Diagnostics

Evaluating a Learning Algorithm

Advice for Applying Machine Learning

Debugging a learning algorithm

- Suppose you have implemented regularized linear regression to predict housing prices

- However you find that it makes large errors in its predictions?
- What's next?

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

Actions

- Some actions
  - Get more training examples
  - Try smaller set of features (a small set of features)
  - Try getting additional features (just the opposite)
  - Try adding polynomial features
  - Try increasing lambda
  - Try decreasing lambda
- People generally randomly choose one and try it, which is waste of time most of the time.

Machine learning diagnostic:

- **Diagnostic**: A test that you can run to gain insight what is/isn't working with a learning algorithm, and gain guidance as to how best to improve its performance.

- Diagnostics can take time to implement, but doing so can be a very good use of your time

Exercise

- Which of the following statements about diagnostics are true? Check all that apply.
  - It’s hard to tell what will work to improve a learning algorithm, so the best approach is to go with gut feeling and just see what works.
  - Diagnostics can give guidance as to what might be more fruitful things to try to improve a learning algorithm.
  - Diagnostics can be time-consuming to implement and try, but they can still be a very good use of your time.
  - A diagnostic can sometimes rule out certain courses of action (changes to your learning algorithm) as being unlikely to improve its performance significantly.

Diagnosing Bias vs Variance

with Polyomial Degree

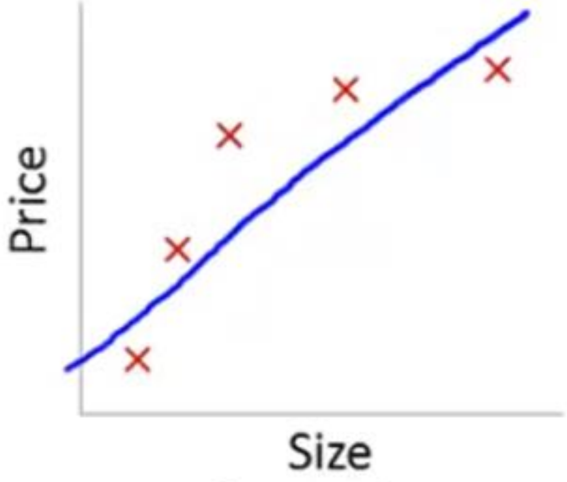
Bias and Variance

Advice for Applying Machine Learning

Introduction

- Most of the time you will have
  - High variance (overfitting)
  - High bias (underfitting)

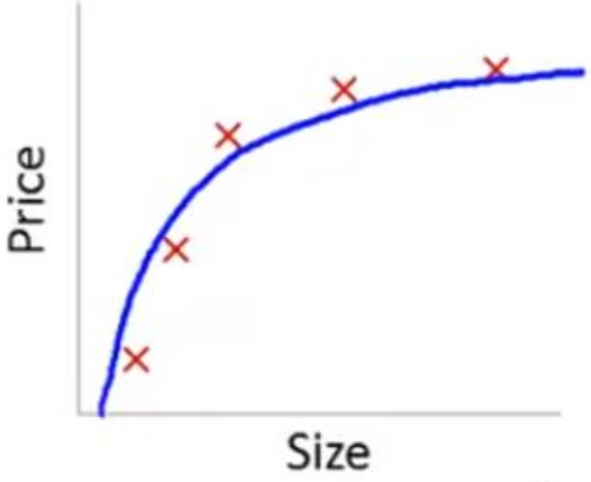
Bias/variance



$\theta_0 + \theta_1 x$

High bias
(underfit)

