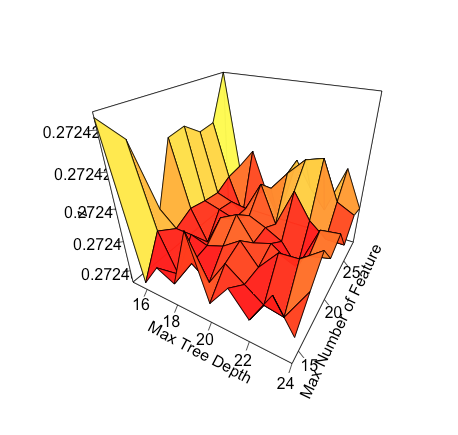
Parameter Selection:

1: Important Features:

Features that could be extracted from molecules can be used as an excellent basis for advanced property prediction. The nature of these features is diverse and it could include some properties such as merck molecular force field energy and molecular mass. There are many features available from a SMILES string but only a subset of them are numeric. The feature used in this model is binary which has already been classified to 0 and 1. The prediction accuracy could be improved by using the numeric form of the data instead of binary form. For random forest regression, count the number of votes cast for the correct class and then randomly permute the values of variable m in the oob cases and put down these cases. Make a subtraction of the number of votes for the correct class in the variable-m-m-permuted oob data from the number of votes for the correct class in the untouched oob data. The average of all this number over all the trees is the raw importance score for variable. By using the algorithm of random forest, important features were selected for the model.

2: Other Tunning Parameters:

In this model, RandomForestRegressor in sklearn library is used. There are three parameters, which needs to be changed for the function. The first one is n\_estimators, for this parameter, the larger the better based on the fact that it is under the limit of time and memory. The second parameter would be the max\_depth, which is the maximum depth of the tree. It may cause overfit problem if the depth of the tree is too larger. The third parameter would be max\_features which is the number of features to consider when looking for the best split. It has similar problem with the max\_depth feature. If too many features are considered, there is a large probability that overfit problem may be caused. In order to find the best parameter combination for max\_depth and max\_feature, it calculates the AME for each combination of max\_depth and max\_feature. The depth is from 15 to 24 and the feature number is from 14 to 28. As we could see from the plot, AME gets the lowest value when depth is 19 and feature is 28.