Name: M. Hamid Raza

AG : 2023-9g-7206

Instuctor : Shoaib Sir

Subject : Machine learning

Assignment 2

Q1: K-Mean Clustring.

K = 3

initial centroids:

1. (19, 15, 39)
2. (64, 19, 3)
3. (22, 17, 76)

	Datapoint			1			2			3			cluster	New
				19	15	39	64	19	3	22	17	76		
A	19	15	39	0			57.77			37.18			1	
B	21	15	81	42.05			89.16			5.48			3	
C	20	16	6	33.03			44.2			70.04			1	
D	23	16	77	38.22			84.65			1.73			3	
E	31	17	40	12.21			49.62			37.11			1	
F	22	17	76	37.18			84.24			0.09			3	
G	35	18	96	36.8			29.17			71.2			2	
H	23	18	94	55.23			99.81			18.06			3	
I	64	19	3	57.77			0.0			84.24			2	
J	30	19	72	135.01			76.92			9.17			3	

Next centroids:

1. (23.33, 16, 28.33)
2. (49.5, 18.5, 4.5)
3. (23.8, 17, 80)

	Data Points			23.33	16	28.33	49.5	18.5	4.5	23.8	17	80	cluster	New
A	19	15	39	11.56			46.18			41.33			1	1
B	21	15	81	52.73			81.71			3.58			3	3
C	20	16	6	22.58			29.64			74.1			1	1
D	23	16	77	48.67			77.23			3.26			3	3
E	31	17	40	14			40.6			40.64			1	1
F	22	17	76	47.7			76.62			4.39			3	3
G	35	18	6	25.27			14.59			74.85			2	2
H	23	18	94	65.7			93.34			14.06			3	3
I	64	19	3	48.01			14.59			86.89			2	2
J	30	19	72	44.28			70.26			10.32			3	3

Q2. Decision tree Regression

Average = $\frac{357+196+221+136+255+244+230+212+225+239+359+338}{12}$

Average = 251.17.

Std = $\sqrt{\frac{(196-251.17)^2+(221-251.17)^2+(136-251.17)^2+(255-251.17)^2+(244-251.17)^2+(230-251.17)^2+(212-251.17)^2+(225-251.17)^2+(239-251.17)^2+(239-251.17)^2+(359-251.17)^2+(338-251.17)^2}{12}}$

Std = 65.12

Attribute Vehical Class Assessment.

→ ~~Good~~ Compact.

Average = $(196+221+136+239)/4 = 198$

Std = $\sqrt{\frac{(196-198)^2+(221-198)^2+(136-198)^2+(239-198)^2}{4}}$

Std = $\frac{\sqrt{4+529+3844+1681}}{4}$

std = 38.92

→ Small

Average = $(255+244+212+225+359+338)/6$

Average = 272.17

$$Std = \sqrt{\frac{(255-272.12)^2 + (244-272.12)^2 + (213-272.12)^2 + (225-272.12)^2 + (359-272.12)^2 + (338-272.12)^2}{6}}$$

Std = ~~1880~~ 55.99

→ Mid $(230+359)/2 = 294.5$ = Average

$$Std = \sqrt{\frac{(230-294.5)^2 + (359-294.5)^2}{2}}$$

Std = 64.5

vehical class	S.D	Data instance
compact	38.92	4
small	55.99	6
mid	64.5	2

weight SD for vehical class = $\frac{4}{12}(38.92) + \frac{6}{12}(55.99) + \frac{2}{12}(64.5)$

weight SD for vehical class = 51.72

Std reduction = 65.12 − 51.72

Std reduction = 13.4

Attibute Engine Size.

→ A

Average = (196 + 221 + 225 + 359 + 359) / 5

Average = 272

$$Std = \sqrt{\frac{(196-272)^2 + (221-272)^2 + (359-272)^2 + (359-272)^2}{5}}$$

Std = 71.73

→ B

Average = (136 + 212 + 239 + 338) / 4

Average = 231.25

$$std = \sqrt{\frac{(136-231.25)^2 + (212-231.25)^2 + (239-231.25)^2 + (338-231.25)^2}{4}}$$

std = 72.28

→ C

Average = (225 + 244 + 230) / 3

Average = 233

$$SD = \sqrt{\frac{(225-233)^2 + (244-233)^2 + (230-233)^2}{3}}$$

SD = 10.23

Engine Size	SD	Instance
A	71.73	5
B	72.28	4
C	10.23	3

weight SD for ~~vehicle~~ engine size = $\frac{5}{12}(71.73) + \frac{4}{12}(72.26) + \frac{3}{12}(10.23)$

weight SD for engine size = 56.53

SD reduction = 65.12 − 56.53

SD reduction = 8.59

Attribute Fuel Type.

→ Z

Average = (196 + 255 + 244 + 359 + 338) / 5

Average = 278.40

$$SD = \sqrt{\frac{(196-278.40)^2 + (255-278.40)^2 + (244-278.4)^2 + (359-278.4)^2 + (338-278.4)^2}{5}}$$

SD = 60.94

→ E

Average = (221 + 136 + 212 + 225 + 359) / 5

Average = 230.60

$$\text{Std} = \sqrt{\frac{(221-230.6)^2 + (136-230.6)^2 + (212-230.6)^2 + (225-230.6)^2 + (259-230.6)^2}{5}}$$

Std = 71.99

→ T

Average = (230 + 239) / 2

Average = 234.5

$$SD = \sqrt{\frac{(230-234.5)^2 + (239-234.5)^2}{2}}$$

SD = 4.5

Z	60.94	5
E	71.99	5
T	4.5	2

weight SD for Fuel type = $\frac{5}{12}(60.94) + \frac{5}{12}(71.99) + \frac{2}{12}(4.5)$

weight SD for fuel type = 56.13

SD reduction = 65.12 − 56.13

SD reduction = 8.99

Attribut Fuel Consumption

→ H

Avg = 196

SD = 0

→ M

Avg = (221 + 244 + 212 + 359) / 4

Avg = 259

$$SD = \sqrt{\frac{(221-259)^2 + (244-259)^2 + (212-259)^2 + (359-259)^2}{4}}$$

SD = 58.9

→ L

Avg = (136 + 255 + 230 + 225 + 239 + 259 + 338) / 7

Avg = 254.57

$$SD = \sqrt{\frac{(136-254.57)^2 + (255-254.57)^2 + (230-254.57)^2 + (225-254.57)^2 + (239-254.57)^2 + (259-254.57)^2 + (338-254.59)^2}{7}}$$

SD = 69.33

H	0	1
M	58.9	4
L	69.33	7

weighted SD for fuel consumption = $\frac{4}{12}(58.9) + \frac{7}{12}(69.33)$

weighted SD for fuel consumption = 60.0762

Std Reduction = 65.12 − 60.08

Std reduction = 5.04

Vehicul class = 13.4

engine size = 8.59

fuel type = 8.99

∴ fuel consumption = 5.04

~~engine Vehicul~~ size	~~fuel type Engine size~~ engine size	fuel ~~Consumption~~ type	~~fuel consumption~~ Co2 emission
A	2	H	196
A	E	M	221
B	E	L	136
B	T	L	239

Average = (196 + 221 + 136 + 239 ...) / 4 = 198

$$SD = \sqrt{\frac{(196-198)^2 + (136-198)^2 + (221-198)^2 + (239-198)^2}{4}}$$

SD = 38.92

Attribute column size.

→ A

Average = (196 + 221) / 2

Average = 208.5

$$SD = \sqrt{\frac{(196-208.5)^2 + (221-208.5)^2}{2}}$$

SD = 12.5

→ B

Average = (136 + 239) / 2

Average = 187.5

$$SD = \sqrt{\frac{(136-187.5)^2 + (239-187.5)^2}{2}}$$

SD = 51.5

weighted SD for engine size = $\frac{2}{4}(12.5) + \frac{2}{4}(51.5)$

weighted SD for engine size = 32

Std reduction = 38.92 − 32

Std reduction = 6.92

Attribute Fuel type.

