Evaluation Report

Dataset Format: Tabular

Dataset Size ~ 1500 rows, 80 columns, 1 Target

Problem: Regression; prediction of house sale price

MAE Loss: Random Forest (~24k MAE), Neural Net (~1.8L MAE)

As we can see for the given tabular dataset Random Forest outperformed Neural Network with tremendous change in the prediction loss. Random Forest (RF) and Neural Net (NN) operates differently. Let’s dig in and understand what made RF perform better than NN.

Random Forest is a deterministic and rule-based approach which results in True/False. It also comes with power of multiple decision trees to make predictions. The given problem would be divided into chunks and rules are applied to carry out the following prediction. It is also referred as ensemble of decision trees. Each decision tree ensemble the regression process and predicts the outcome. Whereas Neural Network is a mesh of connected neurons. These neurons are grouped into layers and process the data based on the probabilities. Hence Neural Nets are also referred as probabilistic approach.

The natural input for decision tree is structured or Tabular data. Neural Network will simply overkill prediction and data regression because of its way of operating. The root cause is prediction of probability rather than simple yes/no ruled-based decisions. Choosing ones and zeros over probabilities would be a right choice in this scenario. As we can clearly see in our example deterministic random forest algorithm outperformed probabilistic neural networks.

One more important feature is that neural networks very well with huge datasets rather than small. It is because it’s a statistical learning model, such models cannot learn the trend/pattern very well without any signal-noise with less data set. Regression problem itself is coined as “regression towards mean” and mean here is underlying pattern. If dataset is not big enough that means mean keeps fluctuating, and the model becomes unreliable. As we could see that our dataset is very small for neural network to understand the pattern and cancel the signal-noise with stable mean.