

Concept Quiz Over Week 8 Material

Due Nov 26 at 11:59pm**Points** 1**Questions** 7**Available** Nov 22 at 12am - Nov 26 at 11:59pm**Time Limit** NoneScore for this survey: **1** out of 1

Submitted Nov 26 at 2:09pm

This attempt took 2,519 minutes.

Question 1

Decision trees have a tendency to overfit training data. What are options for reducing overfitting?

You Answered

☒ Limiting the depth of the tree

You Answered

☒ Monitor the validation accuracy and perform early stopping

You Answered

☒ Removing nodes based on validation performance from the tree after training has finished

All three are valid ways to control the complexity of a decision tree and can reduce overfitting.

Question 2

Explain how decision trees can deal with continuous valued attributes.

Your Answer:

we can sort the values of x_i in the dataset, and consider the threshold between consecutive datapoints.

Compare information gained for each and choose the max height.

Even though continuous valued attributes have an infinite number of possible thresholds, only a finite set of thresholds need to be considered during training because thresholds occurring between the same datapoints will have identical purity measures like Information Gain. As such, we can just consider thresholds at the midpoint between sequential datapoints for an attribute.

Question 3

Decision tree learning can terminate whenever all attributes result in zero information gain.

You Answered

☒ True


☐ False

False. We discussed this as a potential base case; however, many functions (like XOR) may not show any information gain in initial variables but be able to be usefully split later only after multiple variables have been considered.


Question 4

Match the terms with their definitions.


You Answered

BiasError due to assumpti 


You Answered

VarianceError due to sensitivity 

You Answered

Weak LearnerModels that tend to h: 

You Answered

Strong LearnerModels that tend to h: 

Bias -> Error due to assumptions in the model not matching the problem (aka modelling error)

Variance -> Error due to sensitivity to changes in the dataset (aka estimation and optimization error)

Weak Learner -> Models that tend to have high bias but low variance.

Strong Learner -> Models that tend to have low bias but high variance.

Question 5

Given a training dataset {a, b, c, d, e}, show three example bootstrap training sets that could be used for training bagged models.

Your Answer:

sample a = $h_1(a)$

sample b = $h_2(b)$

sample d = $h_3(d)$

Each training dataset for bagging is a sample with replacement from the original dataset with the same number of instances. So I drew a random number between 1 and 5 five times to make each dataset below:

{d, d, b, a, d}

{b, d, a, d, b}

{a, c, e, b, a}

Question 6

Boosting is a method to make an ensemble of weak learner stronger -- i.e., a way to reduce bias in weak learners. Boosting typically trains a sequence of models with each model focusing on the errors of the previous models.

You Answered

☒ True

☐ False

True! We talked about two versions of this -- L2 boosting and Adaboost

Question 7

Bagging is an ensemble method for reducing the variance of weak learners by combining multiple models trained on random samples of the dataset.

☐ True

☒ False

You Answered

False. Bagging is used to reduce variance in **strong learners**. Weak learners already have low variance, but high bias.

Survey Score: **1** out of 1