→ 2 =

Average = 196

SD = 0

→ E

Average = (221 + 136) / 2

Average = 178.5

$$SD = \sqrt{\frac{(221-178.5)^2 + (136-178.5)^2}{2}}$$

SD = 42.5

→ T

Average = 239

SD = 0

weight reduction of fultype = $\frac{2}{4}$ (42.5)

weight reduction of fultype = 21.25

Std reduction = 38.92 − 21.25

S.D reduction = 17.67

Attribute Fuel Consumption

→ H

Average = 196

SD = 0

→ M =

Average = 221

SD = 0

→ L = (136 + 239)/2

L = 187.50

$$SD = \sqrt{\frac{(136-187.5)^2 + (239-187.5)^2}{2}}$$

SD = 51.50

weight SD for fuel consumption = $\frac{2}{4}$ (51.5)

weight SD for fuel consumption = 25.75

SD reduction = 38.92 − 25.75

SD reduction = 13.17

engine size = 6.92

fuel type = 17.67 ←

fuel consumption = 13.17

engine size	fuel type	fuel consumption	Co2 emission
C	2	L	255
C	2	M	244
B	E	M	212
A	E	L	225
A	E	L	359
B	2	L	338

Average = 267.17

SD = 58.51

Attribute engine size

→ C

Average = (255 + 244) / 2

Average = 234.5

$$SD = \sqrt{\frac{(255-234.5)^2 + (244-234.5)^2}{2}}$$

SD = 9.5

→ B

Average = (212 + 338) / 2

Average = 275

$$SD = \sqrt{\frac{(212-275)^2 + (338-275)^2}{2}}$$

SD = 63

→ A

Avg = (225 + 359) / 2

Avg = 292

$$SD = \sqrt{\frac{(225-292)^2 + (359-292)^2}{2}}$$

SD = 67

weight SD for engine size = $\frac{2}{6}(9.5) + \frac{2}{6}(63) + \frac{2}{6}(67)$

weight SD for engine size = 46.5

SD reduction = 58.57 − 46.5

SD reduction = 12.07

Attribute Fuel type

→ Z:

Average = (255 + 244 + 338) / 3

Avg = 269.

$$SD = \sqrt{\frac{(255-269)^2 + (244-269)^2 + (338-269)^2}{3}}$$

SD = 49.4

→ E

Avg = (212 + 225 + 359) / 3

Avg = 265.33

$$SD = \sqrt{\frac{(212-265.3)^2 + (225-265.33)^2 + (359-265.33)^2}{3}}$$

SD = 66.44

weight SD fuel type = $\frac{3}{6}(49.4) + \frac{3}{6}(66.44)$

weight SD fuel type = 57.92

SD reduction = 58.57 − 57.92

SD reduction = 0.65

Attribute Fuel consumption

→ L

Avg = (255 + 225 + 359 + 338)/4

Avg = 286.75

$$SD = \sqrt{\frac{(225-286.75)^2 + (225-286.75)^2 + (359-286.75)^2 + (338-286.75)^2}{4}}$$

