Lec 18 Math 390 4/9/19

New problem... given the same X, there are many (a) transformations of x, (b) many $\mathcal{H}$'s and (c) many $\mathcal{A}$'s all producing different models $g_1(\vec{x}), g_2(\vec{x}), \ldots, g_M(\vec{x})$.
Since all models are approximate, there is no concept of a "correct" model. But you still need to choose one to use. This is the fundamental problem of "model selection" in statistics & science!! Since the dawn of time!! Whole ~~te~~ textbooks written on this... e.g. with $p=1$..

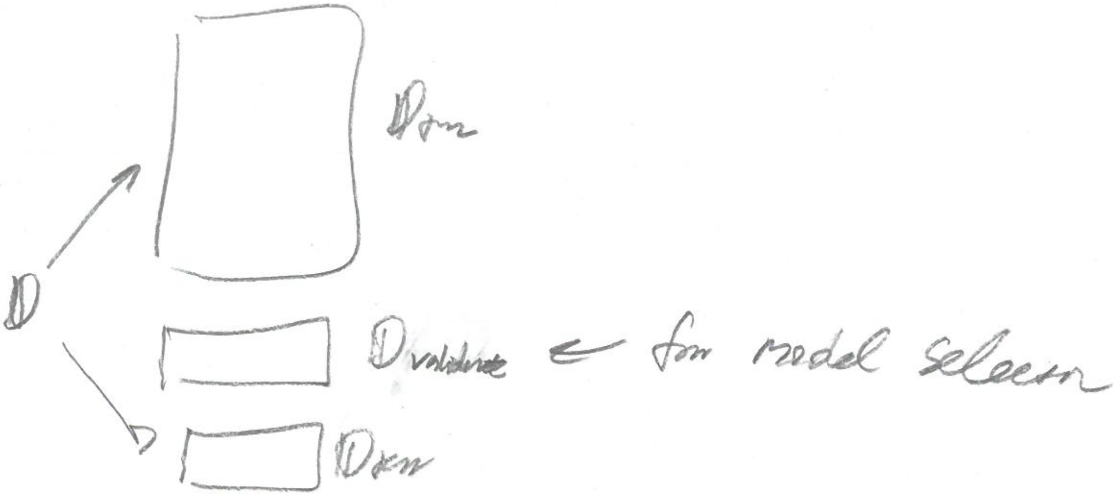
$g_1(x) = b_0 + b_1 x$

$g_2(x) = b_0 + b_1 x + b_2 x^2$

$g_3(x) = b_0 + b_1 \ln(x)$

$g_4(x) = b_0 + b_1 \mathbb{1}_{x \in [0,1)} + b_2 \mathbb{1}_{x \in [1,2)} + \ldots$ X is now ordinal!

In our class, we will do model selection using one technique. This is <u>not</u> the only technique....



Procedure:

① Fit g_1 using D_{train} and get an OOS error est on $D_{validate}$

② Fit g_2 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃

⋮

(M) Fit g_M 〃 〃 〃 〃 〃 〃 〃 〃 〃 〃

(M+1) Locate model m^* with lowest error on $D_{validate}$ (model selection step)

(M+2) Fit g_{m^*} on $D_{train} \cup D_{validate}$ and get OOS error on D_{test}.

(M+3) Fit g_{m^*} on D = all the data and ship.

This also allows for honest model comparison.