Math 390 Lec 13 3/21/19

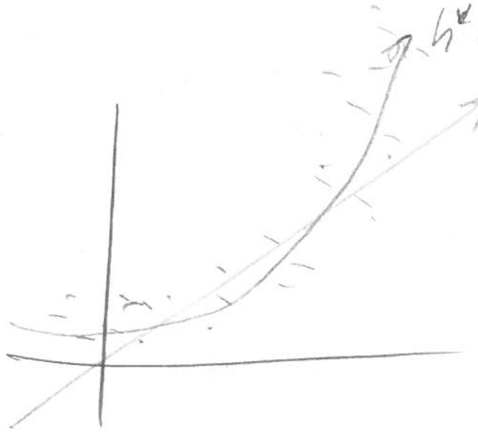
$$y = g(\vec{x}) + \underbrace{(h^*(\vec{x}) - g(\vec{x}))} + \underbrace{(f(\vec{x}) - h^*(\vec{x}))} + \underbrace{(t(\vec{z}) - f(\vec{x}))}$$

Assume $\mathbb{D}$ is fixed with p "raw features" derived from raw from now on.

You cannot ignore error due to ignorance. (need more features) You cannot ignore est. error (need more obs.'s)

You can ignore misspecification error by changing $\mathcal{H}$ and thus changing h^*, more complex model, more d.f.

The risk is you can increase est. error as you will have less data per parameter.



h^* if $\mathcal{H} = \{w_0 + w_1 x_1 + w_2 x_1^2 : w_0, w_1, w_2 \in \mathbb{R}\}$

h_0^* if $\mathcal{H}_0 = \{w_0 + w_1 x : w_0, w_1 \in \mathbb{R}\}$

$\mathcal{H}$ has one more d.f. (one more parameter) than $\mathcal{H}_0$. $\mathcal{H}$ is more complex than $\mathcal{H}_0$. Is this a good tradeoff? Here, yes. Not always...

If $h^* = f$. You can see misspecification error.

$\mathcal{A}$ = OLS and using...

$\mathcal{H}$ is known as "polynomial regression". Is it linear? Yes!

Does it matter if X, X^2 are themselves related? Not for prediction.

It will impact the b_1, b_2 but likely not the $\hat{y}$.

Why is polynomial regression a principled thing to do.

In real analysis, you'll study the Weierstrass Approx. Thm.

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

Any function can be approximated by a polynomial.

How do we do a polynomial regression

$$X = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} \Longrightarrow X = \begin{bmatrix} 1 & x_{11} & x_{21} = x_{11}^2 \\ 1 & x_{21} & x_{22} = x_{21}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{2n} = x_{n1}^2 \end{bmatrix}$$

$p_{raw} = p = 1$

$p = 2$

Is the new X full rank? Yes... why?

Now use OLS as usual $\vec{b} = (X^T X)^{-1} X^T \vec{y}$

$$\vec{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

Oxymoron:

"Non-linear linear model" why??

Can you do $\mathcal{H} = \{w_0 + w_1 x_1 + w_2 x_1^2 + w_3 x^3\}$? Yes... see ...

What if $p+1 = n$

Let $n = 5$

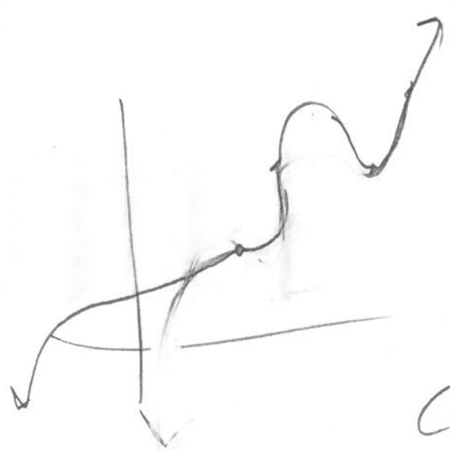
$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & & & \\ 1 & x_3 & & & \\ 1 & x_4 & & & \\ 1 & x_5 & & & x_5^? \end{bmatrix}$$

Is X full rank?

Vandermonde Matrix

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

if all x_i's unique.



If $p+1 = n \Rightarrow \hat{y} = y$. What does this look like?

Can you instead take logs? Sines? Exponentials?

Of course...