

Learning Curves

Training an algorithm on a very few number of data points (such as 1, 2 or 3) will easily have 0 errors because we can always find a quadratic curve that touches exactly those number of points. Hence:

- As the training set gets larger, the error for a quadratic function increases.
- The error value will plateau out after a certain m , or training set size.

Experiencing high bias:

Low training set size: causes $J_{\text{train}}(\Theta)$ to be low and $J_{\text{CV}}(\Theta)$ to be high.

Large training set size: causes both $J_{\text{train}}(\Theta)$ and $J_{\text{CV}}(\Theta)$ to be high with $J_{\text{train}}(\Theta) \approx J_{\text{CV}}(\Theta)$.

If a learning algorithm is suffering from **high bias**, getting more training data will not (**by itself**) help much.

More on Bias vs. Variance

Typical learning curve for high bias (at fixed model complexity):



Experiencing high variance:

Low training set size: $J_{\text{train}}(\Theta)$ will be low and $J_{\text{CV}}(\Theta)$ will be high.

Large training set size: $J_{\text{train}}(\Theta)$ increases with training set size and $J_{\text{CV}}(\Theta)$ continues to decrease without leveling off. Also, $J_{\text{train}}(\Theta) < J_{\text{CV}}(\Theta)$ but the difference between them remains significant.

If a learning algorithm is suffering from **high variance**, getting more training data is likely to help.

More on Bias vs. Variance

Typical **learning curve** for high variance (at fixed model complexity):

