

### Decision Trees in Practice

15 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-6-decision-tree-practical-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 SFrame and NumPy only.
- ☐ I am using other tools, and I understand that I may not be able to complete some of the quiz questions.

1  
point

#### 2. Given an intermediate node with 6 safe loans and 3 risky loans, if the min\_node\_size parameter is 10, what should the tree learning algorithm do next?

- ☒ Create a leaf and return it
  - ☐ Continue building the tree by finding the best splitting feature
- 

1  
point

3. Assume an intermediate node has 6 safe loans and 3 risky loans. For each of 4 possible features to split on, the error reduction is 0.0, 0.05, 0.1, and 0.14, respectively. If the minimum gain in error reduction parameter is set to 0.2, what should the tree learning algorithm do next?

- ☒ Create a leaf and return it
  - ☐ Continue building the tree by using the splitting feature that gives 0.14 error reduction
- 

1  
point

4. Consider the prediction path **validation\_set[0]** with `my_decision_tree_old` and `my_decision_tree_new`. For `my_decision_tree_new` trained with

```
max_depth = 6, min_node_size = 100, min_err  
or_reduction=0.0
```

is the prediction path shorter, longer, or the same as the prediction path using `my_decision_tree_old` that ignored the early stopping conditions 2 and 3?

- ☒ Shorter
  - ☐ Longer
  - ☐ The same
- 

1  
point

5. Consider the prediction path for **ANY** new data point. For `my_decision_tree_new` trained with

```
max_depth = 6, min_node_size = 100, min_err  
or_reduction=0.0
```

is the prediction path for a data point always shorter, always longer, always the same, shorter or the same, or

longer or the same as for my\_decision\_tree\_old that ignored the early stopping conditions 2 and 3?

- ☐ **Always** shorter
  - ☐ **Always** longer
  - ☐ **Always** the same
  - ☒ Shorter or the same
  - ☐ Longer or the same
- 

1  
point

6. For a tree trained on any dataset using parameters

```
max_depth = 6, min_node_size = 100, min_err  
or_reduction=0.0
```

what is the maximum possible number of splits encountered while making a single prediction?

---

1  
point

7. Is the validation error of the new decision tree (using early stopping conditions 2 and 3) lower than, higher than, or the same as that of the old decision tree from the previous assignment?

- ☐ Higher than
  - ☒ Lower than
  - ☐ The same
- 

1  
point

8. Which tree has the smallest error on the validation data?

- ☐ model\_1
- ☐ model\_2
- ☒ model\_3

---

1  
point

9. Does the tree with the smallest error in the training data also have the smallest error in the validation data?

- ☒ Yes
- ☐ No
- 

1  
point

10. Which tree has the largest complexity?

- ☐ model\_1
- ☐ model\_2
- ☒ model\_3
- 

1  
point

11. Is it always true that the tree with the lowest classification error on the training set will result in the lowest classification error in the validation set?

- ☐ Yes, this is ALWAYS true.
- ☒ No, this is NOT ALWAYS true.
- 

1  
point

12. Is it always true that the most complex tree will result in the lowest classification error in the validation\_set?

- ☐ Yes, this is always true.
- ☒ No, this is not always true.
- 

1  
point

13. Using the complexity definition, which model (model\_4, model\_5, or model\_6) has the largest complexity?

- ☒ model\_4

☐ model\_5

☐ model\_6

---

1  
point

14. model\_4 and model\_5 have similar classification error on the validation set but model\_5 has lower complexity. Should you pick model\_5 over model\_4?

☒ Pick model\_5 over model\_4

☐ Pick model\_4 over model\_5

---

1  
point

15. Using the results obtained in this section, which model (model\_7, model\_8, or model\_9) would you choose to use?

☐ model\_7

☒ model\_8

☐ model\_9

---