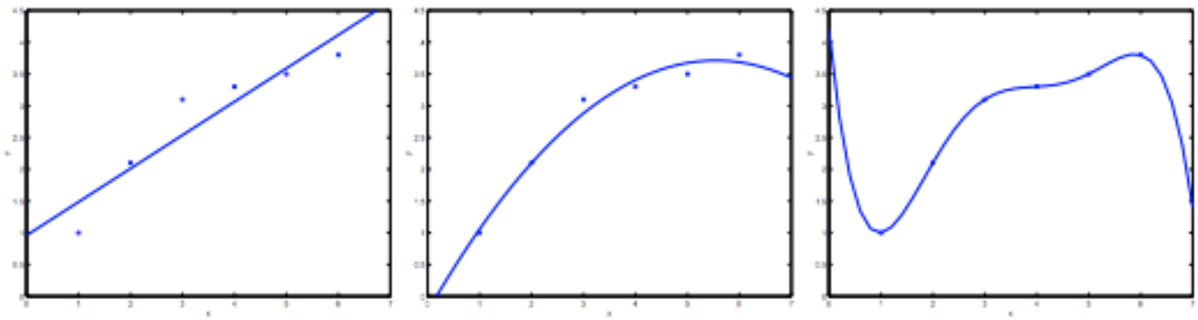
Overfitting

During fitting the parameters, concerns are overfitting or Underfitting. Underfitting, or high bias happens when we try to use very simple model for our complex model. On the other hand, if we try to fit very well on the training examples. In this case we get low training error and high error on testing dataset.



In case of overfitting, we can take these approaches:

- Reduce number of features
  - Manually select what features to keep
  - Use a model selection (learn about it later in a course)
- Regularization
  - Keep all the features, but reduce the magnitude of parameter θ_j
  - *Regularization works well when we have a lot of slightly useful features*

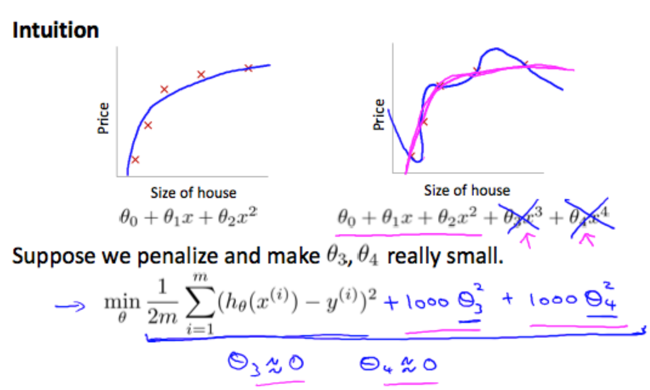
Cost Function

One of the solution to avoid overfitting is to use regularization by reducing the weight of the terms in our function carrying by increasing their cost. Suppose we have this hypothesis:
$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_3 x^4$

We would like to make it more quadratic form and eliminate the effect of last two terms, by without actually getting rid of those terms. We can modify the cost function instead:

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

By adding two extra terms at the end the cost of θ_3 and θ_4 will be inflated. Now in order to get the cost function close to zero, values of θ_3 and θ_4 must reduce to near 0. As a result, we see the hypothesis looks more like quadratic but fits the data better due to extra small values of last two terms.



We can regularize all pf the parameters, using equation below:
$min_\theta \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=0}^{n} \theta_j^2$

Overfitting

Regularized Linear Regression

We may apply gradient descent for linear regression with regularization term of λ.

Repeat {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha [\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}] + \frac{\lambda}{m} \theta_j$$

}

Term $\frac{\lambda}{m}\theta_j$ performs regularization.

Using Regularization in Normal Equation

In equation, we add regularization parameter as follow:

$\theta = (X^T X + \lambda L)^{-1} X^T y$, where L= identity matrix

Matrix L is (n+1) by (n+1) dimension. By adding regularization, the resultant matrix is always invertible.

Regularized Logistic Regression

We can regularize logistic regression in a similar way that we regularize linear regression. As a result, we can avoid overfitting. The following image shows how the regularized function, displayed by the pink line, is less likely to overfit than the non-regularized function represented by the blue line.