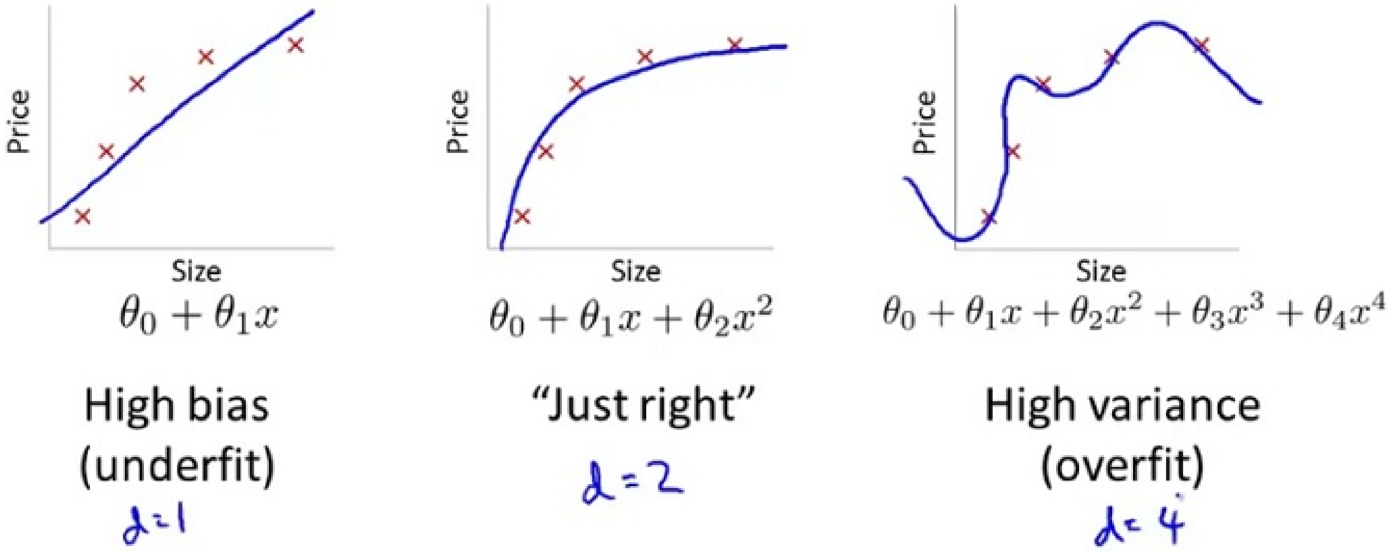
Bias vs. Variance

If we get bad results on our predictions it's usually because of one of:

- **High bias** - underfitting problem
- **High variance** - overfitting problem

Bias/variance shown graphically below:

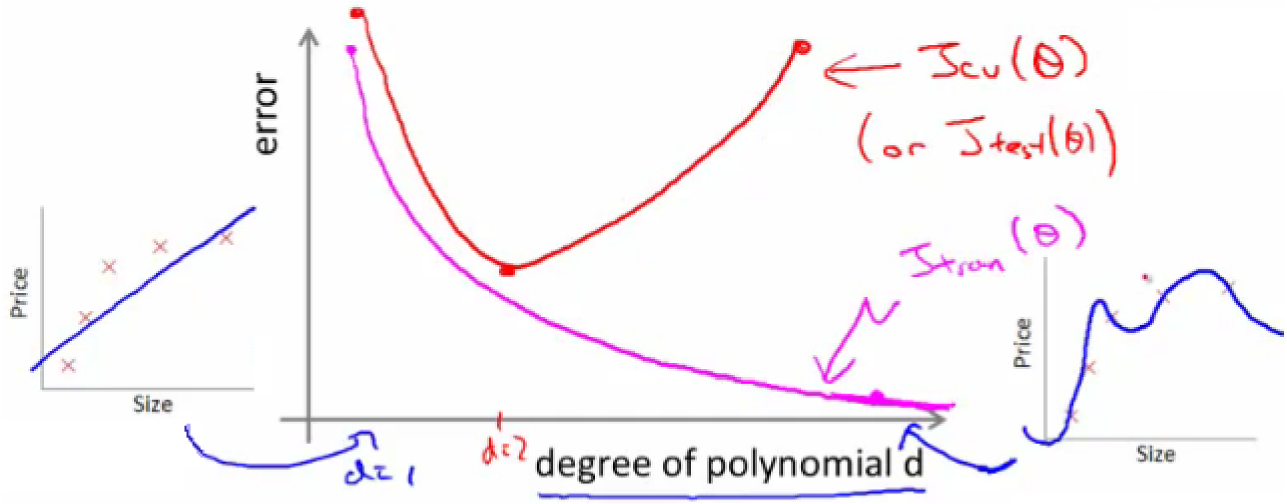


Now that we're armed with the notion of training and validation in test sets, we can understand the concepts of bias and variance a little bit better. Concretely, let's let our training error and cross validation error be defined as we do previously:

Training error: $$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Cross validation error: $$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

Let's look at the training error and the cross validation error and plot them on the figure below (this sort of plot help us to better understand the notions of bias and variance):



Let's start with the training error, as we increase the degree of the polynomial, we're going to be able to fit our training set better and better and so if $d = 1$, then there is high training error, if we have a very high degree of polynomial our training error is going to be really low, maybe even $J_{train}(\Theta) \approx 0$ because will fit the training set really well.

