NLC
National Logistics Cell

①

SM (HR)

After 2nd Sessional.

-: Finals :-

19/12/17.

Lecture# 21

machine Learning:-

100 samples

60 seabass = $P(w_1) = 60/100$

40 salmon = $P(w_2) = 40/100$

→ P(seabass)

→ P(salmon)

→ what is prior probability?

An extra knowledge that what is idea / likely of getting seabass / salmon.

→ if we get any fish out of pond then what we can say it's a S.B or salmon, it will be S.B as it has more probability

$P(x \mid w_i)$

→ $P(\text{lightness} \mid w_1)$

* Training data → extraction of lightness → plot in the form of

* lets we have a test sampl

$x = 14$

$P(14 | w_1) > P(14 / w_2)$

$0.18 > 0.03$

$P(w_i)$ = prior prob

$P(x | w_i)$ = class "

$p(w_i | x)$ = which class we will get. as it is a test data

⇒ we will use both training & prior prob. and based on that we find out a class.

$$p(w_i | x) = \frac{p(x | w_i) P(w_i)}{p(x)}$$

normalizing factor

$$p(x) = \sum_{i=1}^{c=2} P(x | w_i) P(w_i)$$

$P(x) = P(x|w_1)P(w_1) + P(x|w_2)P(w_2)$

$P(x) = 0.18 \times 0.6 + 0.03 \times 0.4$

$P(x) = 0.12$

$\Rightarrow P(w_1|14) = \frac{0.18 \times 0.6}{0.12} = 0.9$

$\Rightarrow P(w_2|14) = \frac{0.03 \times 0.4}{0.12} = 0.1$

$\Rightarrow$ posterior probability of w_1 is more.

FINALS

CLC National Logistics Cell

SM (HR)

20th Dec, 2017 AI Lecture # 22

70 → seabass $= P(w_1) = 0.7$

30 → salmon $= P(w_2) = 0.3$

Class condition Probability

$P(w_i|x) = P(x|w_i) \times P(w_i) / P_x$

$P(14|w_1) = 0.18$

$P(14|w_2) = 0.03$

Baye's Theorem.

$P(x) = \sum_{i=1}^{2} P(x|w_i)P(w_i)$

$= P(x|w_1)P(w_1) + P(x|w_2)P(w_2)$

$= 0.18 \times 2/3 + 0.03 \times 1/3 = 0.13$

$$P(w_1|14) = \frac{P(14|w_1)(P(w_1))}{P(x)}$$

$$= \frac{0.18 \times 2/3}{0.13} = 0.92$$

$$P(w_2|14) = \frac{P(14|w_2)P(w_2)}{P(x)}$$

Sample belongs to w_1.

$$= \frac{0.03 \times 1/3}{0.13}$$

$= 0.076$

$\Rightarrow$ Test samples are given for

Baye's Decision Rule:-

One having greater posterior probability, the sample belongs to that class.

Probability of Error:-

The probability of error other than original class is the probability of error.

$\Rightarrow P(\omega_1 | 14) = 0.2$

$P(\omega_2 | 14) = 0.7$

$P(\omega_3 | 14) = 0.1$

$P(\text{error}) = 1 - 0.7$
$= 0.3$

or

$P(\text{error}) = 0.2 + 0.1 = 0.3$

---

if prior probability same, then class condition.

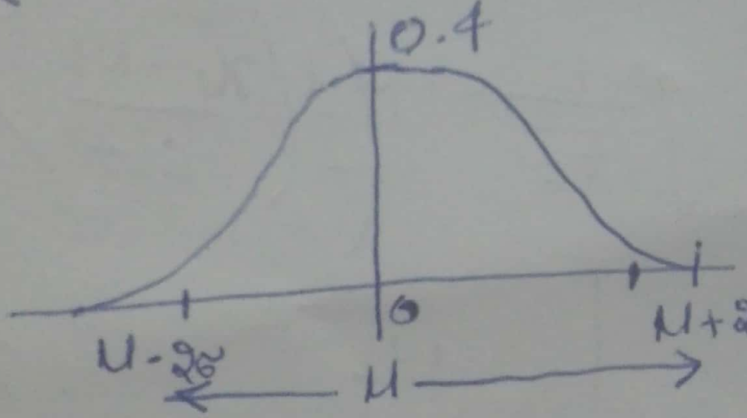
If class condition same, then upon prior prob.

