

Test your knowledge of LO 4.2.3

Free preview
 Quiz 4.2.3
 FDP® Exam Preparation
 www.infrapredictive.com

1. Is the following statement about 'decision trees' generally correct?

Statement:
 "Tree-based methods involve segmenting the predictor space into a number of simple regions. In order to make a prediction for a given observation, we typically use the mean or the mode (majority vote) of the training observations in the region to which it belongs. "

- Not correct
- Correct

2. Which of the following statements about 'advantages and disadvantages of decision trees' is correct?

Statement I:
 "Decision trees are easier to interpret than other classification/regression methods."
Statement II:
 "In terms of prediction accuracy, however, decision trees typically are not competitive with the best supervised learning approaches."
Statement III:
 "We can increase the prediction accuracy of the decision trees using special techniques such as bagging, random forest, and boosting, at the expense of some loss in the interpretation."

- I
- I and II
- I, II, and III

3. Is the following statement about 'Greedy recursive binary splitting' generally correct?

Statement:
 "The recursive approach is top-down since it begins at the top of the tree and then successively binary splits the predictor space; Each split is indicated via two new branches further down on the tree. It is greedy because, at each step of the tree-building process, the best split is made at that particular step, rather than looking ahead and picking a split that will lead to a better tree in some future step."

- Not correct
- Correct

4. Is the following statement about 'tree pruning' generally correct?

Statement:
 "Recursive binary splitting might result in a too complex tree. Therefore, a better strategy is to grow a very large tree T_0 , and then prune it back in order to obtain a subtree, $T \subset T_0$."

- Not correct
- Correct

5. Is the following statement about 'tree pruning and Cost complexity pruning' generally correct?

Statement:
 "Recursive binary splitting might result in a too complex tree. Therefore, a better strategy is to grow a very large tree T_0 , and then prune it back in order to obtain a subtree, $T \subset T_0$. **Cost complexity pruning** – aka weakest link pruning – provides an efficient way to do the pruning. Rather than considering every possible subtree, we can consider a sequence of trees indexed by a nonnegative tuning parameter γ ."

- Not correct
- Correct

6. Is the following statement about 'classification error rate' generally correct?

Statement:
 The classification error rate is the fraction of the training observations in that region that do not belong to the most common class:

$$E = 1 - \max_k (\hat{p}_{mk})$$

- Not correct
- Correct

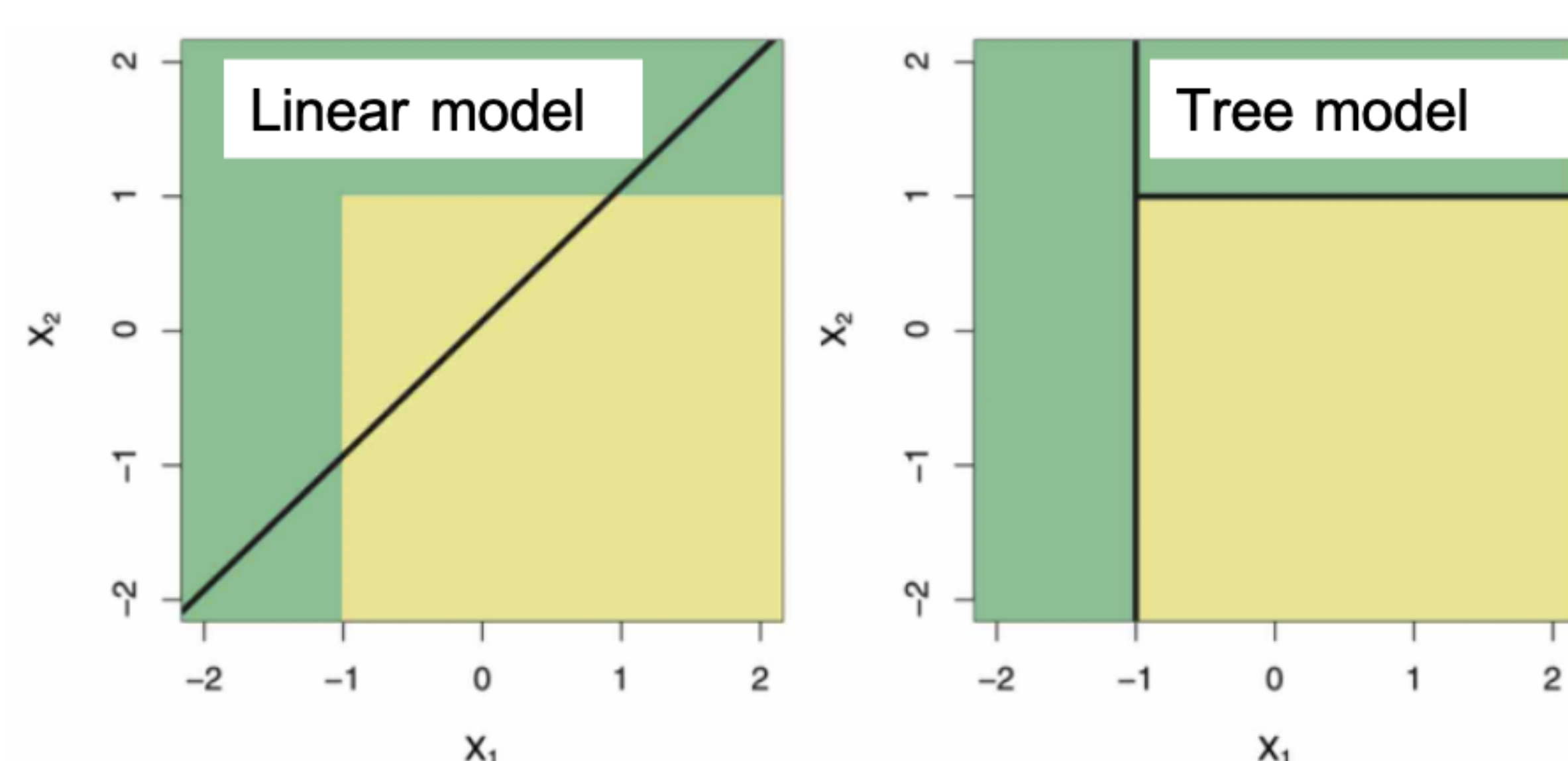
7. Is the following statement about 'Gini index' generally correct?

Statement:
 It is used as a measure of node purity: the lower the Gini index, the purer the node. The Gini Index function is defined by

$$G = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk})$$

- Not correct
- Correct

8. Given the complex relationship between X_1 and X_2 shown in the figure below, which model would be the best performer?



Source: Assigned reading

- Linear model
- Tree model
- Both

9. Is the following statement about 'Bagging' generally correct?

Statement:
 "Bootstrap aggregation or bagging is a general procedure for reducing the variance of a statistical learning method based on the principle that averaging a set of observations reduces variance. In other words, assembling weak predictors/classifiers together makes it a stronger predictor/classifier."

- No correct
- Correct

10. Is the following statement about Out-of-bag (OOB) error generally correct?

Statement:
 "The OOB error is a valid estimate of the test error for the bagged model since the response for each observation is predicted using only the trees that were not fit using that observation."

- No correct
- Correct

11. Is the following statement about Random forests generally correct?

Statement:
 "One of the issues of bagging is that strong predictors will dominate each tree, making the decision trees correlated. Unfortunately, the average of many highly correlated trees does not lead to a significant reduction in variance. **Random forests** provide a small tweak to de-correlate the trees."

- No correct
- Correct

12. Is the following statement about Boosting generally correct?

Statement:
 "Boosting works in a similar way of bagging, except that the trees are grown sequentially. Each tree is grown using information from previously grown trees. Boosting does not involve bootstrap sampling; instead, each tree is fit on a modified version of the original dataset. Unlike fitting a single large decision tree to the data, which amounts to fitting the data hard and potentially overfitting, the boosting approach instead learns slowly, sequentially. Approaches that learn slowly tend to perform well."

- No correct
- Correct

Save View Results

Free preview
 Quiz 4.2.3
 FDP® Exam Preparation
 www.infrapredictive.com