ESL- bias-variance decomposition notes

James Chuang

December 28, 2016

Bias-variance decomposition

From ESL pg. 23:

Suppose we have 1000 training examples x_i generated uniformly on $[-1,1]^p$. Assume that the true relationship between X and Y is

$$Y = f(X) = e^{-8\|X\|^2}$$

, without any measurement error. We use the 1-nearest neighbor rule to predict y_0 at the test-point $x_0 = 0$. Denote the training set by τ. We can then compute the expected prediction error at x_0 for our procedure, averaging over all such samples of size 1000. Since the problem is deterministic, this is the mean squared error (MSE) for estimating $f(0)$:

$$\begin{aligned}
\mathsf{MSE}(x_0) &= \mathsf{E}_\tau \left[f(x_0) - \hat{y}_0\right]^2 \\
&= \mathsf{E}_\tau \left[\hat{y}_0 - f(x_0)\right]^2 \\
&= \mathsf{E}_\tau \left[\hat{y}_0 - \mathsf{E}_\tau\left[\hat{y}_0\right] + \mathsf{E}_\tau\left[\hat{y}_0\right] - \hat{y}_0\right]^2 \\
&= \mathsf{E}_\tau \left[y_0 - \mathsf{E}_\tau[\hat{y}_0]\right]^2 + 2\mathsf{E}_\tau\left[\left(\hat{y}_0 - \mathsf{E}_\tau[\hat{y}_0]\right)\left(\mathsf{E}_\tau\left[\hat{y}_0\right] - f(x_0)\right)\right] + \mathsf{E}_\tau\left[\mathsf{E}_\tau\left[\hat{y}_0\right] - f(x_0)\right]^2 \\
&= \mathsf{Var}_\tau(\hat{y}_0) + 2\mathsf{E}_\tau\left[\left(\hat{y}_0 - f(x_0)\right)\left(f(x_0) - f(x_0)\right)\right] + \mathsf{Bias}^2_\tau(\hat{y}_0) \\
&= \mathsf{Var}_\tau(\hat{y}_0) + \mathsf{Bias}^2_\tau(\hat{y}_0)
\end{aligned}$$

Another way:

$$\begin{aligned}
\mathsf{MSE}(x_0) &= \mathsf{E}_\tau \left[f(x_0) - \hat{y}_0\right]^2 \\
&= \mathsf{E}_\tau \left[f^2(x_0) - 2f(x_0)\hat{y}_0 + \hat{y}_0^2\right] \\
&= \mathsf{E}_\tau \left[f^2(x_0)\right] - \mathsf{E}_\tau\left[2f(x_0)\hat{y}_0\right] + \mathsf{E}_\tau\left[\hat{y}_0\right] \\
&= \mathsf{Var}_\tau(f(x_0)) + \left(\mathsf{E}_\tau\left[f(x_0)\right]\right)^2 - 2f(x_0)\mathsf{E}_\tau\left[\hat{y}_0\right] + \mathsf{Var}_\tau(\hat{y}_0) + \left(\mathsf{E}_\tau\left[\hat{y}_0\right]\right)^2 \\
&= \mathsf{Var}_\tau\left(f(x_0)\right) + \mathsf{Var}_\tau(\hat{y}_0) + \left[\mathsf{E}_\tau(\hat{y}_0 - f(x_0))\right]^2 \\
&= \sigma^2 + \mathsf{Var}_\tau(\hat{y}_0) + \mathsf{Bias}^2(y_0)
\end{aligned}$$