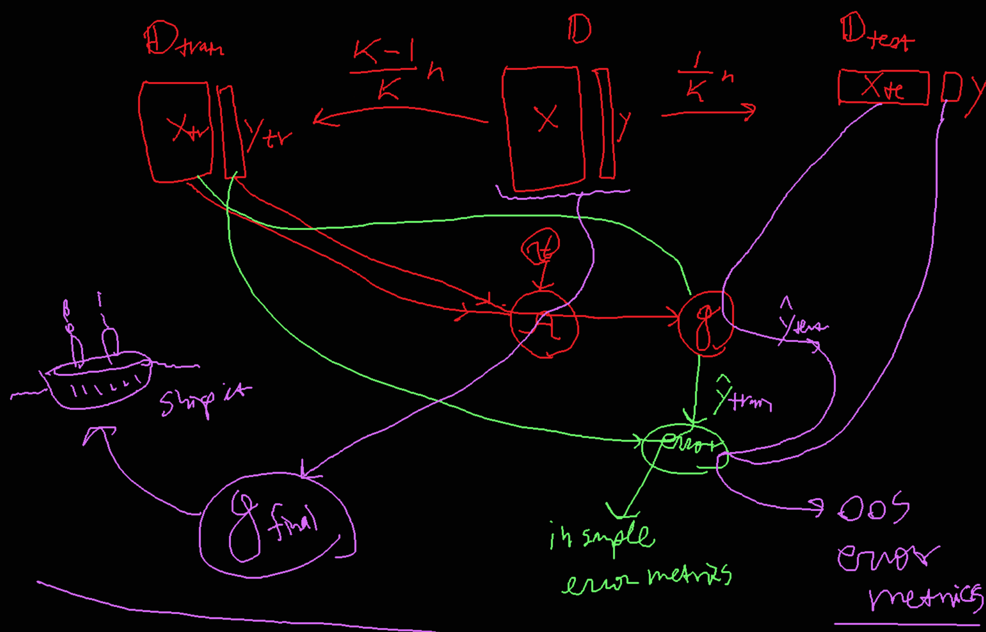


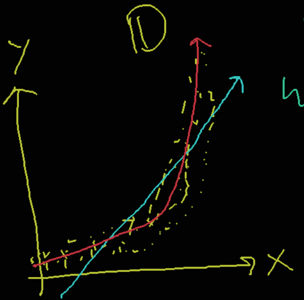
$g \approx g_{final}$?

Yes! But g_{final} is better (in expectation) because $n \uparrow$, generalization error is smaller than g's.

$$y = g + (h^* - g) + (f - h^*) + (t - f)$$

misspecification error which is limited by $\mathcal{H}$. (nature)

nothing we can do



$\mathcal{H}_1 = \{w_0 + w_1 x : \vec{w} \in \mathbb{R}^2\}$ linear models

$\mathcal{H}_2 = \{w_0 + w_1 x + w_2 x^2 : \vec{w} \in \mathbb{R}^3\}$ linear non-linear model

$\mathcal{A}$ = OLS polynomial regression

Why are polynomials a good idea to use as bases for arbitrary function.

Weierstrass Approximation Thm. For any cont. function f, $\exists$ polynomial function p s.t. $\forall \varepsilon > 0\ \forall \vec{x} \in \mathcal{X}$ $|f(\vec{x}) - p(\vec{x})| < \varepsilon$.

$$X_{orig} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ \vdots & \vdots \\ 1 & x_{1n} \end{bmatrix} \rightarrow X = \begin{bmatrix} 1 & x_{11} & x_{11}^2 \\ 1 & x_{12} & x_{12}^2 \\ 1 & x_{13} & x_{13}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{1n}^2 \end{bmatrix}$$

Is this full rank?

$\mathcal{A}$: $(X^T X)^{-1} X^T \vec{y} = \vec{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$

$p_{raw} + 1 = 2$

$p_{raw} = 1$

original measurements

$p + 1 = 3$

$p = 2$

raw + derived

$\mathcal{H}_3 = \{w_0 + w_1 x + w_2 x^2 + w_3 x^3 : \vec{w} \in \mathbb{R}^4\}$ Sure,...

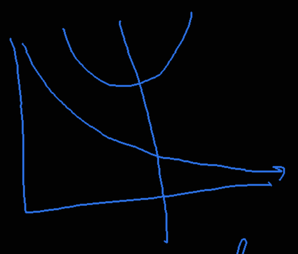
$p + 1 = n \Rightarrow R^2 = 100\%$, $s_e = 0$ perfect fit and overfit.

$X = [\vec{1} \mid \vec{x} \mid \vec{x}^2 \mid \ldots \mid \vec{x}^{n-1}]$ full rank? Thm. Vandermonde Matrix

$n \times n$ matrix

$$\det[X] = \prod_{i=1}^{n} \prod_{j=i+1}^{n} (x_j - x_i) \neq 0 \text{ if}$$

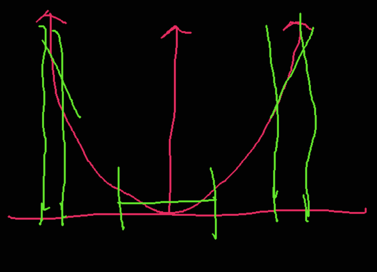
all x's are unique



complexity

optimal

$\text{Cov}[X, X^2]$



QR

$X \rightarrow Q$

$\downarrow \quad \downarrow$ orthog. proj.

$H = H$

$\hat{y} = \hat{y} \Rightarrow R^2 = R^2$

$$y = b_0 + b_1 x + b_2 x^2 = b_0 + (b_1 + b_2 x) x$$