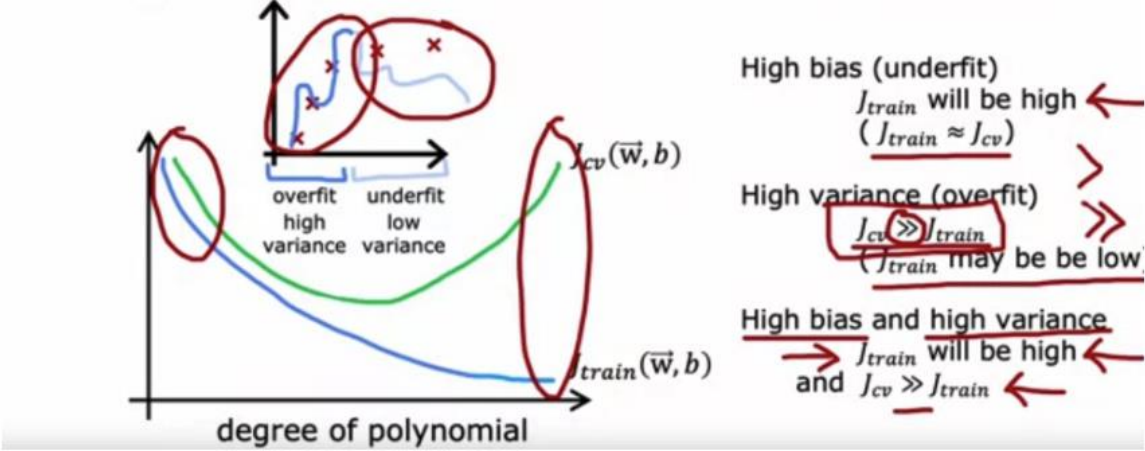


# Diagnosing bias and variance

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



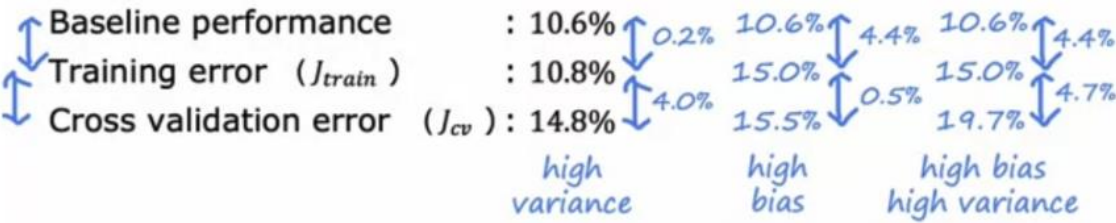
If the model's cross validation error  $J_{cv}$  is much higher than the training error  $J_{train}$ , this is an indication that the model has...

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

✓ Correct

When  $J_{cv} \gg J_{train}$  (whether  $J_{train}$  is also high or not, this is a sign that the model is overfitting to the training data and performing much worse on new examples.

# Bias/variance examples



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

✓ Correct

Correct. If comparing your model's training error to a baseline level of performance (such as human level performance, or performance of other well-established models), if your model's training error is much higher, then this is a sign that the model has high bias (has underfit).

## Debugging a learning algorithm

You've implemented regularized linear regression on housing prices

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

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

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

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.

- ☒ Collect additional features or add polynomial features

✓ **Correct**

Correct. More features could potentially help the model better fit the training examples.

- ☐ Collect more training examples

- ☒ Decrease the regularization parameter  $\lambda$  (lambda)

✓ **Correct**

Correct. Decreasing regularization can help the model better fit the training data.

- ☐ Remove examples from the training set

4.

1 / 1 point

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.

- ☐ Reduce the training set size

- ☒ Collect more training data

✓ **Correct**

Yes, the model appears to have high variance (overfit), and collecting more training examples would help reduce high variance.

- ☐ Decrease the regularization parameter  $\lambda$

- ☒ Increase the regularization parameter  $\lambda$

✓ **Correct**

Yes, the model appears to have high variance (overfit), and increasing regularization would help reduce high variance.