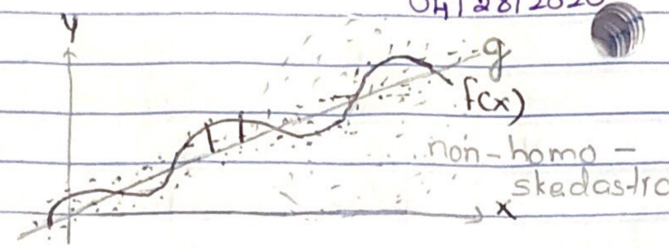
Lecture 22

04/28/2020

$y \subseteq \mathbb{R}$ (regression)

$y = f(\vec{x}) + \delta$



Assume

① δ is a realization from the r.v. Δ which is mean independent of $\vec{x}$.

$E[\Delta \mid \vec{x}] = E[\Delta] = 0 \implies E[y \mid \vec{x}] = E[f(\vec{x}) + \Delta \mid \vec{x}] = f(\vec{x}) + E[\Delta \mid \vec{x}]$
$= f(\vec{x})$

Math 368/621

Conditional expectation function (CEF)

② The second moment of Δ^2 is also independent of $\vec{x}$ and is σ^2. (homo-skedasticity).

$Var[\Delta \mid \vec{x}] = E[\Delta^2 \mid \vec{x}] - E[\cancel{\Delta}^{0} \mid \vec{x}] = E[\Delta^2 \mid \vec{x}] = \sigma^2$ AKA the irreducible error.

$y = g + e \implies e = y - g$ $\quad \implies E = Y - g$ (r.v. for residuals (how much your prediction is wrong))

$y = g + \underbrace{(f-g)}_{\text{misspecification + estimation error}} + \delta \implies e = f - g - \delta \implies E = f - g + \Delta$

$E[E \mid \vec{x}] = E[f - g + \Delta \mid \vec{x}] = f(\vec{x}) - g(\vec{x})$

mean square error of an estimator NOT $SSE/_{n-p-1}$!

$MSE(\vec{x}_*) = E_{\Delta_*}[E^2 \mid \vec{x}_*] = E[(Y_* - g(\vec{x}_*))^2 \mid \vec{x}_*]$

$\vec{x}_*$ is a new unit that we predict on $e = y - \hat{y}$. On avg. What is e^2?

$= E[Y_*^2 - 2g(\vec{x}_*)Y_* + g(\vec{x}_*)^2 \mid \vec{x}_*] = E[Y_*^2 \mid \vec{x}_*] - 2g(\vec{x}_*)E[Y_* \mid \vec{x}] + g(\vec{x}_*)^2$

g is fixed. $g = A(\mathbb{D}) \Rightarrow \mathbb{D}$ fixed
$\langle \vec{x}_1, y_1 \rangle$
$\vdots$
$\langle \vec{x}_n, y_n \rangle$; constants

$$\text{(I)} \triangleq E\left[(f(\vec{x}_*)+\Delta_*)^2 \mid \vec{x}_*\right] - 2g(\vec{x}_*)\,E\left[f(\vec{x}_*)+\Delta \mid \vec{x}_*\right] + g(\vec{x}_*)^2.$$

$$= f(\vec{x}_*)^2 + 2f(\vec{x}_*)\,\underbrace{\cancel{E[\Delta_* \mid \vec{x}_*]}}_{\text{(I)}\to 0} + \underbrace{E[\Delta_*^2 \mid \vec{x}_*]}_{\text{(II)}} - 2g(\vec{x}_*)\left(f(\vec{x}_*) + \underbrace{\cancel{E[\Delta \mid \vec{x}_*]}}_{\text{I}\ 0}\right) + g(\vec{x}_*)^2.$$

$$= \sigma^2 + f(\vec{x}_*)^2 - 2g(\vec{x}_*)f(\vec{x}_*) + g(\vec{x}_*)^2.$$

$$= \sigma^2 + \left(f(\vec{x}_*) - g(\vec{x}_*)\right)^2 \geq \sigma^2.$$

$$\boxed{\mathbb{D}} = \left(\underset{\text{fixed}}{\overset{0,1,\ldots,9}{\boxed{\;X\;}}}\ \boxed{\vec{y}} \right)$$ realization form, $\vec{Y}$ r.v. with X the same.

$y_i = f(\vec{x}_i) + \delta_i \Rightarrow Y_i = f(\vec{x}_i) + \Delta_i \;\forall i$

Assume Δ_i's are independent, but (I), (II) hold; $\mathbb{D}_1, \mathbb{D}_2, \ldots$ etc which are different due to δ's being different.