SD = 62.19

→ M

Avg = (244 + 212)/2

Avg = 228

$$SD = \sqrt{\frac{(244-228)^2 + (212-228)^2}{2}}$$

SD = 16

weight SD fuel consumption = $\frac{4}{6}(62.19) + \frac{2}{6}(16)$

weight SD fuel consumption = 46.8

SD reduction = 58.57 − 46.8

SD reduction = 11.78

engine size = 12.07

fuel type = 0.61

fuel consumption = 11.26

engine size	fuel type	fuel consumption	Co2 emission engine
C	T	L	230
A	Z	M	359

Avg = (230+359)/2 = 294.5

SD = 64.5

engine size

→ A

→ SD = 0

reduction = 64.5

→ B

SD = 0

fuel type

→ T

SD = 0

reduction = 64.5

→ Z

SD 0

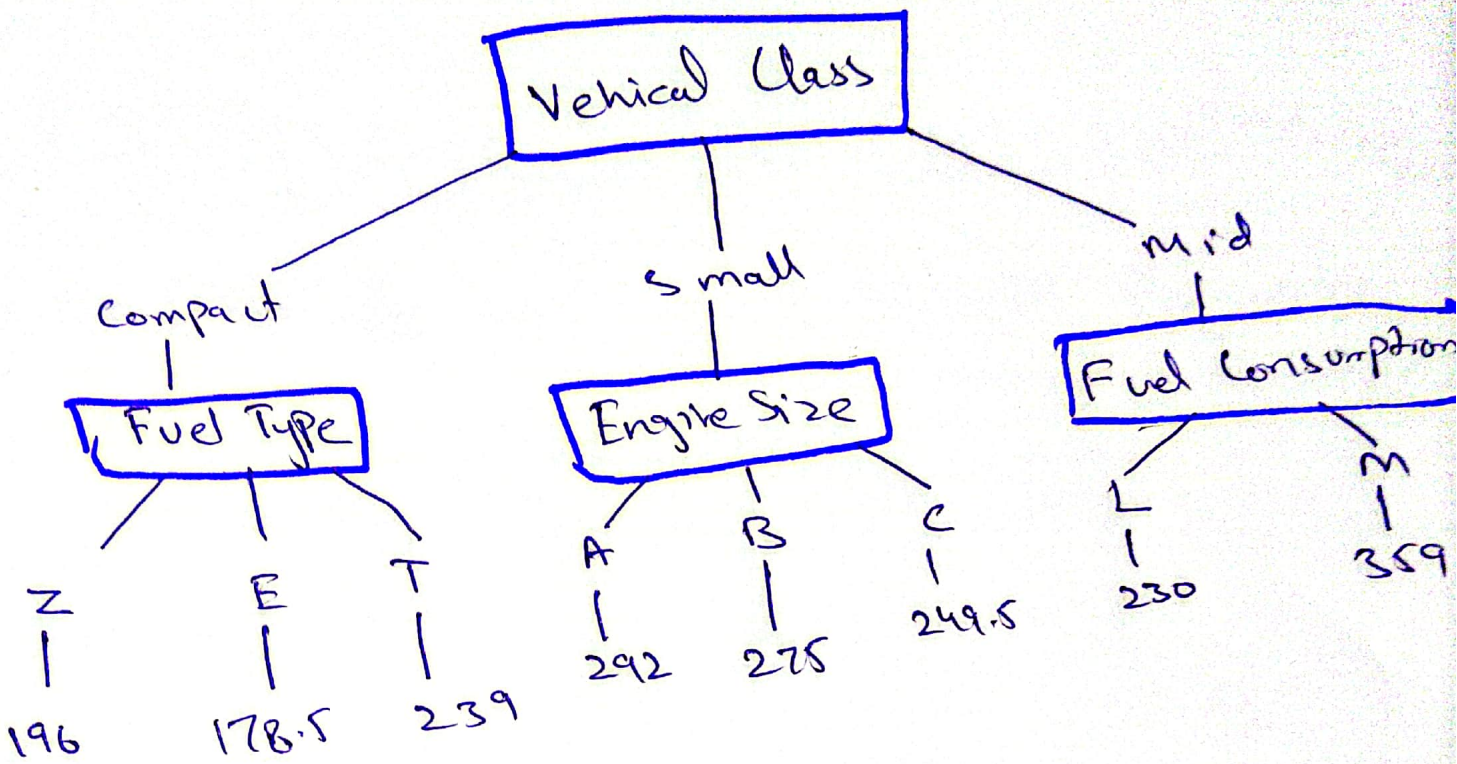
fuel consumption

→ L

SD 0

reduction = 64.5

→ M = 0



Test:

Vehicleclass	engsize	fuel type	consumption	co2 imission
small	A	C	Z	292

Q3: Decision tree Classification.

