Quiz 8

Started: Nov 20 at 7:56pm

Quiz Instructions

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 quiz.

Question 1 — 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 Paragraph B I U A T²

p 0 words </>

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.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words </>

Question 3

4 pts

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

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

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).

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words </>

Question 5

2 pts

Briefly explain how classification of a new observation with features x is determined by a bag of classification trees.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 6

4 pts

Briefly explain the idea behind variable importance summaries produced by a bag of trees (e.g., Random Forest).

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 7

4 pts

Briefly explain the idea behind out-of-bag error estimation in bagging (and/or random forests).

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Quiz saved at 7:56pm

Submit Quiz