

09/11/2024

## Assignment - 2

### Gini Index:-

CGM	Interactive	Practical knowledge	Comm Skills	Yes/No
>9	Yes	v. good	Good	Yes
>8	No	Good	Moderate	Yes
>9	No	Average	Poor	No
<8	No	Average	Good	No
>8	Yes	Good	Moderate	Yes
>9	Yes	Good	Moderate	Yes
<8	Yes	Good	Poor	No
>9	No	v. good	Good	Yes
>8	Yes	Average	Good	Yes
<8	Yes	Good	Poor	No

Impurity a division decision trees:-

Lower gini index value gives the

Pure dataset subset (a) choosing the best split.

$$Gini(D) = 1 - \sum_{i=1}^c P_i^2$$



Step 1:-

Attribute 1 (CAPA)

Ans	Parameters	Yes	No	Total
CAPA	$\geq 9$	3	1	4
	$\geq 8$	3	0	3
	$< 8$	0	3	3

\* calculating the each gini for each parameter

\* sum all the parameter for each Attribute

$$\text{Gini (Parameter Cond)} = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2$$

$$\text{Gini I (CAPA } \geq 9) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$$

$$= 1 - \frac{9}{16} - \frac{1}{16}$$

$$= 1 - 0.5625 - 0.0625$$

$$= 1 - 0.625$$

$$= 0.375$$

$$\text{Gini I (CAPA } \geq 8) = 1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2$$

$$= 0$$

$$\text{Gini I (CAPA } < 8) = 1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2$$

$$= 0$$

\* weighted sum of Gini Index for CAPA for the row

$$= \left(\frac{4}{10}\right) * 0.375 + 0 + 0$$

$$= 0.4 * 0.375$$

$$= 0.15$$

Step 2:-

Attribute 2 (Interactive)

Attribute	Parameter	yes	no	job offer
Interactive	yes	4/6	2/6	6
	no	2/4	2/4	4

$$\begin{aligned}
 \text{GiniF}(\text{Interactive} = \text{yes}) &= 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 \\
 &= 1 - \frac{16}{36} - \frac{4}{36} \\
 &= 1 - 0.444 - 0.111 \\
 &= 1 - 0.55 \\
 &= 0.45
 \end{aligned}$$

$$\begin{aligned}
 \text{GiniF}(\text{Interactive} = \text{no}) &= 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 \\
 &= 1 - \frac{4}{16} - \frac{4}{16} \\
 &= 1 - 0.25 - 0.25 \\
 &= 1 - 0.5 \\
 &= 0.5
 \end{aligned}$$

\* weighted sum of Interactive

$$= \frac{6}{10} * 0.45 + \frac{4}{10} * 0.5$$

$$= 0.6 * 0.45 + 0.4 * 0.5$$

$$= 0.27 + 0.2$$

$$= 0.47$$



step 3 :-

Attribute 3 (Practical knowledge)

Attribute	Parameter	Yes	No	Sub total
Practical knowledge	very good	2	0	2
	good	3	2	5
	average	1	2	3

$$\text{Gini}(P.K = \text{very good}) = 1 - \left(\frac{2}{2}\right)^2 - 0$$

$$= 0$$

$$\text{Gini}(P.K = \text{good}) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$$

$$= 1 - \frac{9}{25} - \frac{4}{25}$$

$$= 1 - 0.36 - 0.16$$

$$= 1 - 0.52$$

$$= 0.48$$

$$\text{Gini}(P.K = \text{average}) = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$$

$$= 1 - \frac{1}{9} - \frac{4}{9}$$

$$= 1 - 0.111 - 0.445$$

$$= 1 - 0.555$$

$$= 0.445$$

\*weighted sum of Practical knowledge

$$= \frac{2}{10} * 0 + 0.48 * \frac{5}{10} + 0.445 * \frac{3}{10}$$

$$= 0 + 0.24 + 0.1335$$

$$= 0.3735$$

Step 4:-  
Attribute 4 (comm skills)

