Lecture 22:

Random Trees:

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

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

Assume

(I) δ is realization from the r.v Δ which is mean-independent of $\vec{X}$ $\Rightarrow Y = f(\vec{X}) + \Delta$

Conditional expectation func (CEF)

$$E[\Delta|\vec{X}] = E[\Delta] = 0 \Rightarrow E[Y|\vec{X}] =$$

368/621.

$$= E[f(x) + \Delta|\vec{X}]$$

$$= f(\vec{X}) + E[\Delta|\vec{X}]$$

$$= f(\vec{X}).$$

(II) The second moment of Δ^2 is also independent of $\vec{X}$ and it's σ^2

(homoskedasticity)

$$Var[\Delta|\vec{X}] = E[\Delta^2|\vec{X}] - \cancelto{0}{E[\Delta|\vec{X}]^2}$$

$$= E[\Delta^2|\vec{X}] = \sigma^2 .$$ AKA irreducable error

$Y = g + e$

$\Rightarrow e = Y - g$

$y = g + \underbrace{(f-g)}_{\substack{\text{misspecification} \\ + \\ \text{estimation} \\ \text{error.}}} + \delta$

r.v for residual. (how much your prediction is wrong.)

$\Rightarrow E = Y - g$

$\Rightarrow e = f - g + \delta \Rightarrow E = f - g + \Delta.$

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

Mean Square error as an estimator NOT SSE / n-p-1

g is fixed. g = A(D), D fixed $\Rightarrow$ $\langle x_1, y_1 \rangle \ldots \langle x_n, y_n \rangle$ constant.

$$MSE(\vec{X}_*) = E_{\Delta_*}[E^2 \mid \vec{X}_*] = E[(y - g(x_*))^2 \mid X_*].$$

$\vec{X}_*$ is a new unit that we predict on, $e = Y - \hat{Y}$.

On avg which is e^2?

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

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

(1)

$\stackrel{1}{=} E\left[\left(f(\vec{X}_*) + \Delta_*\right)^2 \,\middle|\, \vec{X}_*\right] - 2g(\vec{X}_*)\, E\left[f(\vec{X}_*) + \Delta \mid \vec{X}_*\right] + g(\vec{X}_*)^2$

$= f(\vec{X}_*)^2 + 2f(\vec{X}_*)\, E[\Delta_* \mid \vec{X}_*] + \overbrace{E[\Delta_*^2 \mid \vec{X}_*]}^{\sigma^2 \ \text{(II)}}$

$\quad - 2g(\vec{X})\left(f(\vec{X}_*) + \cancel{E[\Delta \mid \vec{X}_*]}^{\text{(I)} \to 0}\right) + g(\vec{X}_*)^2$

$= \sigma^2 + f(\vec{X}_*)^2 - 2g(\vec{X}_*)\, F(Z_*)$

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

$\boxed{\mathbb{D}} = \left\langle \overset{X}{[\;\;]}, \overset{Y}{[\;\;]} \right\rangle$ realization from $\vec{Y}$ rv which X the same.

$y_i = f(\vec{X}_i) + \delta_i \Rightarrow Y_i = f(X_i) + \Delta Y_i$

Assume Δ_i are independent but (I), (II) hold

$\mathbb{D}_1, \mathbb{D}_2, \ldots$, etc are 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 drawn from the r.v G.

r.v's are $\underbrace{\Delta_1, \Delta_2, \ldots, \Delta_n, \Delta_*}_{\text{these rv's create dataset - dataset variation}}$

omiting conditional on x to save time

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

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

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

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

how far from avg model

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

$$= MSE(\vec{X}_*).$$

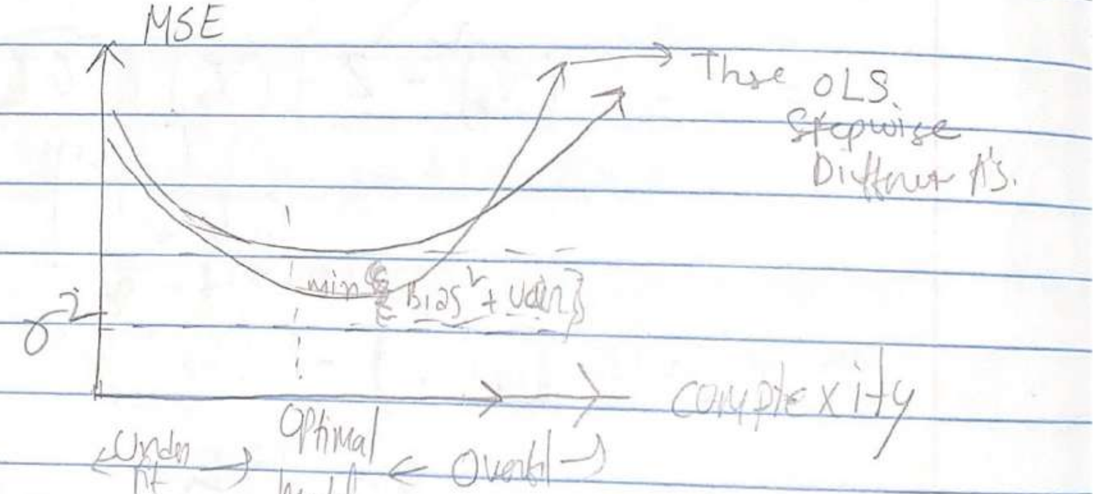
One more set of universe. $\vec{X}_*$ is a realization from X. So are $\vec{X}_1, \ldots \vec{X}_n$;

