Question 20):

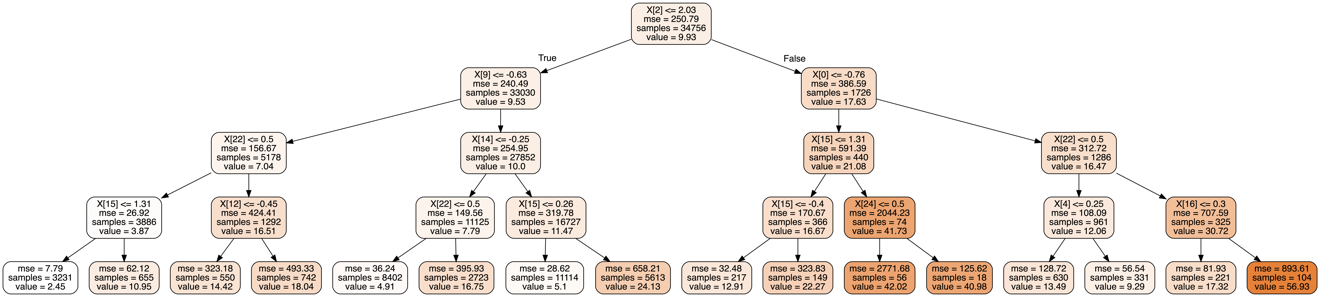
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Case | Max Feature | Number of trees | Depth of each tree | RMSE for train data | RMSE for validate data |
| 1 | 1 | 100 | 20 | 1.5937904191 | 6.3283292068734 |
| 2 | 2 | 100 | 20 | 1.110048485 | 5.018633169864 |
| 3 | 3 | 100 | 20 | 0.8171618290 | 4.3618593311508 |
| 4 | 5 | 100 | 20 | 0.6136713395 | 4.091401723226 |
| 5 | 10 | 100 | 20 | 0.4946636043 | 4.039363402723 |
| 6 | 15 | 100 | 20 | 0.482000093 | 4.10310560786 |
| 7 | 3 | 50 | 20 | 0.8923059169 | 4.485547831422 |
| 8 | 3 | 200 | 20 | 0.798839291 | 4.30222744794 |
| 9 | 3 | 100 | 10 | 5.605842807 | 6.58412778357 |
| 10 | 3 | 100 | 40 | 0.673365491 | 4.31450039854 |

By observing cases 1-6, we find that initially RMSE decreases as the maximum number of features increases, but increases afterwards. By observing cases 3, 7, and 8, we find that RMSE decreases as the number of trees increases in training data, though the validation data increases afterwards, which means it is overfitting. By observing cases 3, 9, and 10, we find that RMSE decreases as depth of each tree increases initially, then RMSE increases as depth of each tree increases.

Question 21): Same as Yunzheng’s

Question 22):

In this part, we pick a tree with depth of the tree that equals to four. We get the result that RMSE for train data of 0.6733, and RMSE for validation data of 4.3145. The tree plot is shown below in Fig. X.



The feature selected for branching at the root node is ‘height. (X[2] represents the 3th feature among the features). The importance of the feature increases as the depth increases. In other words, the more proportion of samples reach that node, the more important that feature is. Thus, the important features are X[15], X[12], X[22], X[4], X[16], X[24] which are correspondingly ‘size, ‘o\_width’, ‘o\_codec\_flv’, ‘framerate’, ‘o\_height’, and ‘o\_codec\_mpeg4’. All three are included in the features explained previously.