

Sahil Rangwala

MSDS_422

Discuss what your models tell you in layman's terms

The linear regression model offers a decent prediction for our SalePrice predictor in comparison to the gradient boosting. Both are still merely okay models after comparing my results to other results of the hidden test case that other users submitted on the leaderboards. The reason I used gradient boosting vs gradient descent is that I believe gradient boosting would offer a better result and I was more comfortable coding through this process. The main difference between gradient descent and gradient boosting is that instead of descending the gradient by introducing more changes to the parameters, I created multiple weak models through an "ensemble" that descended the gradient. I used 3 main metrics to evaluate both of our models as well as plotting these ideas to showcase the results. Mean squared error, square root mean squared error, and mean absolute error. These metrics deal with loss and how poor our model could be, thus the higher the value for these metrics the worse it is for predicting.

The linear regression model had a much higher mean squared error in comparison to the gradient boosting. This could be from the regression model is not the perfect model for this scenario, thus I explored the gradient boosting model. Through this process, it offered overall a lower error in comparison to the linear regression model. Due to these low looking metrics of error, I decided to move with the gradient boosting model to help predict SalePrice.

I initially used the linear regression model to submit my score because I wanted to see how I fared with other submissions. Realizing that this is not the greatest option for the model (score of 0.38331), I decided to move with the gradient boosting alternative. This is the score of 0.20537 you see below. Overall, the gradient boosting through the regression metrics recorded as well as the score associated with the kaggle hidden test case offers the best option between the two models to help predict SalePrice.

submission.csv 4 minutes ago by Sahil Rangwala Gradient descent model	0.20537
submission.csv an hour ago by Sahil Rangwala Linear regression model on test data	0.38331