Univariate Density:- $N(\mu, \sigma^2)$

$$P(x | \omega_j) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_j}{\sigma_j}\right)^2\right]$$

variance = spread.
mean = avg.
Gaussian dist. is centred along mean.

$N(\mu, \sigma^2) \rightarrow N(0,1)$



$\frac{1}{\sqrt{2\pi\sigma^2}} \rightarrow$ amp or peak

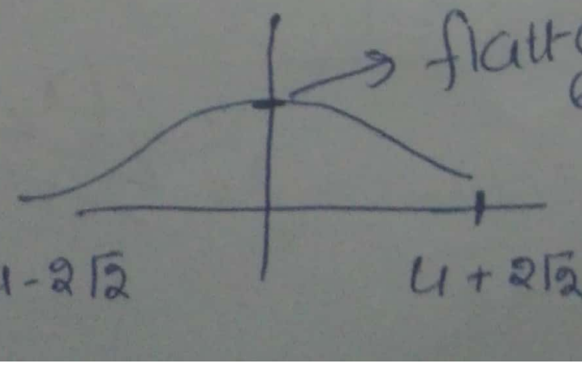
* σ greater ↓ flat
* σ small ↓ heighted

* $\mu + 2\sigma$ & $\mu - 2\sigma$ is the spread.

$N(\mu, \sigma^2) = (3, 2)$

$\mu = 3$

$\mu + 2\sigma = \mu + 2\sqrt{2}$
$= 3 + 2(1.41)$.



→ flatter as σ increased.

$$P(w_i|x) = P(x|w_i)\,P(w_i)$$

class condition probability is not easy so we term this as gaussian distribution.

$$P(x|w_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2\right)$$

ایمن آصف

$$x = \begin{bmatrix} 15 \\ 7 \end{bmatrix}$$

w_1 lightness	w_2
13	11
13.5	11.5
14.5	12.5
14	11.5
13.5	12
$\mu_1 = 13.7$	$\mu_2 = 11.7$
$\sigma^2 = 1.5$	$\sigma^2 = 1$

$$\sigma^2 = \frac{\sum (x_k - \mu)^2}{n}$$

$$\frac{(13-13.7)^2 + (13.5-13.7)^2 + \dots}{5} = \sigma^2$$

$$P(14|\omega_1) = \frac{1}{\sqrt{2\pi(1.5)}} e^{-1/2\left(\frac{14-13.7}{1.5}\right)^2}$$

$= \text{value } 1$

$$P(14|\omega_2) = \frac{1}{\sqrt{2\pi(1)}} e^{-1/2\left(\frac{(14-11.7)^2}{1}\right)}$$

$= \text{value } 2$

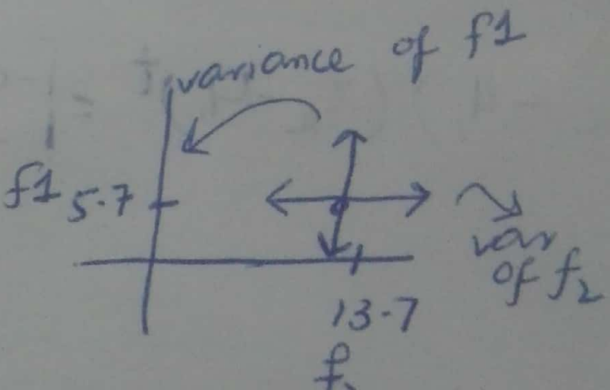
Multivariate :-

$$\omega_1 = \begin{bmatrix} 13 \\ 5 \end{bmatrix} \begin{bmatrix} 13.5 \\ 6 \end{bmatrix} \begin{bmatrix} 14 \\ 5.5 \end{bmatrix} \begin{bmatrix} 14.5 \\ 6.5 \end{bmatrix}$$