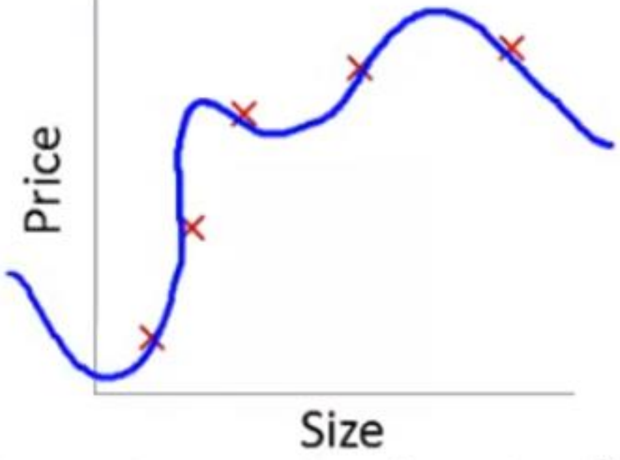
$d=1$



$\theta_0 + \theta_1 x + \theta_2 x^2$

"Just right"

$d=2$



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

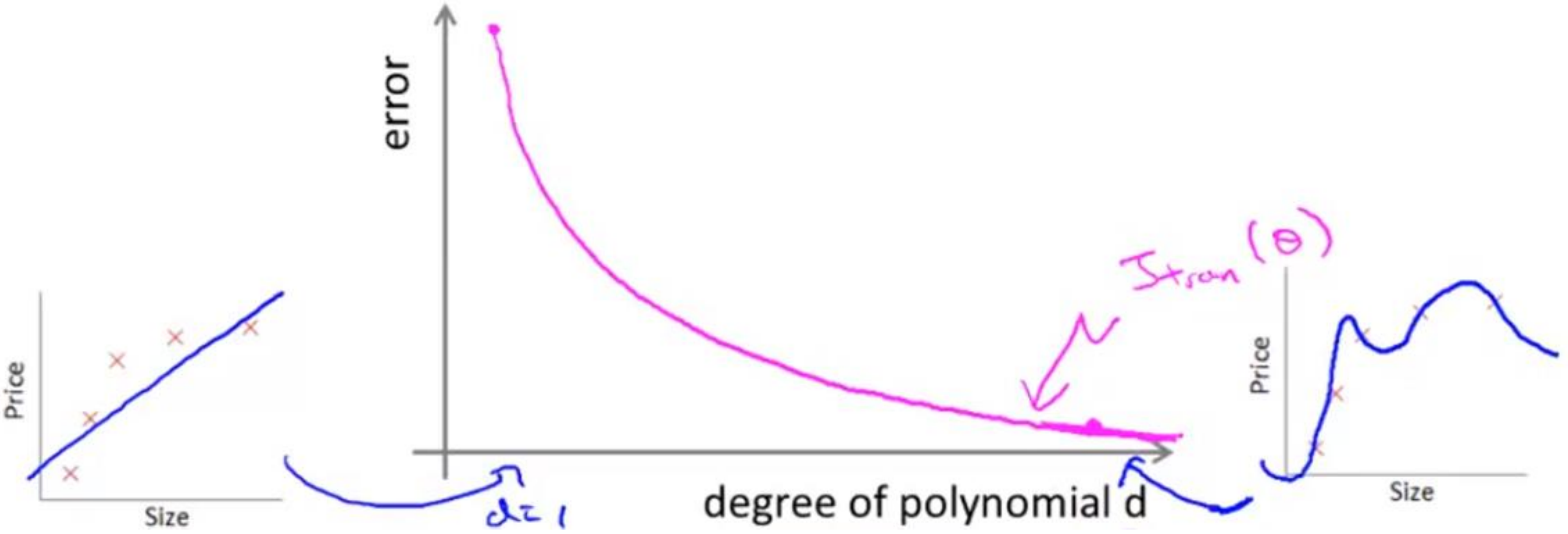
High variance
(overfit)