$$MSE = E_X\left[MSE(\vec{X}_*)\right]$$

$$= \sigma^2 + E_X\left[\text{Bias}\left[G(\vec{X}_*)\right]^2\right] + E_X\left[\text{Var}\left[G(\vec{X}_*)\right]\right]$$

"Bias-Variance Decomposition" or

"Bias-Variance Tradeoff."

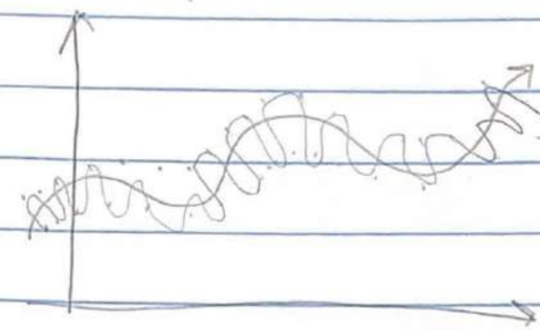


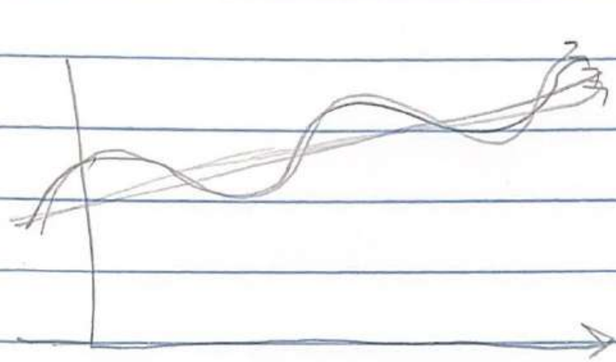
Bias = large

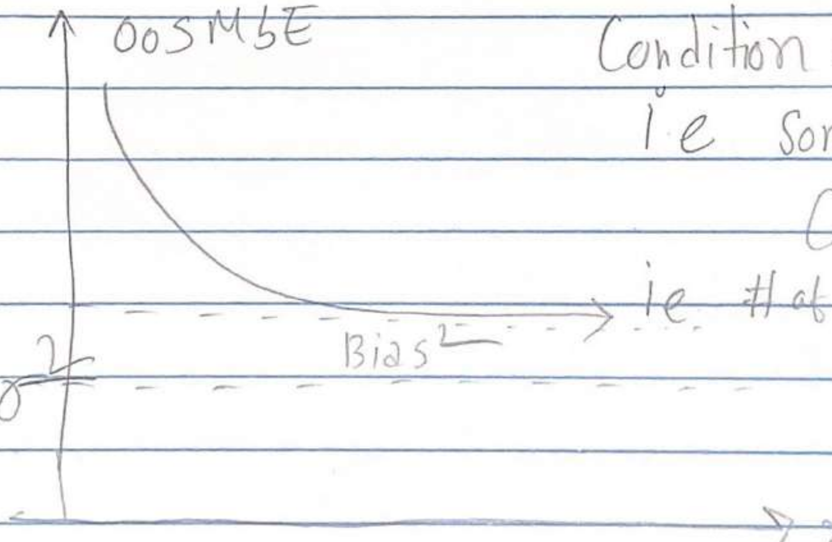
Variance = small

Bias = small

Variance = big







Condition on one model
i.e. Some amount of
Complexity
i.e. # of parameters

$$\lim_{\substack{n \to \infty \\ \downarrow \\ \text{No estimation error}}} Var\left[G(\vec{X})\right] = \lim E\left[\left(\underbrace{G(\vec{X})}_{g(\vec{X})} - \underbrace{E[G(\vec{X})]}_{g(\vec{X})}\right)^2\right]$$

$$= 0.$$