→ S = [4P 8N 2A] Attribute Exposure

$$\text{entropy}(S) = -\frac{4}{14}\log_2\left(\frac{4}{14}\right) - \frac{8}{14}\left(\log_2\left(\frac{8}{14}\right) - \frac{2}{14}\log_2\left(\frac{2}{14}\right)\right)$$

entropy(S) = 1.38

$\because \log_2 = \log_2$

→ $S_{No} = [2P \quad 3N \quad 2A]$

$$\text{entropy}(S)_{No} = -\frac{2}{7}\log\left(\frac{2}{7}\right) - \frac{3}{7}\log\left(\frac{3}{7}\right) - \frac{2}{7}\log\left(\frac{2}{7}\right)$$

entropy $S_{(No)}$ = 1.56

→ $S_{GD} = [0P \quad 3N \quad 0A]$

entropy of S_{GD} = 0

→ $S_{AV} = [2P \quad 2N \quad 0A]$

$$\text{entropy of } S_{AV} = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{0}{4}\log\left(\frac{0}{4}\right)$$

entropy of S_{AV} = 1

Gain:

Gain (S, exposure) = entropy(S) $-\frac{7}{14}$ entropy(S_{NN})

$-\frac{3}{14}$ entropy of Gd $-\frac{4}{14}$ entropy(S_{AN})

Gain(S, exposure) = $1.38 - \frac{7}{14}(1.56) - \frac{3}{14}(0) - \frac{4}{14}(1)$

Gain(S, exposure) = $1.38 - 0.78 - 0.29$

Gain(S, exposure = 0.31)

Attribute Garage Type.

→ $S_{Attchd} = [2p \quad 6_N \quad 0_A]$

entropy $(S)_{Attchd} = -\frac{2}{8}\log\frac{2}{8} - \frac{6}{8}\log\frac{6}{8} - \frac{0}{8}\log\frac{0}{8}$

entropy(S_{Attchd}) = 0.81

→ $S_{Detchd} = [0p \quad 2N \quad 1A]$

entropy (S_{Detchd}) = $-\frac{0}{3}\log\frac{0}{3} - \frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

entropy (S_{Detchd}) = 0.92

→ $S_{Builtin} = [2p \quad 0N \quad 1A]$

entropy($S_{Builtin}$) = $-\frac{2}{3}\log\frac{2}{3} - \frac{0}{3}\log\frac{0}{3} - \frac{1}{3}\log\frac{1}{3}$

entropy ($S_{Builtin}$) = 0.92

Gain (S, Garagetype) = entropy (S) − $\frac{8}{14}$ entropy(S_{Attchd})
− $\frac{3}{14}$ entropy(S_{Detchd}) − $\frac{3}{14}$ entropy($S_{Builtin}$)

Gain (S, Garagetype) = 1.38 − $\frac{8}{14}$ (0.81) − $\frac{3}{14}$ (0.92) − $\frac{3}{14}$ (0.9

Gain (S, Garagetype) = 1.38 − 0.46 − 0.2 − 0.2

Gain (S, Garagetype) = 0.52

Attribute Mass

→ $S_{Brkface}$ = [2P 3N 1A]

entropy ($S_{Brkface}$) = $-\frac{2}{6}\log\frac{2}{6} - \frac{3}{6}\log\frac{3}{6} - \frac{1}{6}\log\frac{1}{6}$

entropy ($S_{Brkface}$) = 1.46

→ S_{stone} = [1P 3N 1A]

entropy(S_{stone}) = $-\frac{1}{5}\log\frac{1}{5} - \frac{3}{5}\log\frac{3}{5} - \frac{1}{5}\log\frac{1}{5}$

entropy(S_{stone}) = 1.37

→ S_{tile} = [1P 2N 0A]

entropy(S_{tile}) = $-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} - \frac{0}{3}\log\frac{0}{3}$

entropy (S_{tile}) = 0.92

$Gain(S, Mass) = entropy(S) - \frac{6}{14} entropy(S_{Bikface}) - \frac{5}{14}(entropy(S_{small}) - \frac{3}{14} entropy(S_{tile})$

$Gain(S, Mass) = 1.38 - \frac{6}{14}(1.46) - \frac{5}{14}(1.37) - \frac{3}{14}(0.92)$

$Gain(S, Mass) = 1.38 - 0.63 - 0.5 - 0.2$

$\boxed{Gain(S, Mass) = 0.05}$

Attribute type I.

