

# Regularization and Bias/Variance

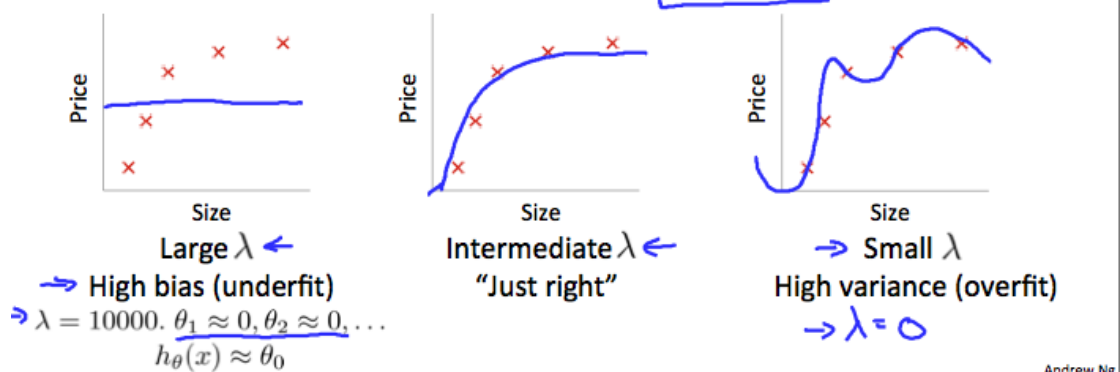
**Note:** [The regularization term below and through out the video should be  $(\lambda/2m)\sum_{j=1}^n \theta_j^2$

and NOT  $(\lambda/2m)\sum_{j=1}^n \theta_j^2$

## Linear regression with regularization

Model: 
$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^m \theta_j^2$$



In the figure above, we see that as  $\lambda$  increases, our fit becomes more rigid. On the other hand, as  $\lambda$  approaches 0, we tend to over overfit the data. So how do we choose our parameter  $\lambda$  to get it 'just right'? In order to choose the model and the regularization term  $\lambda$ , we need to:

1. Create a list of lambdas (i.e.  $\lambda \in \{0, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24\}$ );
2. Create a set of models with different degrees or any other variants.
3. Iterate through the  $\lambda$ s and for each  $\lambda$  go through all the models to learn some  $\Theta$ .
4. Compute the cross validation error using the learned  $\Theta$  (computed with  $\lambda$ ) on the  $J_{cv}(\Theta)$  **without** regularization or  $\lambda = 0$ .
5. Select the best combo that produces the lowest error on the cross validation set.
6. Using the best combo  $\Theta$  and  $\lambda$ , apply it on  $J_{test}(\Theta)$  to see if it has a good generalization of the problem.