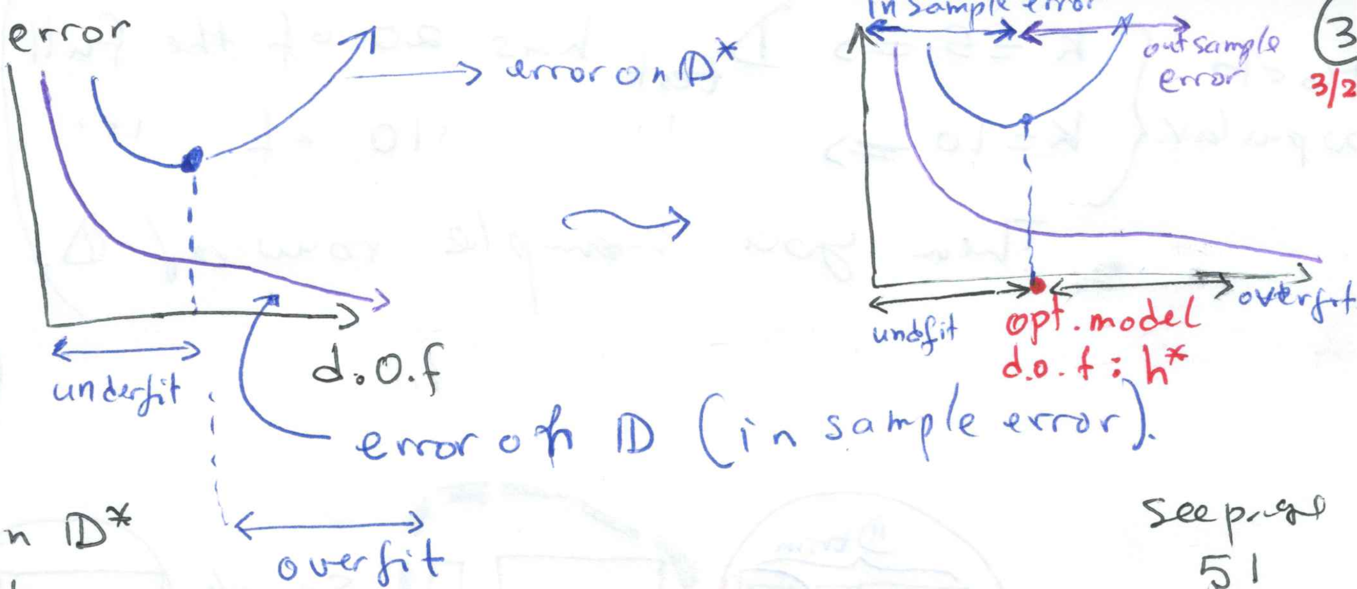
Lect 15

degree of freedom

d.o.f



See page 51 Learning Data.

error on $\mathbb{D}^*$
new data
out of sample on error

$$Y = f(x) + \int$$

To have a new data, get new $\mathbb{D}^*$ different from $\mathbb{D}$
get anew historical data and run new model from the

➤ In-Sample
How do we validate?
Look at.
Need new data $\notin \mathbb{D}$, ie, never seen by $\mathcal{A}_i$
never incorporate in g. and compare y's to $\hat{y}$'s
from g.

