Landing Page... (/)

Variance of the Estimator in Machine Learning

Sep 4, 2020

Bias and variance describes the sources of error in a supervised learning setting, if your model is underfitting it has high bias, if your model is overfitting it has high variance. Its easy to memorise this, but intuition seems a little handwavy, and perhaps we can be a little more concrete on what this means.

Variance from where and of what?

Note that these terms are with respect to the learned model i.e. the estimator given different samples drawn from the training distribution. What is the bias and variance of the estimator f?

The expected error where $l(f, y)$ is the loss associated with the estimator in estimating y using $f(x) = \hat{y}$, and σ^2 is variance of random noise is given by

$$\mathbb{E}_D[l(f, y)] = [\mathbb{E}_D(f(x) - y)]^2 + Var_D[f(x)] + \sigma^2 \qquad (1)$$

Note that this is an expected error with expectation taken over different training sets D and different learned estimators f that have been trained on D.

Sample $D_a, D_b, \cdots D_m$ from the train set, where each $D_a = (x_1, y_1), \cdots (x_n, y_n)$ is a set of training observations. Train $f_a, f_b, \cdots, f_m$ different models on each of this train sets. Now what is the variance in f?

To <u>formally measure variance in the estimator</u>, we fix the dev and test set. If we fix an estimator and vary the dev/test set, that doesn't tell us about the learning algorithm, it tells us more about the dev/test set - let's say my dev/test set has high variance then my output of a trained model would have high variance anyway.

$$Var_D[f] = \mathbb{E}_D[(\mathbb{E}_D[f(x)] - f(x))^2] \tag{2}$$

So with this in mind to ground the intuition:

If all of the estimators are the same, then $\mathbb{E}_D[f(x)] = f(x)$, and the variance of our estimator is exactly 0. That basically means its pretty dumb and not learning anything no matter how we vary the training set. Conversely, if we have an estimator with super high variance, we are learning a different f each time each time we get different training sets, and that's where 'overfitting' happens.

« Some Clustering Papers at ICLR20 (/review/2020/05/08/ClusteringICLR.html)

EMNLP 2020 » (/review/2020/11/21/EMNLP2020.html)

Quality means doing it right when no one is looking - Henry Ford

suzyahyah (https://github.com/suzyahyah)

suzyahyah (https://twitter.com/suzyahyah)

The best time to plant a tree was 20 years ago. The second best time is now. - Japanese proverb

Since October 2017