Previously covered "model select°"

Mod 1: eg: $g_1 = b_0 + b_1 x$ from $A = (\mathcal{H}_1, \mathbb{D})$

Mod 2: eg : $g_2 = b_0 + b_1 x + b_2 x^2$ from $A = (\mathcal{H}_2, \mathbb{D})$

⋮

Mod m : eg: $g_m = b_0 + b_1 x + b_2 x^2 + b \log(x)$ from $A_3 = (\mathcal{H}_3, \mathbb{D}$

$\mathbb{D} =$ D_{train}, D_{select}, D_{test} } Big issue: How to choose or select a model.

Linear models can have: $g_m = b_0 + b_x x$

Iterate fitting Linear models

You start with a huge # of derived predictors. (for exple, polynomial, log, etc...)...

Iterarely is the best predictor. I need to select to model select.

This is called Formal Stepwise linear model construct°: "~~space~~ stepwise regression."

Problems:

① we still need specify predictive set

② model is still linear

③ Computation.

Whenever we specify for exple $\mathcal{H} = \{\vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1}\}$

we made a "parameter assumpt$\equiv$"

i.e the model will be of this ~~form~~ fixed form.

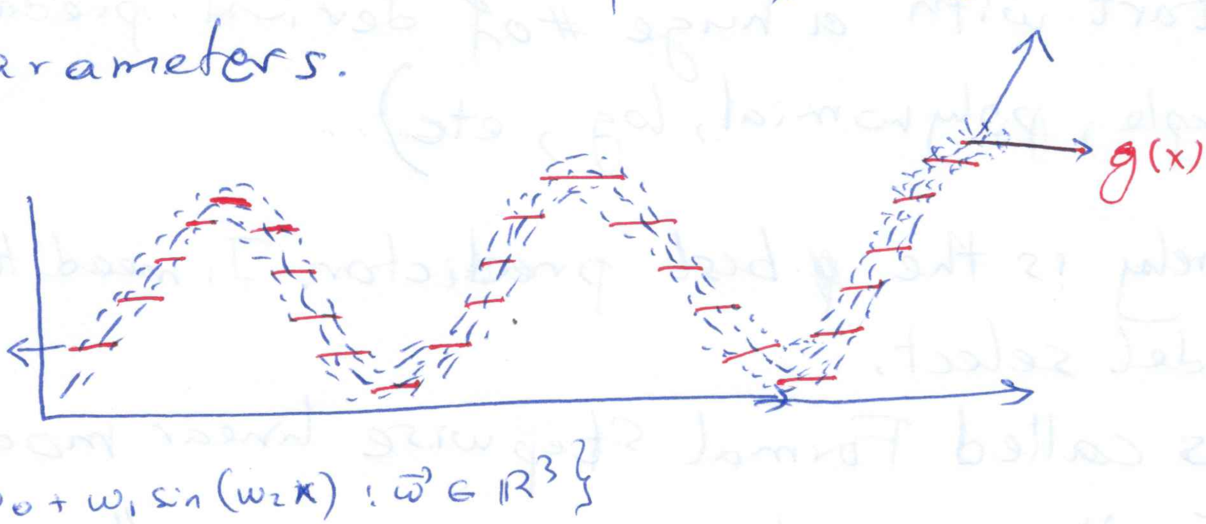
Would is be nice if $\mathcal{H}$ can adjust flexibly

and ~~th~~ iter by th. for a more optimal expreriencefor $\mathbb{D}$ and of $n \nearrow \Rightarrow$ fit a more complex model

This is called "non-parametric regression."

