ISE 2 Component 1 Problem Solving

Good Luck | Page No. | Date

1. Build a decision tree using ID3 algorithm for the given training data in the table and predict the class of the following new example:
age <=30, income = medium, student = yes, credit-rating = fair.

age	income	student	Credit Rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31-40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31-40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31-40	medium	no	excellent	yes

Entropy for entire dataset

$S = [8+, 4-]$ Entropy $(S) = \frac{-8}{12}\log_2\frac{8}{12} - \frac{4}{12}\log_2\frac{4}{12}$

$= 0.918$

Attribute 1 - age
Values (age) = <=30, 31-40, >40

$S_{<=30} = [2+, 3-]$ Entropy $[S_{<=30}] = \frac{-2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$

$= 0.970$

$S_{31-40} = [3+, 0-]$ Entropy $[S_{31-40}] = \frac{-3}{3}\log_2\frac{3}{3} + 0$

$= 0$

GoodLuck | Page No.
Date

$S_{>40} = [3+, 1-]$ Entropy $[S_{>40}] = \frac{-3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$

$= 0.811$

Information Gain (S, age)

$= \text{Entropy}(S) - \sum_{v \in (<=30, 31\text{-}40, >40)} \frac{|S_v|}{S} \text{Entropy}(S_v)$

$= 0.918 - \left[\frac{5}{12} \times 0.970 + \frac{3}{12} \times 0 + \frac{4}{12} \times 0.811\right]$

$= 0.2435$

Attribute 2 - income

values (income) = high, medium, low

$S_{high} = [1+, 2-]$ Entropy $[S_{high}] = \frac{-1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

$= 0.918$

$S_{medium} = [4+, 1-]$ Entropy $[S_{medium}] = \frac{-4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5}$

$= 0.721$

$S_{low} = [3+, 1-]$ Entropy $[S_{low}] = \frac{-3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$

$= 0.811$

Information Gain (S, income)

$= 0.918 - \left[\frac{3}{12} \times 0.918 + \frac{5}{12} \times 0.721 + \frac{4}{12} \times 0.811\right]$

$= 0.117$

Attribute 3 - student

values (student) = yes, no

$S_{yes} = [5+, 1-]$ Entropy $[S_{yes}] = \frac{-5}{6}\log_2\frac{5}{6} - \frac{1}{6}\log_2\frac{1}{6}$

$= 0.65$

GoodLuck | Page No.
Date

S_{NO} = [3+, 3-] Entropy $[S_{NO}] = \frac{-3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6}$

$= 1$

Information Gain (S, student)

$= 0.918 - \left[\frac{6}{12} \times 0.65 + \frac{6}{12} \times 1\right]$

$= 0.093$

Attribute 4 - Credit Rating
Values (Fair, Excellent)

S_{Fair} = [5+, 2-] Entropy $[S_{Fair}] = \frac{-5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7}$

$= 0.863$

$S_{Excellent}$ = [3+, 2-] Entropy $[S_{Excellent}] = \frac{-3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$

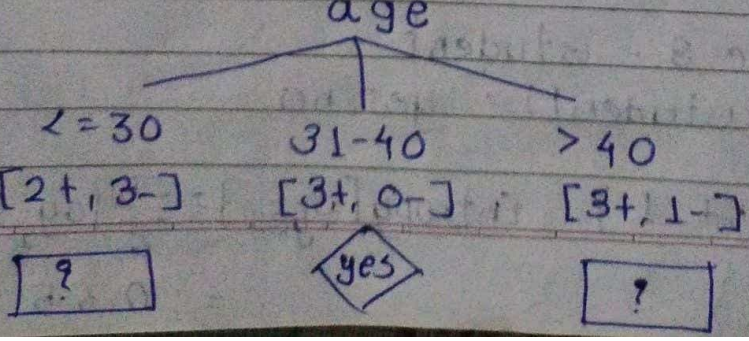
$= 0.97$

Information Gain (S, Credit Rating)

$= 0.918 - \left[\frac{7}{12} \times 0.863 + \frac{5}{12} \times 0.97\right]$

$= 0.0104$

IG(S, age) = 0.2435 ⇒ maximum
IG(S, income) = 0.117
IG(S, student) = 0.093
IG(S, CreditRating) = 0.0104



