SOLUTION MA8701 May 2017

TMA4268 M8

RecEx 2c-f

Problem 1

a) Book Ch. 2

b) Book Ch. 7.2

c) Book Ch. 7.10

d) Book Ch. 7.11

e) Book Ch. 8.7 ~~(Bagging)~~
& "Introduction to" Ch. 8.2

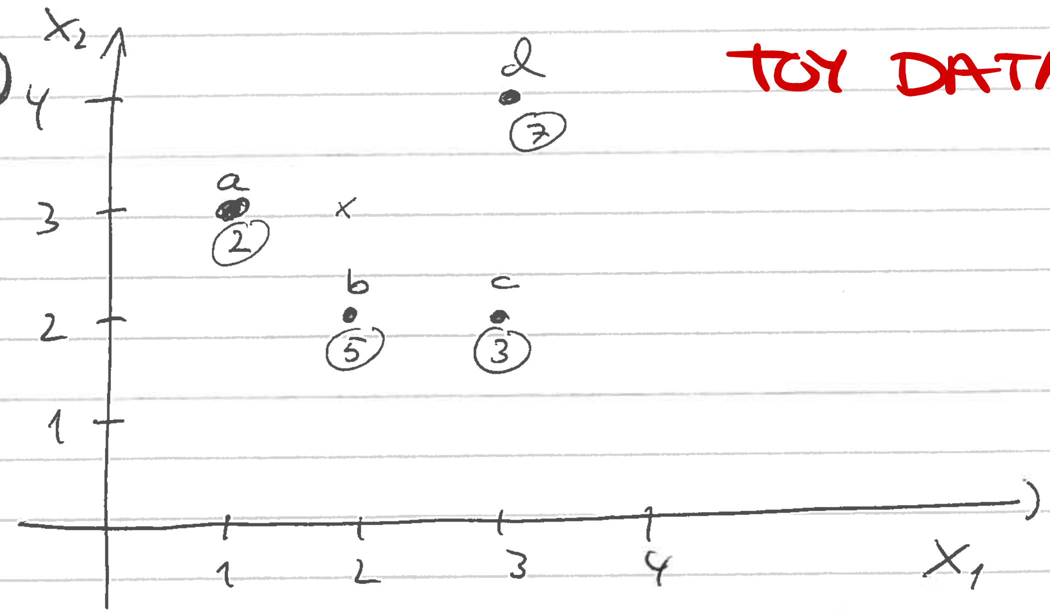
f) Book Ch. 2.5

Problem 2

a) Book Ch. 9.2, Intro Ch. 8.1

b) Book Ch. 9.2.2

c)



TOY DATA

2c) In general the splitting variable and split point s minimizes

$$\min_{c_1} \sum_{i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{i \in R_2(j,s)} (y_i - c_2)^2$$

where c_1, c_2 are averages.

In toy example:

Try x_1 first

- Between a and b:

Left: $(2-2)^2 = 0$

Right: Average is 5, so

$$(5-5)^2 + (3-5)^2 + (7-5)^2 = 8$$

So $\underline{\text{Sum} = 8}$

- Between b and ~~c~~ d:

Left: $\left(2-\frac{7}{2}\right)^2 + \left(5-\frac{7}{2}\right)^2 = \left(\frac{3}{2}\right)^2 + \left(\frac{3}{2}\right)^2 = \frac{18}{4} = 4.5$

Right: $(3-5)^2 + (7-5)^2 = 8$

Sum $\underline{12.5}$

Then try x_2!

- Between a and d:

Upper: $(7-7)^2 = 0$

Lower: $\left(2-\frac{10}{3}\right)^2+\left(5-\frac{10}{3}\right)^2+\left(3-\frac{10}{3}\right)^2$

$$=\left(\frac{4}{3}\right)^2+\left(\frac{5}{3}\right)^2+\left(\frac{1}{3}\right)^2=\frac{42}{9}=\frac{14}{3}=4.67$$

Sum 4.67 (best!)

- Between $\frac{ad}{bc}$:

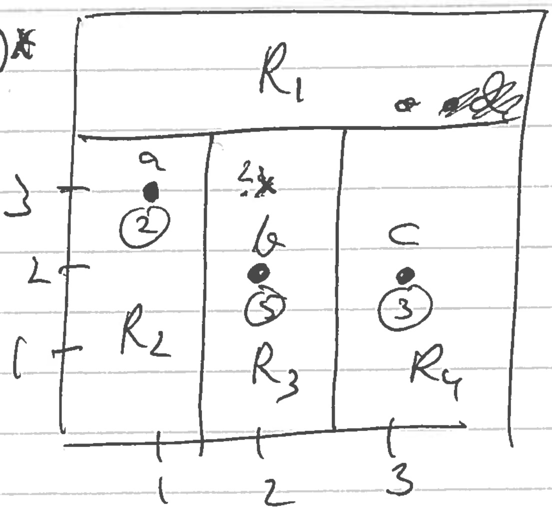
Upper: $\left(2-\frac{9}{2}\right)^2+\left(7-\frac{9}{2}\right)^2=\left(\frac{5}{2}\right)^2+\left(\frac{5}{2}\right)^2=\frac{25}{2}$ $=12.5$

Lower: $(5-4)^2+(3-4)^2 = 2$

Sum 14.5

Thus: First split is for $x_2 < 3.5$ or $x_2 > 3.5$

2d) 1)*



Next split is either for $x_2 < 2.5$ or for $x_1 < 1.5$ (equivalent)

Last split is between the two last points.

[pruning at this level only optional]

For what values of α is T_0 optimal?

2e) In our example, $C_\alpha(T_0) = \alpha |T_0| =$

$= 4\alpha.$

The first pruning would be to add together R_3 and R_4

In that case:

$$\sum_{m=1}^{|T|} \sum_{x_i \in R_m} (y_i - \hat{c}_m)^2$$

would be $(5-4)^2 + (3-4)^2 = 2$

So $C_\alpha(T) = 2 + \alpha \cdot 3$ which is $> 4\alpha$

if $\alpha < 2$

Thus T_0 is optimal if $\alpha < 2$

2f) Predicted with T_0: Either 5 or 2 depending on second division ...

~~2-NN:~~ $\cancel{\frac{2+5}{2} = 3.5}$

~~OLS: Model~~ $\cancel{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon}$

~~Use LS~~