## Decission Tree.

classification

$$\text{Entropy}(S) = \sum_{i=1}^{n} -P_i \log_2 P_i$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum \text{Entropy}(S, A)$$

$\sum$ Prob $\times$ Entropy

focus → higher Information Gain = more entropy remove

Partition -করার আগের এনট্রপি → Target variable.

Partion " পরে এনট্রপি →

than কমবেশি

Information Gain এর বেশি → সেটা root node.

| Gender | Car | travel Cost | income | transport |
|---|---|---|---|---|
| male | 0 | cheap | low | bus |
| male | 1 | cheap | medium | bus |
| female | 1 | cheap | medium | train |
| female | 0 | cheap | low | bus |
| male | 1 | cheap | medium | bus |
| male | 0 | standard | medium | train |
| female | 1 | standard | medium | train |
| female | 1 | expensive | high | car |
| male | 2 | expensive | medium | car |
| female | 2 | expensive | high | car. |

class variable

Entropy Before Partition

$$E_{(T)}(S) = -\left(\frac{4}{10}\log_2\frac{4}{10} + \frac{3}{10}\log\frac{3}{10} + \frac{3}{10}\log\frac{3}{10}\right)$$

$$= -(-.528 - 0.521 - 0.521)$$

$$= 1.57.$$

Now calculate entropy for each attribute:

Gender

$E(\text{Gender\_male}) = -\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{1}{5}\log_2\frac{1}{5} + \frac{1}{5}\log_2\frac{1}{5}\right)$

total male = 5

$= 1.37$

$E(\text{Gende\_female}) = -\left(+\frac{2}{5}\log_2\frac{2}{5} + \frac{1}{5}\log_2\frac{1}{5} + \frac{1}{5}\log_2\frac{1}{5}\right)$

total female = 5

$= 1.52$

$\text{Information Gain(Gender)} = 1.57 - \left(\frac{5}{10}\times 1.37 + \frac{5}{10}\times 1.52\right)$

$= 0.125$

- - - - - - - - - -

- - - - - - - - - -

- - - - - - - - - -

- - - - - - - - - -

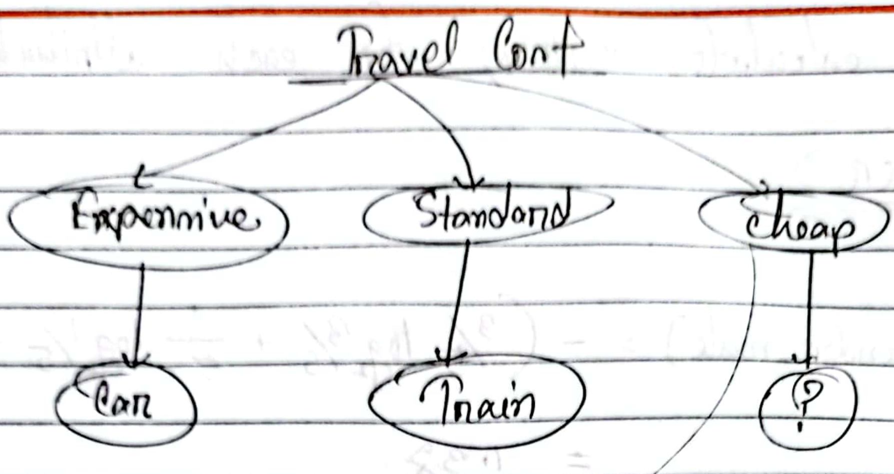
- - - - - - - - - -

Information Gain (Car) = 0.537

Information Gain (Travel Cost) = 1.21 → highest gain

Information Gain (Income) = 0.695

Root node = Travel cost.