GoodLuck | Page No.
Date

Only consider <=30 rows

income	student	credit rating	Buys comp
high	no	fair	no
high	no	excellent	no
medium	no	fair	no
low	yes	fair	yes
medium	yes	excellent	yes

Attribute 1- income

$S_{high} := [0+, 2+]$ Entropy $[S_{income}] = 0 - \frac{2}{2}\log_2\frac{2}{2}$

$= 0$

$S_{medium} = [1+, 1-]$ Entropy $[S_{medium}] = \frac{-1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$

$= 1$

$S_{low} = [1+, 0-]$ Entropy $[S_{low}] = \frac{-1}{1}\log_1\frac{1}{1} - 0$

$= 0$

Info. Gain (S, income)

$= 0.97 - \left[\frac{2}{5} \times 0 + \frac{2}{5} \times 1 + \frac{1}{5} \times 0\right]$

$= 0.57$

Attribute 2 - student

$S_{no} = [0+, 3-]$ Entropy $[S_{no}] = 0 - \frac{3}{3}\log_2\frac{3}{3}$

$= 0$

$S_{yes} = [2+, 0-]$ Entropy $[S_{yes}] = \frac{-2}{2}\log\frac{2}{2} - 0$

$= 0$

Info. Gain (S, student

$= 0.97 - \left[\frac{3}{5} \times 0 + \frac{2}{5} \times 0\right] = 0.97$

Attribut 3 - Credit Rating

$S_{fair} = [1+, 2-]$ Entropy $[S_{fair}] = \frac{-1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

$= 0.918$

$S_{Excellent}$ [1+, 1-] Entropy $[S_{excellent}] = \frac{-1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2$

$= 1$

Info Gain (S, Credit ~~credi~~Rating)

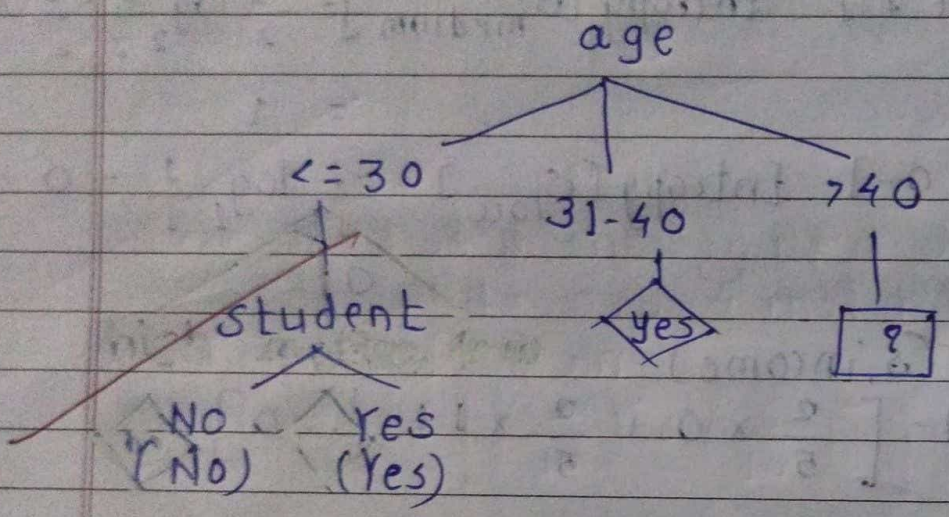
$= 0.97 - \left[\frac{3}{5} \times 0.918 - \frac{2}{5} \times 1\right]$

$= 0.0192$

IG(S, income) = 0.57

IG(S, student) = 0.97 ⇒ maximum

IG(S, creditrating) = 0.0192



Now consider only >40 rows

income	student	credit Rating	Buys comp.
medium	no	fair	yes
low	yes	fair	yes
low	yes	excellent	no
medium	yes	fair	yes

Attribute 1 - income

S_{medium} = [2+, 0-] Entropy $[S_{medium}] = 0$

S_{low} = [1+, 1-] Entropy $[S_{low}] = 1$

IG (S, income) = $0.811 - \left[\frac{2}{4} \times 1\right] = 0.311$

Attribute 2 - student

$S_{no} = [1+, 0-]$ Entropy $[S_{no}] = 0$

$S_{yes} = [2+, 1-]$ Entropy $[S_{yes}] = \frac{-2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$