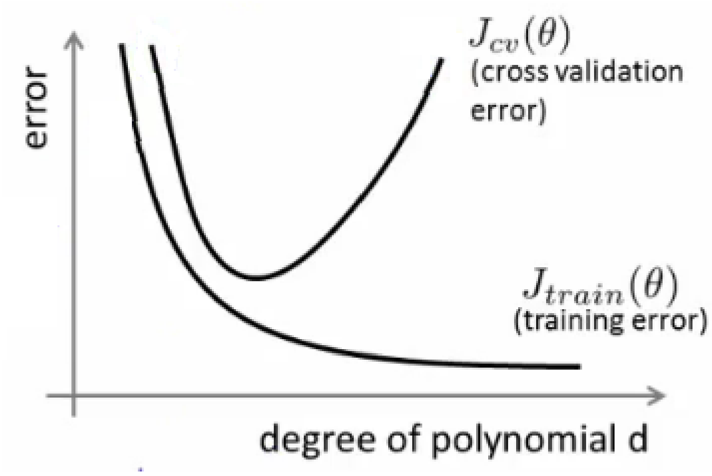
So, as we increase the degree of polynomial, we find typically that the training error decreases.

Now, let's look at the cross-validation error, if we look at the test set error, we'll get a pretty similar result as if we were to plot the cross validation error. So, we know that if $d = 1$, we're fitting a very simple function and so we may be **underfitting the training set** and so it's going to be **very high cross-validation error**.

If we fit an intermediate degree polynomial, we're going to have a much lower cross-validation error because we're finding a much better fit to the data. Conversely, if d were too high. So if d took a high value, then we're again **overfitting**, and so we end up with a **high value for cross-validation error**.

Diagnosing Bias vs. Variance

Suppose our learning algorithm is performing less well than we were hoping. ($J_{CV}(\Theta)$ or $J_{train}(\Theta)$ is high). Is it a bias problem or a variance problem?



- **If d is too small:** This probably corresponds to a **high bias problem**.
- **If d is too large:** This probably corresponds to a **high variance problem**.

For the high bias case, we find both cross validation and training error are high:

- **Bias (underfit):**
  - $J_{train}(\Theta)$ will be high.
  - $J_{train}(\Theta) \approx J_{CV}(\Theta)$.

For high variance, we find the cross validation error is high but training error is low:

- **Variance (overfit):**
  - $J_{train}(\Theta)$ will be low.
  - $J_{CV}(\Theta) >> J_{train}(\Theta)$

Video Question: Suppose you have a classification problem. The (misclassification) error is defined as $\frac{1}{m}\sum_{i=1}^{m} \mathrm{err}(h_\theta(x^{(i)}), y^{(i)})$, and the cross validation (misclassification) error is similarly defined, using the cross validation examples $(x_{\text{cv}}^{(1)}, y_{\text{cv}}^{(1)}), \ldots, (x_{\text{cv}}^{(m_{\text{cv}})}, y_{\text{cv}}^{(m_{\text{cv}})})$. Suppose your training error is 0.10, and your cross validation error is 0.30. What problem is the algorithm most likely to be suffering from?

- High bias (overfitting)
- High bias (underfitting)

High variance (overfitting)

- High variance (underfitting)

Summary

In this section we examine the relationship between the degree of the polynomial d and the underfitting or overfitting of our hypothesis.

- We need to distinguish whether **bias** or **variance** is the problem contributing to bad predictions.
- High bias is underfitting and high variance is overfitting. Ideally, we need to find a golden mean between these two.

The training error will tend to **decrease** as we increase the degree d of the polynomial.

At the same time, the cross validation error will tend to **decrease** as we increase d up to a point, and then it will increase as d is increased, forming a convex curve.

High bias (underfitting): both $J_{train}(\Theta)$ and $J_{CV}(\Theta)$ will be high. Also, $J_{CV}(\Theta) \approx J_{train}(\Theta)$.

High variance (overfitting): $J_{train}(\Theta)$ will be low and $J_{CV}(\Theta)$ will be much greater than $J_{train}(\Theta)$.

The is summarized in the figure below: