$\therefore \quad odd = \frac{P}{1-P}$.

Here we know that odd has range 0 to ∞

→ There is no upper bound for odd.

→ But odd has lower bound.

+ To remove lower bound, to have symmetrical analysis

Ex

odd = 1:6 $\quad 1/6 = 0.167$
odd = 6:1 $\quad 6/1 = 6$
$\Big\}$ not symmetrical

But $\ln(1/6) = \underline{\underline{-1.79}}$
$\ln(6/1) = \underline{\underline{1.79}}$
$\Big\}$ Symmetrical

By taking log we have overcommed the lower bound

also

log(odd)
- will not have upper bound
- will not have lower bound.

$$\underline{\underline{\log(odd)}} = y = \boxed{\underline{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}$$

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n.$$

Let $z = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$.

$$\therefore \quad \log_e\left(\frac{p}{1-p}\right) = z$$

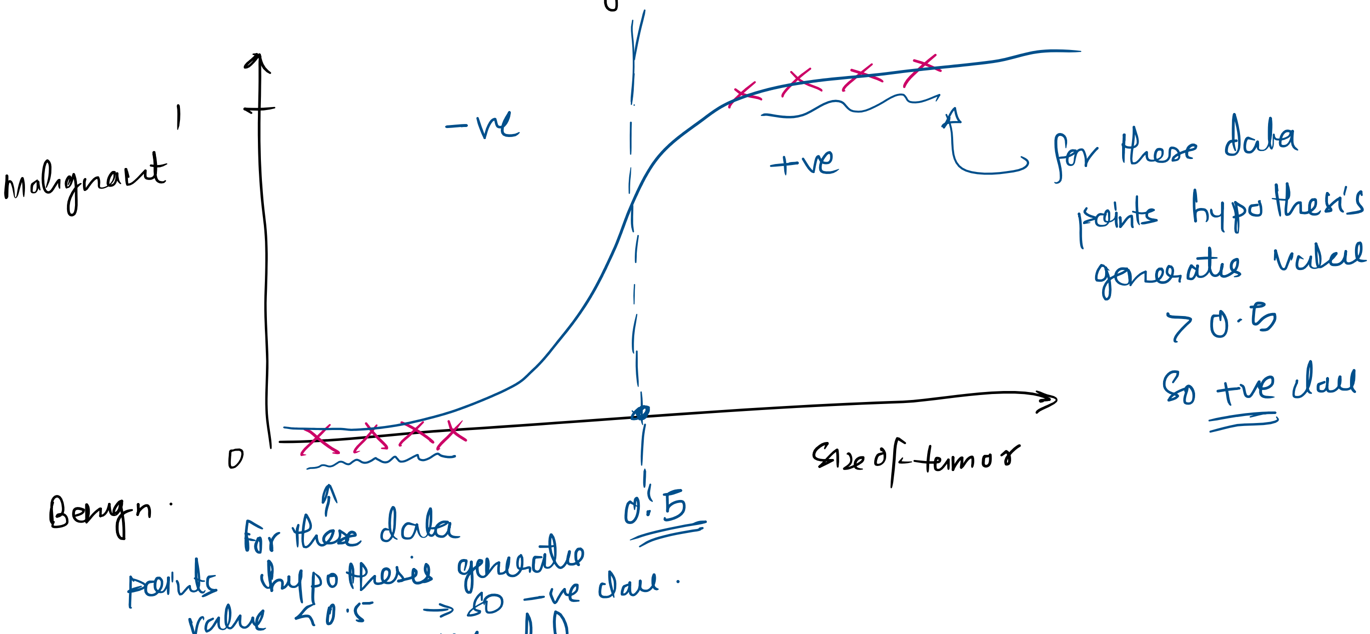
$$\frac{p}{1-p} = e^{z}$$

$$p = -e^{z}p + e^{z}$$

$$p(1+e^{z}) = e^{z}$$

$$p = \frac{e^{z}}{1+e^{z}} = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}$$

Sigmoid f^n.



Logistic Regression Model. →

In linear Reg $h_\theta(x) = \theta^T x$ → range $-\infty$ to $+\infty$

$h_\theta(x)$ ↑ hypothesis (prediction).