$$\omega_2 = \begin{bmatrix} 16.5 \\ 7 \end{bmatrix} \begin{bmatrix} 16 \\ 7.5 \end{bmatrix} \begin{bmatrix} 17.5 \\ 8.5 \end{bmatrix} \begin{bmatrix} 17 \\ 8 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 13.7 \\ 5.7 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 17.7 \\ 8.7 \end{bmatrix}$$



$$\Sigma_1 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

$\sigma_{12} = \sigma_{21} \rightarrow$ correlation

+ve → if value of one feature increases then other's increases auto. (high +ve)

-ve → if one decreasing othere incre. higher -ve value

zero → zero-corelation obtained.

☆ if features are independent then co-variance = 0

"[illegible]

$$x_k - \mu_1 = \begin{bmatrix} 13 - \underline{13.7} & \frac{13.5}{13.7} & 14 - \underline{13.7} & 14.5 - \underline{13.7} \\ 5 - \underline{5.7} & 6 - \underline{5.7} & 5.5 - \underline{5.7} & 6.5 \cdot \underline{5.7} \end{bmatrix}$$

$$x_k - \mu_1 = \begin{bmatrix} -0.7 & -0.2 & 0.3 & 0.8 \\ -0.7 & 0.3 & -0.2 & 0.8 \end{bmatrix}$$

$$(x_k - \mu_1)^t = \begin{bmatrix} -0.7 & -0.7 \\ -0.2 & 0.3 \\ 0.3 & -0.2 \\ 0.8 & 0.8 \end{bmatrix}$$

$$(x_k - \mu_1)(x_k - \mu_1)^t = \begin{bmatrix} -0.7 & -0.2 & 0.3 & 0.8 \\ -0.7 & 0.3 & -0.2 & 0.8 \end{bmatrix} \begin{bmatrix} -0.7 & -0.7 \\ -0.2 & 0.3 \\ 0.3 & -0.2 \\ 0.8 & 0.8 \end{bmatrix}$$

$$\Sigma_1 = \frac{1}{4} \begin{bmatrix} 0.49 + 0.04 + 0.09 + 0.64 & 0.49 - 0.06 - 0.06 + 0.64 \\ \cdot \quad '' & \end{bmatrix}$$

National Logistics Cell

SM (HR

$$\mathcal{E}_2 = \begin{bmatrix} 2\times 2 \text{ matrix} \end{bmatrix}.$$

$$x = \begin{bmatrix} 15 \\ 7 \end{bmatrix}$$

$$P(\omega_i|x) = \frac{P(x|\omega_i)\,P(\omega_i)}{P(x)}$$

$$\begin{array}{cccccc} & f_1 & f_2 & f_3 & f_4 & \\ \omega_1 = & \begin{bmatrix} 13 \\ 5 \end{bmatrix} & \begin{bmatrix} 13.5 \\ 6 \end{bmatrix} & \begin{bmatrix} 14 \\ 5.5 \end{bmatrix} & \begin{bmatrix} 14.5 \\ 6.5 \end{bmatrix} & \begin{matrix} \to \text{length} \\ \to \text{height} \end{matrix} \end{array}$$

2D-dimensional

$$\omega_2 = \begin{bmatrix} 16.5 \\ 7 \end{bmatrix} \begin{bmatrix} 16 \\ 7.5 \end{bmatrix} \begin{bmatrix} 17.5 \\ 8.5 \end{bmatrix} \begin{bmatrix} 17 \\ 8 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 13.7 \\ 5.7 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 16.7 \\ 7.7 \end{bmatrix}$$

$$x_k - \mu_2 = \begin{bmatrix} 16.5-16.7 & 16-16.7 & 17.5-16.7 & 17-16.7 \\ 7-7.7 & 7.5-7.7 & 8.5-7.7 & 8-7.7 \end{bmatrix}$$

$$= \begin{bmatrix} -0.2 & -0.7 & 0.8 & 0.3 \\ -0.7 & -0.2 & 0.8 & 0.3 \end{bmatrix}$$

$$(x_k - \mu_2)^t = \begin{bmatrix} -0.2 & -0.7 \\ -0.7 & -0.2 \\ 0.8 & 0.8 \\ 0.3 & 0.3 \end{bmatrix}$$