$\Rightarrow g_1 = A(\mathbb{D}_1)$, $g_2 = A(\mathbb{D}_2)$, etc. are different models drain from the r.v. G.
r.v.'s are $\underbrace{\Delta_1, \Delta_2, \ldots, \Delta_n, \Delta_*}$
These r.v.'s create dataset-dataset variation

$$MSE(\vec{x}_*) = E_{\Delta_1,\ldots,\Delta_*}\left[(Y_* - G(\vec{x}_*))^2 \mid \vec{x}_*\right]$$ omiting conditional on $\vec{x}_*$ to save time

$$= E_{\Delta_*}[Y_*^2] - 2E_{\Delta_*}[Y_*]\,E_{\Delta_1,\ldots,\Delta_n}[G(\vec{x}_*)] + E_{\Delta_1,\ldots,\Delta_n}[G(\vec{x}_*)^2]$$

$$= \sigma^2 + f(\vec{x}_*)^2 - 2f(\vec{x}_*)E[G(\vec{x}_*)] + E[G(\vec{x}_*)]^2 + Var[G(\vec{x}_*)]$$

$$= \sigma^2 + \big(\underbrace{E[G(\vec{x}_*)] - f(\vec{x}_*)}_{Bias[G(\vec{x}_*)]}\big)^2 + Var[G(\vec{x}_*)]$$

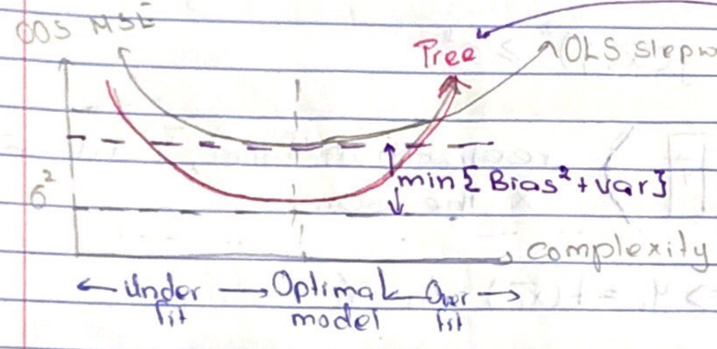
$$= \sigma^2 + Bias[G(\vec{x}_*)]^2 + Var[G(\vec{x}_*)]$$

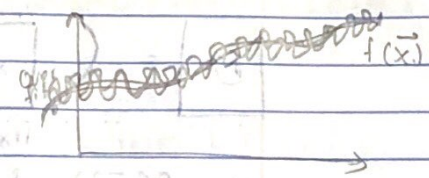
$$= MSE(\vec{x}_*)$$

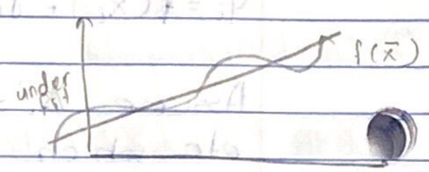
One more set of universes ... $\vec{x}_1$ is a realization from X. So are $\vec{x}_1, \ldots, \vec{x}_n$

$$MSE = E_x[MSE(\vec{x}_1)] = \sigma^2 + E_x\left[\underline{Bias[G(\vec{x}_1)]^2}\right] + E_x\left[\underline{Var[G(\vec{x}_1)]}\right]$$

called - "Bias-Variance Decomposition" or "Bias-Variance Tradeoff"

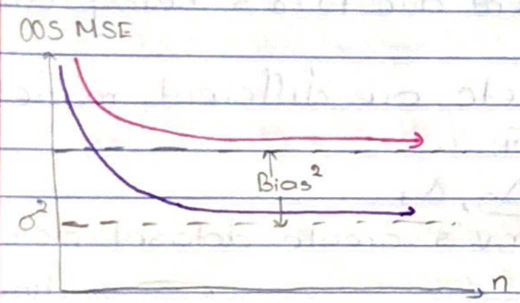






Bias = large — Variance = small

Bias = small — Variance = large (big)



Condition on one model i.e. some amount of complexity i.e. # of parameters.

$$\lim_{n \to \infty} Var[G(\vec{x})] = \lim E\left[(G(\vec{x}) - E[G(\vec{x})])^2\right] = 0$$

$n \to \infty$ ↓ no estimation error; $G(\vec{x})$ ↓ $g(\vec{x})$; $E[G(\vec{x})]$ ↓ $g(\vec{x})$