$d=4$

Bias/variance

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

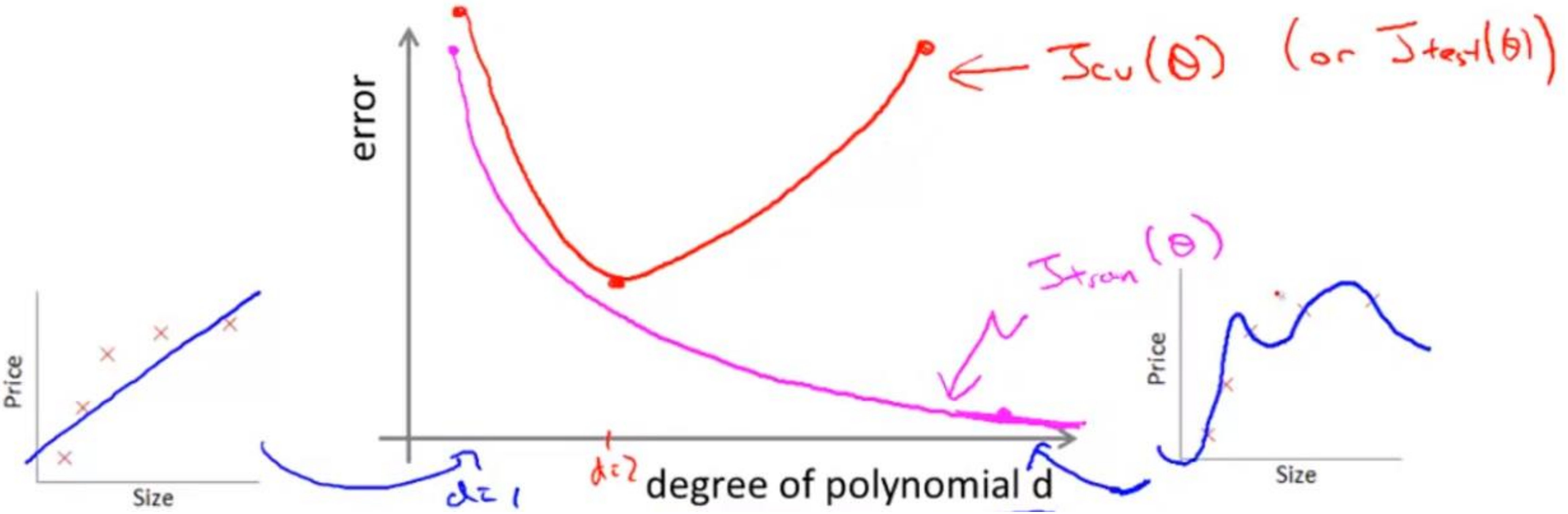
Cross validation error: $J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x^{(i)}_{cv}) - y^{(i)}_{cv})^2$



Bias/variance

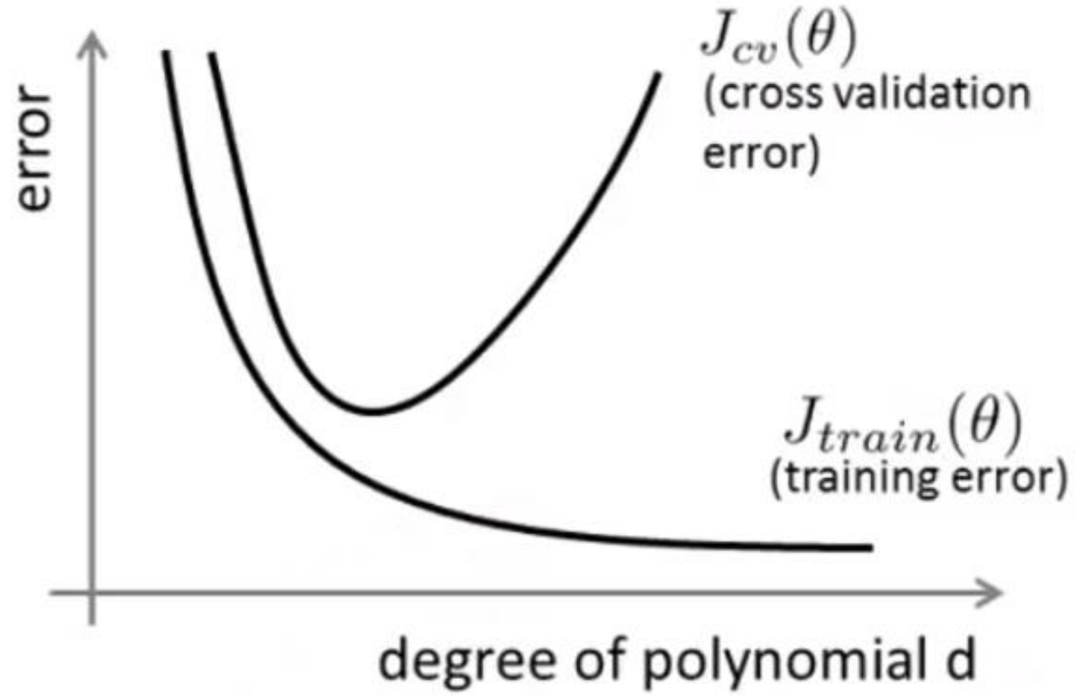
Training error: $J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

