Predictive Analytics Term 4, 2020

Associate Professor Ole Maneesoonthorn

Syndicate Task #3

Using the same data set that you have chosen for Syndicate Task #1 for this task.

1. Use the training data set to construct the following models:
    a. Regression tree
    b. Neural network
    c. K-means clustering
    d. K-NN regression
2. Construct predictions for the test set using these four methods.
3. Comment on how your results from these methods compare to your analysis from Syndicate Task #1. Discuss both the predictability of the models, as well as the relationship between wine quality and the physicochemical properties that these methods reveal.
4. Consider the scenario where you would like to use the predictions for forming pricing and marketing strategies. Wines with quality score of 7 or above are targeted to the premium market segment and typically attracts larger marketing costs. The profit margins for these wines are also much higher than the lower quality wines. If the quality score is underestimated for these wines, the wine maker will miss out on the potential revenue. Construct an asymmetric loss function that is consistent with this usage of wine quality predictions. Comment on the models' predictive performance using this asymmetric loss. Make sure you also consider the models you constructed for Syndicate Task #1 in your discussion.

Produce a 3-page report that summarizes your analysis.

This task is due at 6pm on Sunday 18th October 2020.