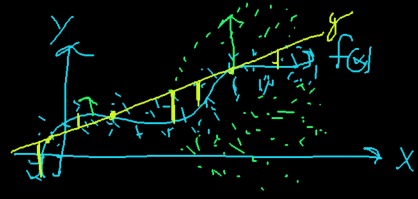
$y \in \mathbb{R}$ (regression)

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



non-homoskedastic.

$Y = f(\vec{x}) + \Delta$

Assume

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

$$E[\Delta \mid \vec{x}] \underset{\text{318/621}}{=} E[\Delta] = 0 \Rightarrow E[Y \mid \vec{x}] = E[f(\vec{x}) + \Delta \mid \vec{x}] = f(\vec{x}) + E[\Delta \mid \vec{x}] = f(\vec{x})$$

conditional expectation function (CEF).

(II) The second moment of Δ^2 is also independent of $\vec{x}$ and is σ^2 (homoskedasticity)

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

$y = g + e \Rightarrow e = y - g$ $\Rightarrow E = Y - g$ (2) r.v. for residual (how much your prediction is wrong)

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

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

mean squared error of an estimator NOT SSE/n-p-1!

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

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

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

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

(2)

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

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

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

$\mathbb{D} = \langle \underbrace{\boxed{X}}_{\text{fixed}}^{0\,1 \ldots p}, \boxed{\vec{Y}} \rangle$ realization from $\vec{Y}$ 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.

$\Rightarrow g_1 = \mathcal{A}(\mathbb{D}_1), g_2 = \mathcal{A}(\mathbb{D}_2)$, etc. are different models drawn from the r.v. G.

r.v.'s are $\Delta_1, \Delta_2, \ldots, \Delta_n, \Delta_*$. these r.v.'s create dataset-dataset variation

omitting conditional on $\vec{x}_*$ to save time

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

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

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

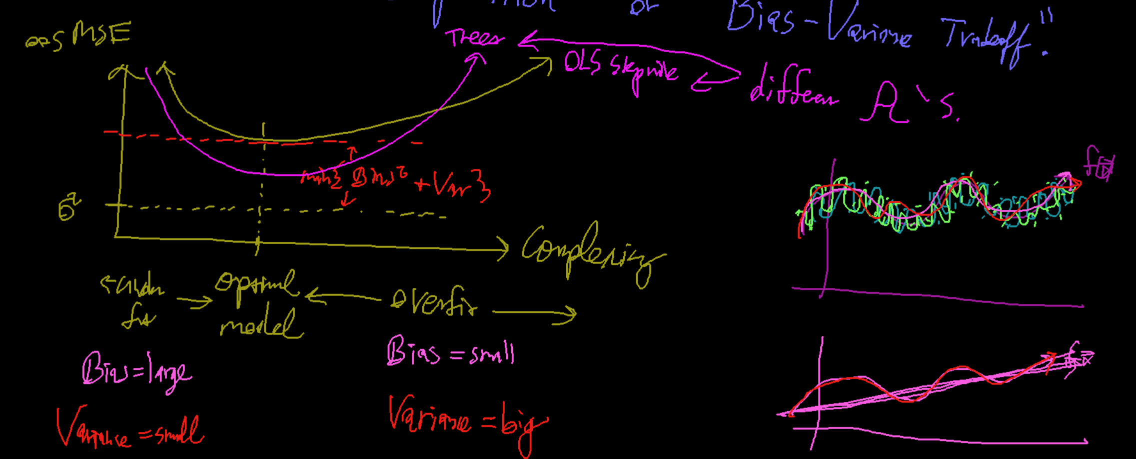
$$= \sigma^2 + \underbrace{(E[G(\vec{x}_*)] - f(\vec{x}_*))^2}_{Bias[G(\vec{x}_*)]} + Var[G(\vec{x}_*)] = \sigma^2 + Bias[G(\vec{x}_*)]^2 + Var[G(\vec{x}_*)]$$

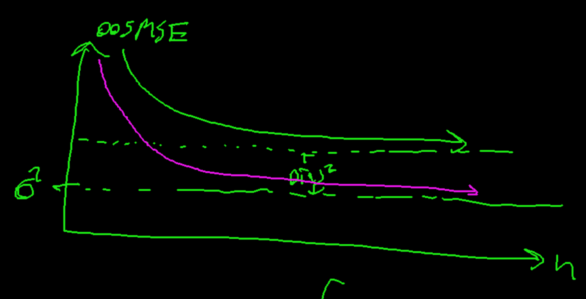
$= MSE(\vec{x}_*)$

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

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

"Bias-Variance Decomposition" or "Bias-Variance Tradeoff".





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

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

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