 penalty is increased. (For example, if the coefficient for 'cat' was more positive than that for 'dog', this remains true as the L2 penalty increases.)

- ( ) True
- (•) False

---

1 point

6. Questions 7, 8, and 9 ask you about the 6 models trained with different L2 penalties.

Which of the following models has the **highest** accuracy on the **training** data?

- (•) Model trained with L2 penalty = 0
- ( ) Model trained with L2 penalty = 4
- ( ) Model trained with L2 penalty = 10
- ( ) Model trained with L2 penalty = 100
- ( ) Model trained with L2 penalty = 1e3
- ( ) Model trained with L2 penalty = 1e5

---

1 point

7. Questions 7, 8, and 9 ask you about the 6 models trained with different L2 penalties.

Which of the following models has the **highest** accuracy on the **validation** data?

- ( ) Model trained with L2 penalty = 0
- ( ) Model trained with L2 penalty = 4
- (•) Model trained with L2 penalty = 10
- ( ) Model trained with L2 penalty = 100
- ( ) Model trained with L2 penalty = 1e3
- ( ) Model trained with L2 penalty = 1e5

---

1 point

8. Questions 7, 8, and 9 ask you about the 6 models trained with different L2 penalties.

Does the **highest** accuracy on the **training** data imply that the model is the best one?

- ( ) Yes
- (•) No