→ $S_G = [3P \ 6N \ 0_A]$

$entropy(S_G) = -\frac{3}{9}\log\frac{3}{9} - \frac{6}{9}\log\frac{6}{9} - \frac{0}{9}\log\frac{0}{9}$

$entropy(S_G) = 0.92$

→ $S_A = [1P \ 2N \ 2A]$

$entropy(S_A) = -\frac{1}{5}\log\frac{1}{5} - \frac{2}{5}\log\frac{2}{5} - \frac{2}{5}\log\frac{2}{5}$

$entropy(S_A) = 1.52$

$Gain(S, Type-I) = entropy(S) - \frac{9}{14} entropy(S_G) - \frac{5}{14} entropy(S_A)$

$Gain(S, Type-I) = 1.38 - \frac{9}{14}(0.92) - \frac{5}{14}(1.52)$

$Gain(S, Type-I) = 1.38 - 0.6 - 0.54$

$\boxed{Gain(S, TypeI = 0.24}$

Gain (S, Exposure) = 0.31

Gain (S, Garage Type) = 0.32 ✓

Gain (S, Mass) = 0.05

Gain (S, TypeI) = 0.24

Exposure	Brief Mass	type I	Sale condition
No	Brkface	G	Partial
Gd	Stone	A	Normal
Gd	Brkface	G	Normal
AV	Brkface	G	Normal
AV	Stone	G	Normal
Gd	Stone	A	Normal
No	Brkface	G	Normal
AV	Stone	A	Partial

Attribute Exposure

→ $S = [2_P \ 6_N \ 0_A]$

$$\text{entropy}(S) = -\frac{2}{8}\log\frac{2}{8} - \frac{6}{8}\log\frac{6}{8} - \frac{0}{8}\log\frac{0}{8}$$

entropy(S) = 0.81

→ $S_{No} = [1_P \ 1_N \ 0_A]$

$$\text{Entropy}(S_{No}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} - \frac{0}{2}\log\frac{0}{2}$$

entropy(S_{No}) = 1

→ $S_{GD} = [0P \quad 3N \quad 0A]$

entropy$(S_{GD}) = 0$

→ $S_{AV} = [1P \quad 2N \quad 0A]$

entropy $(S_{AV}) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} - \frac{0}{3}\log\frac{0}{3}$

entropy $(S_{AV}) = 0.92$

Gain $(S, \text{Exposure}) =$ entropy$(S) - \frac{2}{8}$ entropy$(S_{No}) - \frac{3}{8}$ entropy$(S_{GD}) - \frac{3}{8}$ entropy(S_{AV})

Gain $(S, \text{Exposure}) = 0.81 - \frac{2}{8}(1) - \frac{3}{8}(0) - \frac{3}{8}(0.92)$

Gain $(S, \text{Exposure}) = 0.81 - 0.25 - 0.345$

Gain $(S, \text{Exposure}) = 0.215.$

Attribute Mass

→ $S_{Brkface} = [1P \quad 3N \quad 0A]$

entropy$(S_{Brkface}) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} - \frac{0}{4}\log\frac{0}{4}$

entropy$(S_{Brkface}) = 0.81$

→ $S_{stone} = [1P \quad 3N \quad 0A]$

entropy$(S_{stone}) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{1} - \frac{0}{4}\log\frac{0}{4}$

entropy$(S_{stone}) = 0.81$

Gain (S, Mass) = entropy (S) $-\frac{4}{8}$ entropy ($S_{Big/Apple}$) $-\frac{4}{8}$ entropy (S_{Stone})

