Decision Tree

(Row 2, Row 3, Row 4, Row 5) X train

(Row 6) Y train

(1) Values = 1725 + 4449 + 21984 + 3866

(2-5)

Average = 8006

starting Messiness (variance) = 66,155,786

(2) Finding Best Question

Option A: split by Age (If Age < 30)

Left room (Young) : Row 2 (18), Row 3 (28)

Cost : 1725, 4449

messiness score:

Right Room: Row 4 (33), Row 5 (32)

Cost: 21,984, 3,866

(Room average) 41258459 (verbad)

Messiness score:

Option B (split by BMI) (< 25)

Left room: 0

Right room: 1371588 (very low) (win)

Room Averages

(3) Building leaf nodes

Bucket (1)
(skinny and expensive gp)

Prediction: 21984

Bucket (2)
(the high BMI Gp)

Prediction: $\frac{1725 + 4449 + 3866}{3}$

: 3347

(4) Test Row 6

Age: 31

BMI: 25.74

Y test Actual Cost: $3756

Predicted Path

1. Is your BMI < 25
2. go right
3. $3347 (model)

* off only by $409

Messiness Score is get?

1. Find mean : $\frac{1726+4449+2984+3867}{4}$

: 8006.59

2. mean Squared error for each person

Person(1) = $(1726-8007)^2$ = 39451372

⋮

Person(4) = $(3867-8007)^2$ = 17137363

3. Averge mean Squard error = $\frac{\text{result}}{4\text{people}}$

= 6655786

* High variance = Bad
* low variance = Good

How to find winner

messiness score

Left room: (1 out of 4 people) = 25.1.0

Right room: (3 out of 4 people) = 754.13771588

Total Score = (0.25 x 0) + (0.75 x 1371588)

= 1028691

* starting score: 66155786
* Age split score: 41958459
* BMI split Score: 1028691 ✓

Y test : 3756

The error: 409

the math: 409 ÷ 3756 = 0.108 (10.8%)

0% - 10% (Excellent)

10% - 20% (Good)

20% - 50% (okey)

> 50% (Bad)

MAE = (20 Xtest Prediction error) / 20 = final average

(actual - predicted)

20xtest