DECISION TREE QUIZES

Which of the following is NOT a major benefit of a decision tree model?

a. Fast training

b. Easy to interpret

c. Fast prediction

**d. Predictive** power

What, ultimately, are we after when constructing a decision tree model?

a. Interpretation of the data

b. A more precise rendering of the data

**c. A decision boundary**

d. Explanation of the past

With a binary decision tree, asking a single question…

a. Splits the feature space into four

b. Splits the feature space into eighths

**c. Splits the feature space in half**

d. Splits the feature space into sixteenths

A decision tree partitions the feature space into cells. What are these cells?

a. Our data points

b. Spaces containing one data point each

c. Our features

**d. Our categories or classes**

The partitioning of the feature space by a binary decision tree is always a diagonal line

TRUE

**FALSE**

Which of the following is NOT a criterion for identifying the purity of a node?

a. Entropy

b. Gini index

c. Misclassification index

**d. Depth**

What’s the aim of looking at node purity gain?

**a. To work out how pure splitting on that feature and threshold is**

b. To work out how deep the tree should be

c. To work out how pure the tree is

d. None of the above

Consider the following algorithm. i. Check if already finished ii. For each feature xi: a. Calculate the gain from splitting on xi b. Let xbest be the feature with the lowest gain iii. Create a decision node that splits on xbest Which step is mistaken?

a. (i)

b. (ii)(a)

**c. (ii)(b)**

d. (iii)

The value for misclassification and gini gain can come apart, for a given node

**TRUE**

FALSE

Which of the following is not a relevant consideration mentioned by the lecturer for when the decision tree algorithm should terminate?

**a. The tree is too shallow**

b. node contains less than x data points

c. max depth is reached

d. max purity is sufficient

How does extreme overfitting typically manifest itself in a decision tree?

a. A very deep tree

b. A tree with high impurity

**c. A little cell around every data point in the set**

d. A very shallow tree

What is the way of treating overfitting?

**a. Pruning**

b. Using entropy

c. Using an SVM instead

d. Comparing pre- and post-split entropy

Which of the following is NOT a disadvantage of decision trees:

a. Overfitting

**b. Slow execution**

c. Sensitive to small changes in the data

d. Only axis-aligned splits

Let’s compare decision trees and SVMs. SVMs are better at handling missing values

TRUE

**FALSE**

Decision trees are better at extracting linear combinations of features

TRUE

**FALSE**

What is the main purpose of ensemble methods?

**a. To augment the predictive power of decision trees**

b. To make decision trees execute faster

c. To make decision trees better at handling missing values

d. To make decision trees better at extracting linear combinations of features

With ensemble methods such as random forest, what’s the name of the method we use to ensure that all the trees in the ensemble don’t just uselessly learn in exactly the same way?

a. Permutation test

b. SVM

c. Cross-validation

**d. Bootstrap**

Which of the following claims is NOT true about bootstrapping?

a. It’s a resampling method for statistics

**b. It’s not a useful way to get the error bars on estimates**

c. It draws with replacement

d. It’s used to make the trees in an ensemble method learn in different ways

Bagging

**a. Normally uses one type of classifier**

b. Is an ensemble method where you fit many models to all your data.

c. Generates models that are easy to interpret.

We should use simple bootstrap to generate train and test data from the same dataset

TRUE

**FALSE**

**Ensemble Method Quizzes**

Ensemble methods:

a. Combine many learners, provided each is individually strong

**b. Combine many learners, each of which could be individually weak**

c. Come in just one variety

d. None of the above

Which of the following is not an ensemble method?

a. Random forest

b. Bagging

c. Boosting

**d. Clustering**

In the movie recommendations analogy to random forest, what was our leaving out certain films we like/dislike to one friend, but leaving out other films to a different friend, the analog of?

a. Bagging

b. Gradient descent

**c. Bootstrapping**

d. Entropy

Suppose in the movie recommendation analogy I like both Wild at Heart and Blue Velvet, but not because they’re directed by David Lynch. How is this accounted for in random forest?

a. By pruning the tree

**b. By making randomness partly determine which attribute to split on at each node**

c. By overfitting

d. By gradient descent

Why is the wisdom of crowds dubious?

**a. Because there’s no reason to suspect that positive bias always counterbalances negative bias**

b. Because there’s no reason to suspect that there is a General Will

c. Because human beings’ beliefs converge to falsehood, due to falsifying cognitive faculties

d. None of the above

The principal difference between boosting and bagging is that with boosting, weak learners evolve over time, and votes are weighted

**TRUE**

FALSE

Which of the following is NOT a tuning parameter for boosting:

a. Number of trees

b. Number of splits in each tree (often stumps work well)

c. Parameters controlling how weights evolve

**d. Splitting criterion**

Consider the AdaBoost reweighting step.

**a. If the indicator is 0, then we’re not changing the weight. If the indicator is 1, then we misclassified the observation, so we put more weight on it.**

b. If the indicator is 0, then we misclassified the observation, so we put more weight on it. If the indicator is 1, then we’re not changing the weight.

c. If the indicator is 0, then we misclassified the observation, so we square its weight. If the indicator is 1, then we’re not changing the weight.

d. None of the above.

Boosting can be prone to overfitting .

**TRUE**

FALSE

A complicated decision tree can give you complex boundaries but will overfit. The bagging process applied properly will smooth it out.

**TRUE**

FALSE