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



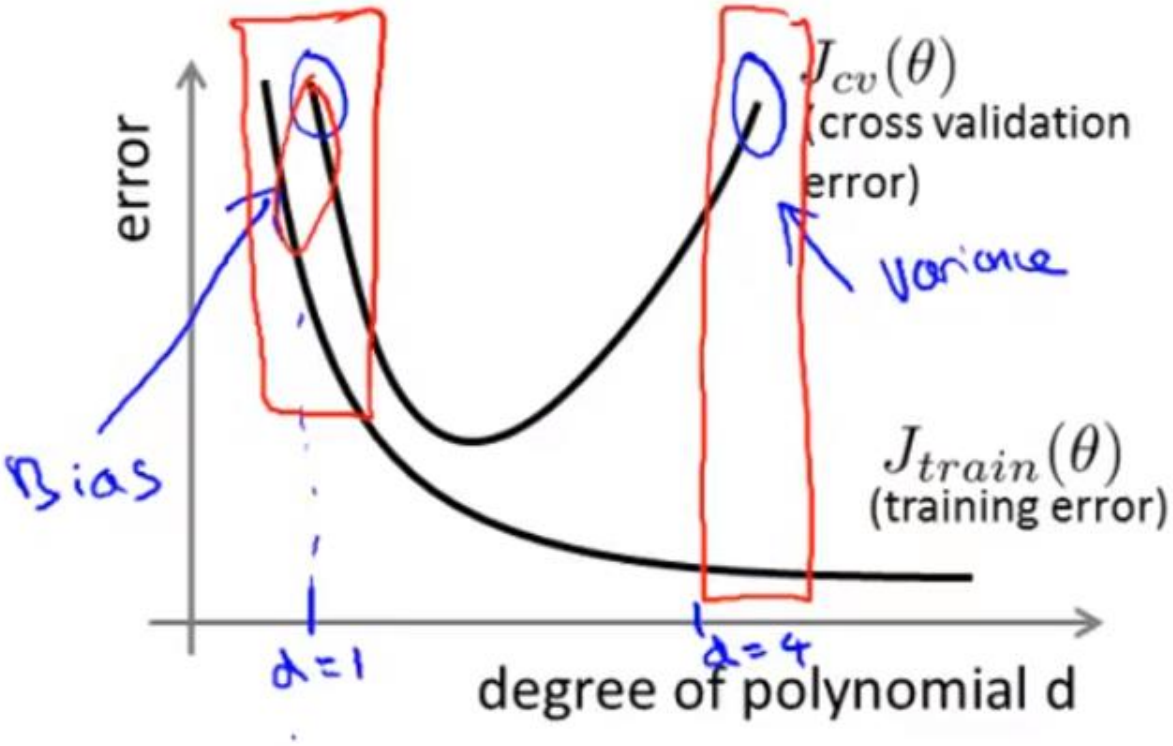
Diagnosing bias vs. variance

Suppose your learning algorithm is performing less well than you were hoping. ($J_{cv}(\theta)$ or $J_{test}(\theta)$ is high.) Is it a bias problem or a variance problem?



Diagnosing bias vs. variance

Suppose your learning algorithm is performing less well than you were hoping. ($J_{cv}(\theta)$ or $J_{test}(\theta)$ is high.) Is it a bias problem or a variance problem?