$= 0.918$

$IG(S, student) = 0.811 - \left[\frac{3}{4} \times 0.918\right] = 0.1225$

Attribute 3 - credit scoring

$S_{fair} = [3+, 0-]$ Entropy $[S_{fair}] = 0$

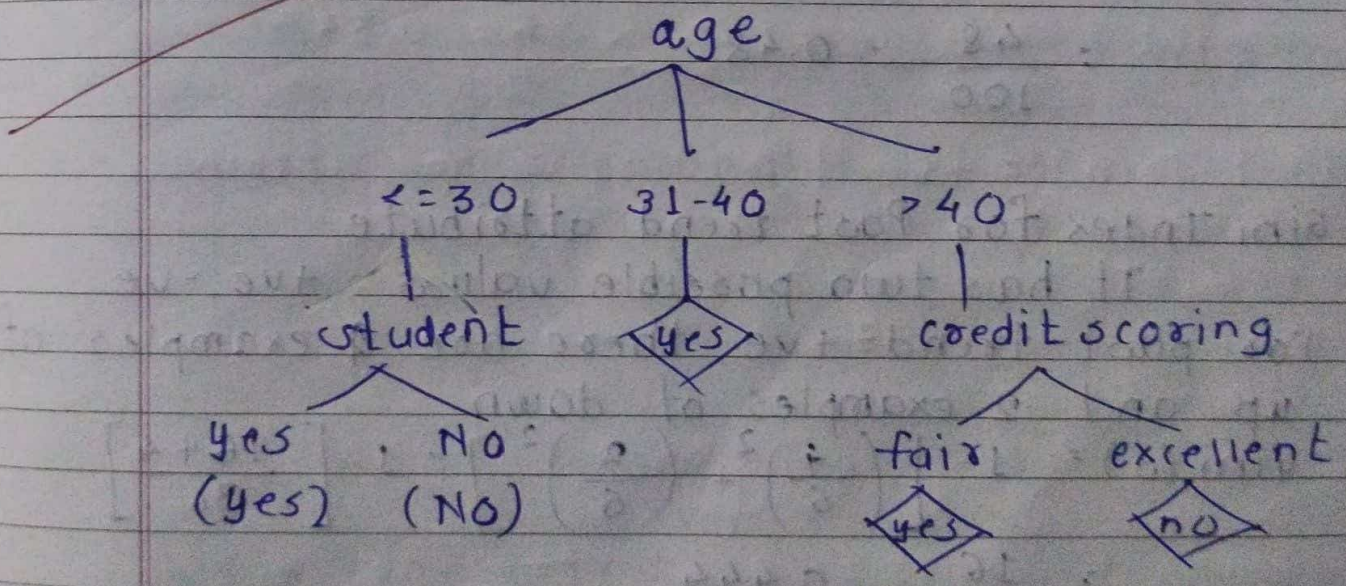
$S_{excellent} = [0+, 1-]$ Entropy $[S_{excellent}] = 0$

$IG(S, credit\ scoring) = 0.811 - [0] = 0.811$

$IG(S, income) = 0.311$

$IG(S, student) = 0.1225$

$IG(S, credit\ scoring) = 0.811 \Rightarrow$ maximum



for age <= 30, income = medium, student = yes
credit rating = fair
$\Rightarrow$
B - Computer = yes

GoodLuck | Page No.
Date

2. Build a decision tree using CART algorithm for the given training data in the table and predict the class of the following new example:
Past Trend = -ve, Open interest = Low, Trading vol. = low

Past Trend	Open Interest	Trading Volume	Return
+ve	Low	High	Up
-ve	High	Low	Down
+ve	Low	High	Up
+ve	High	High	Up
-ve	Low	High	Down
+ve	Low	Low	Down
-ve	High	High	Down
-ve	Low	High	Down
+ve	Low	Low	Down
+ve	High	High	Up

There are two possible output variables - up, down
The data has 4 instances of up and 6 instances of down.

