Page No.
Date

Maximum Likelihood Estimation

YYYNNNNN

what is apriori probability of Y that maximizes this likelihood of data for P(Y).

we denote probability by symbol π

i.e. $P(Y) = \pi$

Incoming $\rightarrow$ $P(Y_i) = \begin{cases} \pi & , \text{if } Y_i = Y \\ 1-\pi & , \text{if } Y_i = N \end{cases}$

YYY NN NNN

$\boxed{11100000}$

i.e. $P(Y_i) = \pi^{Y_i} (1-\pi)^{1-Y_i}$

if Y_i becomes 1 then $(1-\pi)^{1-Y_i}$ becomes 0.

if Y_i becomes 0 then π^{Y_i} becomes 0

Assuming data independence, then joint probability

$$P(\text{data}) = \prod_{i=1}^{n} P(Y_i)$$

$$P(\text{data}) = \pi^{\text{count}(Y_i=1)} (1-\pi)^{\text{count}(Y_i=0)}$$

$$= \pi^3 (1-\pi)^5$$

Find π that maximizes the expression i.e. that maximizes ~~for~~ the logarithm of expression (monotonic increasing)

$$\log P(\text{data}) = 3\log\pi + 5\log(1-\pi)$$

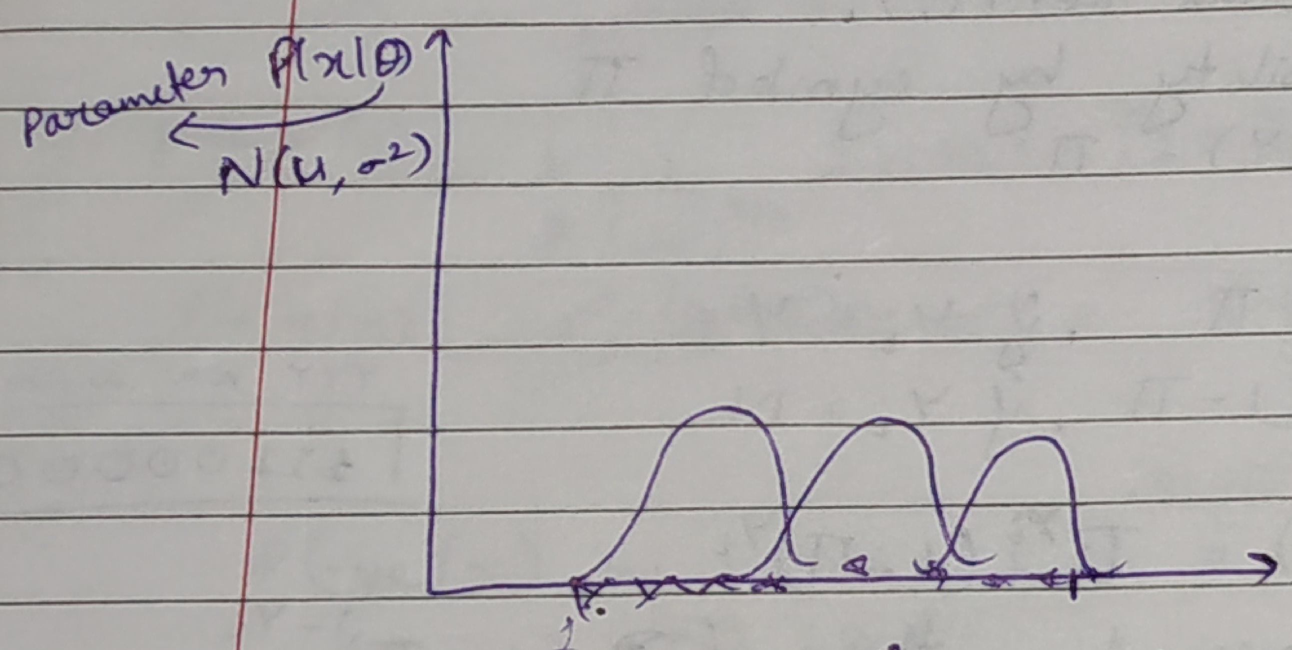
$$\nabla_\pi \log P(\text{data}) = 0$$

$\nabla \log \pi$

$$\Rightarrow \frac{3}{\pi} - \frac{5}{1-\pi}$$

$$\Rightarrow \frac{3}{\pi} = \frac{5}{1-\pi}$$

$$\Rightarrow \pi = \frac{3}{8}$$



Considering $\sigma^2 = 1$ (Shape is fixed)

