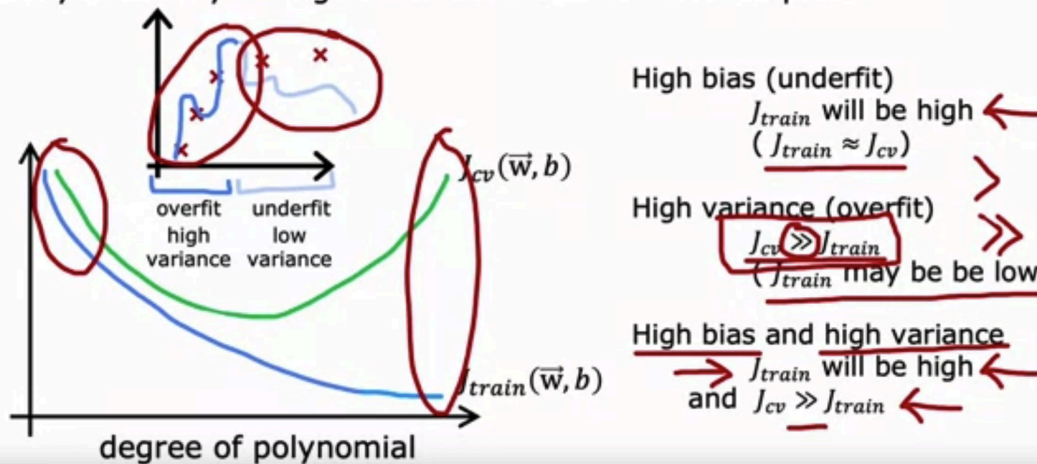


Diagnosing bias and variance

How do you tell if your algorithm has a bias or variance problem?



If the model's cross validation error J_{cv} is much higher than the training error J_{train} , this is an indication that the model has...

- ☐ high bias
- ☒ high variance
- ☐ Low bias
- ☐ Low variance

Bias/variance examples

Baseline performance	: 10.6%		10.6%		10.6%
Training error (J_{train})	: 10.8%	0.2%	15.0%	4.4%	15.0%
Cross validation error (J_{cv})	: 14.8%	4.0%	15.5%	0.5%	19.7%
		high variance	high bias	high bias	high variance

2. Which of these is the best way to determine whether your model has high bias (has underfit the training data)?

- ☐ Compare the training error to the cross validation error.
- ☐ See if the training error is high (above 15% or so)
- ☒ Compare the training error to the baseline level of performance
- ☐ See if the cross validation error is high compared to the baseline level of performance

Debugging a learning algorithm

You've implemented regularized linear regression on housing prices

$$J(\vec{w}, b) = \frac{1}{2m} \sum_{i=1}^m (f_{\vec{w}, b}(\vec{x}^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^n w_j^2$$

But it makes unacceptably large errors in predictions. What do you try next?

- Get more training examples
 - Try smaller sets of features x, x^2, x^3, x^4, \dots
 - Try getting additional features $(x_1^2, x_2^2, x_1 x_2, \text{etc})$
 - Try adding polynomial features
 - Try decreasing λ
 - Try increasing λ
- fixes high variance
fixes high variance
fixes high bias
fixes high bias
fixes high bias
fixes high variance

3.

You find that your algorithm has high bias. Which of these seem like good options for improving the algorithm's performance? Hint: two of these are correct.

- ☐ Remove examples from the training set
- ☒ Decrease the regularization parameter λ (lambda)
- ☐ Collect more training examples
- ☒ Collect additional features or add polynomial features

4. You find that your algorithm has a training error of 2%, and a cross validation error of 20% (much higher than the training error). Based on the conclusion you would draw about whether the algorithm has a high bias or high variance problem, which of these seem like good options for improving the algorithm's performance? Hint: two of these are correct.

- ☒ Increase the regularization parameter λ
- ☒ Collect more training data
- ☐ Reduce the training set size
- ☐ Decrease the regularization parameter λ