Chapter 1 Advice for Applying Machine Learning

This chapter is about how to implement powerful algorithm.

Sometimes, a better algorithm is better than more data. But usually, we should try :

- more training examples
- smaller sets of features
- additional features
- polynomial features
- changing λ

We need **Machine Learning Diagnostic** to test whether we can gain insight that if it works on it. A diagnostic will take time to implementm, but it will be worth the time.

1.1 Evaluating a Hypothesis

Less error is not always meaning better, for it's possible to generalize to new examples.

We can divide dataset into training set and test set. Normally training set is of 70%.

Learn from training set and compute via test set. You can apply the cost or misclassification error.

$$\text{err}(h_\theta(x), y) = \begin{cases} 1, \text{if } h_\theta(x) \geq 0.5, y = 0 \\ 1, \text{if } h_\theta(x) \leq 0.5, y = 1 \\ 0, \text{else} \end{cases}$$

And the total cost is

$$\text{Test Err} = \frac{1}{m_{test}} \sum_{i=1}^{m_{test}} \text{err}(h_\theta(x_{test}^{(i)}), y^{(i)})$$

Via this, we can simply check whether the algorithm is proper or not.

1.1.1 Model Selection and Train/Validation/Test Sets

How to decide the degree of polynomial or regularized parameters?

Once the parameters are overfitting a dataset, the error will be lower than it should be.

We can do model selection over the degree of the polynomial functions. But the cost in training is not a fair estimate of the model generalization. We should divide the dataset into 3 parts rather than 2: training set(60%), cross validation set(20%), test set(20%). And get error for each part, then treat the loss of cross validation set as a quality of generalization.

1.2 Bias vs. Variance

1.2.1 Diagnosing Bias vs. Variance

High bias turns to underfitting and high variance turns to overfitting. It's important to determine which kind of problem we are facing.

We can see them from the training error and cross validation error. As degree of polynomial increasing, the error will decrease. And the cross validation error will be higher than it, like **Figure** 1.1.
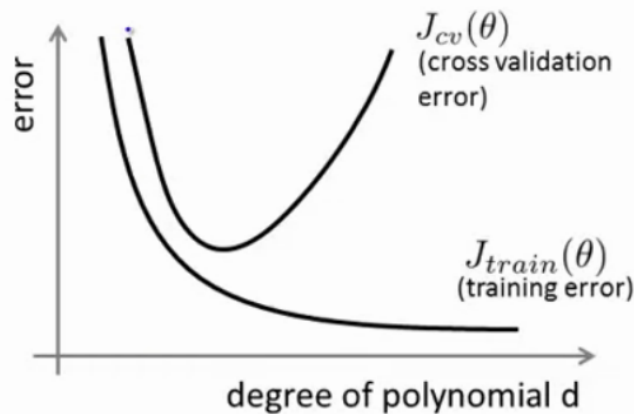

Figure 1.1: The error with degree of polynomial

If the bias is too big, the $J_{train}(\theta) \approx J_{cv}(\theta)$ will be high; If the variance is too big, $J_{train}(\theta) \ll J_{cv}(\theta)$

1.2.2 Regularization and Bias/Variance

If we have applied regularization to our algorithm, we will get an extra λ. If the λ is too large, the parameters will be rather small and then the bias gets too big then underfitting. If too small, similarly, overfitting.

So, we need to try different λ, and observe the relationship between loss and λ.

Word

sad Avenues