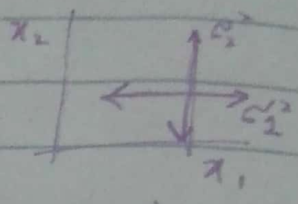
$$= \begin{bmatrix} -0.2 & -0.7 & 0.8 & 0.3 \\ -0.7 & -0.2 & 0.8 & 0.3 \end{bmatrix} \begin{bmatrix} -0.2 & -0.7 \\ -0.7 & -0.2 \\ 0.8 & 0.8 \\ 0.3 & 0.3 \end{bmatrix}$$

$$= \begin{bmatrix} \cancel{-0.2}\ +0.04+0.49+0.64+0.09 & +0.14+0.14+0.64+0.09 \\ +0.14+0.14+0.64+0.09 & +0.49+0.04+0.64+0.09 \end{bmatrix}$$

$$\Sigma = \frac{\begin{bmatrix} 1.26 & 1.01 \\ 1.01 & 1.26 \end{bmatrix}}{n}$$

$$\Sigma = \begin{bmatrix} 0.315 & 0.2525 \\ 0.2525 & 0.315 \end{bmatrix}$$

same spread
forms a circle.



σ^2

spread

$$P(x \mid w_j) = \left\{ \frac{1}{(2\pi)^{2/2} \, |\Sigma_j|^{1/2}} \quad \cdots \right\}$$

test sample $x = \begin{bmatrix} 15 \\ 7 \end{bmatrix}$

$-\frac{1}{2}(x_k - \mu_1)(\text{inv. cov})(x - \mu)^t$

$\frac{1}{(2\pi)^{d/2} |\text{cov}|^{1/2}} e$

$$= \frac{1}{(2\pi)^{2/2} \begin{vmatrix} 0.315 & 0.25 \\ 0.25 & 0.315 \end{vmatrix}^{1/2}} e^{-\frac{1}{2}\begin{bmatrix} 15-16.7 & 7-7.7 \end{bmatrix} \begin{bmatrix} 8.75 & -6.94 \\ -6.94 & 8.75 \end{bmatrix}}$$

$$= \frac{1}{(2\pi)^1 (0.1913)} e^{-\frac{1}{2}\begin{bmatrix} -1.7 & -0.7 \end{bmatrix} \begin{bmatrix} 8.75 & -6.94 \\ -6.94 & 8.75 \end{bmatrix} \begin{bmatrix} -1.7 \\ -0.7 \end{bmatrix}}$$

$$= \frac{1}{0.201} e^{-\frac{1}{2}[\ \cdots\ ]}$$

= value. → likelihood

$\det|\Sigma_2| =$

$$\begin{bmatrix} 8.75 & -6.94 \\ -6.94 & 8.75 \end{bmatrix}$$

$\det|\Sigma_2| = 0.036$

$d \times d$
covariance matrix

Naive Baye's Classifier:-

$$P(C \mid X) = \frac{P(X \mid C)\, P(C)}{P(X)}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

MAP: Maximum A Posterior

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$P(C|X) = P(X|C)P(C)$

$P(C|X) = P(x_1, x_2, x_3 \ldots x_n | C) P(C)$

All variables are independeduet.

posterior // $= [P(x_1|C) P(x_2|C) \ldots P(x_n|C)] P(C)$

multivariate $\rightarrow$ univariate (easy).