Bias (underfit):

$J_{train}(\theta)$ will be high
$J_{cv}(\theta) \approx J_{train}(\theta)$

Variance (overfit):

$J_{train}(\theta)$ will be low
$J_{cv}(\theta) \underline{\gg} J_{train}(\theta)$

Exercise

- Suppose you have a classification problem. The (misclassification) error is defined as

$$\frac{1}{m}\sum_{i=1}^{m} err\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

- and the cross validation (misclassification) error is similarly defined, using the cross validation examples $(x_{CV}^{(1)}, y_{CV}^{(1)}), \ldots, (x_{CV}^{m_{cv}}, y_{CV}^{m_{cv}})$

- 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)

Diagnosing Bias vs Variance

with Regularization Parameter

Bias and Variance

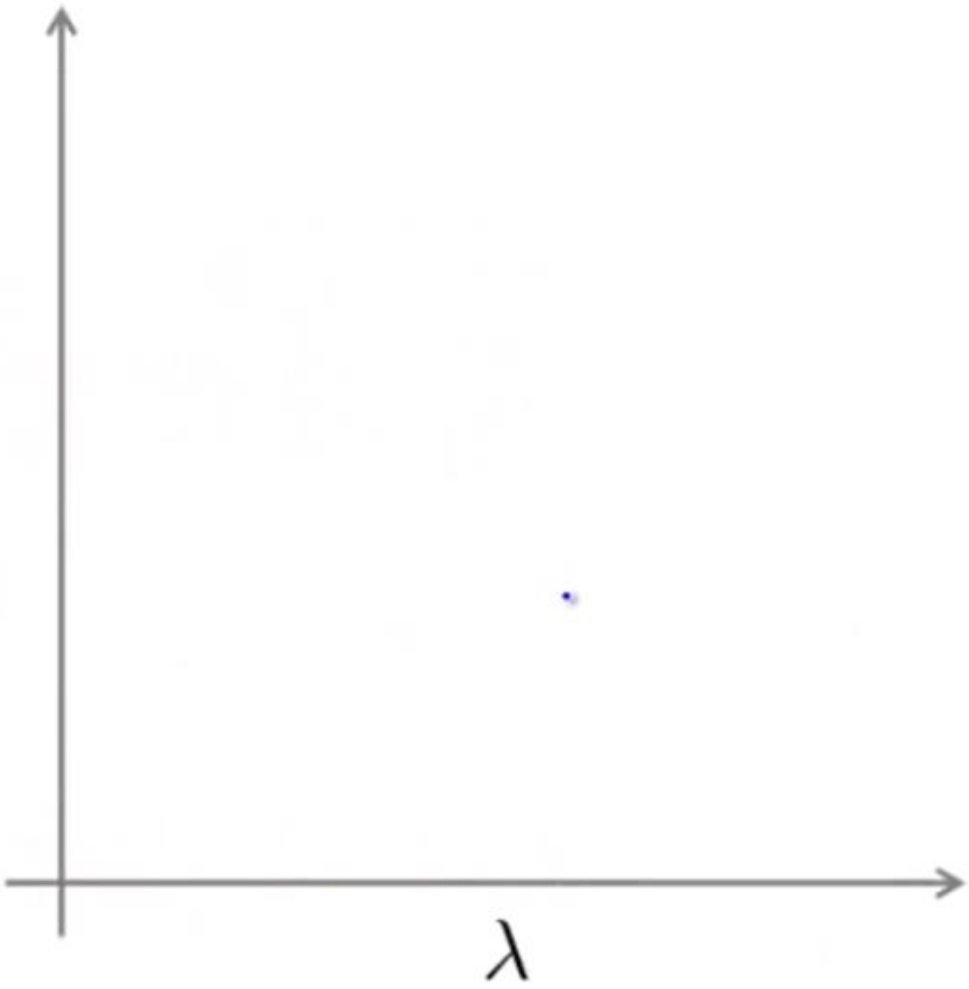
Advice for Applying Machine Learning

