**2. In the function feature\_derivative\_with\_L2, was the intercept term regularized?**

Yes

**No**

**3. Does the term with L2 regularization increase or decrease the log likelihood ℓℓ(w)?**

Increases

**Decreases**

**4. Which of the following words is not listed in either positive\_words or negative\_words?**

love

disappointed

great

money

**quality**

**5. Questions 5 and 6 use the coefficient plot of the words in positive\_words andnegative\_words.**

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

**True**

False

**6. Questions 5 and 6 use the coefficient plot of the words in positive\_words andnegative\_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**

**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 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

**8. 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

**9. 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**