

We have:

$$\begin{aligned}
 E((Y - \hat{Y})^2) &= E\left(\underbrace{(Y - \hat{f}(x))}_{= f(x) + \varepsilon}\right)^2 \\
 &= E\left((f(x) + \varepsilon - \hat{f}(x))^2\right) \\
 &= E\left((f(x) - \hat{f}(x))^2 + \varepsilon^2 + 2\varepsilon(f(x) - \hat{f}(x))\right) \\
 &= E\left((f(x) - \hat{f}(x))^2\right) + E(\varepsilon^2) + E\left(2\varepsilon(f(x) - \hat{f}(x))\right)
 \end{aligned}$$

We have:

$$E(\varepsilon^2) = \text{Var}(\varepsilon) + \underbrace{(E(\varepsilon))^2}_{=0} = \text{Var}(\varepsilon)$$

Note that Y and $\hat{f}(x)$ are independent since $\hat{f}(x)$ depends on $Y_1, \dots, Y_N$, which are independent of Y , so ε and $\hat{f}(x)$ are independent. This means that

$$E\left(2 \cdot \varepsilon(f(x) - \hat{f}(x))\right) \stackrel{\text{ind.}}{=} 2 \underbrace{E(\varepsilon)}_{=0} \cdot E(f(x) - \hat{f}(x)) = 0$$

We obtain

$$E((Y - \hat{Y})^2) = E((Y - \hat{f}(x))^2) = \underbrace{E((f(x) - \hat{f}(x))^2)}_{\text{Reducible}} + \underbrace{\text{Var}(\varepsilon)}_{\text{Irreducible}}$$

We have :

$$\begin{aligned} E((f(x) - \hat{f}(x))^2) &= E\left((f(x) - E(\hat{f}(x))) + (E(\hat{f}(x)) - \hat{f}(x))^2\right) \\ &= E\left((f(x) - E(\hat{f}(x)))^2 + (E(\hat{f}(x)) - \hat{f}(x))^2\right. \\ &\quad \left.+ 2(f(x) - E(\hat{f}(x)))(E(\hat{f}(x)) - \hat{f}(x))\right) \\ &= E\left((f(x) - E(\hat{f}(x)))^2\right) + E\left((\hat{f}(x) - E(\hat{f}(x)))^2\right) \\ &\quad + 2(f(x) - E(\hat{f}(x))) \cdot E(\hat{f}(x) - E(\hat{f}(x))) = \text{Var}(\hat{f}(x)) \\ &= (f(x) - E(\hat{f}(x)))^2 + \text{Var}(\hat{f}(x)) \\ &\quad + 2(f(x) - E(\hat{f}(x))) \cdot \underbrace{(E(\hat{f}(x)) - E(\hat{f}(x)))}_{=0} \\ &= \underbrace{(f(x) - E(\hat{f}(x)))^2}_{\text{Bias}} + \text{Var}(\hat{f}(x)) \\ &= \text{Bias}^2 + \text{Variance} \end{aligned}$$