$$P(data) = \cancel{e^{-\frac{(x-2)^2}{2}}}\ 3e^{-\frac{(x-1)^2}{2}} + 5e^{-\frac{x^2}{2}}$$

$$\cancel{P(data) = 3e^{-\frac{(1-\mu)^2}{2}} + 5e^{-\frac{\mu^2}{2}}}$$

$$P(data) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}}$$

$$P(data) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}}$$

$$\log P(data) = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2} + \left(\frac{-n}{2}\right)\log(2\pi)$$

$$\frac{d \log P(data)}{d\mu} = -\sum_{i=1}^{n} \frac{2(x_i-\mu)(-1)}{2} + 0$$

$$0 = \sum_{i=1}^{n} (x_i - \mu)$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu$$

$$\int_1^n x\,dx = \int_1^n \mu\,dx$$

$\frac{n(n+1)}{2}$ $n^2!$

$$\frac{n^2-1}{2} = \mu[n-1]$$

$$\frac{(n-1)(n+1)}{2} = \mu[n-1]$$

$$\mu = \frac{n+1}{2}$$

Nearest Neighbor Classifier:

Ex	Shape	Size	Vol.	Color	Shade	Class	Diff
x		unknown		data		?	
Ex1		Training data				Yes	3
Ex2							4
						Yes	4
							1
⋮						Yes	2
							3
						⋮	5
Ex8						No	5
Ex9						No	6

K-Nearest Neighbor → Consider min. K values
↑ Odd values

$$d_E(x,y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

$x = [1,4,3]$
$ex_1 = \{[1,3,2], pos\}$
$ex_2 = \{[3,6,2], pos\}$
$ex_3 = \{[2,3,5], neg\}$
$ex_4 = \{[5,2,4], neg\}$

Normalizing

$$x = \frac{x - MIN}{MAX - MIN}$$

Eg. $[7, 4, 25, -1, 10]$

⇓

$[8/26, 5/26, 26/26, 0, 11/26]$

$MIN = -1$, $MAX = 25$

Normalized → $[0.3077, 0.1923, 1, 0, 0.423]$

Page No. | Date

~~Principal Component Analysis~~:
~~(PCA)~~

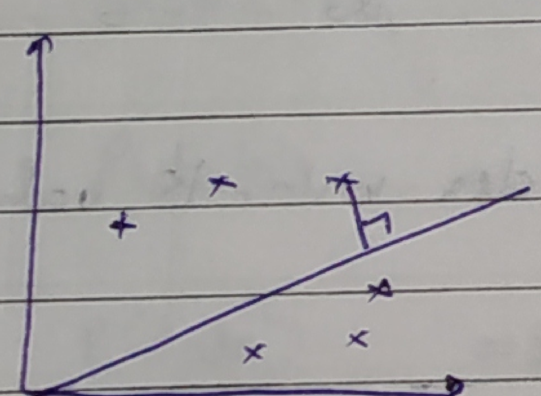
Bayesian Classifier is better than nearest neighbor classifier as error rate is less than the nearest neighbor

~~x — {1, 0.2, 2 6 4}~~

Weighted Nearest Neighbor:

Principal Component Analysis (PCA):

The projected vector represent the best in terms of least squared error.



$x_k = m + a_k e$ (eqn of line)

If $a = 0$, then $x = m$

Here, e is unit vector passing through m.

a_k represent a point x_k, where it met on the line. It will incorporate some amount of error & error is $(m + a_k e) - x_k$.