$P(Yes) = \frac{9}{14} = 0.64$ $P(No) = 0.36$

Outlook	Play = Yes	Play = No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5
Total	P = 1	P = 1

$P(Outlook / Yes) = 4/9$ $P(Outlook/No) = 0/5$

$P(S/Y) = 2/9$ $P(S/N) = 3/5$

$P(R/Y) = 3/9$ $P(R/N) = 2/5$

Temp	Play = Yes	Play = No
Hot	2/9	2/5
mild	4/9	2/5
Cool	3/9	1/5
Total	1	1

outlook = Sunny
Temp = Cool
Humidity = High
wind = strong

Sunny Hot High strong.

$P(Yes|x) = P(O=S/Y)P(T=C/Y)P(H=H/Y)P(W=S|Y)P(Yes)$

$= (2/9 \times 3/9 \times 3/9 \times 3/9)(9/14)$

$= 0.0053$

$P(No|x) = P(O=S|N)P(T=C|N)P(H=H|N)P(W=S|N)P(N)$

$= 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14$

$= 0.0206$

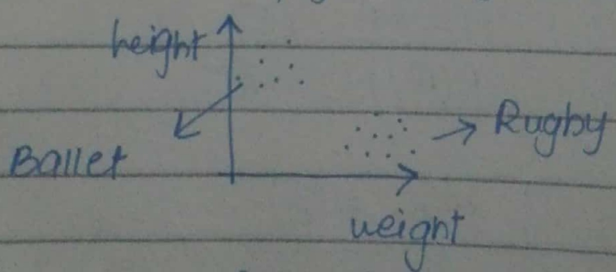
$P(Yes|x) < P(No|x)$

Therefore, No play.

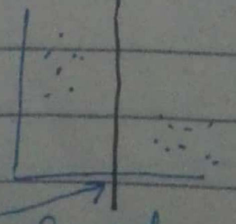
3rd January, 2018. Lecture # 24

Perceptron:-

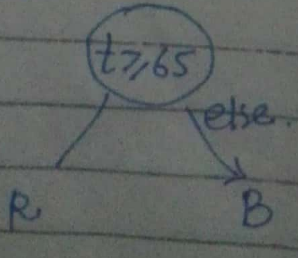
Used to classify only two classes.



Keep on moving the line until there's no misclassification



Threshold value t found.

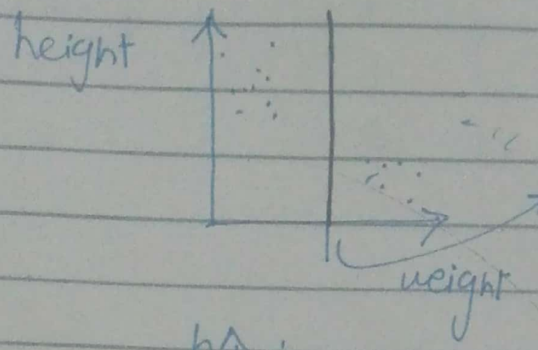


Bullet
Belly.

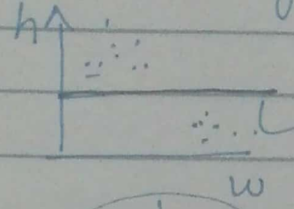
Model:- parameter setting. Any test sample passes through that model.

Error function:: which finds error.

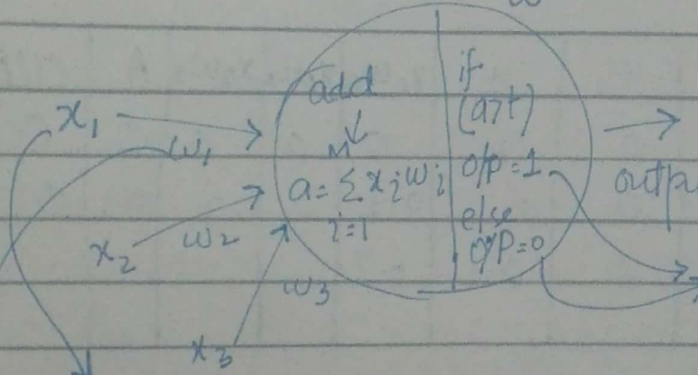
Learning Algo:: Optimal value of parameter.



we are considering ~~value of only~~ one feature i.e, weight



gives value of height.



tell which is the class.

