Lecture 18 04/14/2080

$\mathcal{H} = \{ w_1 b_1(\vec{x}) + w_2 b_2(\vec{x}) + \ldots + w_B b_B(\vec{x}) : \vec{w} \in \mathbb{R}^B \}$ $y = \mathbb{R}$
and $b_1, b_2, \ldots, b_B$ are known function that attempt to span the function space of $f : \mathbb{R}^p \to \mathbb{R}$.

Example set of function: Set of all first order interactions

$x_1, x_2, \ldots, x_p, x_1^2, x_2^2, \ldots, x_p^2, x_1 x_2, x_1 x_3, \ldots, x_{p-1} x_p$

$B = 2p + \binom{p}{2} =$,, ,, ,, ,, ,, ,, ,, ,, $, x_1^3,$
$x_2^3, \ldots, x_p^3, x_1^2 x_2, \ldots, x_{p-1}^2 x_p, x_1 x_2 x_3, \ldots, x_{p-2} x_{p-1} x_p$

Set of all second order interactions.
B is exponensially large.

A: Likely ... w's will be spark i.e. most w's = 0.

Forward stepwise OLS.

⓪ Let $g(\vec{x}) = g_0(\vec{x}) = \bar{y}$, $X = [\vec{1}]$

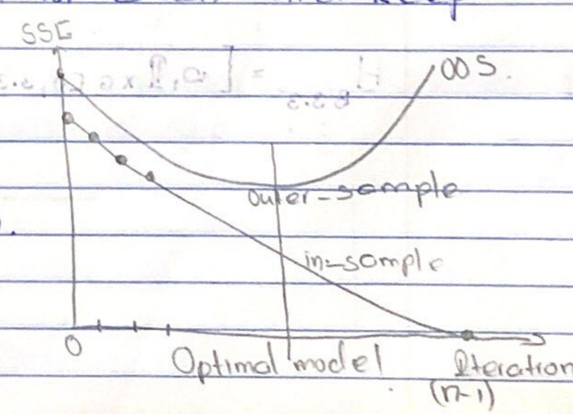
OLS regression

① Try all B individually. $y \sim b_1(\vec{x}), y \sim b_2(\vec{x}), \ldots, y \sim b_B(\vec{x})$ and compute SSE reduction for each $X = [\vec{1}\ b_k(\vec{x})]$. and keep best one b_*.

② Try all B-1 remaining individually; $y \sim [\vec{1}\ b_{*_1}[\vec{x}]\ b_k[\vec{x}]]$ and update SSE reduction for each and keep best one b_{*_2}.

repeat
⋮

⑤ Stop if ..... OOS SSE goes up.



$\mathbb{D}_{train}$

$\mathbb{D}_{select}$

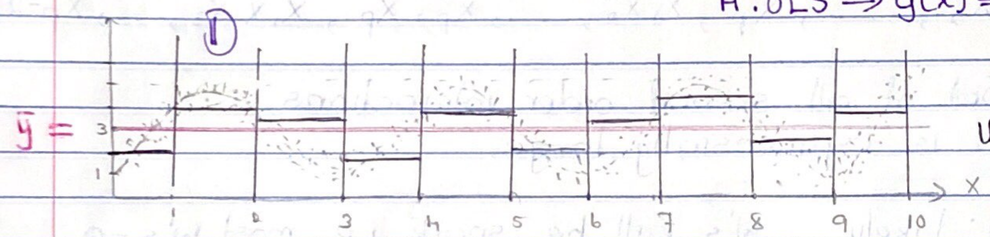
$\mathbb{D}_{test}$

II Model selection procedure $g_1, g_2, \ldots, g_m$
$m = 2^B$

The algorithm could be modified to use K-fold CV (inner & outer)

$\underbrace{\text{Classification \& Regression}}_{y=\mathbb{R}}$ Tree (CART) Algorithm (1984)

$A: OLS \Rightarrow g(x) = \bar{y}$



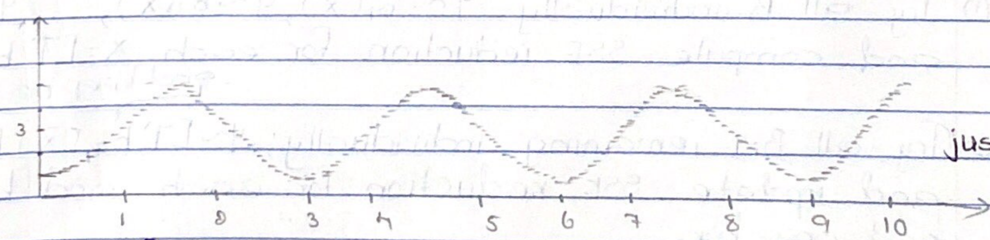
underfit bc all bunch of misspecification error

Underfit.

bin size = 1

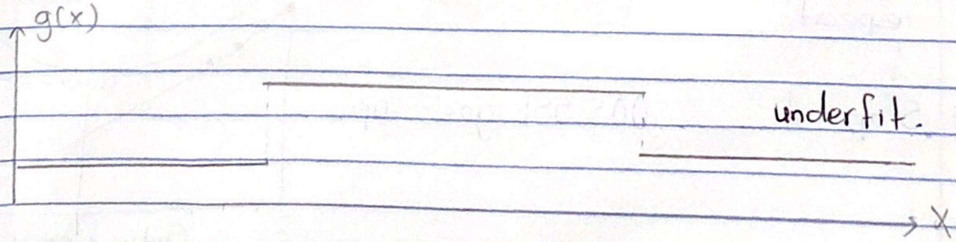
$\mathcal{H}_{B1} = \{ w_1 \mathbb{1}_{x \in [0,1]} + w_2 \mathbb{1}_{x \in [1,2]} + \ldots + w_9 \mathbb{1}_{x \in [9,10]} : \vec{w} \in \mathbb{R}^9 \}$

