Identifying safe loans with decision trees

8 questions

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 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-5-decision-tree-assignment-1-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. What percentage of the predictions on sample_validation_data did decision_tree_model get correct?

- ( ) 25%
- (●) 50%
- ( ) 75%
- ( ) 100%

1 point

3. Which loan has the highest probability of being classified as a safe loan?

- ( ) First
- ( ) Second
- ( ) Third
- (●) Fourth

1 point

4. Notice that the probability preditions are the exact same for the 2nd and 3rd loans i.e 0.472267584643798. Why would this happen?

- (●) During tree traversal both examples fall into the same leaf node.
- ( ) This can only happen with sheer coincidence?

1 point

5. Based on the visualized tree, what prediction would you make for this data point?

- ( ) +1
- (●) -1

1 point

6. What is the accuracy of decision_tree_model on the validation set, rounded to the nearest .01 (e.g. 0.76)?

0.64

1 point

7. How does the performance of big_model on the validation set compare to decision_tree_model on the validation set? Is this a sign of overfitting?

- ( ) big_model has higher accuracy on the validation set than decision_tree_model. This is overfitting.
- ( ) big_model has higher accuracy on the validation set than decision_tree_model. This is not overfitting.
- (●) big_model has lower accuracy on the validation set than decision_tree_model. This is overfitting.
- ( ) big_model has lower accuracy on the validation set than decision_tree_model. This is not overfitting.

1 point

8. Let us assume that each mistake costs money:

- Assume a cost of $10,000 per false negative.
- Assume a cost of $20,000 per false positive.

What is the total cost of mistakes made by decision_tree_model on validation_data?

50300000

Submit Quiz