θ^T ↑ parameter vector

x ← feature vector.

In logistic Regression $0 \le h_\theta(x) \le 1$

↑ predicted value

$$h_\theta(x) = g(\theta^T x)$$

g ↑ Sigmoid

Sigmoid

let $z = \theta^T x$

$$\boxed{h_\theta(x) = g(z) = \frac{1}{1+e^{-z}}}$$

Estimated probability that $y=1$ on given input x.

prediction.

If $h_\theta(x) = 0.7$

this means there is 70% chance that tumor is malignant.

Since $h_\theta(x) = 0.7 > 0.5$

so $\boxed{y=1}$

If $h_\theta(x) = 0.3$

this means there is 30% chance of tumor being malignant

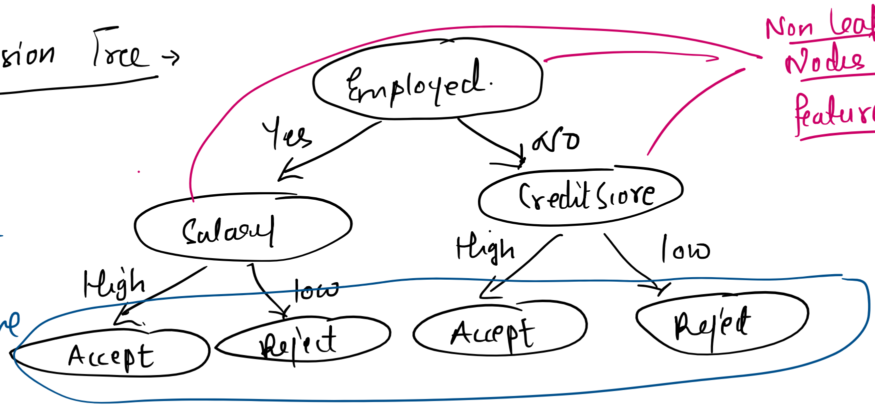
Since $h_\theta(x) = 0.3 < 0.5$

so $\boxed{y=0}$

Decision Tree :→

Consider → Bank $\xrightarrow{\text{Details}}$ Accept / Reject
Of Loan (1) (0)
Applicant
(Input)

A Sample Decision Tree →



Non leaf Nodes are features.

Outcomes are leaf Node

Here
Independent → Employed
Features → Salary
→ Credit Score

Outcome → Accept
→ Reject

Q) Create Decision Tree for following Using Gini Index
(Classification & Regression Tree) → CART ✓

Weekend	Weather	Parent	Money	Decision (y)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

P

Solution → Independent Features:
- Weather
- Parent
- Money

Decision/Outcome
- Cinema
- Tennis
- StayIn
- Shopping.

\ shopping.

step 1)

We will calculate Gini Index for Overall collection of Outcomes of Training Examples.

There are 4 possible outcomes for decision
- Cinema — 6 instances
- Tennis — 2 instances
- StayIn — 1 instance
- Shopping — 1 instance.

$$\underset{(\text{decision})}{\text{Gini}} = 1 - \left(\left(\tfrac{6}{10}\right)^2 + \left(\tfrac{2}{10}\right)^2 + \left(\tfrac{1}{10}\right)^2 + \left(\tfrac{1}{10}\right)^2\right)$$

$$= 1 - \left(\frac{42}{100}\right) = \underline{0.58}$$

[Note → In machine learning, Gini index/coefficient is utilized as an impurity measures in decision tree for classification.

$$\text{Gini} = 1 - \sum_{i=1}^{n} (P_i)^2$$ where P_i probability of outcome of specific class]

step 2: To find Gini Index for money

$\underset{(\text{money})}{\text{Gini}}$ → 2 possible value
- Rich
- poor

① $\underset{(\text{money = Rich})}{\text{Gini}}$ → 7 instance

possible Decision
- Cinema → 3 time
- Tennis → 2 time
- StayIn → 1 time
- Shopping → 1 time.

$$\underset{(\text{money = Rich})}{\text{Gini}} = 1 - \left(\left(\tfrac{3}{7}\right)^2 + \left(\tfrac{2}{7}\right)^2 + \left(\tfrac{1}{7}\right)^2 + \left(\tfrac{1}{7}\right)^2\right)$$

$$= \underline{0.694}$$

(2) Gini → 3 instance → Cinema.