Taken all the lines together, error function

$$J(a_1, a_2, \ldots a_n, e) = \sum_{i=1}^{n} ||(m + a_k e) - x_k||^2$$

Assuming e to be fixed & minimizing square error

Page No. | Date

$$= \sum_{k=1}^{n} ||a_k e - (x_k - m)||^2$$

$$= \sum_{k=1}^{n} a_k^2 ||e||^2 - 2 \sum a_k e^t (x_k - m) + \sum ||x_k - m||^2$$

$$\frac{\partial J}{\partial a} = 2 \sum a_k - 2 \sum e^t (x_k - m) = 0$$

$\Rightarrow a_k = e^t (x_k - m) \Rightarrow$ Projections are orthogonal

Direction of line

J, the criterion function as a function of e

$$J(e) = \sum a_k^2 - 2 \sum a_k e^t (x_k - m) + \sum ||x_k - m||^2$$

$\because a_k = e^t (x_k - m)$

$$\therefore J(e) = \sum a_k^2 - 2 \sum a_k^2 + \sum ||x_k - m||^2$$

$$= - \sum [e^t (x_k - m)]^2 + \sum ||x_k - m||^2$$

$$= - \sum \underbrace{[e^t (x_k - m)(x_k - m)^t . e]}_{\text{Matrix representation / Scatter matrix}} + \sum ||x_k - m||^2$$

e is independent of k.

$$J(e) = - e^t \underbrace{[\sum (x_k - m)(x_k - m)^t]}_{\text{covariance matrix (scaled version)}} e + \sum ||x_k - m||^2$$

$$\Rightarrow J(e) = - e^t S e + \sum ||x_k - m||^2$$

$u = e^t S e - \lambda (e^t e - 1)$ Lagrangian eqn

$$\frac{\partial u}{\partial e} = 2 S e - 2 \lambda e = 0$$

$$S e = \lambda e$$

$S \to$ Scatter matrix

$e \to$ Eigen vector corresponding to the largest eigen value

a_k are principle component

also called K-L transform

$u \to$ unit vector

$m \to$ mean of x

$a_k \to$ projections of x_k on line

2019/4/11 11:26

Ques

x_1	1	3	4	5	7	6	9	10	12	13
x_2	2	5	3	6	5	2	4	1	3	6

$\mu_1 = \frac{70}{10} = 7$

$\mu_2 = \frac{37}{10} = 3.7$

~~mea~~ mean subtracted,

$(x_1 - \mu_1) = [-6 \;\; -4 \;\; -3 \;\; -2 \;\; 0 \;\; -1 \;\; 2 \;\; 3 \;\; 5 \;\; 6]_{1\times 10}$

$(x - \mu_1)^T = \begin{bmatrix} -6 \\ -4 \\ -3 \\ -2 \\ 0 \\ -1 \\ 2 \\ 3 \\ 5 \\ 6 \end{bmatrix}$

~~$(x-\mu_1)\cdot(x-\mu_1)^T = 140$~~

$x - \mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 7 \\ 3.7 \end{bmatrix}$

$x - \mu_1$	-6	-4	-3	-2	0	-1	2	3	5	6
$x - \mu_2$	-1.7	1.3	-0.7	2.3	1.3	-1.7	0.3	-2.7	-0.7	2.3

$(x-\mu) \quad (x-\mu)^T$

$M_1 = \begin{bmatrix} -6 \\ -1.7 \end{bmatrix}_{2\times1} [-6 \;\; -1.7]_{1\times2} = \begin{bmatrix} 36 & 10.2 \\ 10.2 & 2.89 \end{bmatrix}$

$M_2 = \begin{bmatrix} -4 \\ 1.3 \end{bmatrix} [-4 \;\; 1.3] = \begin{bmatrix} 16 & -5.2 \\ -5.2 & 1.69 \end{bmatrix}$

$M_3 = \begin{bmatrix} -3 \\ -0.7 \end{bmatrix} [-3 \;\; -0.7] = \begin{bmatrix} 9 & +2.1 \\ +2.1 & 0.49 \end{bmatrix}$

$M_4 = \begin{bmatrix} -2 \\ 2.3 \end{bmatrix} [-2 \;\; 2.3] = \begin{bmatrix} 4 & -4.6 \\ -4.6 & 5.29 \end{bmatrix}$

$M_5 = \begin{bmatrix} 0 \\ 1.3 \end{bmatrix} [0 \;\; 1.3] = \begin{bmatrix} 0 & 0 \\ 0 & 1.69 \end{bmatrix}$

$M_6 = \begin{bmatrix} -1 \\ -1.7 \end{bmatrix} [-1 \;\; -1.7] = \begin{bmatrix} 1 & 1.7 \\ 1.7 & 2.89 \end{bmatrix}$

$M_7 = \begin{bmatrix} 2 \\ 0.3 \end{bmatrix} [2 \;\; 0.3] = \begin{bmatrix} 4 & 0.6 \\ 0.6 & 0.09 \end{bmatrix}$