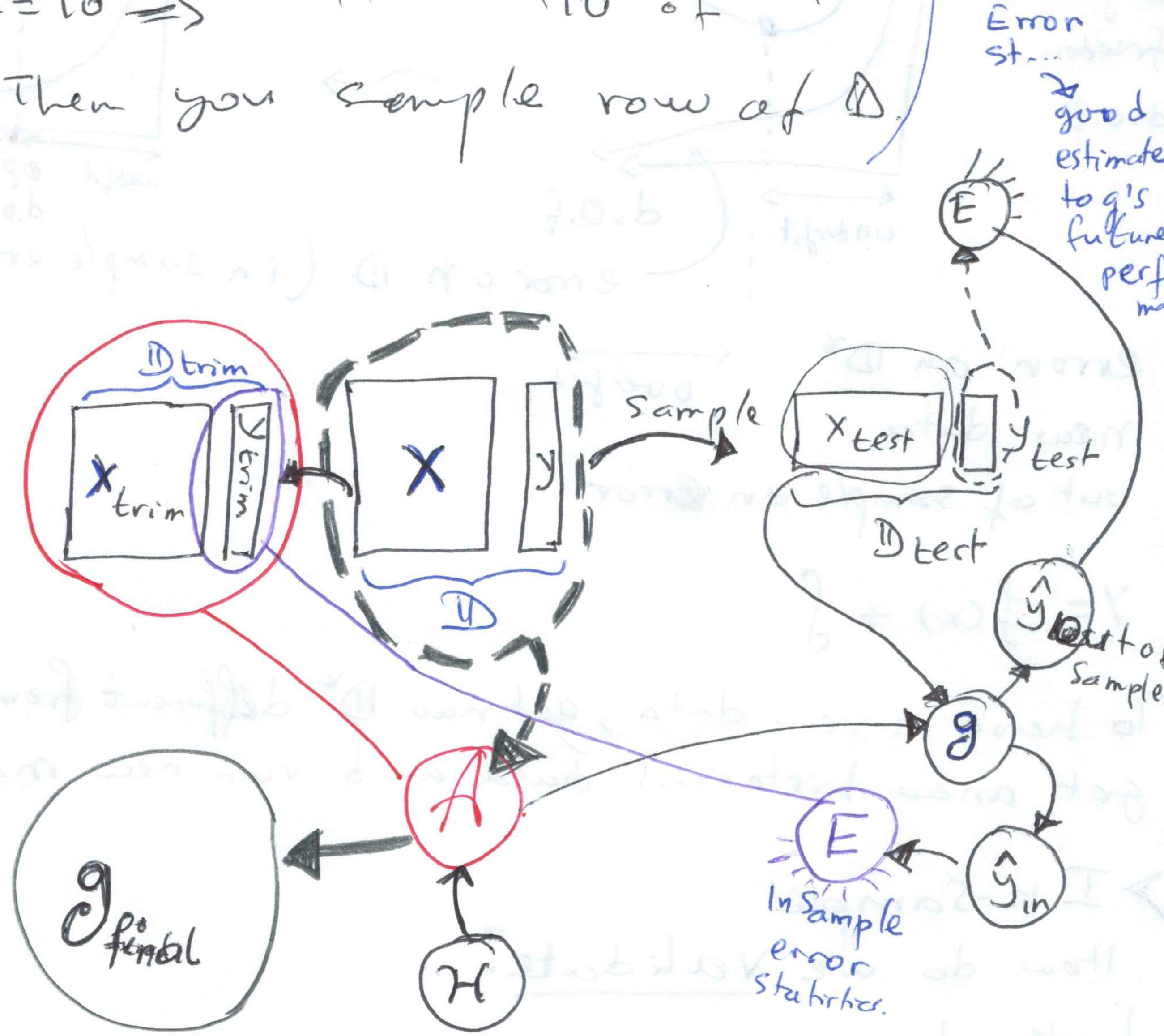
Pretend : $\mathbb{D} = \underbrace{\mathbb{D}_{trim}}_{\text{use this to build } g} \cup \underbrace{\mathbb{D}_{test}}_{\text{use this to validate } g}$

This is one possible strategy.

We select K
where $\frac{1}{K} :=$ prop of $\mathbb{D}$ is $\mathbb{D}_{test}$.

Most popular. { $k = 5 \Rightarrow D_{test}$ has 20 of the full
$k = 10 \Rightarrow$ " 10 of "

Then you sample row of $\mathbb{D}$.



