

Run the code:

My Python version is 3.6(64 bit). After opening the file by IDLE and running the module, the scatter plot of the Gr Liv Area feature vs the sale price will be shown. After closing the scatter plot figure, the RMSE based on single variable model and all variables model for validation set will be printed on the screen.

Then, plot of the RMSE on the train and validation set for each value of alpha will be shown.

After closing the RMSE plot figure, the number of non-zero coefficients in the model of the values of alpha will be shown.

After closing the number plot, the RMSE value based on single variable model and all variables model for test set will be printed on the screen.

The last figure shows the RMSE on the test set for each value of Alpha.

Code Content:

line 63 starts to separate data into training, validation and test set based on requirements.

Line 87 starts to build model based on equation one variable OLS model, which is the function named as best fit slope and intercept.

Line 119 starts to build all features least squares model based on imported package, which is from sklearn.linear_model to import LinearRegression.

Line 123, we employ panda get_dummies to transform categorical features. Then we train all features least squares model with the training set, validate it with the validation set.

Line 151 starts to build Lasso regression model based on imported package, which is from sklearn.linear_model to import Lasso.

Line 153 starts to normalize those data sets.

Line 164 starts to employ cross validation and train Lasso regression model.

Line 226 applies the single variable model, the all features least squares model and the regularized model to the test set. The RMSE for each condition for both the validation set and the test set are listed in the table