$M_8 = \begin{bmatrix} 3 \\ -2.7 \end{bmatrix} [3 \;\; -2.7] = \begin{bmatrix} 9 & -8.1 \\ -8.1 & 7.29 \end{bmatrix}$

$M_9 = \begin{bmatrix} 5 \\ -0.7 \end{bmatrix} [5 \;\; -0.7] = \begin{bmatrix} 25 & -3.5 \\ -3.5 & 0.49 \end{bmatrix}$

$M_{10} = \begin{bmatrix} 6 \\ 2.3 \end{bmatrix} [6 \;\; 2.3] = \begin{bmatrix} 36 & 13.8 \\ 13.8 & 5.29 \end{bmatrix}$

Covariance matrix, $\Sigma (x-\mu)(x-\mu)^T = \begin{bmatrix} 140 & 7 \\ 7 & 28.1 \end{bmatrix}$

$|A - \lambda I| = 0 \Rightarrow \begin{bmatrix} 140 & 7 \\ 7 & 28.1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$

$\Rightarrow \begin{vmatrix} 140-\lambda & 7 \\ 7 & 28.1-\lambda \end{vmatrix} = 0$

Page No.
Date

$(140-\lambda)(28.1-\lambda) - 49 = 0$

$3934 + \lambda^2 - \lambda(168.1) - 49 = 0$

$\lambda^2 - 168.1\lambda + 3885 = 0$

$\cancel{\lambda_1 = \quad , \lambda_2 =}$

$$\lambda = \frac{168.1 \pm \sqrt{(168.1)^2 - 4(1)(3885)}}{2}$$

$$\lambda = \frac{168.1 \pm 112.77}{2}$$

$\lambda_1 = 140.435, \quad \lambda_2 = 27.665$

$\lambda = \max(\lambda_1, \lambda_2) = 140.435$

#

$AX = \lambda X$

$$\begin{bmatrix} 140 & 7 \\ 7 & 28.1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 140.43 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} 140x_1 + 7x_2 \\ 7x_1 + 28.1x_2 \end{bmatrix} = \begin{bmatrix} 140.43x_1 \\ 140.43x_2 \end{bmatrix}$$