Lets say $t = 0.5$

feature values

contribution of weights from learning algo.

randomly initialized.

original label

$$w_{new} = w_{old} + 0.1 \times (\text{target} - \text{output}) \times \text{input}\,(x_1)$$

learning rate → through which weight update (normal update).

also a label

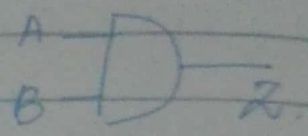
if (target - output) = 0 → it means weights are correct.

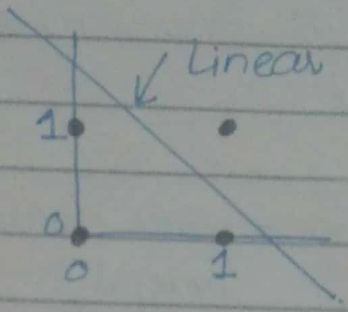
@ same target & output, no updation of weights.

Example NAND Gate:

Bias input. can be +ve/ -ve

x_0	x_1	x_2	z = label.
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0





threshold = 0.5

x_0	x_1	x_2	z	w_0	w_1	w_2	x_0w_0	x_0w_1	x_0w_2	A	O/P
1	0	0	1	0	0	0	0	0	0	0	0
1	0	1	1	0.1	0	0	0.1	0	0	0.1	0
1	1	0	1	0.2	0	0.1	0.2				
1	1	1	0								

If (A > t)
O/P = 1
else O/P = 0

1st Iteration:

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0.1 \times (1-0) \times \begin{bmatrix} 1 & x_0 \\ 0 & x_1 \\ 0 & x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0.1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix} + 0.1 \times (1-0) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0 \\ 0.1 \end{bmatrix}$$

Lab #14 (Perceptron).

Single Layered Feed forwarded/Perceptron
feature / input x weight = activation (single value).
~~actior~~ activation = threshold.
if a>t then fire output.
weights increase. iteratively through an algo
unless output = target.
This is learning phase.
* always generates a linear boundary which continuously adjusts itself.
* This is solvable if the problem is linearly separable.

9th Nov, 2018. Lecture # 25

stop when error = 0
or weights become equal.

-: Decision Trees:-

if data is not linearly separable.
in b/w input and O/P there is a hidden layer.
Another way of making decision
- metric method. (numeric values).
- Non-method (non-numeric values)

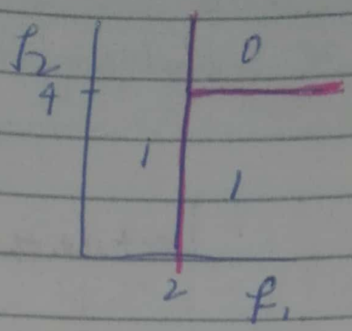
→ This is used.
→ Just like naive Bayes, can work on both numeric & non-numeric data
→ Fruit classification / weather / Credit.

Color	Size	Shape	Taste	Class.
Green	-	-	-	Apple
-	-	-	-	

→ Queries based on feature
Classes /

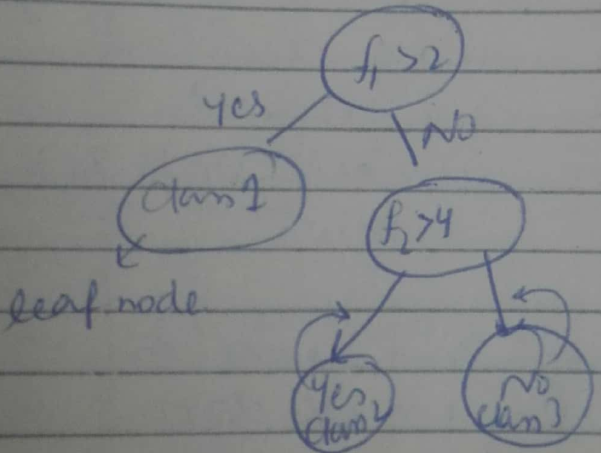
Training data → build a decision tree
Binary tree will be here: BF = 2

