Topic4. Bias-Variance decomposition

Simply the note from http://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote12.html

$$E(x_0) = E\left[\left(f(x_0) - \hat{f}(x_0)\right)^2 | X = x_0\right]$$

$$= E\left[f(x_0)\text{-}\overline{\hat{f}(x_0)} + \overline{\hat{f}(x_0)} - \hat{f}(x_0)\right]^2$$

$$=E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right) + \left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)\right]^2$$

$$= E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)^2\right] + E\left[\left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)^2\right] + 2E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right) \times \left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)\right] \ldots (1)$$

$$E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right) \times \left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)\right]$$

$$= E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)\right] \times E\left[\left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)\right]$$

$$= E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)\right] \times \left(E\left[\overline{\hat{f}(x_0)}\right] - E\left[\hat{f}(x_0)\right]\right)$$

$$= E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)\right] \times \left(\overline{\hat{f}(x_0)} - \overline{\hat{f}(x_0)}\right)$$

$$= E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)\right] \times 0 = 0 \cdots (2)$$

$$E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)^2\right]$$

$$= E\left[\left(f(x_0)\text{-}\overline{f(x_0)} + \overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)^2\right]$$

$$= E\left[\left(\left(f(x_0)\text{-}\overline{f(x_0)}\right) + \left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)\right)^2\right]$$

$$= E\left[\left(f(x_0)\text{-}\overline{f(x_0)}\right)^2\right] + E\left[\left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)^2\right] + 2E[\left(f(x_0)\text{-}\overline{f(x_0)}\right) \times \left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)] \cdots (3)$$

$$E\left[\left(f(x_0)\text{-}\overline{f(x_0)}\right) \times \left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)\right]$$

$$= E\left[f(x_0)\text{-}\overline{f(x_0)}\right] \times E\left[\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right]$$

$$= \left(E[f(x_0)] - E\left[\overline{f(x_0)}\right]\right) \times E\left[\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right]$$

$$= \left(\overline{f(x_0)} - \overline{f(x_0)}\right) \times E\left[\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right]$$

$$= 0 \times E\left[\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right] = 0 \cdots (4)$$

Combining (3) & (4), we get

$$E\left[\left(f(x_0)\text{-}\overline{\hat{f}(x_0)}\right)^2\right] = E\left[\left(f(x_0)\text{-}\overline{f(x_0)}\right)^2\right] + E\left[\left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)^2\right] \cdots (5)$$

Combining (1), (2) & (5), we get

$$E(x_0) = E\left[\left(f(x_0) - \hat{f}(x_0)\right)^2 | X = x_0\right]$$

$$= E\left[\left(f(x_0)\text{-}\overline{f(x_0)}\right)^2\right] + E\left[\left(\overline{f(x_0)} - \overline{\hat{f}(x_0)}\right)^2\right] + E\left[\left(\overline{\hat{f}(x_0)} - \hat{f}(x_0)\right)^2\right]$$

$$= \text{Noise} + \text{Bias}^2 + \text{Variance}$$