$140x_1 + 7x_2 = 140.43x_1$

$7x_1 + 28.1x_2 = 140.43x_2$

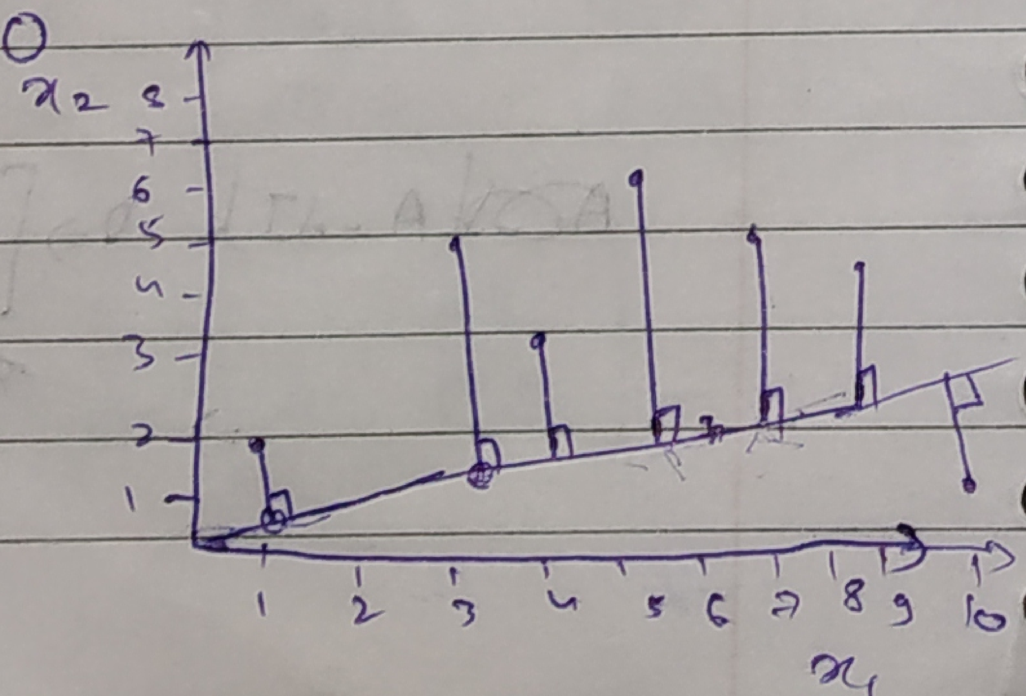
~~Taking $x_1 = 1$~~

$7x_2 - 0.43x_1 = 0$

$-112.33x_2 + 7x_1 = 0$

Taking $x_1 = 1$

$x_2 = \frac{0.43}{7}$



Page No.
Date

Decision Trees :

	Volume	Shape	Size	Class
ex1	Big	Circle	Small	+ve
ex2	small	Circle	Small	+ve
ex3	Big	Square	Small	neg
ex4	Big	Triangle	Small	neg
ex5	Big	Square	Big	+ve
ex6	Small	Square	Small	neg
ex7	Small	Square	Big	+ve
ex8	Big	Circle	Big	+ve
ex9				
ex10				