Model form is not pre-specified but construct

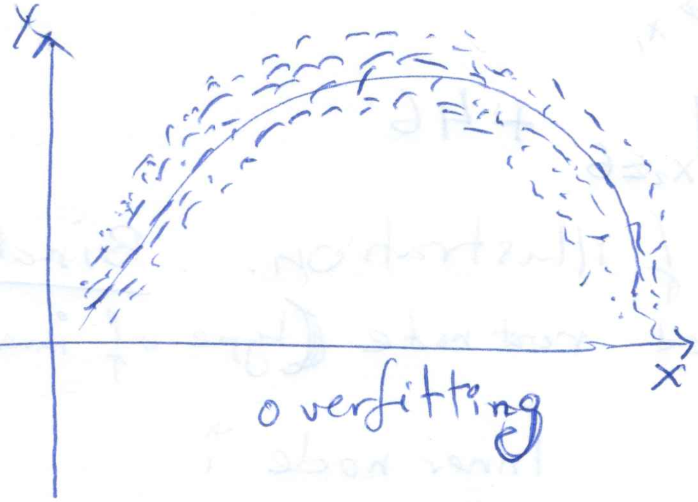
$\mathbb{D}$ provides ① model space and ② estimate of the parameters.



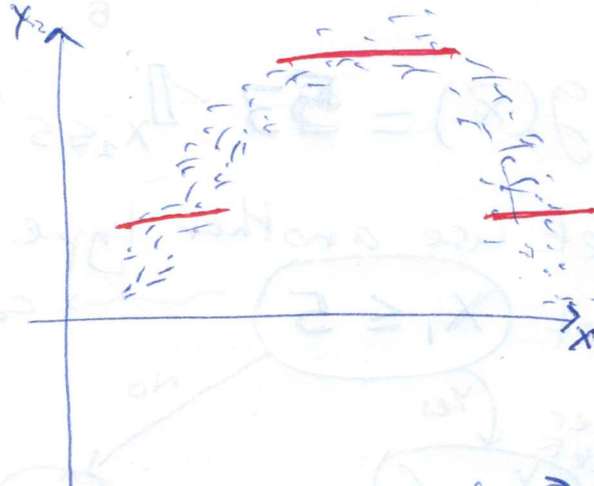
$\mathcal{H}_1 = \{w_0 + w_1 \sin(w_2 x) : \vec{w} \in \mathbb{R}^3\}$

$\mathcal{H}_2 = \{w_0 + w_1 x + w_2 x^2 : \vec{w} \in \mathbb{R}^3\}$

g(x) is a collect° of subf°s (constants)
at small intervals whe the uni° of all intervals =
intervals = $\mathcal{X} \overset{e.g}{=} \mathbb{R}$

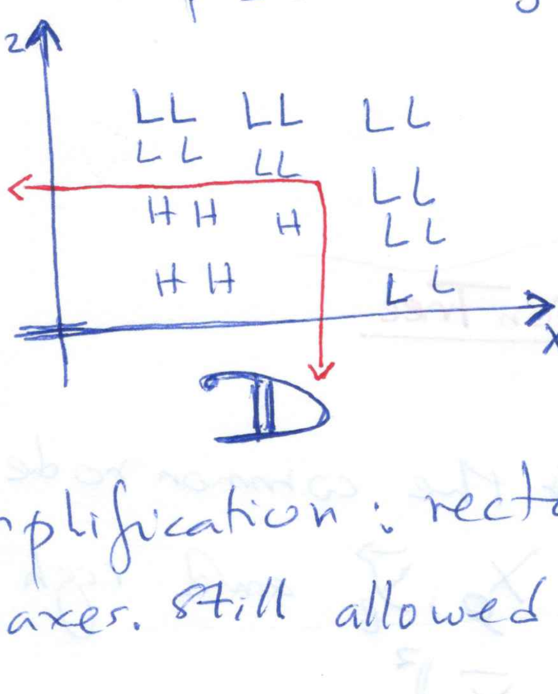


overfitting



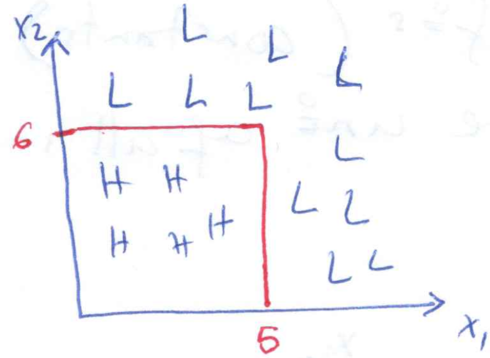
Why this is not a parametric $\Rightarrow$ b/c no specific $\mathcal{H}$
The values are estimated.
In large intervals $\Rightarrow$ underfitting. Non-specification.