Bias/variance as a function of the regularization parameter λ

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

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

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

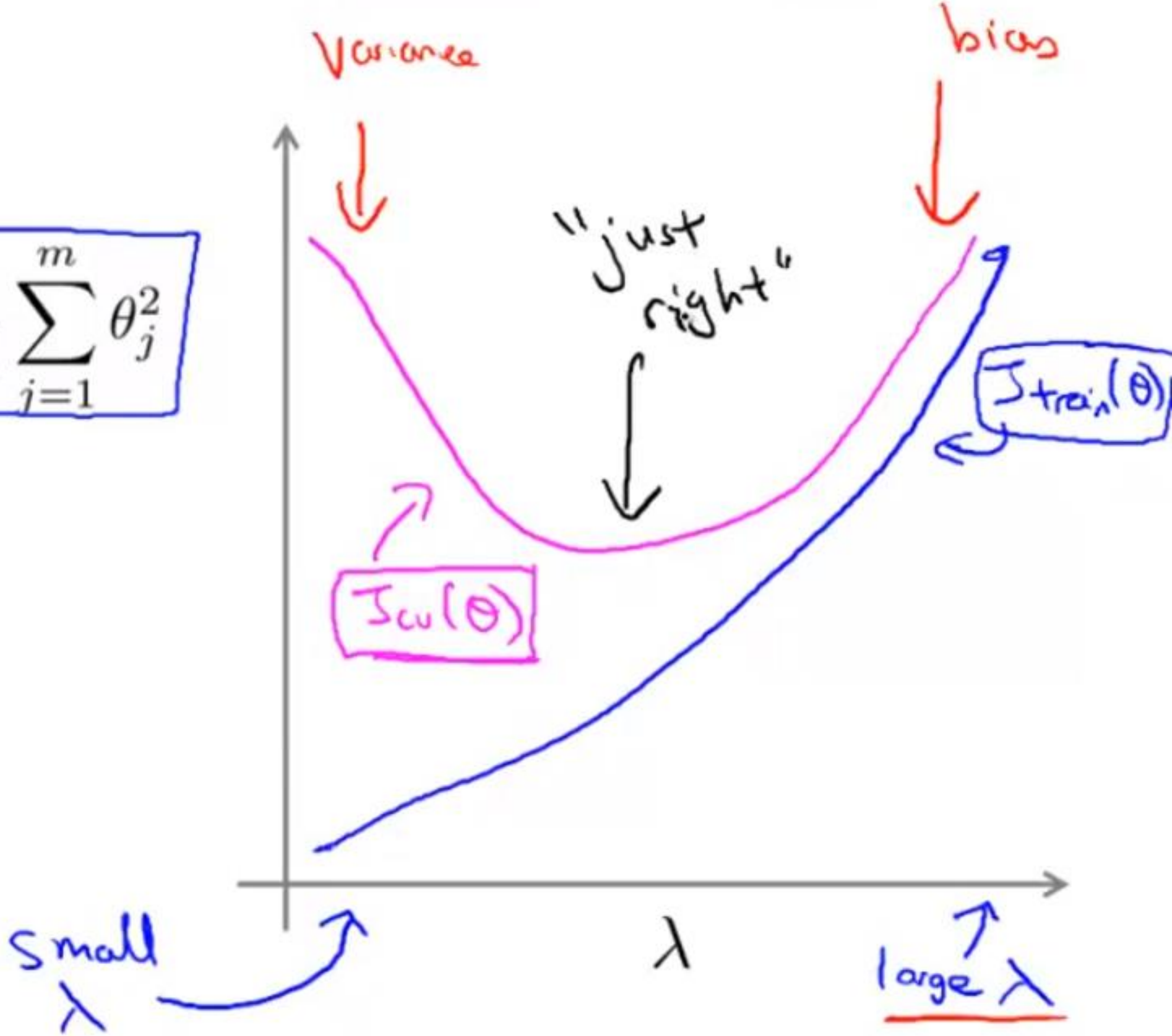


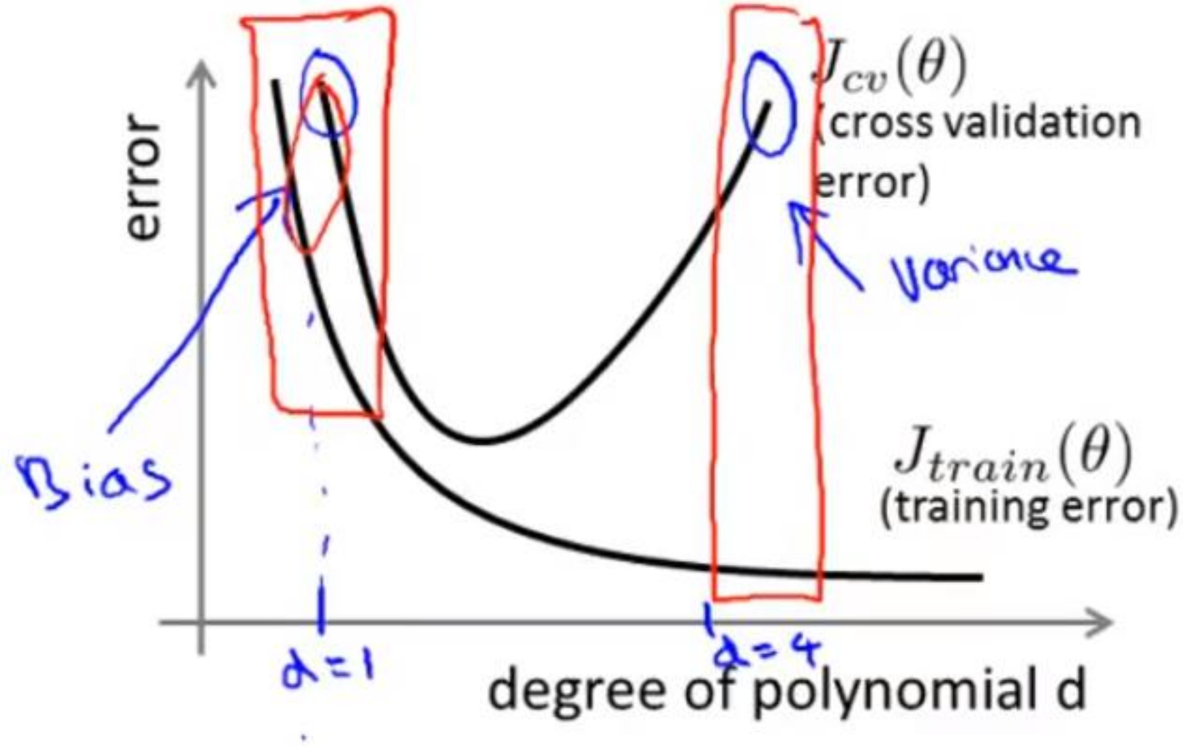
Bias/variance as a function of the regularization parameter λ