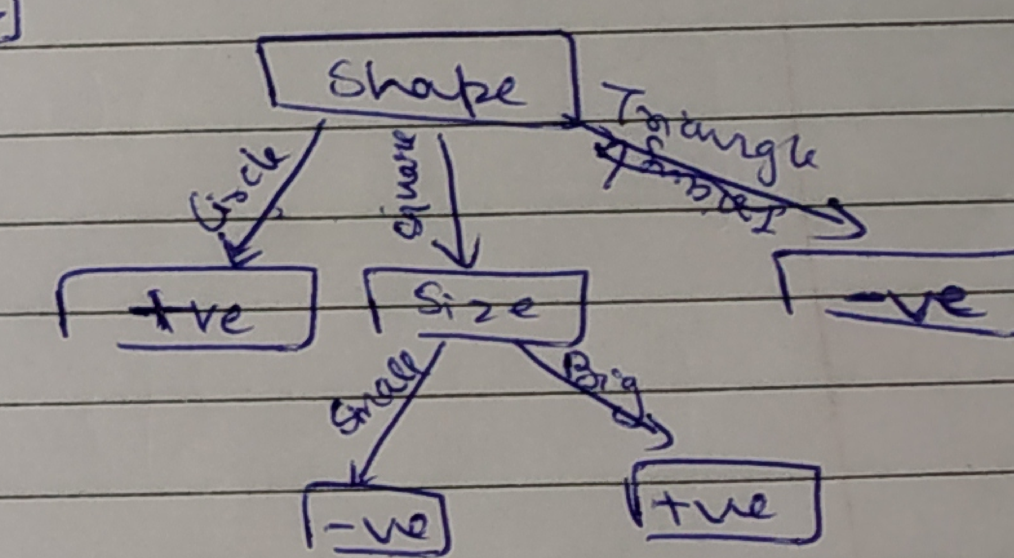
Accuracy

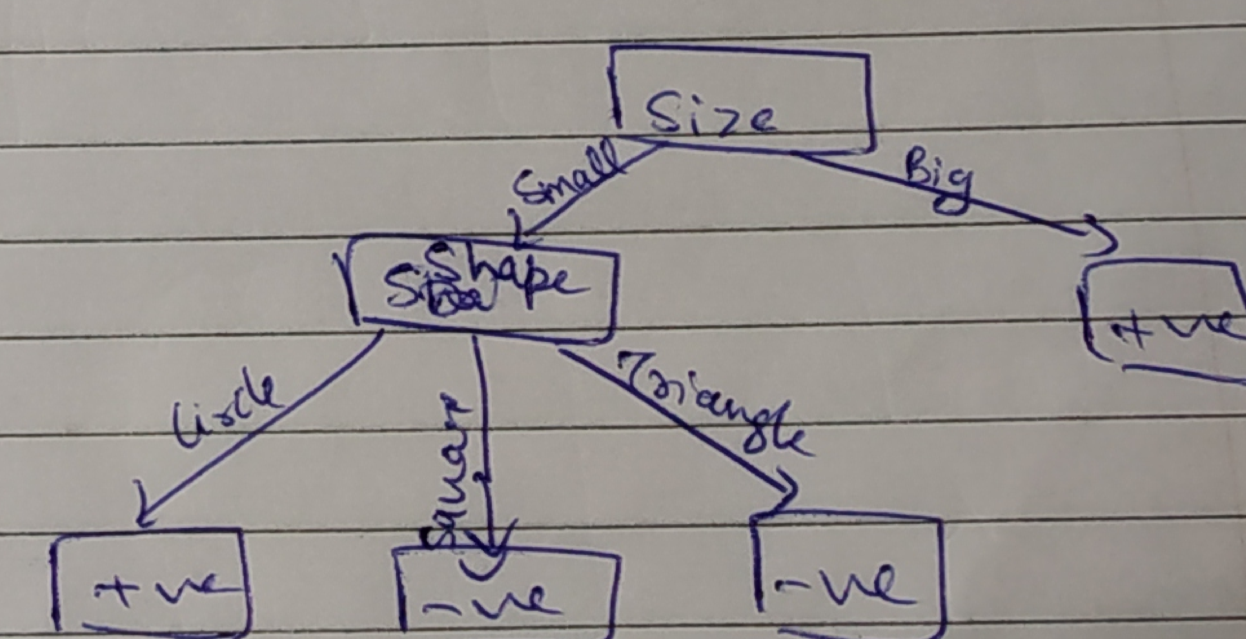
1st ~~Error~~

Volume → error = 3
Shape → error = 2
Size → error = 2

Volume → error = 2
Size → error = 0

2nd
Volume = 1
shape error = 0





Regression → Normal
Polynomial

Page No.
Date

<u>Rules:</u>

1. If shape = circle, predict +ve
else if shape = triangle, predict -ve
else if shape = square,
{ if size = small , predict -ve
else if size = Big , Predict +ve
}

OR

Regression
Polynomial · Normal.

Page No. Date: / /

Rules :-

If shape = Circle, predict +ve
else if shape = triangle, predict −ve
else if shape = Square,
{ if size = Small, predict −ve
else if size = Big, predict +ve
}

11/4/19

Age	Req.	Type	Class
Old	Y	Software	−ve
O	N	S	−ve
O	N	Hardware	−ve
Mid	Y	S	−ve
M	Y	H	−ve
M	N	H	+ve
M	N	S	+ve
New	Y	S	+ve
N	N	H	+ve
N	N	S	+ve

Entropy of Class : $I(P_i, N_i)$.

Gain = Entropy of class − Entropy of attribute

$P \to 1 \Rightarrow I = 0$

Information is directly related to uncertainity or inversely proportional to the probability of Occurance of that event.

$$I(E) = \log \frac{1}{P(E)}$$

Entropy

individual probability

$$H(S) = \sum_{i=1}^{q} I(S_i)P(S_i) = \sum_{i=1}^{q} P(S_i) \log \frac{1}{P(S_i)}$$

$$= -\sum_{S} P(S_i) \log P(S_i).$$

Page No. Date: / /

Ent of Attri = 0.8127

$$\text{Entropy of Attribute} = \frac{\sum P_i + N_i}{P+N}\left(I(P_i, N_i)\right)$$

Entropy of class :- $I(P_i, N_i)$.

Gain = Entropy of Class − Entropy of attribute.

So, Entropy of Class (−ve) = $-\sum_{S}\left(\frac{5}{10}\right)\log$

$= 1.$

for Age

	P_i	N	$I(P_i, N_i)$
Old	0	3	0
Mid	2	2	1
New	3	0	0

→ If any of is 0 than always $I(P_i, N_i) = 0$.

→ If both are same than $I(P_i, N_i) = 1$ always

Entropy of attribute = $\frac{\left(\frac{3}{10} + \frac{4}{10} + \frac{3}{10}\right) + 10}{1+10} = \frac{\frac{10}{10} + 10}{11}$

$= \frac{11}{11} = 1.$

$= \frac{0+3}{10}(0)$

If in table we have
type
P_i	N	$I(P_i, N_i)$
1	3	?

$-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$
$= 0.8127.$

Root node
at 6.

for Requirement

Entropy of class =

	P_i	N	$I(P_i, N_i)$
Y	0	4	
N	1	6	0.0127

Gain
~~Entropy of attribute~~ = 0.12 (3) + 0.16 (5).
= 0.812.

