

# Decision Tree

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

(Row 6) Y train

$$(1) \text{ Values} = 1725 + 4449 + 21984 + 3866$$

(2-5)

$$\text{Average} = 8006$$

$$\text{starting Messiness (variance)} = 66,155,786$$

(2) Finding Best Question  
Option A: split by Age (If Age < 30)

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

(Young)

Cost : 1725, 4449

messiness score

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

Cost : 21,984, 3,866

(Room average)  
messiness score : 41,958, 459 (verbad)

Option B (split by BMI) ( $< 25$ )

Left room: 0

Room Averages

Right room: 1371588 (very low) (win)

(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 :  $\frac{1726 + 4449 + 2184 + 3867}{4}$

: 8006.59

2. Mean Squared error For each person

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

⋮

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

3. Average mean Squared error =  $\frac{\text{result}}{4 \text{ people}}$

= 6655786

\* High variance = Bad

\* Low variance = Good

How to find winner

Left room: (1 out of 4 people) = 25.1%, 0

Right room: (3 out of 4 people) = 75.1%, 171588

$$\text{Total Score} = (0.25 \times 0) + (0.75 \times 1371588) \\ = 1028691$$

\* starting score 66155786

\* Age split score 41958459

\* BMI split score 1028691 ✓

