Toronto CSC321
Notes - Lecture 9

Tony Chan

20 October 2018

1 What is Generalization?

"Generalization" means decreasing test error due to decreasing training error in model training. We always want to achieve more generalization, as we only target the test error. The difference between training error and test error is called "generalization gap".

2 Error Decomposition

Assume everything conditioned on x, t is the output variable.

$$\mathbb{E}[(y-t)^2] = (y_* - \mathbb{E}[y])^2 + Var(y) + Var(t) \tag{1}$$

where y is a random variable, corresponding to the output of a model trained on a sample of training set, $y_* := \mathbb{E}[t]$. The terms on the right are "bias", "variance", and "Bayes error".

3 Bag of Tricks

Though the best way to improve generalization is to collect more data, In practice, data are hard to get. Winning machine learning systems have used some or all of the following tricks.

1. Data Augmentation
2. Reducing the number of parameters
3. Weight decay
4. Early stopping
5. Ensembles
6. Stochastic regularization, e.g. dropout.