Attributes	Parameters	Yes	No	Total
comm skills	Good	$\frac{3}{4}$	$\frac{1}{4}$	4
	Moderate	$\frac{3}{3}$	0	3
	Poor	0	$\frac{3}{3}$	3

$$\begin{aligned}
 \text{Gini}_I(\text{CS} = \text{Good}) &= 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 \\
 &= 1 - 0.5625 - 0.0625 \\
 &= 1 - 0.625 \\
 &= 0.375
 \end{aligned}$$

$$\begin{aligned}
 \text{Gini}_I(\text{CS} = \text{moderate}) &= 1 - \left(\frac{3}{3}\right)^2 - 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{Gini}_I(\text{CS} = \text{Poor}) &= 1 - 0 - \left(\frac{3}{3}\right)^2 \\
 &= 0
 \end{aligned}$$

\* weighted sum of comm skills

$$= \frac{4}{10} * 0.375 + 0 + 0$$

$$= 0.15$$

Step 5:- overall Gini index

$$\begin{aligned}
 \text{Gini}_I(D) &= 1 - \left(\frac{4}{10}\right)^2 + \left(\frac{6}{10}\right)^2 \\
 &= 1 - (0.16 + 0.36) \\
 &= 1 - 0.52 \\
 &= 0.48
 \end{aligned}$$



Step 6 :-

→ overall gini index = 0.48

→ CAPA = 0.15

→ Interactive = 0.57

→ Practical knowledge = 0.37

→ comm skills = 0.15

superior

capa

unsuperior

k means

by selecting minimum gini index to calculate  
and among the parameters of attributes

\* CAPA → 0.375, 0.70 (choose lowest)

calculate gini :-

$$\begin{aligned} \text{gini}(CAPA) &= \text{gini}(T) - \text{gini}(T, CAPA) \\ &= 0.48 - 0 \\ &= 0.48 \end{aligned}$$

\* Interactive → 0.45, 0.5

calculate gini :-

$$\begin{aligned} \text{gini}(\text{Interactive}) &= 0.48 - \text{Interactive} \\ &= 0.48 - 0.45 \\ &= 0.03 \end{aligned}$$

\* P.K. → 0.48, 0.555

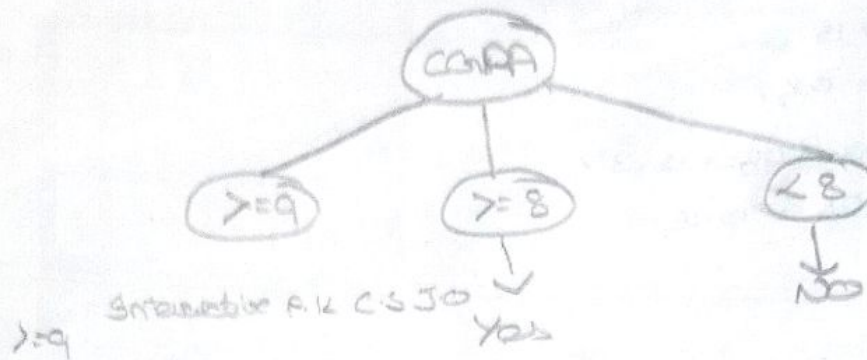
$$\begin{aligned} \text{gini}(P.K.) &= 0.48 - 0 \\ &= 0.48 \end{aligned}$$

\* C.S. → 0.375, 0.75

$$\begin{aligned} \text{gini}(C.S.) &= 0.375 - 0 \\ &= 0.375 \end{aligned}$$



# Regression tree



CGPA	Interactive	Practical knowledge	C.S	J.O
>=9	Yes	v. good	Good	Yes
>=9	No	Avg	Poor	No
>=9	Yes	Good	Med	Yes
>=9	No	v. good	Good	Yes

