Assignment 4

**Project 1**

Compete in the Kaggle.com House Prices: Advanced Regression Techniques competition located here. [https://www.kaggle.com/c/house-prices-advanced-regression-techniques (Links to an external site.)](https://www.kaggle.com/c/house-prices-advanced-regression-techniques).

Investigate many variables.

Employ at least two regression modelling methods:

* linear regression,
* stochastic gradient descent,
* ridge regression,
* lasso regression, and
* elastic net.

Also employ random forests to the regression problem, following methods described in Géron (2017) Chapter 7.

Evaluate these methods within a cross-validation design, using root mean-squared error (RMSE) as an index of prediction error.

Submit at least two models to Kaggle.com for evaluation.

Try alternative versions of random forests and gradient boosting.

Select a best modeling method .

Employ that method on the full data set, obtaining results that you can report to management.

Regarding the management problem, imagine that you again are advising a real estate brokerage firm in its attempt to employ machine learning methods. The firm wants to use machine learning to complement conventional methods for assessing the market value of residential real estate. Of the modeling methods examined in your study, which would you recommend to management and why? Reviewing the results of the random forests and gradient boosting model you have selected to present to management, which explanatory variables are most important in predicting home prices?