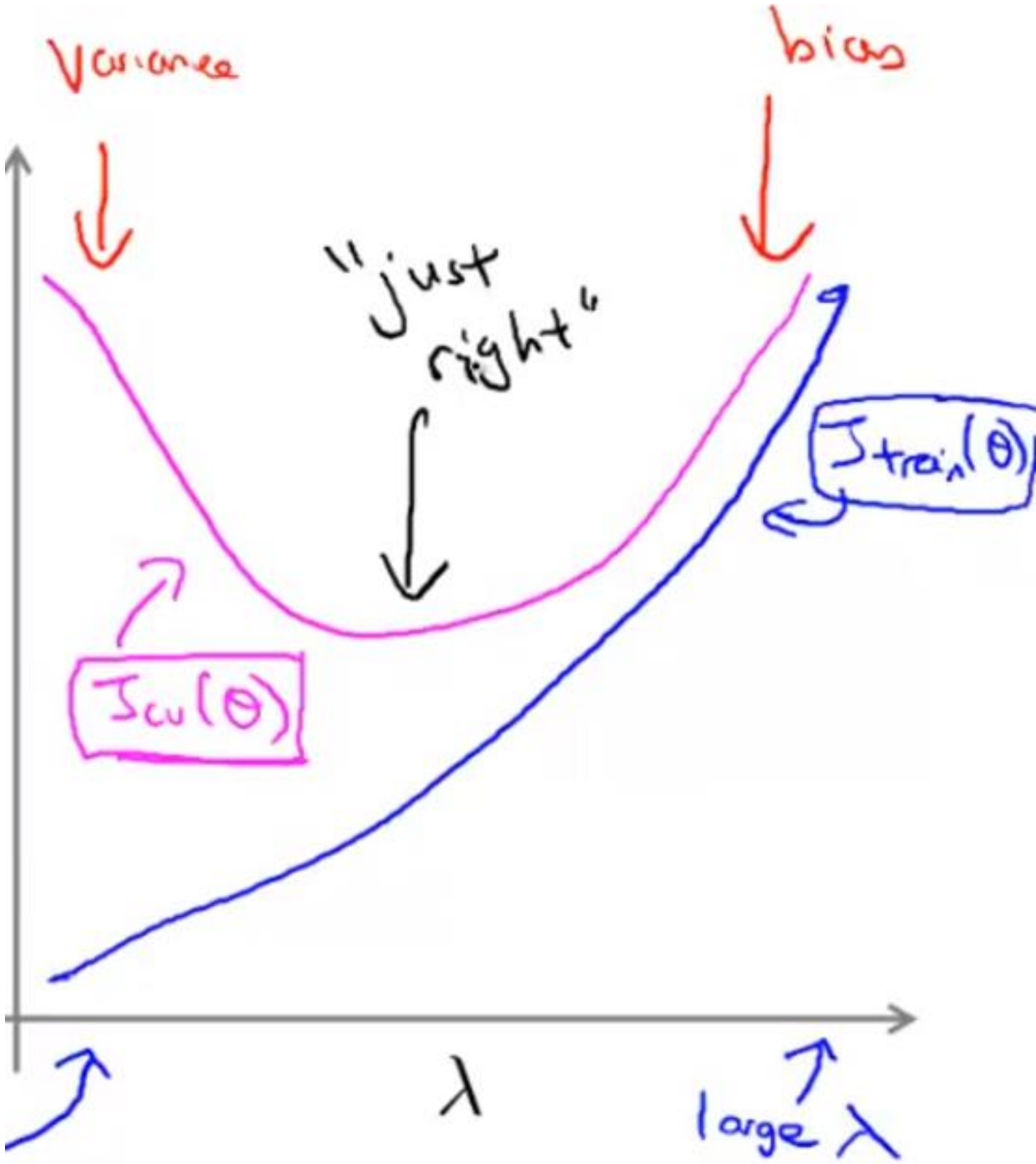
$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \boxed{\frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2}$$

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

$$\boxed{J_{cv}(\theta)} = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x^{(i)}_{cv}) - y^{(i)}_{cv})^2$$




error
$J_{cv}(\theta)$
(cross validation error)
$J_{train}(\theta)$
(training error)
degree of polynomial d
Bias
variance
d=1
d=4


variance
bias
"just right"
$J_{train}(\theta)$
$J_{cv}(\theta)$
λ
large λ

Summary

- We understood that
  - as λ increases, our fit becomes more rigid.
  - as λ approaches 0, we tend to overfit the data.
- So how do we choose our parameter λ to get it 'just right' ?
- In order to choose the model and the regularization term λ, we need to:
  - Create a list of lambdas (i.e. $\lambda \in \{0,0.01,0.02,0.04,0.08,0.16,0.32,0.64,1.28,2.56,5.12,10.24\}$);
  - Create a set of models with different degrees or any other variants.
  - Iterate through the λs and for each λ go through *all the models* to learn some Θ.
  - Compute the CV error using the learned Θ (computed with λ) on the $J_{CV}(\Theta)$ **without** regularization or λ = 0.
  - Select the best combo that produces the lowest error on the cross validation set.
  - Using the best combo Θ and λ, apply it on $J_{test}(\Theta)$ to see if it has a *good generalization* of the problem.