Attribute	Parameter	Yes	No	T
Interactive	Yes	2	0	2
	No	1	1	2

$$\text{Gini I (Yes)} = 1 - \left(\frac{2}{2}\right)^2 - 0$$

$$= 0$$

$$\text{Gini I (No)} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4}$$

$$= 1 - 0.25 - 0.25$$

$$= 0.5$$

weighted sum of interactive

$$= \frac{2}{4} * 0 + \frac{2}{4} * 0.5$$

G.I (Amount) 0.25

Attribute 2

Attribute	Parameter	yes	no	<del>F</del>
P.H	0. good	2	0	2
	Aug	0	1	1
	good	1	0	1

$$G.I (0. good) = 1 - \left(\frac{2}{2}\right)^2 = 0$$

$$= 1 - 1$$

$$= 0$$

$$G.I (Aug) = 1 - \left(\frac{1}{1}\right)^2 = 0$$

$$0 * \frac{1}{2} + 0 * \frac{1}{2} = 1 - 1 = 0$$

$$G.I (good) = 1 - \left(\frac{1}{1}\right)^2 = 0$$

$$= 0$$

weighted sum:-

$$G.I (P.H) = \frac{2}{4} * 0 + \frac{1}{4} * 0 + \frac{1}{4} * 0$$

$$= 0$$



Attributes :-

Attribute	Parameters	Yes	No	T
Common skills	Good	2	0	2
	mod	1	0	1
	Poor	0	1	1

$$G.I (Good) = 1 - \left(\frac{2}{2}\right)^2 = 0$$

$$= 0$$

$$G.I (mod) = 1 - \left(\frac{1}{1}\right)^2 = 0$$

$$= 0$$

$$G.I (Poor) = 1 - 0 - \left(\frac{1}{1}\right)^2 = 0$$

$$= 0$$

Weighted sum :-

$$G.I (CS) = \frac{2}{4} * 0 + \frac{1}{4} * 0 + \frac{1}{4} * 0$$

$$= 0$$

$$= 0$$

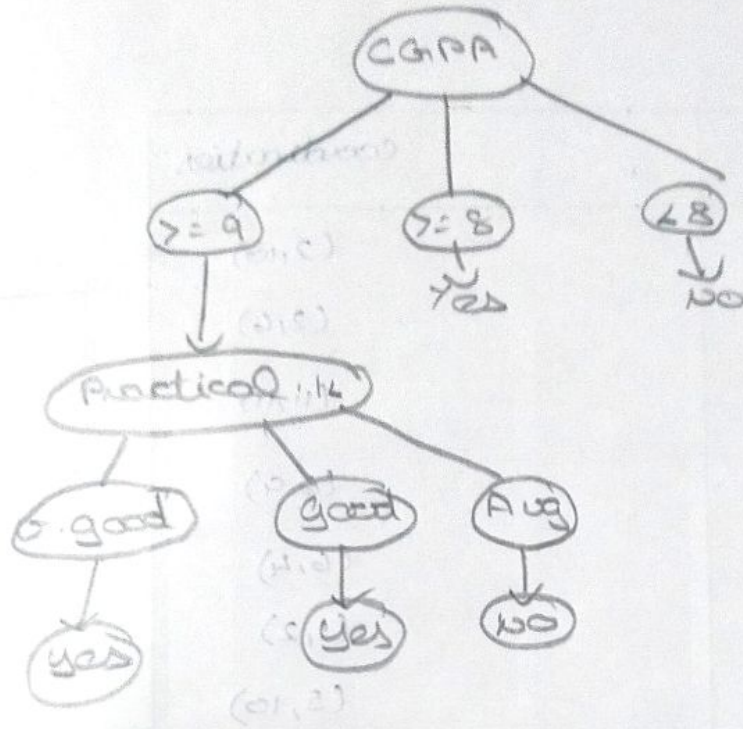
Gini Index

$$1) \text{ Interactive} = 0.25$$

$$2) P.I.K = 0 \quad \checkmark$$

$$3) C.S = 0$$

Regression tree



(0, 2)

(0, 0)

(2, 0)

starting at output

(0, 0) → 0

(0, 1) → 1