$A: OLS \Rightarrow g(\vec{x}) =$
$\mathcal{H}_{B0.1} = \{ w_1 \mathbb{1}_{x \in [0,0.1]} + \ldots + w_{99} \mathbb{1}_{x \in [9.9,10]} : \vec{w} \in \mathbb{R}^{99} \}$

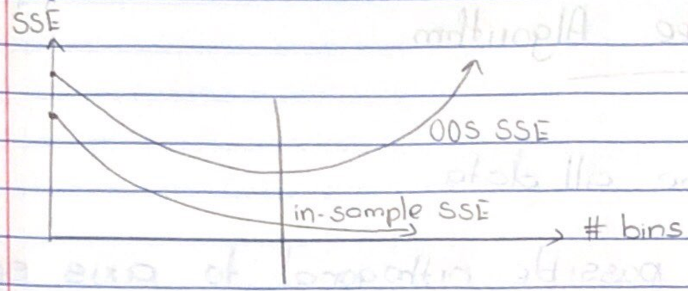


just right

$\mathcal{H}_{B3.3} = \{ w_1 \mathbb{1}_{x \in [0,3.3]} + w_2 \mathbb{1}_{x \in [3.3,6.6]} + w_3 \mathbb{1}_{x \in [6.6,10]} : \vec{w} \in \mathbb{R}^3 \}$



underfit.



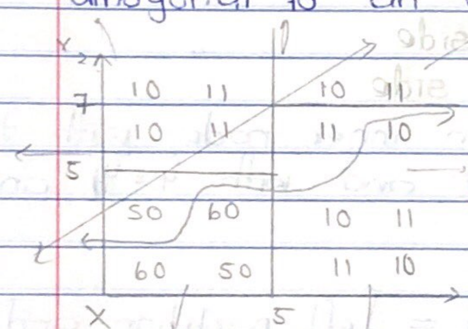
Select bin size using the OOS SSE curve

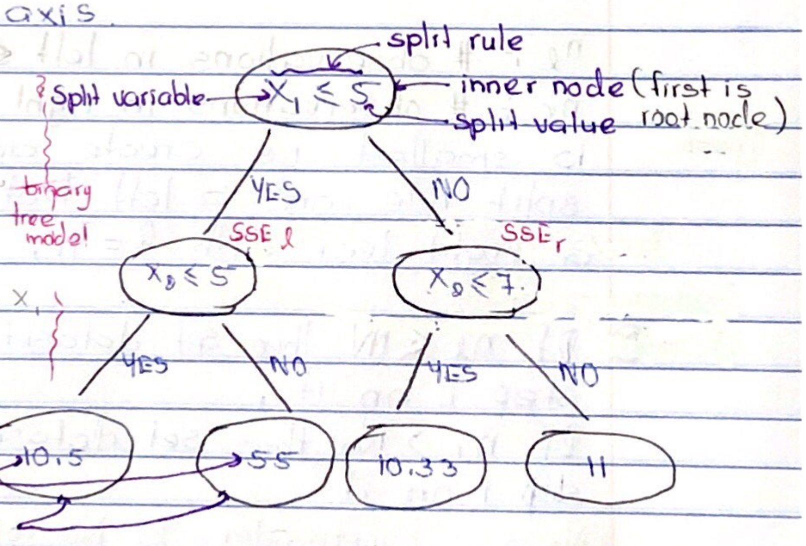
P=1 ↑

In two dimentions (p=2), the bins are squares.
B = # bins/dim $\Longrightarrow B_{tot} = B^2$

In p dimentions $B_{tot} = B^p$ which > n ; very fast.

A solution: "Binary trees" with split rules that are othogonal to an axis.





= Union of mutually exclusive, collectively exhausser possibly infinitely large hyperrectangles.

Regression Tree Algorithm

⓪ Let dataset be all data

① Consider every possible orthogonal to axis split
$x_j \leq x_{rj}$ $j = 1 \ldots p$, $l \in 1, \ldots, n-1$

↑ ordered/ sorted values

And calculate SSE_l, SSE_r the SSE's in the putative left node and right node. Select the rule where

$$SSE_{weighted} := \frac{n_l\, SSE_l + n_r\, SSE_r}{n_l + n_r}$$

n_l := # observations in left set.
n_r := # observations in right set
is smallest. i.e create an inner node with that split rule and a left leaf and with $\hat{y} = \bar{y}_l$ and a right leaf with $\hat{y} = \bar{y}_r$.

② If $n_l > N_0$ then set dataset = left partition and run step 1 on it,
If $n_r > N_0$ then set dataset = right partition and run step 1 on it,
(N_0 is a hypermater. If N_0 is small e.g. $N_0 = 1 \Rightarrow$ g is overfit.)
If N_0 large $\Rightarrow$ Underfit model.
How to pick N_0? Use 3-fold selection.