$$Gini(s) = 1 - \left[\left(\frac{4}{10}\right)^2 + \left(\frac{6}{10}\right)^2\right] = 1 - \left[\frac{16+36}{100}\right]$$

$$= \frac{48}{100} = 0.48$$

Gini Index for Past trend attribute
It has two possible values = +ve, -ve
For past trend = +ve, there are 4 examples of up and 2 examples of down.

$$Gini(s) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 1 - \left[\frac{16+4}{36}\right]$$

$$= \frac{16}{36} = 0.444$$

For past trend = +ve, there are 4 examples of down and 0 examples of up

$$Gini(s) = 1 - \left[\frac{4}{4}\right]^2 = 0$$

weighted average (Past trend) = $\frac{6}{10} \times 0.444 + \frac{4}{10} \times 0$

$= 0.2664$

Gini Index for Open Interest attribute

It has two possible values = low, high

For open interest = low, there are 2 examples of up and 4 examples of down

$Gini(s) = 1 - \left[\left(\frac{2}{6}\right)^2 + \left(\frac{4}{6}\right)^2\right] = 0.444$

For open interest = high, there are 2 examples of up and 2 examples of down

$Gini(s) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 1 - \left[\frac{4}{16} + \frac{4}{16}\right] = 0.5$

weighted average (Open Interest)

$= \frac{6}{10} \times 0.444 + \frac{4}{10} \times 0.5$

$= 0.4664$

Gini Index for Trading volume attribute

It has two possible values = high, low

For trading volume = high, there are 4 examples of up and 3 examples of down

$Gini(s) = 1 - \left[\left(\frac{4}{7}\right)^2 + \left(\frac{3}{7}\right)^2\right] = 1 - \left[\frac{16}{49} + \frac{9}{49}\right]$

$= \frac{24}{49} = 0.4897$

For trading volume = low, there are 3 examples of down and 0 examples of up

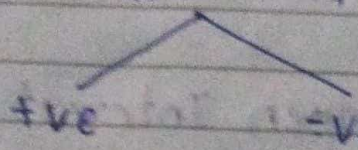
$Gini(s) = 1 - \left[\left(\frac{3}{3}\right)^2\right] = 0$

weighted average (Trading volume)

$= \frac{7}{10} \times 0.4897 + 0 = 0.3427$

GoodLuck | Page No.
Date

Past trend has minimum Gini Index



Past Trend	O.I	T.V	Return
+ve	low	High	up
+ve	low	High	up
+ve	High	High	up
+ve	low	Low	Down
+ve	Low	Low	Down
+ve	High	High	up

Past Trend	Return
-ve	Down
-ve	Down
-ve	Down
-ve	Down

Past Trend = +ve | Open Interest attribute

O.I = low

$$Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 1 - \frac{8}{16} = 0.5$$

O.I = high

$$Gini(S) = 1 - \left[\left(\frac{2}{2}\right)^2\right] = 0$$

Weighted average = $\frac{4}{6} \times 0.5 = 0.333$

Past Trend = +ve | Trading volume attribute

T.V = High

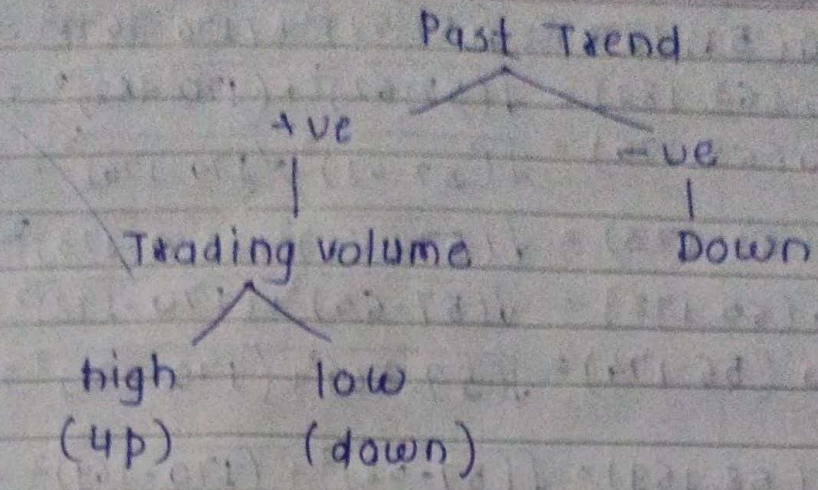
$$Gini(S) = 1 - \left[\left(\frac{4}{4}\right)^2\right] = 0$$

