# EE 2211 Lecture 8

## Review of Bias-Variance Tradeoff

Test Error = Bias$^2$ + Variance + Variance of Irreducible Error

$\shortparallel$

MSE: mean squared error

Suppose $y = f(x) + \varepsilon$ ← noise  $f$: deterministic

E.g., $f(x) = w_0 + w_1 x + w_2 x^2$

$\varepsilon$: random. mean 0 & var $\sigma^2$

Sample 3 training datasets $D_1, D_2, D_3$ from $p(D)$. $= \{(\underline{x}_i, y_i) : i = 1, \ldots, m\}$

Learn 3 regression functions $\hat{f}_{D_1}, \hat{f}_{D_2}, \hat{f}_{D_3}$

$$\hat{y}_1 = \hat{f}_{D_1}(\underline{x}) = [\underbrace{(X^T X)^{-1} X^T y}_{\underline{w}}]^T \underline{x}$$

$$\hat{y}_2 = \hat{f}_{D_2}(\underline{x}) = [\underbrace{(X^T X)^{-1} X^T y}_{\underline{w}}]^T \underline{x}$$

Given a new data point $(\underline{x}, y)$, $y = f(\underline{x}) + \varepsilon$

Average prediction: $f_{avg}(\underline{x}) = \frac{1}{3}(\hat{f}_{D_1}(\underline{x}) + \hat{f}_{D_2}(\underline{x}) + \hat{f}_{D_3}(\underline{x}))$

$$\text{Bias} = f_{avg}(x) - f(x)$$

[Note that this is not y].

$$\text{Bias}^2 = (f_{avg}(\underline{x}) - f(\underline{x}))^2$$

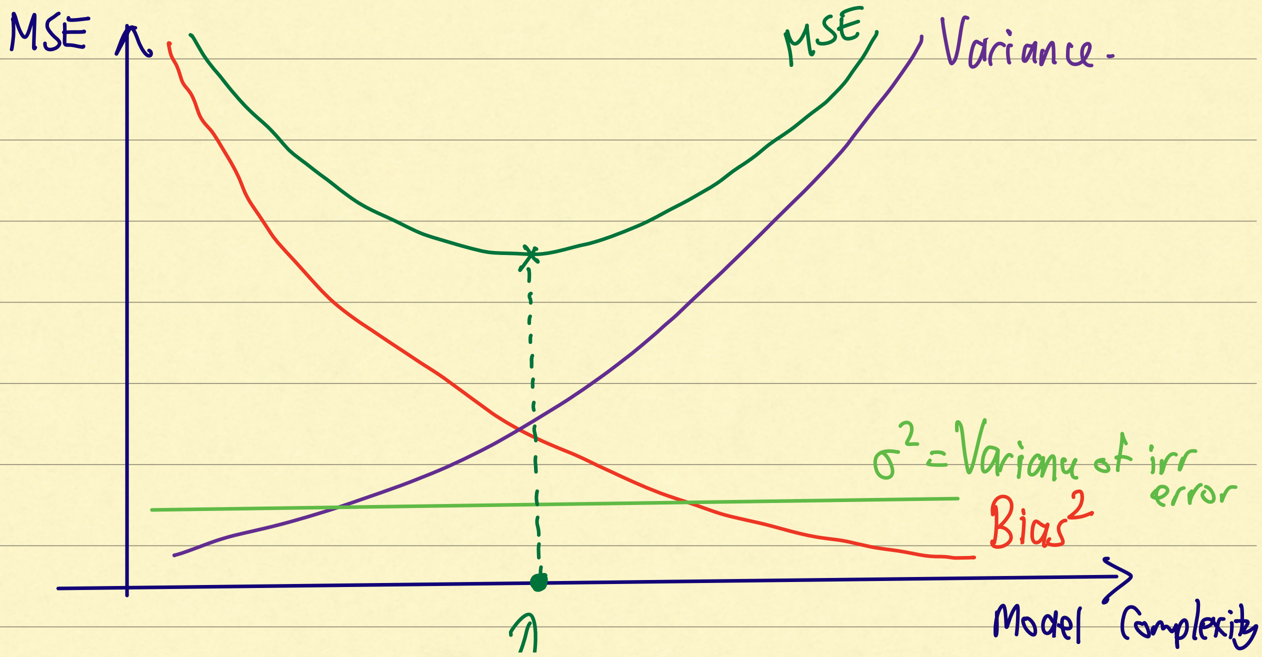
$$\text{Variance} = E[(\hat{f}_D(x) - f_{avg}(x))^2]$$
$$= \frac{1}{3}[(\hat{f}_{D_1}(x) - f_{avg}(x))^2 + (\hat{f}_{D_2}(x) - f_{avg}(x))^2 + (\hat{f}_{D_3}(x) - f_{avg}(x))^2]$$

$$\text{MSE} = E[(\hat{f}_D(x) - y)^2]$$
$$= \frac{1}{3}[(\hat{f}_{D_1}(x) - y)^2 + (\hat{f}_{D_2}(x) - y)^2 + (\hat{f}_{D_3}(x) - y)^2]$$

[Note that this is subtracting y, and not f(x)].

Var of $\varepsilon$, irr noise.

Claim: $\text{MSE} = \text{Bias}^2 + \text{Variance} + \sigma^2$.



Sweet spot

Polynomial order ↑

# of features ↑