Bus/train

এটার জন্য আবার IG বের করতে হবে। যার IG বেশি সেটা Root হবে।

only for cheap

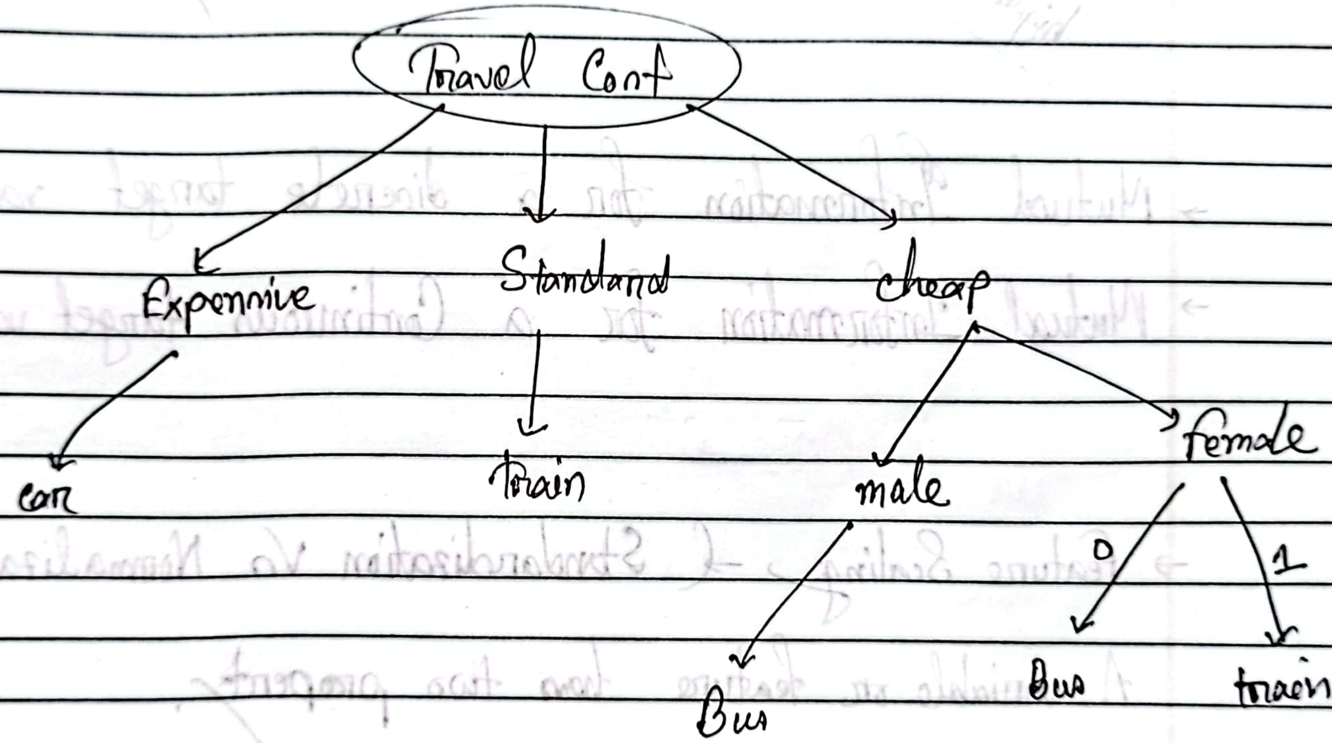
| Gender | Car | Income | transport |
|---|---|---|---|
| male | 0 | low | bus |
| male | 1 | medium | bus |
| female | 1 | medium | train |
| female | 0 | low | Bus |
| Male | 1 | medium | Bus |

According to apply Entropy and IG law.

Gender (IG) → 0.322 → high.

Car (IG) → 0.120

Income (IG) → 0.120




final tree

Entropy value Range

0 - 1

## Mutual Information

→ two random non-negative values of variable.

Feature variable    target variable

higher dependency    or

measure the dependancy

independent

→ Mutual Information for a discrete target variable.

→ Mutual Information for a Continuous target variable.

→ Feature Scaling → (Standardization Vs Normalization)

A variable or feature has two property.

Unit and Magnitude.

value.

hight 175 cm

ft

m

25 years

value unit

or

magnitude.

→ Normalization help's you to scale your data 0 to 1.