(2) Gini (money = poor) → 3 instance —Decision→ Cinema.

$$= 1 - \left((3/3)^2 \right) = 0$$

Weighted Average Gini (money) = (Gini (money = Rich) * Proportion of rich) + (Gini (money = poor) * Proportion of poor)

$$= (0.694 * 7/10) + (0 * 3/10)$$

$$\boxed{\text{Gini (money)} = 0.485}$$

Step 3 Gini Index on Parent

For Parent feature —Possible Values→ Yes / NO

Gini (parent = Yes) = 5 instances —Possible Decision— Cinema

$$= 1 - \left((5/5)^2 \right) = 0$$

Gini (parent = NO) = 5 instance —possible decision→ Tennis → 2 times, Stay In → 1 times, Shopping → 1 time, Cinema → 1 time.

$$= 1 - \left((2/5)^2 + (1/5)^2 + (1/5)^2 + (1/5)^2 \right)$$

$$= 1 - (7/25) = 0.72$$

Weighted Average of Gini (parent) = $(0 * 5/10) + (0.72 * 5/10) = 0.36$

$$\boxed{\text{Gini (parent)} = 0.36}$$

$$\boxed{Gini_{(parent)} = 0.36}$$

Step 4) Gini Index for Weather.

Weather $\xrightarrow[\text{values}]{\text{Possible}}$

- Sunny = 3 instances
- Windy = 4 instances
- Rainy ⇒ 3 instances.

$Gini_{(weather = Sunny)}$ ⇒ $\xrightarrow{\hspace{2em}}$ 3 instances — possible outcomes:

- Cinema → 1 time
- Tennis ⇒ 2 time

$$= 1 - \left(\left(\tfrac{1}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 \right) = \underline{0.444}$$

$Gini_{(weather = Windy)}$ $\xrightarrow{\hspace{2em}}$ 4 instances — possible outcome:

- 3 time Cinema
- 1 time Shopping

$$= 1 - \left(\left(\tfrac{3}{4}\right)^2 + \left(\tfrac{1}{4}\right)^2 \right) = \underline{0.375}$$

$Gini_{(weather = Rainy)}$ $\xrightarrow{\hspace{2em}}$ 3 instance — possible outcomes:

- 2 cinema
- 1 Stay In

$$= 1 - \left(\left(\tfrac{2}{3}\right)^2 + \left(\tfrac{1}{3}\right)^2 \right) = \underline{0.444}$$

$$\text{Weighted Average } Gini_{(Weather)} = \left(0.44 * \tfrac{3}{10}\right) + \left(0.375 * \tfrac{4}{10}\right) + \left(0.44 * \tfrac{3}{10}\right)$$

$$= \underline{0.414}$$

= 0.486

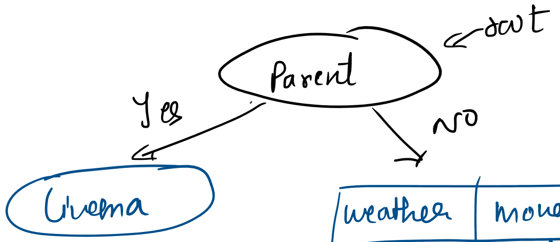
$\therefore$

$$\boxed{\begin{array}{l} \text{Gini (money)} = 0.486 \\ \text{Gini (parent)} = 0.36 \\ \text{Gini (weather)} = 0.416 \end{array}}$$

Minimum Gini $\longrightarrow$ Minimum Impurity in Decision.

