# Feedback Decision Trees In-Class

Jack Jewson • 10:27 AM

Dear Students,

I have just sent you all feedback from the Decision trees In-Class test.

Students generally did excellently. Generally the data was minimally process and at the very least a well tuned decision Tree and  random forests were considered, tuning some of their inputs through Grid Search.

Notes:
+ You should not need to worry about creating polynomial features. The decision tree already has the ability to capture this structure and adding more correlated features will just slow down the ability of random forest algorithms to minimise variance.
+ A few students are still imputing their testing data independently of the training data. You should use training set summaries (i.e. .fit to the training data) to impute the testing data (.transform the testing data)
+ target encoding was very effective for the diagnoses, I wouldn't recommend doing this for classes with only a few categories though, dummies will be better for low cardinality categories.
+ GridSearches: unlike for kNN and SVM, the decision tree models have many inputs you could think about tuning. I wouldn't recommend trying to run a massive grid search over everything (particularity as many of them do similar things), fix a few things and then pick maybe 2 or 3 of the inputs to play around with e.g. max_leaf_nodes
and min_impurity_decrease are things I would recommend. For Random Forests you would usually consider n_estimators to be fixed (e.g. 50 or 100) for boosting you can think about maybe adding more helps

Please let me know if you have any questions,

Jack

Class comments

Add class comment...