

1
point

1. How many reviews in **amazon_baby_subset.gl** contain the word **perfect**?

2955

1
point

2. Consider the **feature_matrix** that was obtained by converting our data to NumPy format.

How many features are there in the **feature_matrix**?

194

1
point

3. Assuming that the intercept is present, how does the number of features in **feature_matrix** relate to the number of features in the logistic regression model? Let x = [number of features in feature_matrix] and y = [number of features in logistic regression model].

- ☐ $y = x - 1$
- ☒ $y = x$
- ☐ $y = x + 1$
- ☐ None of the above

1
point

4. Run your logistic regression solver with provided parameters.

As each iteration of gradient ascent passes, does the log-likelihood increase or decrease?

- ☒ It increases.
- ☐ It decreases.
- ☐ None of the above

1
point

5. We make predictions using the weights just learned.

How many reviews were predicted to have positive sentiment?

25126

1
point

6. What is the accuracy of the model on predictions made above? (round to 2 digits of accuracy)

0.75

1
point

7. We look at "most positive" words, the words that correspond most strongly with positive reviews.

Which of the following words is **not** present in the top 10 "most positive" words?

- ☐ love
- ☐ easy
- ☐ great
- ☐ perfect
- ☒ cheap

1
point

8. Similarly, we look at "most negative" words, the words that correspond most strongly with negative reviews.

Which of the following words is **not** present in the top 10 "most negative" words?

- ☒ need
- ☐ work
- ☐ disappointed
- ☐ even
- ☐ return