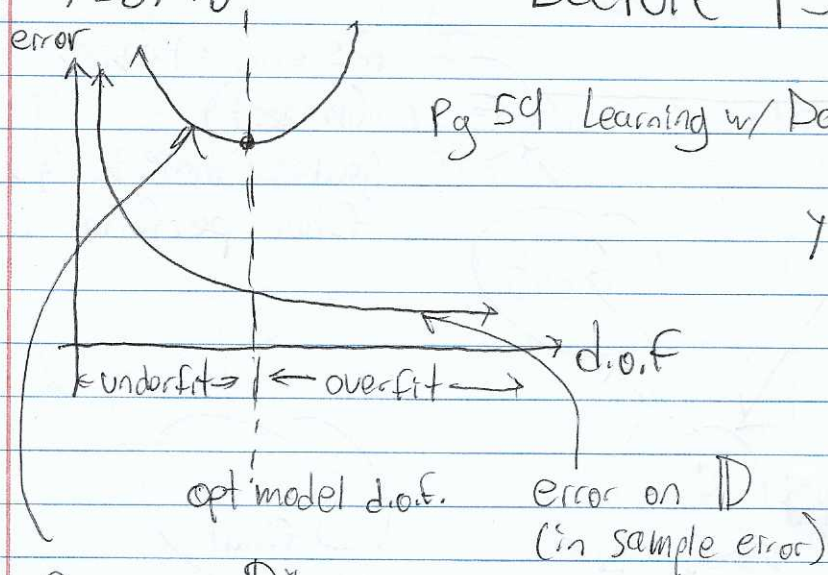


3/28/18

# Lecture 15

Pg 54 Learning w/ Data



$$y = f(w) + \delta$$

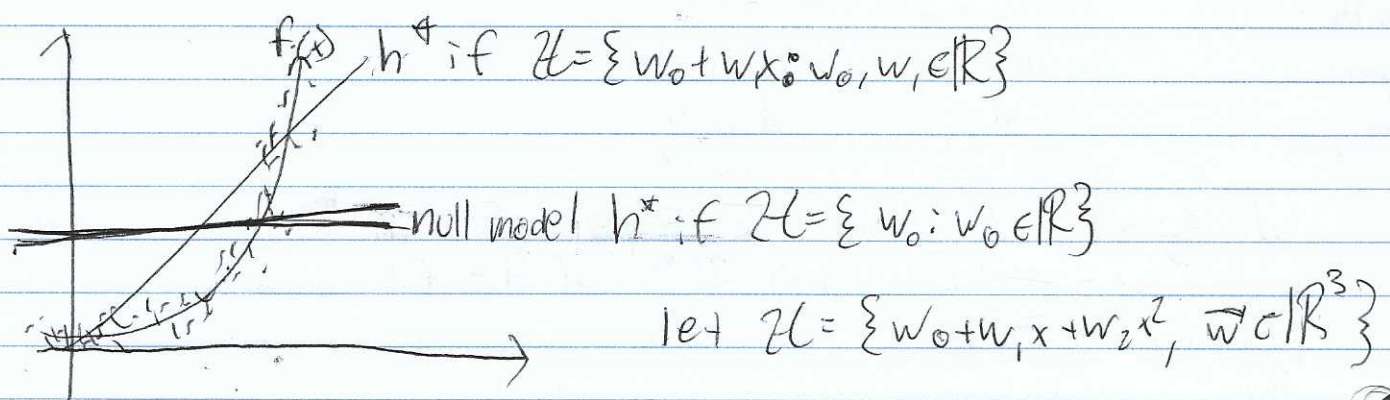
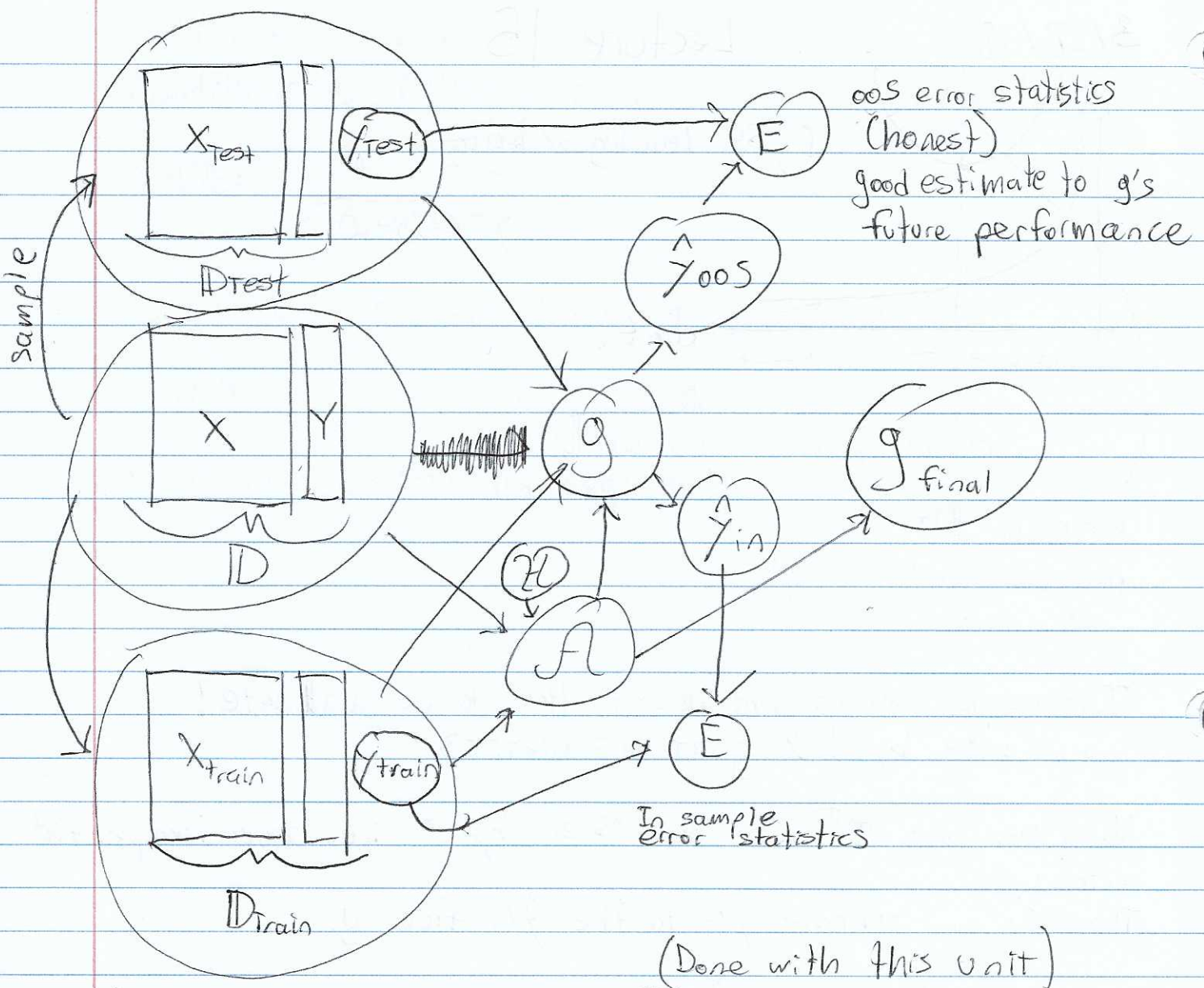
error on  $D^*$   
new data, out-of-sample  
OOS error

In-sample metrics not honest. How do we validate?  
We validate by looking at oos metrics

Need new data  $\notin D$  i.e. never "seen" by  $f$  i.e. never incorporated into  $g$ .  
Then we can compare  $y$ 's to the  $\hat{y}$ 's from  $g$ .

Pretend  $D = \underbrace{D_{\text{train}}}_{\text{use this to build } g} \cup \underbrace{D_{\text{test}}}_{\text{use this to validate } g}$

We select a  $k$  where  $\frac{1}{k} := \text{prop. of } D \text{ in } D_{\text{test}}$   
most popular  $\begin{cases} k=5 \Rightarrow D_{\text{test}} \text{ has } 20\% \text{ of the full dataset} \\ k=10 \Rightarrow 10\% \end{cases}$   
then... you randomly sample rows of  $D$



I "see" misspecification error.  
I need to make  $\mathcal{H}$  "rich" i.e. more d.o.f.



## Weierstrass Approx Thm

For every continuous function  $f$  where its domain is  $x \in [a, b]$ ,  
[a polynomial function  $p$  s.t.  $\forall \epsilon > 0, \forall x \in [a, b] |f(x) - p(x)| < \epsilon$ ]

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

$$p=1$$

$$p+1$$

$$p+1+1 \\ = 2p+1$$

$$b = (X^T X)^{-1} X^T \vec{y}$$

$$\text{let } n=5 \quad d=4$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 \end{bmatrix}$$

Vandermonde matrix