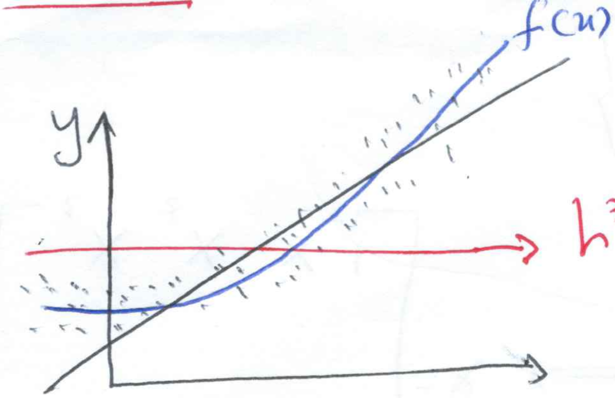
Would you agree g_{final} is better than g?
sample error will be less than other error.

H

END

POLYNOMIAL FUNCTIONS

POLYNOMIAL REGRESSIONS



h^* if $\mathcal{H} = \{ w_i : w_0 \in \mathbb{R} \}$

h^* if $\mathcal{H} = \{ w_0 + w_1 x : w_0, w_1 \in \mathbb{R} \}$

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

Weierstrass Approximation Theorem

For every continuous f: f ~~ℝ~~ where its domain is $x \in [a, b]$, $\exists$ a polynomial function P such that

$$\forall \varepsilon > 0,\ \forall x \in [a,b]\ \ |f(x) - p(x)| < \varepsilon$$

First Polynomial be degree 2

Let $\mathcal{H} = \{ w_0 + w_1 x + w_2 x_1^2,\ \vec{w} \in \mathbb{R}^3 \}$

$$X = \underset{p=1}{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}} \xrightarrow{\text{use Linear Model}} X = \underset{p+1}{\begin{bmatrix} 1 & x_1 \\ & \\ 1 & x_n \end{bmatrix}} \Longrightarrow X = \underset{p+1+1 = 2p+1}{\begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}}$$

How to get the best from this model?

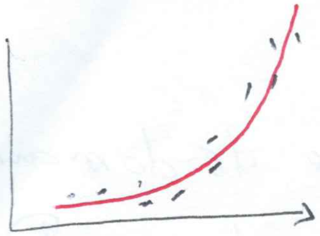
<u>Ans</u>: it is a linear model: $\vec{b} = (X^T X)^{-1} X^T \vec{y}$

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

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x \end{bmatrix} \Longrightarrow x = \begin{bmatrix} 1 & x_1 & \cancel{x_1^3}\, x_1^2 \\ \vdots & & \\ 1 & x_n & \cancel{x_n^3}\, x_n^2 \end{bmatrix} \longrightarrow X = \begin{bmatrix} 1 & x & x^2 & x^3 \\ \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

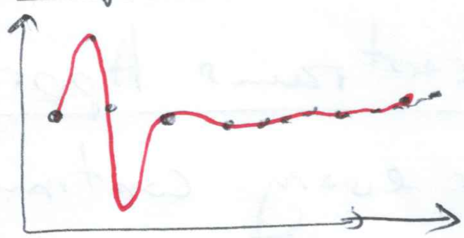
Linear model $\Longrightarrow$ estimate error is very small

Degree 3



Degree 8

Degree 13



Let $n=5$ and $d=4$

$$x = \begin{bmatrix} \ \end{bmatrix} \Longrightarrow \underset{p+1}{\begin{bmatrix} 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}} \Longrightarrow \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_4^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 \end{bmatrix}}_{X}$$

Vandermonde matrix

$$\text{determinant} = \prod_{i=1}^{n} \prod_{j=i+1}^{n} (x_j - x_i) \neq 0$$

if $x_1, \ldots, x_n$ are unique.

I could have this:

$$x = \begin{bmatrix} \ \end{bmatrix} \Longrightarrow x = \begin{bmatrix} \ \end{bmatrix} \Longrightarrow X = \begin{bmatrix} 1 & x_1 & \log(x_1) \\ \vdots & & \\ 1 & x_n & \log(x_5) \end{bmatrix}$$

$\longrightarrow$ other functions / transformation of the predict?.