

Contribution	Chih-Hsiang Wang	Jui-Hung Lu	Webster Cheng
Percentage (%)	33.3	33.3	33.4

## General Introduction:

This report analyzes the data set of house sold in Seattle. By the training data, we can use gradient descent to achieve some models with different weights on different features. These weights can be further applied on validation data to find the smallest SSE of price. The techniques we use are data analysis before modeling, choosing different learning rate, experiment with different regularization coefficient, and discussion about normalization. Hope after the effort on each part, we can get the desirable weights to predict the test data with small SSE.

Part 0 : Preprocessing and simple analysis. Perform the following preprocessing of the your data.

(a) Remove the ID feature. Why do you think it is a bad idea to use this feature in learning?

Because the ID feature has no impact on the price, it is just a series of numbers to label data.

(b) Split the date feature into three separate numerical features: month, day, and year. Can you think of better ways of using this date feature?

We calculated the days from selling the house to 5/31/2015, trying to figure out the relation between price and the date when the house was sold. It is better to merge three separate numerical features together because they can be presented as a feature showing date to avoid the different units of year, month, and day.

(c) Build a table that reports the statistics for each feature. For numerical features, please report the mean, the standard deviation, and the range. For categorical features such as waterfront, grade, condition (the later two are ordinal), please report the percentage of examples for each category.

training	mean	std	range
dummy	1	0	0
date	1311.1734	113.193376	390
bedrooms	3.3752	0.943246	32
bathrooms	2.118875	0.765128	7.25
sqft_living	2080.2232	911.334358	9520
sqft_lot	1.51E+04	4.12E+04	1650787
floors	1.5037	0.542647	2.5































































