Models Comparison

Dingchao Zhang, 01/28/2016

Gradient Boosting Regressor

Advantages:

1. Ensemble methods like boosting, random forest empirically produce best performance (https://www.cs.cornell.edu/~caruana/ctp/ct.papers/caruana.icml06.pdf) , so the gradient boosting allows modeler to discover the performance ceiling for the data set.
2. Gradient Boosting Regressor like other Tree-based and rule-based models deals can effectively handle many types of predictors (sparse, skewed, continuous, categorical, etc.) without the need to pre-process them(e.g. encoding and creating dummy variables for categorical data)
3. Gradient Boosting Regressor is very flexible and does not require the user to specify the form of the predictors' relationship to the response like, for example, a linear regression model requires.
4. It can effectively handle missing data and implicitly conduct feature selection, thus not requiring any missing data removal or permutation work like other models
5. It can deal with near-zero variance predictors and non-informative predictors naturally, as it automatically performs feature selection when in each iteration looking for the best predictor and its value to split to minimax SSE.
6. It does not require data centering and scaling, which is required in other models like Linear Regression.

Disadvantages:

1. Gradient Boosting is less interpretable compared to Linear Regression Models.
2. It has more parameters to be tuned like learning_rate, n_estimators, subsample, min_samples_split and it is slower to train.
3. It is not stable or has could have a large variance there are multi- collinearity relationships exist in predictors.
4. If the relationship between predictors and the response cannot be adequately defined by rectangular, subspaces of the predictors, then Gradient Boosting like other tree-based or rule-based models will have larger prediction error than other kinds of models.

Lasso Regularized Ordinary Linear Regression

Advantages:

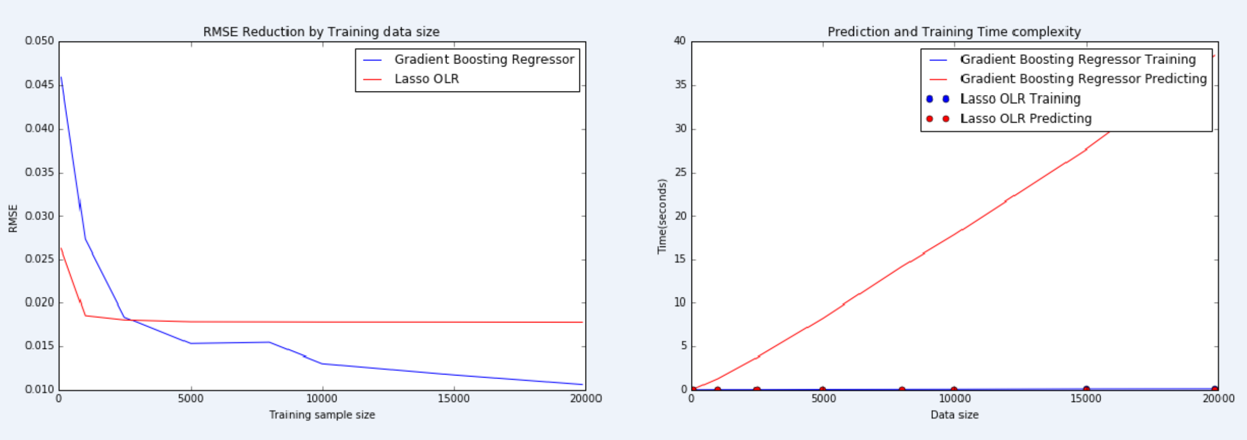
1. It is highly interpretable since the absolute values of coefficients of each predictors directly reflect the predictor's importance in influencing the response. For example, if the estimated coefficient of a predictor is 2.5, then a 1 unit increase in that predictor's value would, on average, increase the response by 2.5 units. Furthermore, relationships among predictors can be further interpreted through the estimated coefficients
2. Lasso regularization has performs automatically feature selection as it drives down less-important predictors' coefficients to 0.

3. It is quick to predict since fewer predictors were used compared to other Ordinary Linear Regressions.
4. It is also quick to train quicker to train and mathematically the best parameters estimate plane can be calculated if certain conditions are met(no strong collinearity among predictors, and the number of predictors is smaller than the number of samples).
5. Regularized regression has lower variance.

Disadvantages:

1. Linear regression model is appropriate when the relationship between the predictors and response falls along a hyperplane, this constraint may not be held true if there is any curvy/ non-linear relationship exists in the data.
2. A lot of data pre-processing is required for Linear Regression Mode, such as addressing skewness, different scales, different means, missing values, collinearity among predictors.

Performance comparison in State-Farm Interest Rate Prediction:



Models Performance: When training sample size is small, Lasso Ordinary Linear Regression produces smaller RMSE than Gradient Boosting Regressor, but Lasso OLR's RMSE reduction rate quickly flattens out when sample size is beyond 1000, while Gradient Boosting Regressor quickly drops it RMSE when training sample size increases and flattens out more smoothly.

Training time complexity: Gradient Boosting Regressor training time linearly increases as training sample size increases; while Lasso OLR training time is almost constantly small.

Prediction time complexity: Gradient Boosting Regressor training time and Lasso OLR training time are both constantly small as data size increases.