T.V = low

$$Gin(S) = 1 - \left[\left(\frac{2}{2}\right)^2\right] = 0$$

Weighted average = 0

Trading volume has minimum attribute



For Past Trend = -ve ; Open Interest = low, Trading volume = low

8. Consider a dataset that contains two variables: height (cm) & weight (kg). Each point is classified as normal or underweight using the KNN algorithm. New data point (x_1, y_1) and we need to determine its class

57kg	170 cm	?

Weight(x_2)	Height(y_2)	class
51	167	underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	underweight
58	169	Normal
57	173	Normal
55	170	Normal

Given condition : weight: 57kg height: 170 cm
Calculate the euclidean distance between each instance and new instance.

Distance = $\sqrt{(x-a)^2+(y-b)^2}$

GoodLuck | Page No. | Date

Distance to (51, 167) = $\sqrt{(57-51)^2 + (170-167)^2}$ = 6.7082

Distance to (62, 182) = $\sqrt{(57-62)^2 + (170-182)^2}$ = 13

Distance to (69, 176) = $\sqrt{(57-69)^2 + (170-176)^2}$ = 13.4164

Distance to (64, 173) = $\sqrt{(57-64)^2 + (170-173)^2}$ = 7.6157

Distance to (65, 172) = $\sqrt{(57-65)^2 + (170-172)^2}$ = 8.2462

Distance to (56, 174) = $\sqrt{(57-56)^2 + (170-174)^2}$ = 4.1231

Distance to (58, 169) = $\sqrt{(57-58)^2 + (170-169)^2}$ = 1.4142

Distance to (57, 173) = $\sqrt{(57-57)^2 + (170-173)^2}$ = 3

Distance to (55, 170) = $\sqrt{(57-55)^2 + (170-170)^2}$ = 2

Weight	Height	distance	Rank	class
51	167	6.7082	5	Underweight
62	182	13	8	Normal
69	176	13.4164	9	Normal
64	173	7.6157	6	Normal
65	172	8.2462	7	Normal
56	174	4.1231	4	Underweight
58	169	1.4142	1	Normal
57	173	3	3	Normal
55	170	2	2	Normal

If k=1 ⇒ Normal
If k=2 ⇒ Normal
If k=3 ⇒ Normal
If k=4 ⇒ Normal
If k=5 ⇒ Normal

4. Consider the car theft problem with attributes color, type, origin, and the target, stolen can be either Yes or No. Based on the dataset you need to classify the following new car (Red SUV domestic) is getting stolen or not using Naive Bayes algorithm.

Example NO	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	NO
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	NO
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	NO
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	NO
10	Red	Sports	Imported	Yes

Prior Probability

$P(\text{stolen} = \text{Yes}) = 5/10 = 0.5$

$P(\text{stolen} = \text{NO}) = 5/10 = 0.5$

Conditional / current Probability of each attribute

color	Yes	NO
Red	3/5	2/5
Yellow	2/5	3/5

Type	Yes	NO
Sports	4/5	2/5
SUV	1/5	3/5

Origin	Yes	NO
Domestic	2/5	3/5
Imported	3/5	2/5

$P(\text{Yes} \mid \text{New Instance}) = P(\text{Yes}) * P(\text{color} = \text{Red} \mid \text{Yes}) * P(\text{Type} = \text{SUV} \mid \text{Yes}) * P(\text{Origin} = \text{Domestic} \mid \text{Yes})$

$= 0.5 \times \frac{3}{5} \times \frac{1}{5} \times \frac{2}{5}$

$= 0.024$

$P(\text{No} \mid \text{New Instance}) = P(\text{No}) * P(\text{color} = \text{Red} \mid \text{No}) * P(\text{Tye} = \text{SUV} \mid \text{No}) * P(\text{Origin} = \text{Domestic} \mid \text{No})$

$= 0.5 \times \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} = 0.072$

GoodLuck | Page No.
Date

P(Yes | New Instance) < P(No | New Instance)

The new car is not getting stolen.

SP
31/10/23