**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:

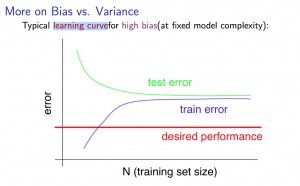
* As the training set gets larger, the error for a quadratic function increases.
* The error value will plateau out after a certain m, or training set size.

**Experiencing high bias:**

**Low training set size**: causes Jtrain(Θ) to be low and JCV(Θ) to be high.

**Large training set size**: causes both Jtrain(Θ) and JCV(Θ) to be high with Jtrain(Θ) ≈ JCV(Θ).

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



**Experiencing high variance:**

**Low training set size**: Jtrain(Θ) will be low and JCV(Θ) will be high.

**Large training set size**: Jtrain(Θ) increases with training set size and JCV(Θ) continues to decrease without leveling off. Also, Jtrain(Θ) < JCV(Θ) but the difference between them remains significant.

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