Gain (S, Mass) = $0.81 - \frac{4}{8}(0.81) - \frac{4}{8}(0.81)$

Gain (S, Mass) = $0.81 - 0.405 - 0.405$

Gain (S, Mass) = 0

Attribute type-I

→ $S_G = [1P\ 4N\ 0_A]$

entropy (S_G) = $-\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5} - \frac{0}{5}\log\frac{0}{5}$

entropy (S_G) = 0.72

→ $S_A = [1P\ 2N\ 0_A]$

entropy (S_A) = $-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$

entropy (S_A) = 0.92

Gain (S_1 typeI) = entropy (S) $-\frac{5}{8}(0.72) - \frac{3}{8}(0.92$

Gain (S, typeI = 0.81 − 0.45 − 0.345

Gain (S, TypeI) = 0.015

Gain (S, Exposure) = 0.215

Gain (S, Mass) = 0

Gain (S, type1) = 0.015

Mass	type1	Sale condition
Bikface	G	Partial
Bikeface	G	Normal

$S = [1P\ 1N\ 0A]$

entropy (S) = 1

Attribute Mass.

→ $S_{Bikface} = [1P\ 1N\ 0A]$ Gain = 1−1 = 0

entropy(S) 1

Attribute type1

→ $S_{G} = [1P\ 1N\ 0A]$

entropy (S_G) = 1

Gain = 1−1 = 0

~~Gain = 1 − 0 − 1/2 − 1/2~~

~~No~~ Bikface. store A

No tive G

No tive G

Mass	type]	sale condition
BrkFace	G	Normal
Stone	G	Normal
Stone	A	partial

$S = [1P\ 2N\ 0A]$

entopy = 0.92

Attribute Mass

$S_{BrkFace} = (0P\ 1N\ 0A]$

entopy = 0

$S_{Stone} = [1P\ 1N\ 0A]$

$S_{Stone} = 1$

$Gain = 0.92 - \frac{1}{3}(0) - \frac{2}{3}(1)$

Gain = 0.67

Attribute type].

$S_G = [0P\ 2A\ 0A)$

entopy = 0

$S_A = (1P\ 0N\ 0A]$

entopy = 0

Gain = 0.92

Exposure	Mass	types	Sale Condition
No	Stone	A	Abnormal
No	tile	G	Normal
No	tile	G	Abnormal

$S = [Op\ 2N\ 1A]$

entropy(S) = 0.92

Attribute Exposure.

$S_{No} = [Op\ 2N\ 1A]$

entropy = 0.92

Gain = 0.92 − 0.92 = 0

Attribute. Mass

$S_{stone} = [Op\ ON\ 1A]$

entropy(S_{stone}) = 0

$S_{tile} = [0\ 2N\ 0A]$

entropy(S_{tile}) = 0

Gain = 0.92

Attribute. type.

$S_A = [Op\ ON\ 1A]$ entropy = 0

$S_G = [Op\ 2N\ 0A]$ entropy = 0.

Gain = 0.92.

Exposure	Mass	Type	Condition
Av	Tile	G	Partial
No	Surface	A	Absorbed
No	Surface	G	Partial

$S = [2P\ 0N\ 1A]$

entropy = 0.92

Attribute Exposure:

→ $S_{AV} = [1P\ 0N\ 0A]$

entropy $S(AV) = 0$

→ $S_{NO} = [1P\ 0N\ 1A]$

entropy$(S_{NO}) = 1$

$$\text{Gain} = 0.92 - \frac{2}{3}(1)$$

Gain = 0.25

Attribute Mass.

Gain = 0.25

Attribute Type.

