

Homework1 Report

Qi Liu

March 16, 2016

1 Data Normalization

In machine learning tasks, the data may contain different dimensions of features. And the scalar of these dimensions may differ. This will make training hard to converge. Thus we normalized the data before training.

The method we used is simple. For one dimension data x , calculate its mean μ and standard deviation σ . The normalized data

$$x' = \frac{x - \mu}{\sigma}.$$

This makes the data to a distribution with mean equals zero and variance equals one.

Since the normalization method is linear. If we get the parameters of linear regression for the normalized data, the parameters of linear regression for the unnormalized data can be calculated easily. For example, let $x_i (1 \leq i \leq n)$ be the unnormalized features and y be the unnormalized label. The normalized data $x'_i (1 \leq i \leq n)$ and y' satisfy

$$x'_i = \frac{x_i - \mu_i}{\sigma_i} (1 \leq i \leq n)$$

and

$$y' = \frac{y - \mu_y}{\sigma_y}.$$

The parameters for normalized data θ' satisfy

$$y' = \theta'_0 + \sum_{i=1}^n \theta'_i x'_i,$$

the parameters for unnormalized data θ can be calculated as

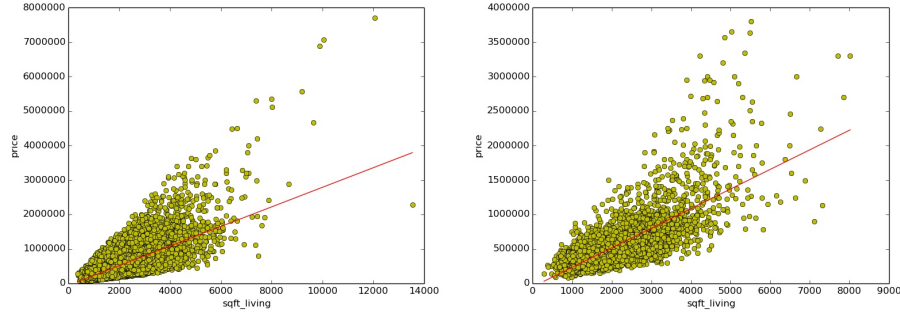
$$\theta_i = \frac{\theta'_i \sigma_y}{\sigma_i} (1 \leq i \leq n)$$

$$\theta_0 = \mu_y + \sigma_y (\theta'_0 - \sum_{i=1}^n \frac{\theta'_i \mu_i}{\sigma_i}).$$

2 Normal Equation

Let $X = (x_0, x_1, \dots, x_n)$ be the feature matrix and y be the label vector. Normal equation shows parameter

$$\theta = (X^T X)^{-1} X^T y.$$



The above scatter-plots show the results. Left one is training data and right one is test data. The fitting function is

$$\text{price} = 283.842 \times \text{sqft_living} - 49800.406$$

and RMSE is 264054.500.

3 Newton's Method

Let X and y be the feature and label. Newton's method sets the original parameter θ_0 manually and gets

$$\theta_i = \theta_{i-1} - H^{-1} \nabla J(\theta).$$

Here $H = \nabla^2 J(\theta)$ is the Hessian matrix. For linear regression,

$$\nabla J(\theta) = X^T X \theta - X^T y$$

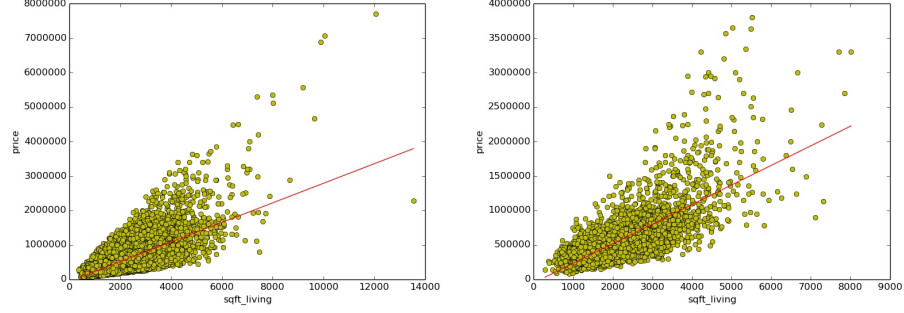
and

$$H = X^T X.$$

The Newton's method converges in 10 iterations. In fact, it will converge in one iteration. This is because

$$\theta_1 = \theta_0 - H^{-1} \nabla J(\theta) = \theta_0 - (X^T X)^{-1} (X^T X \theta_0 - X^T y) = (X^T X)^{-1} X^T y.$$

Which is the same as normal equation.



The above scatter-plots show the results. Left one is training data and right one is test data. The fitting function is

$$\text{price} = 283.842 \times \text{sqft_living} - 49800.406$$

and RMSE is 264054.500. The result is the same as normal equation.

4 Gradient Descent

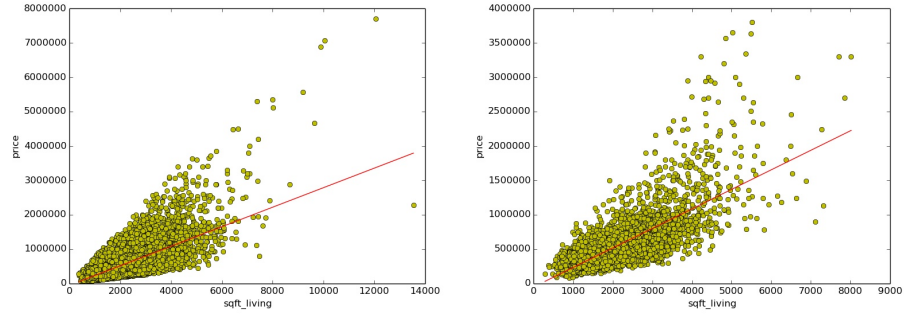
Let X and y be the feature and label. Gradient Descent sets the original parameter θ_0 manually and gets

$$\theta_i = \theta_{i-1} - \alpha \nabla J(\theta).$$

Here α is the learning rate. For linear regression,

$$\nabla J(\theta) = X^T X \theta - X^T y.$$

We set $\alpha = 0.1/n$ which n is the number of samples. The gradient descent converges in 1000 iterations.



The above scatter-plots show the results. Left one is training data and right one is test data. The fitting function is

$$\text{price} = 283.842 \times \text{sqft_living} - 49800.406$$

and RMSE is 264054.500. The result is the same as normal equation and Newton's method.

5 Optional Task

We use **sqft_living** and **bedrooms** to predict **price**. We use all the three methods and get the same results. The fitting function is

$$\mathbf{price} = 314.850 \times \mathbf{sqft_living} - 53259.436 \times \mathbf{bedrooms} + 65820.347$$

and RMSE is 259160.225.