$p = 2$ $\quad$ $y = \mathbb{R}$ $\quad$ Basic L : low value
H : High value



$\mathbb{D}$

$\Rightarrow$ Simplification: rectangle whose sides are parallel in the axes. still allowed to be infinity large.



$y = \mathbb{R}$

$g(\vec{x}) = ?$

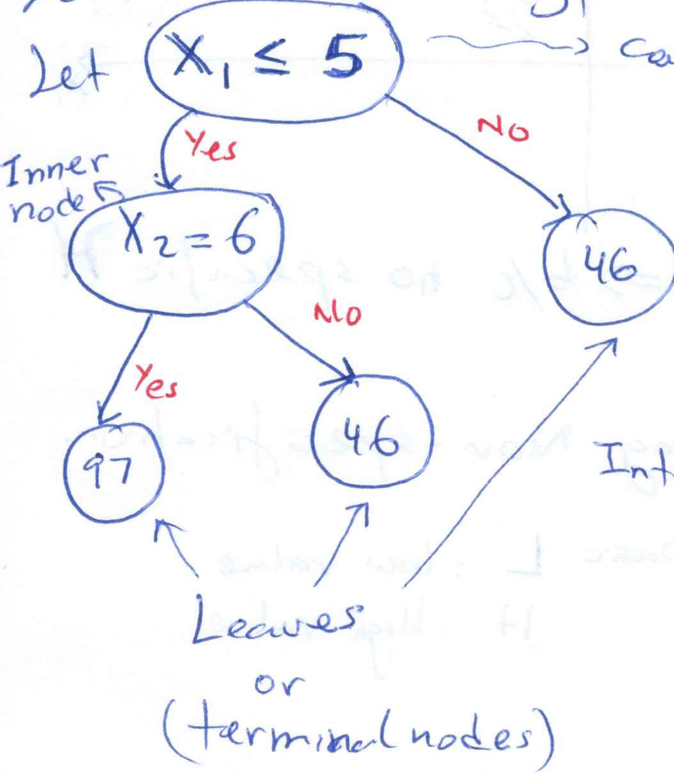
L: Low value = 96

H: High value = 97

$$g(\vec{x}) = 53\, \mathbb{1}_{x_1 \leq 5}\, \mathbb{1}_{x_2 \leq 6} + 46$$

Let use another type of illustration. <u>Binary Tree</u>

Let



called root node (type of inner node)

<u>Inner node:</u>

have "split nodes"

① split value X_j

② split values m

$\Big\}\ X_j \leq m$

Interpretable

<u>One</u> such algorithm : <u>Regression Tree</u>

① Begin with all $X, \vec{y}$

② For every possible split at the common node, divide the data into left $X_\ell, \vec{y}_\ell$ and righ $X_r, \vec{y}_r$

and compose $SSE_\ell = \sum (y_{\ell,i} - \bar{Y}_\ell)^2$

$SSE_r = \sum (y_{r,i} - \bar{Y}_r)^2$

③ Find the split which minimizes the "tot SSE"

(3) ~~continue~~ Find the split..

$$SSE_{avg} = \frac{n_l SSE_l + n_r SSE_r}{n_l + n_r}$$

(4) Create the split assign $\cancel{X}_l, \vec{Y}_l$ and $X_r, \vec{Y}_r$ to the daughter nodes

(5) Reuse a ~~sp~~ step 2-4 for both daughter nodes until "STOP".