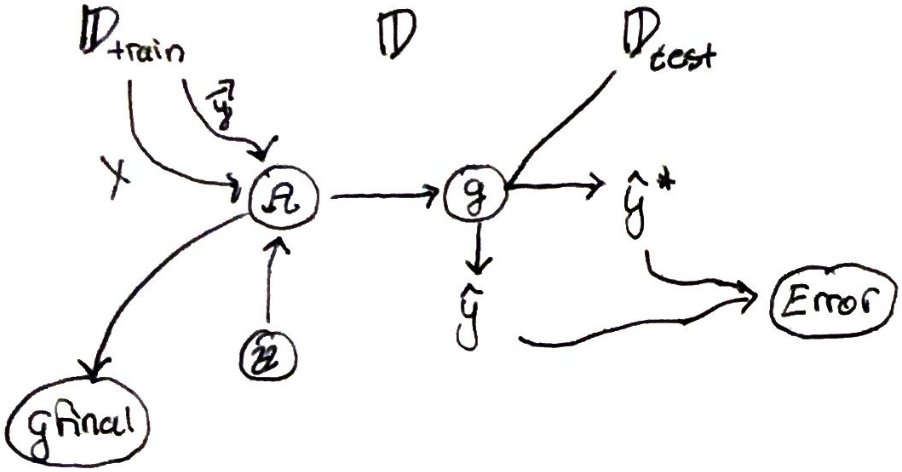
Math 342W Lecture 13



oos error metrics are honest estimates of future performance of g and the future performance on gfinal is expected to be better than g

$n = n_{train} + n_{test}$

$k = \frac{n}{n_{test}} \Rightarrow \frac{1}{k} =$ proportion test data to full data set

If k is large... $\Rightarrow$ $\underbrace{k \uparrow, \|\vec{\beta} - \vec{b}\|^2 \downarrow}$ (estimation error)

Typical $k = 10$ or 5

$\hookrightarrow |\bar{E}^*| \sim N\left(\mu_{E^*}, \frac{\sigma^2_{E^*}}{n^*}\right)$

$y = g(\vec{x}) + \underbrace{(h^*(\vec{x}) - g(\vec{x}))}_{\text{Estimation Error}} + \underbrace{(f(\vec{x}) - h^*(\vec{x}))}_{\text{Misspecification Error}} + \underbrace{(t(\vec{z}) - f(\vec{x}))}_{\text{Ignorance error}}$

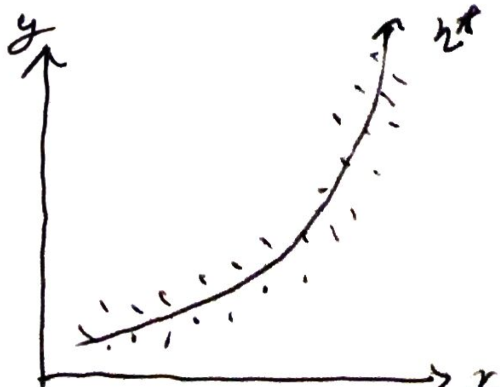
$k \uparrow$, $\mu_{E^*} \downarrow$

$k \uparrow$, Variance of oos R^2 & RMSE $\uparrow$

Polynomial Regression

$\mathcal{H} = \{w_0 + w_1 x_1 + w_2 x_2^2\}$

$w_0, \dots, w_2 \in \mathbb{R}$



$$X_{row} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{22} \\ 1 & x_{31} \\ 1 & x_{41} \\ 1 & x_{51} \end{bmatrix} \xrightarrow{\text{Expand}} \begin{bmatrix} 1 & x_{11} & x_{21}^2 \\ 1 & x_{21} & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_{51} & x_{51}^2 \end{bmatrix} \Rightarrow \mathcal{A} : OLS$$

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

Using OLS to fit curves to data is a "non-linear linear" model

Polynomial Regression is "principle" (theoretically just)

<u>Weierstrauss Aproximation THM (1885)</u>

For any continuous function f whose domain is $\mathcal{X} = [a,b]$
$\exists\, p$, a polynomial function st. $\forall \varepsilon > 0$, $\forall x \in \mathcal{X}$ $|f(x) - p(x)| < \varepsilon$

Stone (1937) → Generalized to any # dimensions