Here Minimum Gini Value = Gini (parent) = 0.36

So the root of Decision is <u>Parent</u>



weather	money	Decision
Sunny	Rich	Tennis
Rainy	Rich	Stay In
Windy	Poor	Cinema
Windy	Rich	Shopping
Sunny	Rich	Tennis

We need to find Gini (parent = No and weather)

Also Gini (parent = No and money)

steps) Gini (parent = No and weather) $\longrightarrow$ Possible value
<u>5 instances</u>

- Sunny – 2 time
- Windy – 2 time
- Rainy – 1 time

Gini (parent = No and weather = Sunny) $\longrightarrow$ 2 instance — Tennis Possible Outcome $= 1 - \left(\left(\frac{2}{2}\right)^2\right) = 0$

Gini (parent = No and weather = Windy) $\longrightarrow$ 2 instances — Cinema, Shopping possible Outcome $= 1 - \left(\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right)$
$= 0.5$

$Gini_{(parent = No \text{ and } Weather = Rainy)} \xrightarrow[\text{1 instance}]{} \underset{\text{possible outcome}}{Stayin} = 1 - ((1/1)^2) = 0$

$$\boxed{\text{Weighted Average } Gini_{(parent = No \text{ and } Weather)} = 0.5 \times 2/5 = \underline{0.2}}$$

Step 6) $Gini_{(parent = No \text{ and } Money)}$ → possible values: Rich = 4 time, poor = 1 time

5 instance

$Gini_{(parent = No \text{ and } Money = Rich)} \xrightarrow[\text{possible outcome (4)}]{}$ Tennis − 2, StayIn − 1, Shopping − 1

$$= 1 - \left((2/4)^2 + (1/4)^2 + (1/4)^2 \right) = \underline{0.625}$$

$Gini_{(parent = No \text{ and } Money = Poor)} \xrightarrow[\text{possible outcome}]{}$ Cinema $= 1 - ((1/1)^2)$

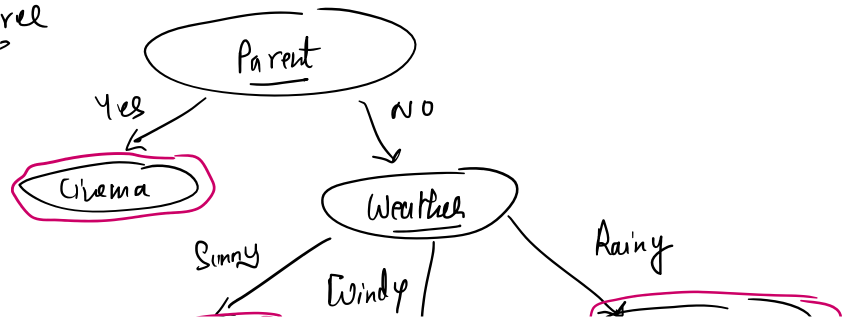
$= 0$

$$\boxed{\text{Weighted Average } Gini_{(parent = No \text{ and } Money)} = 0.625 \times 4/5 = 0.5}$$

Here $Gini_{(parent = No \text{ and } Weather)}$ has smallest value

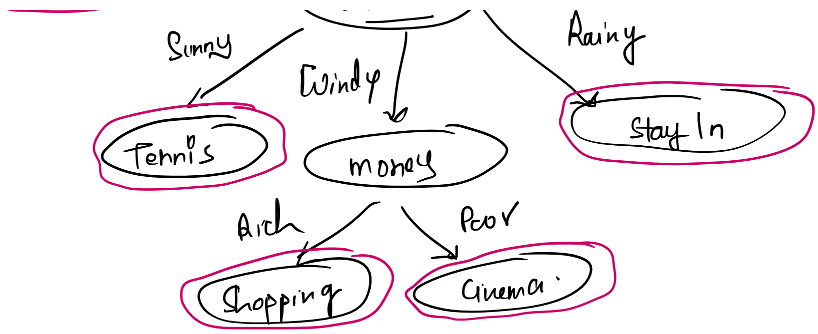
So now Next Node = <u>Weather</u>

* <u>Updated Decision Tree</u>



Ans

Ans



HW

Construct an optimal Decision Tree for following.

Outlook	Temperature	Humidity	Windy	Play (Decision)
Sunny	Hot	High	False	NO
sunny	Hot	High	True	NO
Overcast	Hot	High	F	Y
Rainy	mild	High	F	Y
Rainy	cool	Normal	F	Y
Rainy	cool	Normal	T	N
overcast	cool	Normal	T	Y
sunny	mild	High	F	N
Sunny	cool	Normal	F	Y
Rainy	mild	Normal	F	Y
sunny	mild	Normal	T	Y
overcast	mild	High	T	Y
overcast	Hot	Normal	F	Y
Rainy	mild	High	T	N