→ Keep on asking question until you reach the end.
→ Yes to the left & no to the right.
→ leaf nodes @ extreme end.
→ CART: Classification & Regression Trees
→



f_1	f_2	class
1	3	1
4	2	1
4	5	0

$f_1 > val$ → threshold $f_2 > val$



Entropy:- Randomness / disorder.
⇒ impure data → all class mixed → more Entropy
you purse your data and classify to make entropy = 0
Best Case: Pure data.
Entropy Impurity:-

$$i(N) = -\sum_{i \in c} P(w_i) \log_2 P(w_i)$$

① $\omega_1 \to 5$ $\qquad \omega_2 \to 5$

$i(N) = -(P(\omega_1)\log_2(\omega_1) + P(\omega_2)\log_2 P(\omega_2))$

$= -\left(\frac{5}{10}\log_2\frac{5}{10} + \frac{5}{10}\log_2\frac{5}{10}\right)$

E.I = 1

② $\omega_1 = 0$ $\qquad \omega_2 = 10$

$i(N) = -\left(\frac{0}{10}\log_2\frac{0}{10} + \frac{10}{10}\log_2\frac{10}{10}\right)$

E.I = 0 = data is pure/ properly classified

—— x —— x ——

if one class maximum and other min or null → none impurity

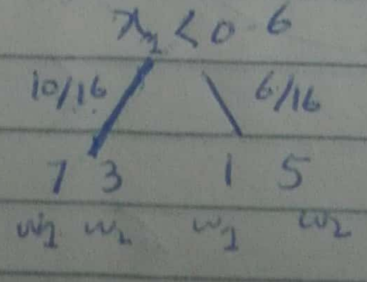
Query selection criterion

Example:

$i(N) = -(8/16\log_2 8/16 + 8/16\log_2 8/16)$

= 1 maximum impurity

$\max(\Delta i(N)) \to \max \to$ root node query



$\Delta i N = i(N) - (P_L\, i(N_L) + P_R\, i(N_R))$

$$i(N_L) = -\left(\frac{7}{10}\log_2\frac{7}{10} + \frac{3}{10}\log_2\frac{3}{10}\right)$$

$$i(N_L) = 0.88$$

$$i(N_R) = \left(\frac{1}{6}\log_2\frac{1}{6} + \frac{5}{6}\log_2\frac{5}{6}\right)$$

$$= 0.65$$

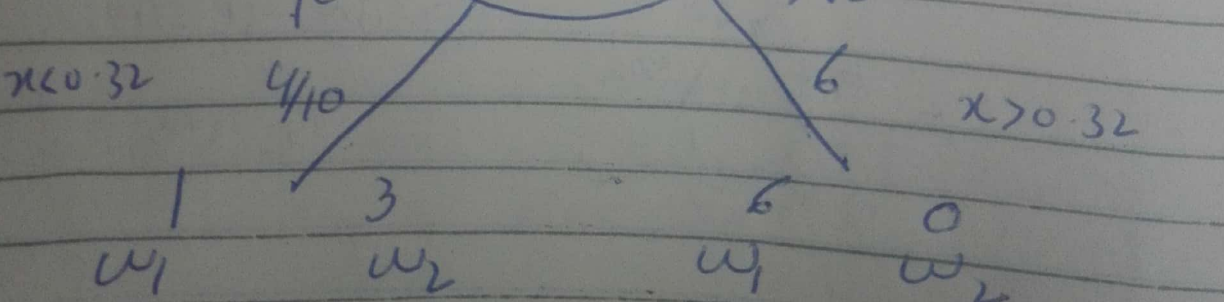
$$\Delta i_N = 1 - \left(10/16 \times 0.88 + 6/16 \times 0.65\right)$$

$$= 0.206$$

If three values are given then reach to this step and choose that value which gives maximum Δi_N.

