2018-12-12-From-Statistics-To-Lack-of-fit-test

Xinrui Hu

December 12, 2018

Contents

1 Background

According to Alan Turing, a system, which probably looks pretty complicate, can be described with simple mathmatical equation. In machine learning domain, the models can be usually interpreted with equations, and the parameters of "equations" is what we want to learn. Usually, the parameter we finally get is an estimator of the real value. But how do we describe how good the estimator is?

2 Bias and Variance

Let's start with the **Estimator.**

The parameter is a quantity of a real model e.g. the mean or the variance of random variable, the parameter of a probability distribution, etc. An estimator, according to Wikipedia is a rule for calculating an estimate of parameter based on observed data. Because the estimator is calculated by

a function of observed data, so the estimator is itself also random variable. (This concept is also the main idea of Bayesian Model in machine learning).

Suppose we want to know a paramter θ, which can be estimated by a sample set D_S drawn from X. $\hat{\theta}_S(X)$ is denoted as the estimator of θ based on observed data of random variable X. Let's denote x as a particular obseved data $X = x$. Sampling is an approximation of the real world. So it could happen that for some reasons the sample set doesn't represent the real issuse very well. There are some important quantified properties which are helpful before we dive deeper.

2.1 Sampling Deviation

For a given sample $X = x$, the sampling deviation is denoted as:

$$d(x) = \hat{\theta}_S(x) - E(\hat{\theta}_S(X)) = \hat{\theta}_S(x) - E(\hat{\theta}) \tag{1}$$

where $E(\hat{\theta})$ is the expected value of the estimator. As mentioned, the estimator is itself a random variable.

2.2 Estimator Bias

We randomly itteratively sample several times. Each time we give an estimate of θ. The bias then can be denoted as:

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta = E(\hat{(}\theta) - \theta) = E(error) \tag{2}$$

The distance btw. expected value of θ and the real theta being estimated. Bias can describe how far the average of estimates be off-target.

2.3 Estimator Variance

The expected value of the squared sampling deviation:

$$Var(\hat{\theta}) = E(\sqrt{(}(\hat{\theta} - E(\hat{\theta})))) \tag{3}$$

It describes the scatterness or clusterness of the estimates.

2.4 Mean Square Error

The expected value of the squared error:

$$MSE(\hat{\theta}) = E[\sqrt{(}(\hat{\theta}(X) - \theta))] \tag{4}$$

It is proven that $MSE = \sqrt{(}Bias) + Var$

3 *Estimator Variance

measures how "scatter" our estimator is to sampling, e.g. if we observe the stock price every 100ms instead of every 10ms would the estimator change a lot?

In statistics and machine learning, the **bias–variance tradeoff** is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Ideally, we hope to simultaneously minimize these two sources of error. However, the all-known delimma is that we can't get the "whole" data at all. In oder to optimal the performance we have to consider the trade-off btw. bias and variance. Firsty, what do they actually mean? According to Wikipedia:

- **Bias** is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
- **Variance** is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs(overfitting)