STATS 7022 – Data Science PG Assignment 1

Dang Thinh Nguyen

2024-06-13

Question 2: Variance-Bias Trade-Off

We have:

- General form: $Y = f(\boldsymbol{X}) + \epsilon$
- Mean square error (MSE): $MSE = E[(y - \hat{f}(x))^2]$
- Bias: $b_f\left(\hat{f}(x)\right) = E\left[\hat{f}(x)\right] - f(x)$ (1)
- Variance: $\text{var}\left(\hat{f}(x)\right) = E[\left(\hat{f}(x) - E\left[\hat{f}(x)\right]\right)^2]$ (2)

From a single new observation (x_0, Y_0), MSE becomes:

$$E\left[\left(Y_0 - \hat{f}(x_0)\right)^2\right]$$

$$= E[(f(x_0) + \epsilon - \hat{f}(x_0))^2]$$

$$= E[(f(x_0) - \hat{f}(x_0) + \epsilon)^2]$$

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

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

$$= E\left[\left(f(x_0) - \hat{f}(x_0)\right)^2\right] + 2E[f(x_0)\epsilon] - 2E\left[\hat{f}(x_0)\epsilon\right] + E[\epsilon^2] \quad (3)$$

By definition, ϵ is independent of **X**, $E[\epsilon] = 0$, $E[\epsilon^2] = var(\epsilon)$, (3) becomes:

$$E\left[\left(Y_0 - \hat{f}(x_0)\right)^2\right]$$

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

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

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

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

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

$$-2E\left[\left(f(x_0) - E\left[\hat{f}(x_0)\right]\right)\right]E\left[\left(\hat{f}(x_0) - E\left[\hat{f}(x_0)\right]\right)\right] + \text{var}(\epsilon) \quad (4)$$

Substitute (1) and (2) into (4), we have:

$$E\left[\left(Y_0 - \hat{f}(x_0)\right)^2\right]$$

$$= E\left[b_f\left(\hat{f}(x_0)\right)^2\right] + \text{var}\left(\hat{f}(x_0)\right) + \text{var}(\epsilon)$$

$$-2E\big[\big(f(x_0) - E\big[\hat{f}(x_0)\big]\big)\big]E\big[\big(\hat{f}(x_0) - E\big[\hat{f}(x_0)\big]\big)\big] \quad (5)$$

We have: $E\big[\hat{f}(x_0)\big]$ and $f(x_0)$ are constant, $b_f\left(\hat{f}(x_0)\right)$ is also a constant. Thus:

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

The equation (5) becomes:

$$E\left[\left(Y_0 - \hat{f}(x_0)\right)^2\right]$$

$$= b_f\left(\hat{f}(x_0)\right)^2 + \text{var}\left(\hat{f}(x_0)\right) + \text{var}(\epsilon)$$

$$-2E\big[\big(f(x_0) - E\big[\hat{f}(x_0)\big]\big)\big]E\big[\big(\hat{f}(x_0) - E\big[\hat{f}(x_0)\big]\big)\big] \quad (6)$$

Because $E\big[\hat{f}(x_0)\big]$ is a constant: $E\left[E\big[\hat{f}(x_0)\big]\right] = E\big[\hat{f}(x_0)\big]$

So:

$$E\big[\big(\hat{f}(x_0) - E\big[\hat{f}(x_0)\big]\big)\big]$$

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

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

$$= 0 \quad (7)$$

Substitute (7) into (6), we have:

$$E\left[\left(Y_0 - \hat{f}(x_0)\right)^2\right] = b_f\left(\hat{f}(x_0)\right)^2 + \text{var}\left(\hat{f}(x_0)\right) + \text{var}(\epsilon)$$