Exploring precision and recall

14 试题

1 point

1.

Are you using GraphLab Create? Please make sure that

1. You are using version 1.8.3 of GraphLab Create. Verify the version of GraphLab Create by running

```
graphlab.version
```

inside the notebook. If your GraphLab version is incorrect, see this post (https://www.coursera.org/learn/ml-classification/supplement/LgZ3I/installing-correct-version-of-graphlab-create) to install version 1.8.3. **This assignment is not guaranteed to work with other versions of GraphLab Create.**

2. You are using the IPython notebook named module-9-precision-recall-assignment-blank.ipynb obtained from the associated reading.

This question is ungraded. Check one of the three options to confirm.

- I confirm that I am using the right version of GraphLab Create and the right IPython notebook.
- I am using scikit-learn.
- I am using tools other than GraphLab or scikit-learn, and I understand that I may not be able to complete some of the quiz questions.

1 point

2.
Consider the logistic regression model trained on **amazon_baby.gl** using GraphLab Create.

Using accuracy as the evaluation metric, was our **logistic regression model** better than the **majority class classifier**?

- ○ Yes
- ○ No

---

1 point

3. How many predicted values in the **test set** are **false positives**?

1443

---

1 point

4.
Consider the scenario where each false positive costs $100 and each false negative $1.

Given the stipulation, what is the cost associated with the logistic regression classifier's performance on the **test set**?

- ○ Between $0 and $100,000
- ○ Between $100,000 and $200,000
- ○ Between $200,000 and $300,000
- ○ Above $300,000

---

1 point

5.

Out of all reviews in the **test set** that are predicted to be positive, what fraction of them are **false positives**? (Round to the second decimal place e.g. 0.25)

0.05

1 point

6.

Based on what we learned in lecture, if we wanted to reduce this fraction of false positives to be below 3.5%, we would:

- Discard a sufficient number of positive predictions
- Discard a sufficient number of negative predictions
- Increase threshold for predicting the positive class (y_hat = +1)
- Decrease threshold for predicting the positive class (y_hat = +1)

1 point

7.

What fraction of the positive reviews in the **test_set** were correctly predicted as positive by the classifier? Round your answer to 2 decimal places.

0.95

1 point

8.

What is the recall value for a classifier that predicts **+1** for all data points in the **test_data**?

1

1 point

9.
What happens to the number of positive predicted reviews as the threshold increased from 0.5 to 0.9?

- More reviews are predicted to be positive.
- Fewer reviews are predicted to be positive.

1 point

10.
Consider the metrics obtained from setting the threshold to 0.5 and to 0.9.

Does the **recall** increase with a higher threshold?

- Yes
- No

1 point

11.
Among all the threshold values tried, what is the **smallest** threshold value that achieves a precision of 96.5% or better? Round your answer to 3 decimal places.

0.838

1 point

12.
Using threshold = 0.98, how many **false negatives** do we get on the **test_data**? (**Hint**: You may use the graphlab.evaluation.confusion_matrix function implemented in GraphLab Create.)

5826

1 point

13.
Questions 13 and 14 are concerned with the reviews that contain the word **baby**.

Among all the threshold values tried, what is the **smallest** threshold value that achieves a precision of 96.5% or better for the reviews of data in **baby_reviews**? Round your answer to 3 decimal places.

0.864

1 point

14.
Questions 13 and 14 are concerned with the reviews that contain the word **baby**.

Is this threshold value smaller or larger than the threshold used for the entire dataset to achieve the same specified precision of 96.5%?

- Larger
- Smaller

提交测试