















Announcements

Assignments

Fall 2022

Home

Discussions

Grades People



Files

Syllabus

Quizzes

Collaborations

Chat

Office 365

Library Research

GatorEvals

Zoom Conferences

Quiz 8 At

Started: Nov 22 at 9:08am

Quiz Instructions

Question 1

This quiz is open book, open notes, "open R". The expected duration is 60 minutes. Two attempts are allowed. If both attempts are taken, the score for the second attempt will "overwrite" that from the first attempt (regardless if it is higher or lower). Even though the quiz has 22 points, it will be graded out of 20 points (i.e., 2 points bonus).

You are allowed to use any of the class materials from our SML class, but no other materials (no internet browsing or communication with other parties online/offline).

Even if a question is asking for a numerical value or True/False answer, in order to receive full credit (if your "guess" is correct) or partial credit (if appropriate, if your "guess" is incorrect), please provide your rationale as comments in the uploaded file requested at the end of the

2 pts

Suppose you are given definitions for regions R1, ..., Rk (i.e., a partition of the predictor space) corresponding to leaves of a regression tree T that was built using a vector of (quantitative) responses Y and design matrix X. Explain how to predict the (unobserved) response for a new vector of features x (i.e., we are given a new pair (x,y), where x is known and y is unknown). Your answer should include the expression for the predicted value.

Edit View Insert Format Tools Table

12pt \vee Paragraph \vee B $I \cup \underline{A} \vee \underline{\mathscr{D}} \vee \mathsf{T}^2 \vee$:

$$f(X) = \sum_{m=1}^{M} c_m \cdot 1_{(X \in R_m)}$$
 (8.9)

A decision tree is used in a top down fashion. It essentially is a directed acyclic graph. The output is generated by a sequence of thresholds splitting the data into partitions (here it is R_1 ... R_m). The test observation belongs to the partition of the feature space that is assigned by the leaf of the tree. For regression the predicted value is often the mean of the of the training observations in the partition the test data falls into.

p





Question 2

2 pts

Consider the same setup as in Q1 above, but one has fitted a classification tree for a categorical response with K categories. Briefly explain how a prediction for a new x is made.

Each category will be associated with a partition or a number of partitions. Unlike regression the mean of the training data is not taken and the most commonly occurring training category is enough to assign to the new prediction of x. The rest is the same where the tree is

Questions

- Question 1
- ✓ Question 2
- ✓ Question 3
- Ouestion 4
- ✓ Question 5
- ✓ Question 6
- ✓ Question 7

Time Running: Hide Time Attempt due: Nov 22 at 10:30am 18 Minutes, 57 Seconds



considered from a top down approach. **★** 56 words </> ✓ **★** р

Question 3

Consider a single regression tree model. Briefly explain the idea behind the "grow-and-prune" strategy, as opposed to the purely "grow" strategy.

A purely grow strategy is when a tree is generated and allowed to grow continuously without penalty. This produces a lot of branches and leaves for the tree. These types of trees are often more sensitive to overfitting which leads to poor generalization of the model.

The grow-and-prune strategy applies some penalty to continuous growth of the tree. Pruning reduces overfitting and lowers the variance of the tree by reducing its size. However, pruning is difficult to perform at the same time as growing. By stopping growth early when the RSS stops decreasing we stand the chance of falling into local optima and miss the global optima with the model. It is a better approach to grow first and then prupe away the branches after







4 pts

Question 4 4 pts

Briefly explain the idea behind bagging regression trees (either the naïve version with m=p presented in class or the Random Forest); its merits and drawbacks (relative to a single tree model).

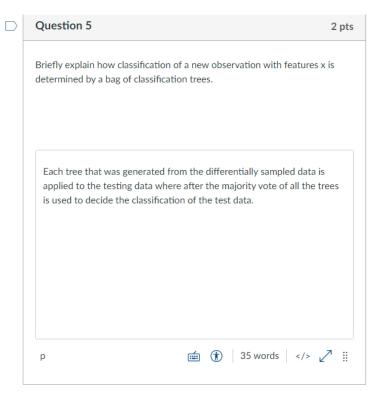
approach lowers the variance and improves the accuracy. bagging takes the concept of bootstrapping (sampling form the training data with replacement) and applies it to decision trees. The training data is used and sampled multiple times to generate a host of weaker performing trees than a single tree generated off of all the training data. All of these smaller trees are then aggregated and used as an ensemble of models to generate a full tree. The aggregation is often done by just averaging all the trees. This approach is best used in cases where the bias is not too high. It also makes the resulting model less explainable than a single tree model. This approach is also more computationally expensive so the tradeoffs should be considered on a contextual basis.



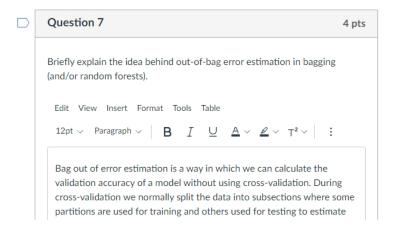












accuracy or performance of the model. During bagging the data is already split and partitioned during the bootstrapping process. During sampling the data points that are not used to train a tree can then be used as a validation set for that specific tree as it was not used during training. The aggregation of the performance of all these bagged trees can then give us an estimate of how well the model would perform on out of sample data.

p









Quiz saved at 9:49am

Submit Quiz