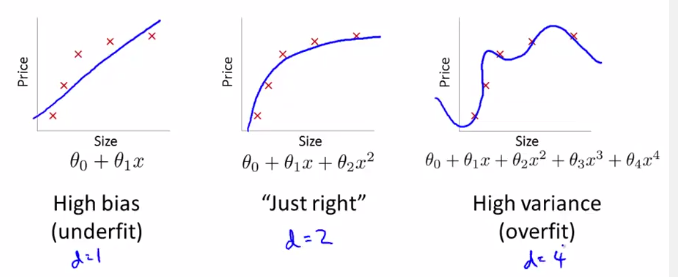
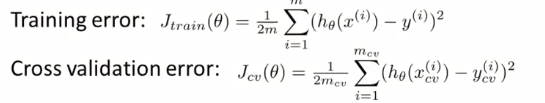
***Bias vs. Variance***

1. **DIAGNOSING BIAS VS. VARIANCE**

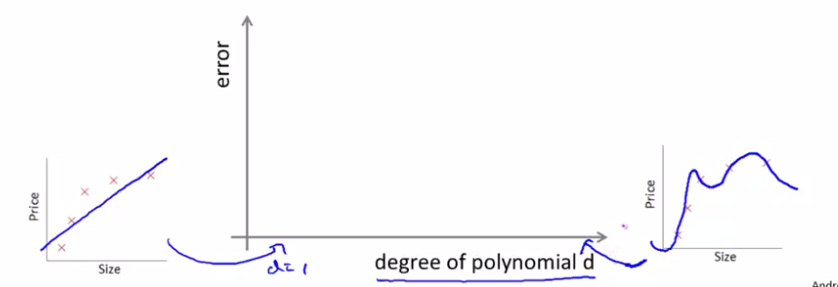
* If running a learning algorithm + it doesn't do as well as you were hoping, almost all the time it will be b/c you have either a **high bias problem** or a **high variance problem** (either an **underfitting** problem or an **overfitting** problem)
* In this case, it's very important to figure out which of these 2 problems it is, bias or variance, or a bit of both
* Knowing which of these 2 things is happening gives a very strong indicator for whether there are
* If you fit a too-simple hypothesis, it looks like a straight line through data that that underfits it.
* If you fit a too-complex hypothesis, it might fit a training set perfectly, but overfit new data
* We want a hypothesis of some intermediate level of complexity, maybe a 2 degree polynomials, not too low and not too high of a degree
* i.e. want a degree that gives you best generalization error.



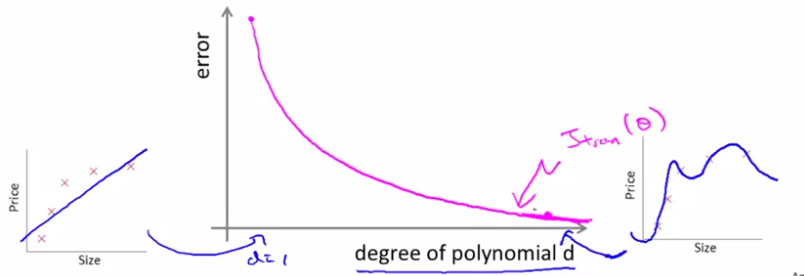
* Remember the **training error** and **cross validation error**



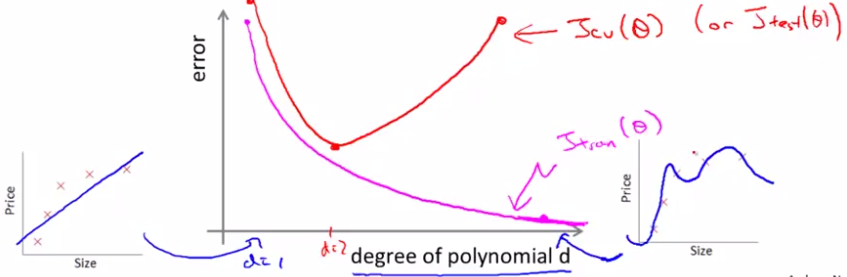
* i.e. the **average squared error** as measured on either set
* Now let's plot the hypothesis error as a function of the degree of polynomial d



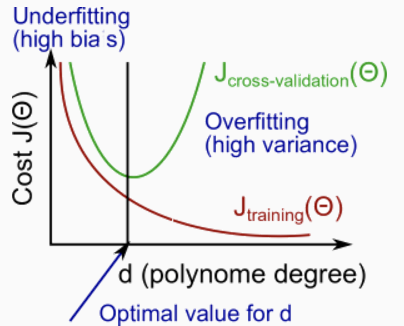
* Where maybe d = 1, we’re fitting very simple functions, when d = 4 or more, we’re fitting very complex, high-order polynomials that might fit the training set w/ much more complex functions
* Let's start with the training error.
* As we increase the degree of the polynomial, we're going to fit our training set better + better and so, if d = 1, we’ll have a very high training error.
* If we have a very high-degree polynomial, our training error is going to be very low, maybe even 0 if it fit the training set that well.
* As we increase of d, typically the training error, J\_train(Ө), decreases



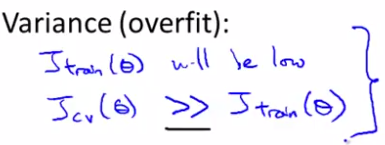
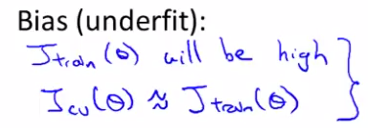
* Next, let's look at the CV error (if we were to plot the test set error, we'd get a pretty similar result to the CV error)
* If we fit an intermediate-degree polynomial, d= 2, we’re going to have a much lower CV error b/c we find a much better fit to the data.
* Conversely if d were too high, d = 4, then we're again *overfitting* + end up w/ a high value for J\_cv(Ө)



* This plot also helps us to better understand the notions of bias + variance.
* If you have a learning algorithm + the CV or test set error is high, we need to figure out if the algorithm suffering from high bias or high variance.
* When CV error is high, this corresponds to 1 of 2 regions
* The region w/ the lower degree d polynomial corresponds to a **high bias problem** (fitting an overly-low order polynomial when we really needed a higher-order polynomial to fit the data)
* The region w/ the higher degree d polynomial corresponds to a high **variance problem** (if d was too large for the data set that we have)



* For the high-bias case (under-fitting), what we find is that both the CV *AND* training error are going to be high.
* If you see this combo, that's a sign your algorithm may be suffering from high bias.
* If your algorithm is suffering from high variance, the training error is going to be low (fitting training set very well) but the CV error is much bigger
* The key that distinguishes these 2 cases is:
* If you have a **high bias** problem, the training set + CV set error will *both* be
* If you have a **high variance** problem, the training set error will usually be lower than the CV error



* By diagnosing whether a learning algorithm is suffering from high bias or high variance, we get much better guidance for what might be promising things to try in order to improve the performance of the learning algorithm

**II. REGULARIZATION AND BIAS/VARIANCE**

**III. LEARNING CURVES**

**IV. DECIDING WHAT TO DO NEXT REVISITED**