

Machine Learning Week-5

ramay7

March 7, 2017

Contents

1	Evaluating a Learning Algorithm	1
1.1	Evaluating a Hypothesis	1
1.2	Model Selection and Train/Validation/Test Sets	1
2	Bias vs Variance	2
2.1	Learning Curves	2
2.2	Deciding What to Do Next Revisited	3

1 Evaluating a Learning Algorithm

1.1 Evaluating a Hypothesis

A hypothesis may have a low error for the training examples but still be inaccurate (because of overfitting). Thus, to evaluate a hypothesis, given a dataset of training examples, we can split up the data into two sets: a **training set** and a **test set**. Typically, the training set consists of 70% of your data and the test set is the remaining 30%.

The new procedure using these two sets is then:

1. Learn Θ and minimize $J_{train}(\Theta)$ using the training set
2. Compute the test set error $J_{test}(\Theta)$

1.2 Model Selection and Train/Validation/Test Sets

One way to break down our dataset into the three sets is:

- Training set: 60%
- Cross validation set: 20%
- Test set: 20%

We can now calculate three separate error values for the three different sets using the following method:

1. Optimize the parameters in Θ using the training set for each polynomial degree.
2. Find the polynomial degree d with the least error using the cross validation set.

3. Estimate the generalization error using the test set with $J_{test}(\Theta^{(d)})$, ($d = \text{theta from polynomial with lower error}$);

This way, the degree of the polynomial d has not been trained using the test set.

2 Bias vs Variance

In order to choose the model and the regularization term λ , we need to:

- Create a list of lambdas (i.e. $\lambda \in \{0, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24\}$);
- Create a set of models with different degrees or any other variants.
- Iterate through the λ s and for each λ go through all the models to learn some Θ .
- Compute the cross validation error using the learned Θ (computed with λ) on the $J_{CV}(\Theta)$ without regularization or $\lambda = 0$.
- Select the best combo that produces the lowest error on the cross validation set.
- Using the best combo Θ and λ , apply it on $J_{test}(\Theta)$ to see if it has a good generalization of the problem.

2.1 Learning Curves

Experiencing **high bias**:

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

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

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**:

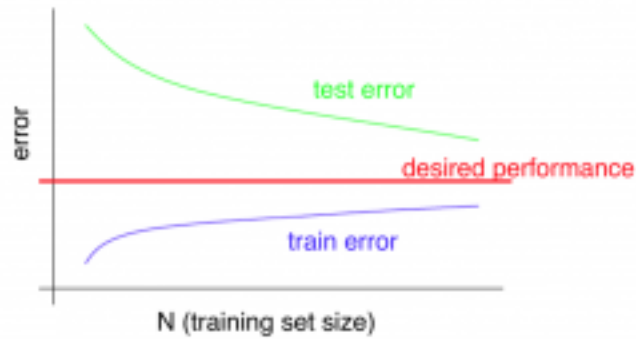
Low training set size: $J_{train}(\Theta)$ will be low and $J_{CV}(\Theta)$ will be high.

Large training set size: $J_{train}(\Theta)$ increases with training set size and $J_{CV}(\Theta)$ continues to decrease without leveling off. Also, $J_{train}(\Theta) < J_{CV}(\Theta)$ 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):



2.2 Deciding What to Do Next Revisited

Our decision process can be broken down as follows:

- **Getting more training examples:** Fixes high variance
- **Trying smaller sets of features:** Fixes high variance
- **Adding features:** Fixes high bias
- **Adding polynomial features:** Fixes high bias
- **Decreasing λ :** Fixes high bias
- **Increasing λ :** Fixes high variance.