

## Cost Function

**Note:** [5:18 - There is a typo. It should be  $\sum_{j=1}^n \theta_j^2$  instead of  $\sum_{i=1}^n \theta_j^2$ ]

If we have overfitting from our hypothesis function, we can reduce the weight that some of the terms in our function carry by increasing their cost.

Say we wanted to make the following function more quadratic:

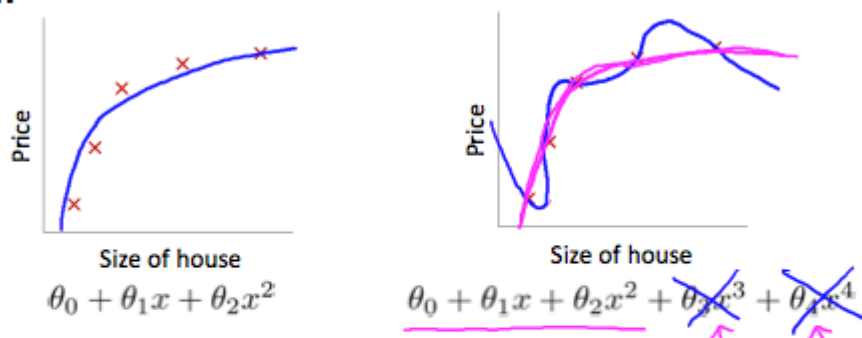
$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

We'll want to eliminate the influence of  $\theta_3 x^3$  and  $\theta_4 x^4$ . Without actually getting rid of these features or changing the form of our hypothesis, we can instead modify our **cost function**:

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + 1000 \cdot \theta_3^2 + 1000 \cdot \theta_4^2$$

We've added two extra terms at the end to inflate the cost of  $\theta_3$  and  $\theta_4$ . Now, in order for the cost function to get close to zero, we will have to reduce the values of  $\theta_3$  and  $\theta_4$  to near zero. This will in turn greatly reduce the values of  $\theta_3 x^3$  and  $\theta_4 x^4$  in our hypothesis function. As a result, we see that the new hypothesis (depicted by the pink curve) looks like a quadratic function but fits the data better due to the extra small terms  $\theta_3 x^3$  and  $\theta_4 x^4$ .

### Intuition



Suppose we penalize and make  $\theta_3, \theta_4$  really small.

$$\rightarrow \min_{\theta} \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + 1000 \theta_3^2 + 1000 \theta_4^2$$

$\theta_3 \approx 0$        $\theta_4 \approx 0$

We could also regularize all of our theta parameters in a single summation as:

