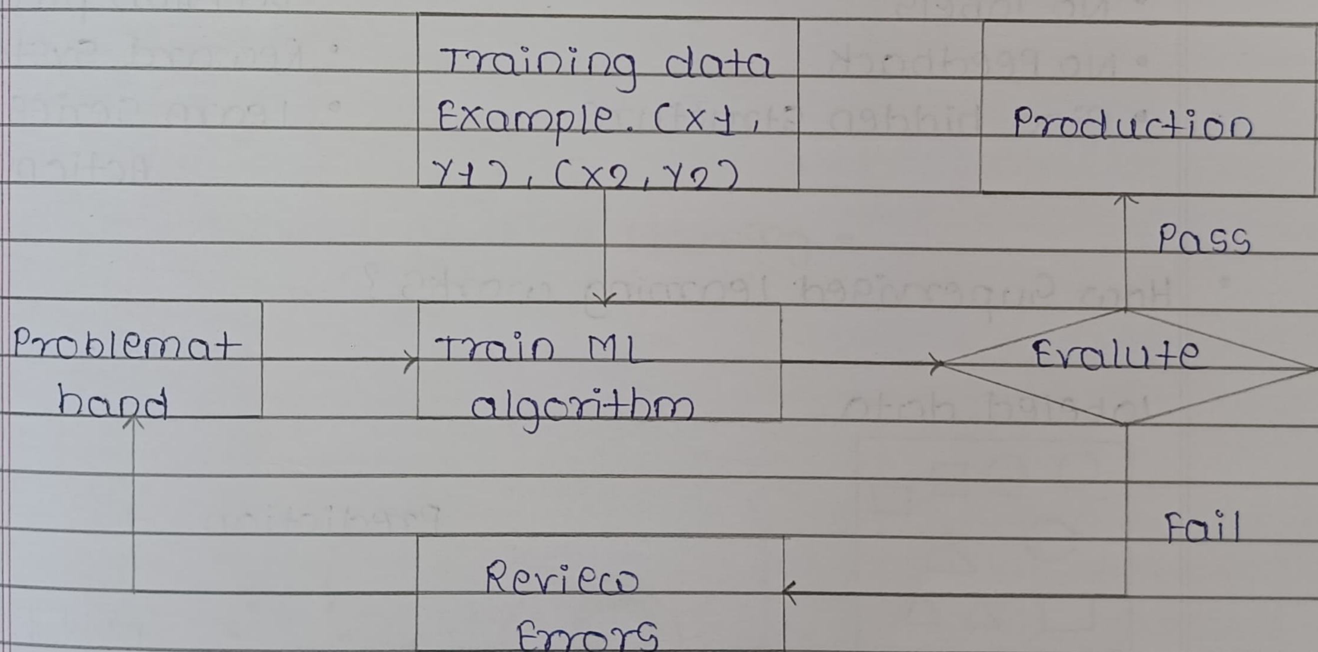
UNIT - 1

Introduction to Machine Learning

- Machine Learning -
  The capability of Artificial intelligence systems to learn by extracting patterns from data is known as machine learning.

- Machine learning Approach -
  machine learning relies on learning patterns based on simple data.



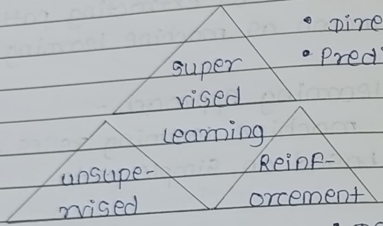
- Types of machine learning -

1. Supervised Learning -
   learning with a labeled training set.

2. Unsupervised learning -
Discover patterns in unlabled data.

3. Reinforcement learning -
learn to act based on feedback or reward.

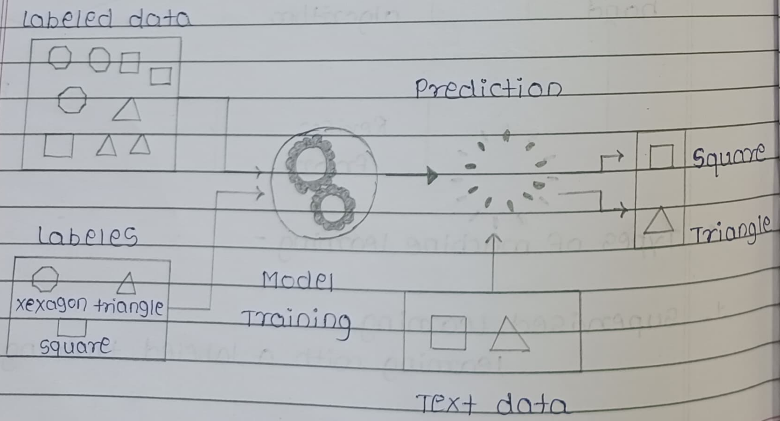
• labeled data
• Direct feedback
• Predict outcome / future



• No labels
• No feedback
• find hidden structure

• Decision process
• Reward system
• learn series of Actions

• How Supervised learning works ?



• Unsupervised learning - (Goals)

- The goal of the unsupervised learning is :
  i) To find the underlying structure of dataset.
  ii) group that data according to similarities.
  iii) represent that dataset in compressed format.

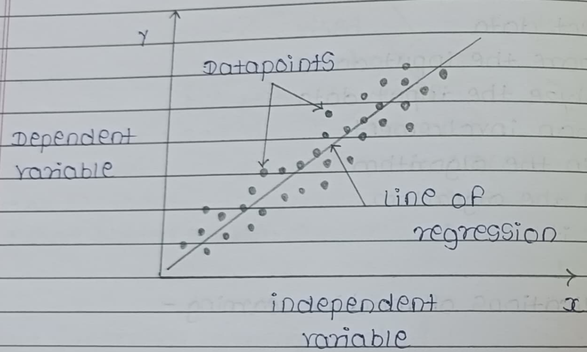
• Steps in developing a machine learning application

1. collect data
2. prepare the input data
3. Analyze the input data
4. human involvement
5. Train the algorithm
6. Test the algorithm
7. use it.

• Applications of machine learning -

- Regression in machine learning -

machine learning regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. it used as a method for predictive modelling in machine learning, in which an algorithm is used to predict continuous outcomes.



- Simple linear regression -

- mathematically, we can represent linear regression:

$$Y = \beta_0 + \beta_1 x + \epsilon \quad \text{or} \quad Y = bx + a$$

Y = dependent variable (Target variable)
x = independent variable (input variable)
B_0 / a = intercept of the line
B_1 / b = linear regression coefficient (angle)
ϵ = random error

- multiple linear regression -

- The form of the model is:

$$Y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_n xn$$

Here,
Y is a dependent variable
$x_1, x2 \ldots\ldots xn$ are independent variable
$B_0, B_1 \ldots\ldots B_n$ are the regression coefficients
$B_i \, (1 <= i <= n)$ is the slope or weight that specifies the factor by which x_i has an impact on Y.

- Formulas -

$$Y = B_0 + B_1 x + \epsilon$$

$$B_0 = \frac{(\Sigma Y)(\Sigma x^2) - (\Sigma x)(\Sigma xY)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$B_1 = \frac{n(\Sigma xY) - (x)(\Sigma Y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

- Simple linear regression -

- Example -

1. Find a linear regression equation for the following two sets of data.

ii) 37 hours

$z = -64 + 2 * 37$

$z = -64 + 74$

$z = 10$

$$p = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-10}} = 0.99$$

There are 99% of chances if student studies 37 hours.

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

$0.74 = \frac{1}{1+e^{-10}}$

$0.74 (1 + e^{-10}) = 1$

$0.74 + 0.74 e^{-10} = 1$

$0.74 \cdot e^{-10} = 1 - 0.74$

$0.74 e^{-10} = 0.26$

$p = 0.99 \Rightarrow 99\%$

$\ln (0.74 \cdot e^{-10}) = \ln (0.26)$

$\ln (e^{-10}) = \ln \left(\frac{0.26}{0.74}\right)$

$e^{-10} = \ln (0.3513)$

$z = -0.454$

$z = 0.454$

$z = -64 + 2 * \text{hours}$

$0.454 = -64 + 2 * \text{hours}$

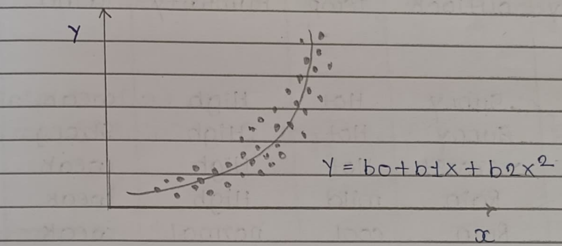
$64 + 0.454 = 2 * \text{hours}$

$\therefore \text{hours} = \frac{64.454}{2} = 32.227$

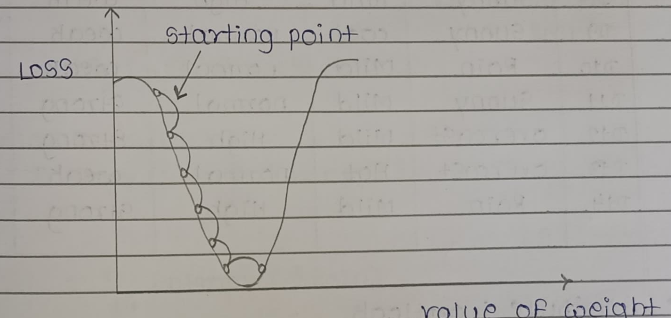
$\therefore$ For getting 74 % student should study 32.227 hours.

- Polynomial Regression -

$$Y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_n x^n$$



- How does Gradient Descent work?



point of convergence i.e. where the cost function.

1. The dataset of on a set in an exam of I student this given in table below if we used logistic regression as the classifier on assume the model suggested by the optimizer will become the following for act of passing.

Hr Study	Result
29	0
15	0
33	1
28	1
39	1

Pass = 1
Fail = 0

optimizer ⟶ log (odds) = $-64 + 2 * \text{hours}$

i) calculate the probability of pass if student study for 33 hours.

ii) To have to 95% of passing chance for how many hours student should study.

i) 33 hours

$z = -64 + 2 * \text{hours}$

$z = -64 + 2 * 33$

$z = -64 + 66$

$z = 2$

$$P = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-2}} = 0.88$$

There are 88% of chances if student studient 33 Hours.

ii) $P = \frac{1}{1+e^{-z}}$

$0.95 = \frac{1}{1+e^{-2}}$

$0.95(1+e^{-2}) = 1$

$0.95 + 0.95e^{-2} = 1$

$0.95 \cdot e^{-2} = 1 - 0.95$

$0.95e^{-2} = 0.05$

$P = 0.88 \Rightarrow 88\%$

$ln(0.95 \cdot e^{-2}) = ln(0.05)$

$ln(e^{-2}) = ln\left(\frac{0.05}{0.95}\right)$

$e^{-2} = ln(0.0526)$

$z = -2.94$

$z = 2.94$

$z = -64 + 2 * \text{hours}$

$2.94 = -64 + 2 * \text{hours}$

$64 + 2.94 = 2 * \text{hours}$

$\therefore \text{hours} = \frac{66.94}{2} = 33.47$

For getting 95% student should study 33.47 hours.

2. Example -

i) Find out the probability of pass if student study of 37 hours.

ii) To have to 74% of passing chance for how many hours student should study.

3. predict the missing value using linear regression.

x	1	2	3	4	5	15	18
y	1	2	1.3	3.75	2.25	10.3	?

x	y	x^2	xy
1	1	1	1
2	2	4	4
3	1.3	9	3.9
4	3.75	16	15
5	2.25	25	11.25
15	10.3	225	154.5
$\Sigma x = 30$	$\Sigma y = 20.6$	$\Sigma x^2 = 280$	$\Sigma xy = 189.65$

$$\beta_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{20.6 \times 280 - 30 \times 189.65}{6 \times 280 - 900} = \frac{78.5}{780} = 0.100$$

$$\beta_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{6 \times 189.65 - 30 \times 20.6}{6 \times 280 - 900} = \frac{519.9}{780} = 0.666$$

The value of $a = 0.100$ and $b = 0.666$ put in the equation,

$$y = 0.100 + 0.666x$$

put the value $x = 18$

$$y = 0.100 + 0.666 \times 18$$
$$y = 12.08$$

x	1	2	3	4	5	15	18
y	1	2	1.3	3.75	2.25	10.3	12.08

- Logistic model/Regression -

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

- The log term in on the LHS can be removed by reasing the RHS as a power of e :

$$\frac{p}{1-p} = e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}$$

now, we can easily simplify to obtain the value of p :

$$p = \frac{e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}}{1 + e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}}$$

OR

$$p = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)}}$$

$$F(x) = \frac{1}{1 + e^{-(x)}}$$

- Example -

Page No.
Date

x	2	4	6	8
Y	3	7	5	10

x	Y	x^2	xY
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\Sigma x = 20$	$\Sigma Y = 25$	$\Sigma x^2 = 120$	$\Sigma xY = 144$

$$\beta_0 = \frac{(\Sigma Y)(\Sigma x^2) - (\Sigma x)(\Sigma xY)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400} = \frac{120}{80} = 1.5$$

$$\beta_1 = \frac{n(\Sigma xY) - (\Sigma x)(\Sigma Y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400} = \frac{76}{80} = 0.95$$

Hence, we got the value of $a = 1.5$ and $b = 0.95$ The linear equation is given by,

$$Y = a + bx$$

put the value a and b in the equation linear regression is,

$$Y = 1.5 + 0.95x$$

Page No.
Date

2.

x	1	2	3	4
Y	3	4	5	7

x	Y	x^2	xY
1	3	1	3
2	4	4	8
3	5	9	15
4	7	16	28
$\Sigma x = 10$	$\Sigma Y = 19$	$\Sigma x^2 = 30$	$\Sigma xY = 54$

$$\beta_0 = \frac{(\Sigma Y)(\Sigma x^2) - (\Sigma x)(\Sigma xY)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{19 \times 30 - 10 \times 54}{4 \times 30 - 100} = \frac{30}{20} = 1.5$$

$$\beta_1 = \frac{n(\Sigma xY) - (\Sigma x)(\Sigma Y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{4 \times 54 - 10 \times 19}{4 \times 30 - 100} = \frac{26}{20} = 1.3$$

Hence, we got the value of $a = 1.5$ and $b = 1.3$

The linear equation is given by,

$$Y = a + bx$$

put the value a and b in the equation linear regression is,

$$Y = 1.5 + 1.3x$$

- Decision Tree –

1. Example – Draw a decision tree for a given dataset using ID3 algorithm or using information gain and entropy.

Day	outlook	Temp	Humidity	wind	Play Tennis
D1	Sunny	Hot	High	weak	NO
D2	Sunny	Hot	High	Strong	NO
D3	overcast	Hot	High	weak	Yes
D4	Rain	mild	High	weak	Yes
D5	Rain	cool	normal	weak	Yes
D6	Rain	cool	normal	Strong	NO
D7	overcast	cool	normal	Strong	Yes
D8	Sunny	Mild	High	weak	NO
D9	Sunny	cool	normal	weak	Yes
D10	Rain	Mild	normal	weak	Yes
D11	Sunny	Mild	normal	Strong	Yes
D12	overcast	Mild	High	Strong	Yes
D13	overcast	Hot	normal	weak	Yes
D14	Rain	Mild	High	Strong	NO

Attribute : outlook

values (outlook) = Sunny, overcast, Rain

$S = [9+, 5-]$

by default Ans

$+2, -2 = 1.0$

$$\text{Entropy}(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

- $S_{sunny} \leftarrow [2+, 3-]$

$$\text{Entropy}(S_{sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$$

- $S_{overcast} \leftarrow [4+, 0-]$

$$\text{Entropy}(S_{overcast}) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0$$

- $S_{Rain} \leftarrow [3+, 2-]$

$$\text{Entropy}(S_{Rain}) = -\frac{3}{5}\log_2\frac{2}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$$

- Information Gain –

- $$\text{Gain}(S, \text{outlook}) = \text{Entropy}(S) - \sum_{v \in (\text{Sunny, overcast, Rain})} \frac{|S_v|}{S}\,\text{Entropy}(S_v)$$

- Gain(S, outlook)

$$= \text{Entropy}(S) - \frac{5}{14}\text{Entropy}(S_{sunny}) - \frac{4}{14}\text{Entropy}(S_{overcast}) - \frac{5}{14}\text{Entropy}(S_{Rain})$$

- $$\text{Gain}(S, \text{outlook}) = 0.94 - \frac{5}{14}0.971 - \frac{4}{14}0 - \frac{5}{14}0.971$$

$$= 0.2464$$

Page No. Date

$$\text{Entropy}(S_{sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$$

- S_{Hot} [0+, 2-]

$$\text{Entropy}(S_{Hot}) = -\frac{0}{2}\log_2\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0$$

- S_{mild} [1+ 1-]

$$\text{Entropy}(S_{mild}) = -\frac{1}{1}\log_2\frac{1}{1} - \frac{1}{1}\log_2\frac{1}{1} = 1.0$$

- S_{cool} [1+, 0-]

$$\text{Entropy}(S_{cool}) = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0$$

- ~~S~~ Information Gain -

- Gain (S, Sunny, Temp) = Entropy (S) $- \sum_{(Hot, mild, cool)} \frac{|S_v|}{S}$ Entropy (S_v)

- Gain (S_{Sunny}, Temp)

$$= \text{Entropy}(S) - \frac{2}{5}\text{Entropy}(S_{Hot}) - \frac{2}{5}\text{Entropy}(S_{mild}) - \frac{1}{5}\text{Entropy}(S_{cool})$$

- Gain (S_{Sunny}, Temp) $= 0.97 - \frac{2}{5}0.0 - \frac{2}{5}1 - \frac{1}{5}0.0$

$= 0.570$

Page No. Date

- Attribute : Humidity

value (Humidity) = High, normal

- S_{Sunny} [2+, 3-]

$$\text{Entropy}(S_{Sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$$

- S_{High} [0+, 3-]

$$\text{Entropy}(S_{High}) = -\frac{0}{3}\log_2\frac{0}{3} - \frac{3}{3}\log_2\frac{3}{3} = 0$$

- S_{normal} [2+, 0-]

$$\text{Entropy}(S_{normal}) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2} = 0$$

- Information Gain -

- Gain (S_{Sunny}, Humidity) = Entropy (S) $- \sum_{(High, normal)} \frac{|S_v|}{S}$ Entropy (S_v)

- Gain (S_{Sunny}, Humidity)

$$= \text{Entropy}(S) - \frac{3}{5}\text{Entropy}(S_{High}) - \frac{2}{5}\text{Entropy}(S_{normal})$$

- Gain (S_{Sunny}, Humidity)

$= 0.97 - \frac{3}{5}0.0 - \frac{2}{5}0.0 = 0.97$

- $Gain(S, Humidity) = 0.94 - \frac{7}{14} 0.9852 - \frac{7}{14} 0.5916$

$= 0.1516$

- Attribute - Wind
  values (wind) = weak, strong

- $S[9+, 5-]$

$Entropy(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.94$

- $S_{weak} \leftarrow [6+, 2-]$

$Entropy(S_{weak}) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8} = 0.8112$

- $S_{strong} \leftarrow [3+, 3-]$

$Entropy(S_{strong}) = -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6}$

$= 1.0$

- Information Gain

- $Gain(S, wind) = Entropy(S) - \sum_{(weak, strong)} \frac{|S_v|}{S} Entropy(S_v)$

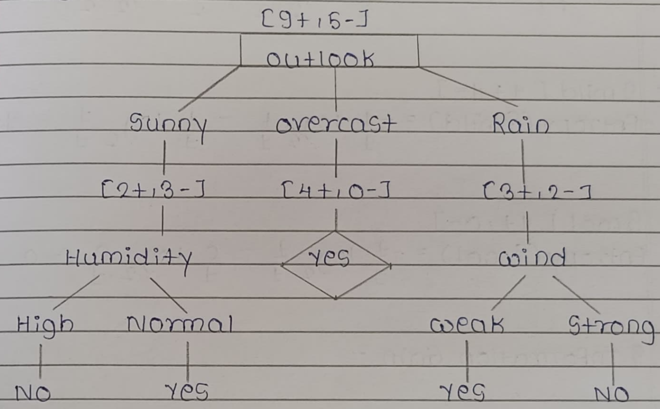
- $Gain(S, wind)$

$= Entropy(S) - \frac{8}{14} Entropy(S_{weak}) - \frac{6}{14} Entropy(S_{strong})$

- $Gain(S, wind) = 0.94 - \frac{8}{14} 0.8112 - \frac{6}{14} 0.1$

$= 0.4336$

- Tree



Day	Temp	Humidity	wind	Play Tennis
D1	Hot	High	weak	NO
D2	Hot	High	strong	NO
D8	Mild	High	weak	NO
D9	cool	normal	weak	Yes
D11	Mild	normal	strong	yes

- Attribute : Temp
  values (temp) = Hot, Mild, cool

- $S_{sunny} [2+, 3-]$

- Attribute : outlook
  values (Temp) = Hot, Mild, cool

- $S = [9+, 5-]$
  $\text{Entropy}(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

- $S_{Hot} \leftarrow [2+, 2-]$
  $\text{Entropy}(S_{Hot}) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1.0$

- $S_{mild} \leftarrow [4+, 2-]$
  $\text{Entropy}(S_{mild}) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = 0.9183$

- $S_{cool} \leftarrow [3+, 1-]$
  $\text{Entropy}(S_{cool}) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.8113$

Information Gain -

$$\text{Gain}(S, \text{Temp}) = \text{Entropy}(S) - \sum_{(Hot, Mild, cool)} \frac{|S_v|}{S}\,\text{Entropy}(S_v)$$

Gain (S, Temp)

$$= \text{Entropy}(S) - \frac{4}{14}\text{Entropy}(S_{Hot}) - \frac{6}{14}\text{Entropy}(S_{mild}) - \frac{4}{14}\text{Entropy}(S_{cool}).$$

- $\text{Gain}(S, \text{Temp}) = 0.94 - \frac{4}{14}1.0 - \frac{6}{14}0.9183 - \frac{4}{14}0.8113$
  $= 0.0289$

- Attribute : Humidity
  values (Humidity) = High, Normal

- $S = [9+, 5-]$
  $\text{Entropy}(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

- $S_{High} \leftarrow [3+, 4-]$
  $\text{Entropy}(S_{High}) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.9852$

- $S_{Normal} \leftarrow [6+, 1-]$
  $\text{Entropy}(S_{Normal}) = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.5916$

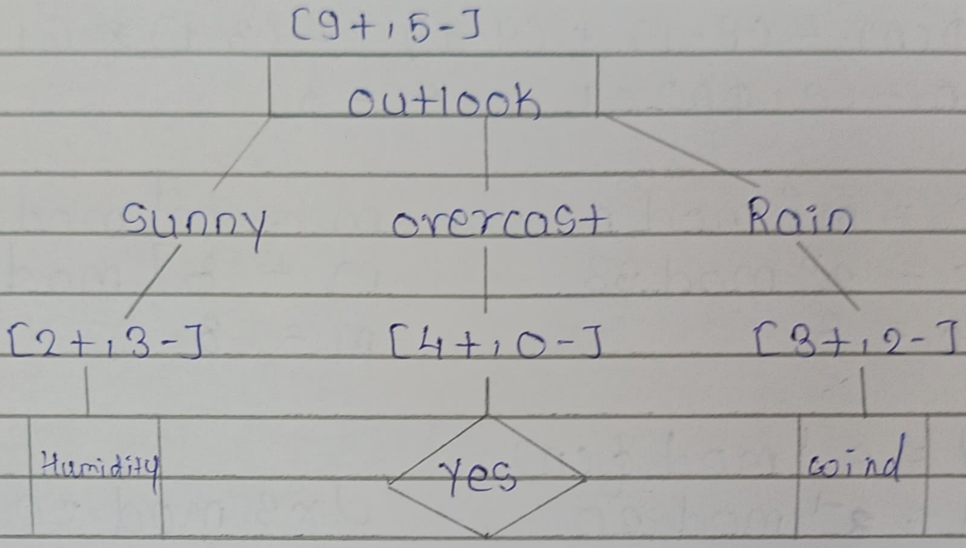
- Information Gain -

- $$\text{Gain}(S, \text{Humidity}) = \text{Entropy}(S) - \sum_{(High, Normal)} \frac{|S_v|}{S}\,\text{Entropy}(S_v)$$

- Gain (S, Humidity)
  $$= \text{Entropy}(S) - \frac{7}{14}\text{Entropy}(S_{High}) - \frac{7}{14}\text{Entropy}(S_{normal})$$

- Gain(S, wind) $= 0.94 - \frac{8}{14}\, 0.8112 - \frac{6}{14}\, 0.1 = 0.4336$

- Tree



Day	Temp	Humidity	wind	Play Tennis
D1	Hot	High	weak	NO
D2	Hot	High	Strong	NO
D8	Mild	High	weak	NO
D9	cool	normal	weak	yes
D11	Mild	normal	Strong	yes

- Attribute : Temp
  values (Temp) = Hot, Mild, cool

- S_{sunny} = [2+, 3-]