Mean-Varaince Tradeoff Assignment

2021313075 백경인

November 20, 2024

Overleaf을 사용해서 증명을 유도했습니다.

Theorem 1

$$\textit{Expected Prediction Error} = \sigma^2 + [bias(f_{\hat{w}}(x_t))]^2 + var(f_{\hat{w}}(x_t)),$$

where:

1. σ^2 *represents the irreducible error due to noise in the data.*
2. $bias(f_{\hat{w}}(x_t))$ *measures how much the average prediction of the model deviates from the true function.*
3. $var(f_{\hat{w}}(x_t))$ *captures the variability in the model's predictions due to randomness in the training process.*

Proof 1 ***Proof of the Decomposition***

Let the true target value be y*, the true function be* $f(x)$*, and the model's prediction be* $f_{\hat{w}}(x)$*. Assume* $y = f(x) + \epsilon$*, where* $\epsilon \sim \mathcal{N}(0, \sigma^2)$ *represents the noise. Like what the professor said in class, the existence of noise is inevitable.*

Step 1. ***Expected Squared Prediction Error****:*

$$E[(y - f_{\hat{w}}(x))^2]$$

Substitute $y = f(x) + \epsilon$*:*

$$E[(f(x) + \epsilon - f_{\hat{w}}(x))^2].$$

Step 2. ***Expand the Squared Term****:*

$$E[(f(x) - f_{\hat{w}}(x))^2 + 2\epsilon(f(x) - f_{\hat{w}}(x)) + \epsilon^2].$$

Since ϵ ***has mean 0 and is independent with*** $f(x)$***-*** $f_{\hat{w}}(x)$***, the cross-term*** $E[2\epsilon(f(x) - f_{\hat{w}}(x))]$ ***vanishes because if two variable A and B are independent,*** $E(AB) = E(A)E(B)$***.*** *The equation simplifies to:*

$$E[(f(x) - f_{\hat{w}}(x))^2] + E[\epsilon^2].$$

Since $E[\epsilon^2] = \sigma^2$*, we get:*

$$E[(y - f_{\hat{w}}(x))^2] = \sigma^2 + E[(f(x) - f_{\hat{w}}(x))^2].$$

Step 3. ***Decomposition*** $E[(f(x) - f_{\hat{w}}(x))^2]$*:*
Add and subtract the expected prediction $E[f_{\hat{w}}(x)]$*:*

$$E[(f(x) - f_{\hat{w}}(x))^2] = E[(f(x) - E[f_{\hat{w}}(x)] + E[f_{\hat{w}}(x)] - f_{\hat{w}}(x))^2].$$

Let

$$A = (f(x) - E[f_{\hat{w}}(x))$$

$$B = E[f_{\hat{w}}(x)] - f_{\hat{w}}(x)$$

The cross-term vanishes because A is a constant that does not depend on the choice of the training dataset.

$$E[(c + B)^2] = c^2 + E(B^2)$$

This leaves:

$$E[(f(x) - f_{\hat{w}}(x))^2] = (f(x) - E[f_{\hat{w}}(x)])^2 + E[(E[f_{\hat{w}}(x)] - f_{\hat{w}}(x))^2].$$

Step 4. ***Interpretation****:*

- *The first term* $(f(x) - E[f_{\hat{w}}(x)])^2$ *is the squared bias of the model.*
- *The second term* $E[(E[f_{\hat{w}}(x)] - f_{\hat{w}}(x))^2]$ *is the variance of the model's predictions.*

Thus, the total error is:

$$E[(y - f_{\hat{w}}(x))^2] = \sigma^2 + [bias(f_{\hat{w}}(x))]^2 + var(f_{\hat{w}}(x)).$$