Bias variance decomposition

Let $y = f(x) + \varepsilon$ where ε is iid noise with zero mean and variance σ^2.

We estimate f using $\hat{f}$. Then the expected MSE for a new y at x_0 will be equal to

$$E[(y - \hat{f}(x_0))^2] = [\text{Bias}(\hat{f}(x_0))]^2 + \text{Var}(\hat{f}(x_0)) + \sigma^2$$

where

$$\text{Bias}(\hat{f}(x_0)) = E[\hat{f}(x_0)] - f(x_0)$$

$$\text{and} \quad \text{Var}(\hat{f}(x_0)) = E\left[\left(\hat{f}(x_0) - E[\hat{f}(x_0)]\right)^2\right].$$

Proof

We will abbreviate $f = f(x_0)$ and $\hat{f} = \hat{f}(x_0)$.

Since f is deterministic, $E[f] = f$ and $\text{Var}[f] = 0$.

$$
\begin{aligned}
E[(y - \hat{f})^2] &= E[(y \underline{-f + f} - \hat{f})^2] \\
&= E[(y - f)^2 + (f - \hat{f})^2 + 2(y - f)(f - \hat{f})] \\
&= \sigma^2 + E[(\hat{f} - f)^2] + 2E[(y - f)(f - \hat{f})]
\end{aligned}
$$

Now

$$
\begin{aligned}
E[(\hat{f} - f)^2] &= E[(\hat{f} \underline{-E[\hat{f}] + E[\hat{f}]} - f)^2] \\
&= E[(\hat{f} - E[\hat{f}])^2] + (E[\hat{f}] - f)^2 + 2E[(\hat{f} - E[\hat{f}])(E[\hat{f}] - f)] \\
&= \text{Var}[\hat{f}] + \text{Bias}^2[\hat{f}] + 2E[(\hat{f} - E[\hat{f}])(E[\hat{f}] - f)]
\end{aligned}
$$

Both cross-product terms are equal to zero as can be shown by expansion:

$$
\begin{aligned}
E[(y - f)(f - \hat{f})] &= E[yf - f^2 - y\hat{f} + f\hat{f}] \\
&= f^2 - f^2 - E[y\hat{f}] + fE[\hat{f}] \\
&= -E[(f + \varepsilon)\hat{f}] + fE[\hat{f}] \\
&= -E[f\hat{f}] - E[\varepsilon\hat{f}] + fE[\hat{f}] \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
E[(\hat{f} - E[\hat{f}])(E[\hat{f}] - f)] &= E\left[\hat{f}E[\hat{f}] - E[\hat{f}]E[\hat{f}] - \hat{f}f + E[\hat{f}]f\right] \\
&= E[\hat{f}]E[\hat{f}] - E[\hat{f}]E[\hat{f}] - E[\hat{f}]f + E[\hat{f}]f \\
&= 0
\end{aligned}
$$