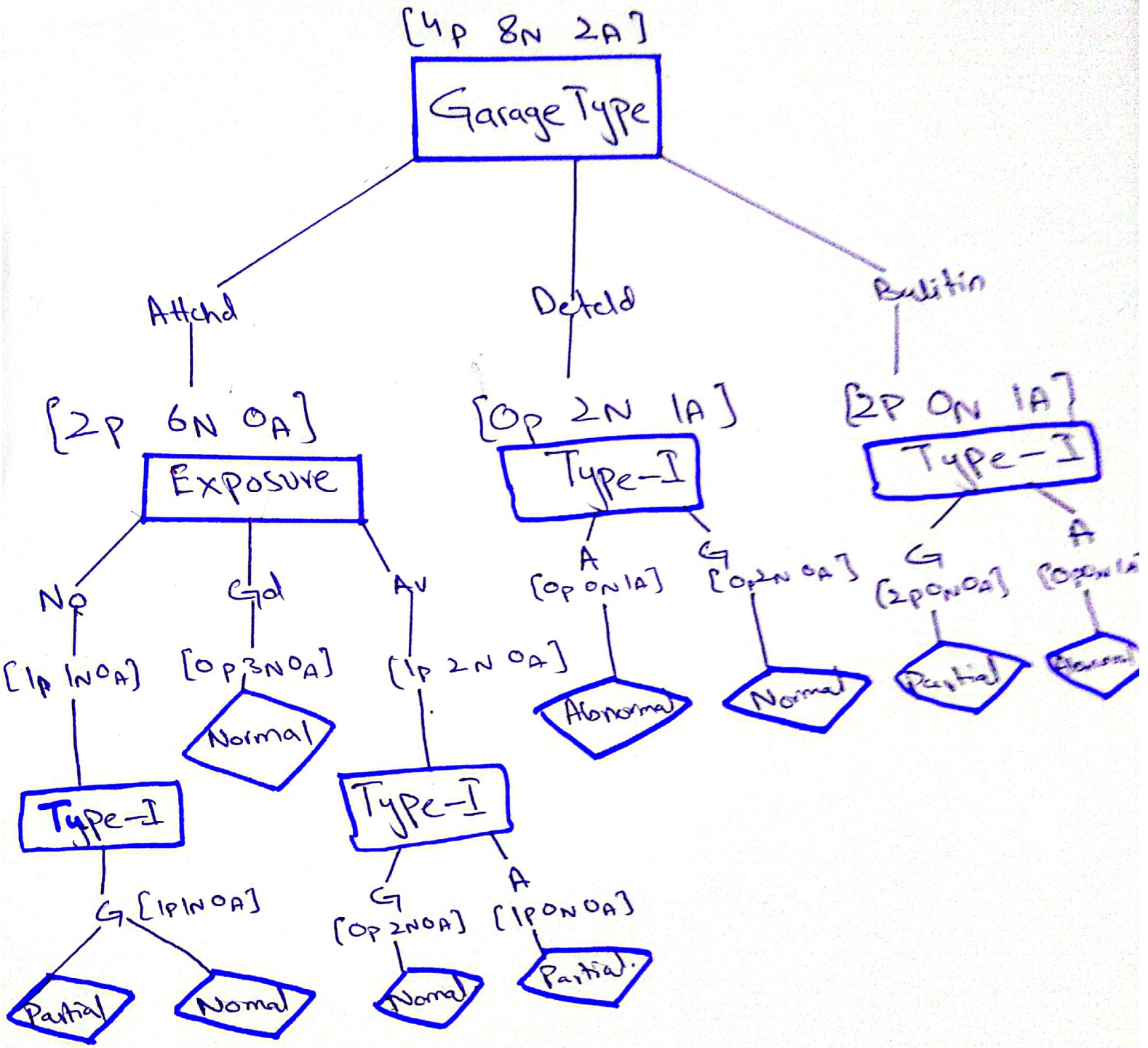
$S_G = [2P\ 0N\ 0A]$

entropy $S_G = 0$

$S_A = [0P\ 0N\ 1A]$

entropy $S_A = 0$

Gain = 0.92 ✓



Test set:

Exposure	Garage type	Mass	type-I	Condition
Gd	detched Tq.	Tile	A	Abnormal