Standardization help's you to scale your feature based on Standard Normal Distribution.

mean = 0
σ median = 1

→ library → standardscaler. (Standardization) — z-score.
→ minmaxscaler. (normalization)

"when we use what"?

check also
Drill

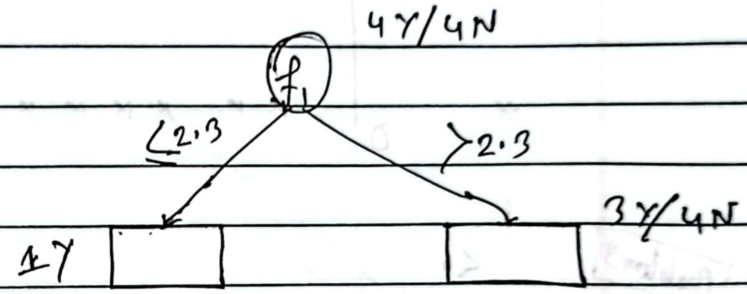
→ Decision tree Split for Numerical Variable.

| $f_1$ | o/p |
|---|---|
| 2.3 | Y |
| 3.6 | Y |
| 4 | N |
| 5.2 | N |
| 6.7 | Y |
| 8.9 | N |
| 10.5 | Y |
| 14.2 | N |

① Sorting All the value.

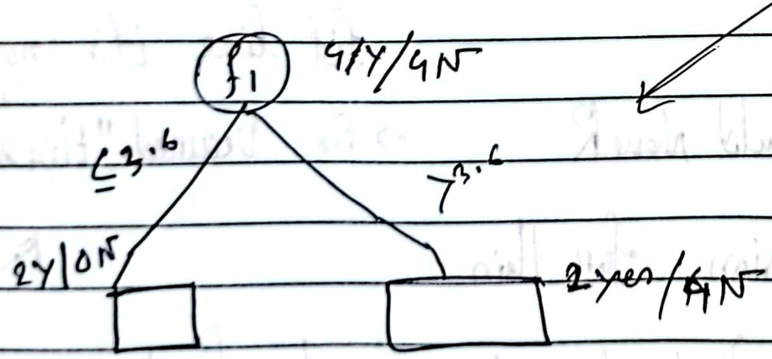
② Consider threshold value.

Lets TH = 2.3 $x_i \le 2.3$




→ Next th = 3.6

$x_1 \le 3.6$




Compare ① Entropy ② Information gain

Disadvantage

High time Complexity

→ Decision tree classification          Entropy [0-1]

→ Decision tree Regression

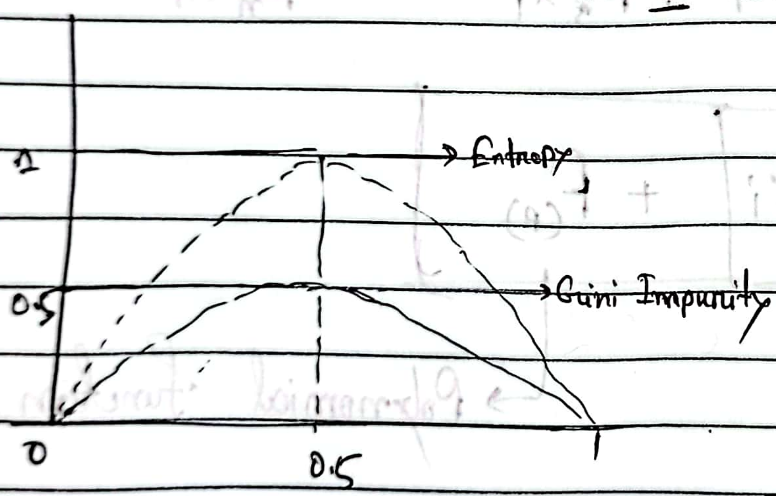
→ Highest Info. Gain and Small Entropy. [Root node selection]

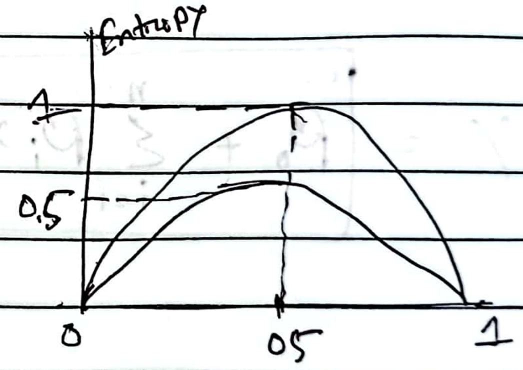
→ Information Gain.

→ Gini Impurity.

$$G.I = 1 - \sum_{i=1}^{n} (P_i)^2$$

$$= 1 - [P_{+ve}^2 + P_{-ve}^2]$$




Computationally efficient.

Easy to Implement.