2019/4/11 11:27

Impurity $= -\frac{n_{class_1}}{n_{Total}} \log\left(\frac{n_{class_1}}{n_{Total}}\right) - \frac{n_{class_2}}{n_{Total}} \log\left(\frac{n_{class_2}}{n_{Total}}\right)$

$= \underline{0.2878}$

11.4.19

Age	Req	Type	Class
Old	Y	Software	-ve
Old	N	S	-ve
Old	N	Hardware	-ve
Mid	Y	S	-ve
M	~~N~~ Y	H	-ve
M	N	H	+ve
M	N	S	+ve
New	Y	S	+ve
New	N	H	+ve
New	N	S	+ve

log is base 2 here

Entropy of class : $I(P_i, N_i)$

Gain = Entropy of class − Entropy of attribute

$P \to 1 \implies I \to 0$

Information is directly related to uncertainty or inversely proportional to the probability of occurrence of that event

$$I(E) = \log\left(\frac{1}{P(E)}\right)$$

Entropy

$$H(S) = \sum_{i=1}^{2} I(S_i) P(S_i)$$

$$= \sum_{i=1}^{2} P(S_i) \log\left(\frac{1}{P(S_i)}\right)$$

$$= -\sum_{S} P(S_i) \log(P(S_i))$$

Entropy of attribute $= \frac{\Sigma (P_i + N_i)}{(P + N)} (I(P_i, N_i))$

Entropy of class : $(I(P_i, N_i))$

$\boxed{\text{Gain} = (\text{Entropy of class}) - (\text{Entropy of attribute})}$

Entropy of class $= -\frac{5}{10}\log\frac{5}{10} - \frac{5}{10}\log\frac{5}{10}$

$= -\log_2 \frac{1}{2}$

$= \log_2 2 = 1$

always if $P_+ = P_-$

Now, we find the rest, so make table for each attr.

Age	P_i → +ve	N_i → -ve	$I(P_i, N_i)$
Old	0	3	0
Mid	2	2	1
New	3	0	0

→ calc. using $\Sigma P_i \log_2 P_i$

$\implies$ Entropy of ~~Attrib~~ age $= \frac{(0+3)}{10} \times 0 + \frac{4 \times 1}{10} + \frac{(3+0) \times 0}{10}$

$= 0.4$

$\implies$ Gain$_{age}$ = 0.6

Req	P_i	N_i	$I(P_i, N_i)$
Y	1	3	0.815
N	4	2	0.916

$\to -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$

$-\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = \frac{2}{4} + \frac{3}{4}(2 - \log_2 3)$

$= 0.389 + 0.527 = 0.916 \cdot 0.815$

2019/4/11 11:28

for Mid as a class.
Original ta...

Age	
M	
M	
M	
M	

Entropy of attribut...

Than Derive...

0.12

Page No. Date

$$\text{Entropy} = \frac{4 \times 0.815}{10} + \frac{6 \times 0.916}{10}$$

$$= \frac{3.26 + 5.496}{10}$$

$$= 0.8756$$

Gain $= 0.1244$

Type	P_i	N_i	$I(P_i, N_i)$
S	3	3	1
H	2	2	1

$-\frac{3}{6}$

$$\text{Entropy} = \frac{6 \times 1}{10} + \frac{4 \times 1}{10}$$

$$= 1$$

$\text{Gain}_{\text{Type}} = 0$

⇒ Max. gain attribute = age

⇒ Root is age

age → We explore mid only as others have 0 info".

Now table for mid,

~~Entropy of mid~~

	Req.	Type	
Mid	Y	S	−
Mid	Y	H	−
Mid	N	H	+
Mid	N	S	+

Do prooning on ur own.

2 ~~...~~ theory + mathematics + ~~eg~~ Example

Regression, Logistic + Linear

Draw trees & write rules on ipl

Page No. Date

Req.	P_i	N_i	$I(P_i, N_i)$
Y	0	1	0
N	1	0	0

⇒ Entropy = 0

Gain = 1 − 0 = 1

calc. new → $-\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4}$

Type	P_i	N_i	I
S	1	1	1
H	1	1	1

$$\text{Entropy} = \frac{2 \times 1}{4} + \frac{2 \times 1}{4}$$

$$= 1$$

Gain = 0

Also, stop exploring an attribute when its I for all i is equal to 0.

Pay attention to class part

Also write rules