<u>2nd Iteration:</u>

already selected samples then go for further selection



-tPw $-(\omega_1 P(\omega_1) + \omega_2 P(\omega_2))$

$i(N_L) = (\frac{1}{4} \log \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4})$

$i(N_L) = 0.81$

$\Delta(i(N_L)) = \underset{\text{previous}}{\underline{0.88}} - \left(\frac{4}{16} \times 0.81 + \frac{6}{10} \times 0\right)$

$\Delta(i(N_L)) = 0.88 -$ ——

$\Delta(i(N_L)) = 0.55$

Make it complete.

$i(w_L) = \left(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right)$

$i(N_L) = 0.81$

$\Delta(i(N_L)) = \underset{\downarrow \text{previous}}{\underline{\underline{0.88}}} - \left(\frac{4}{16} \times 0.81 + \frac{6}{10} \times 0\right)$

$\Delta(i(N_L)) = 0.88 - \text{\_\_\_\_}$

$\Delta(i(N_L)) = 0.55$

Make it complete.

—— x —— x —— x ——

Decision Trees:-

Lec#
16th Jan, 2017.

$x_1 < 0.35$

$x_2 < 0.32$

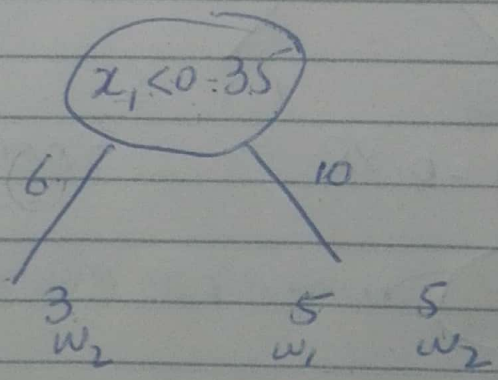
$x_1 < 0.6$

$x. < 0.61$

$x_1 < 0.69$

w_1	w_2
8	8

Taking 0.35:



$\Delta i_N = i(N) - (P_L\, i(N_L) + P_R\, i(N_R))$

$i(N) = 1$

$= 1 - (6/10$

$i(N_L) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{5}{10}\log_2\frac{5}{10}\right)$

$= -$

$(i N_L) = -\left(\frac{3}{6}\log_2 \frac{3}{6} + \frac{3}{6}\log_2 \frac{3}{6}\right)$

$= 1$

$(i N_R) = -\left(\frac{5}{10}\log_2 \frac{5}{10} + \frac{5}{10}\log_2 \frac{5}{10}\right)$

$= 1$

$\Delta i(N) = i(N) - (P_L\, i(N_L) + P_R\, i(N_R))$

$= 1 - \left(\frac{6}{16}(1) + \frac{10}{16}(1)\right)$

$= 1 - 1 = 0.$

0.31
$\Delta i(N) = 0$
$\Delta i(N) = 0.$
$\downarrow$
0.6

Now taking for 0.6

$\Delta(i_N) = 1$

$x_1 < 0.6$

10/16   6/16

7   3   1   5

$i(N_L) = 0.88$ $i(N_R)$

$\Delta i(N) = 0.88$

$x_2 < 0.32$ → Now rest for these 10

4/10   6/10

1   3   6   0

ω_1 entropy=0

$\Delta i(N) = i(N) - (P_L\, i(N)$
$+ P_R\, i(N_R))$

$i(N) = -\left(\frac{1}{4}\log \frac{1}{4} + \frac{3}{4}\log_2 \frac{3}{4}\right)$ $i(N_R) = 0$

$= (-0.5 + (-0.311))$

$= 0.81$

$$i(N) = \left(\frac{4}{10}\log_2 \frac{4}{10} + \frac{6}{10}\log_2 \frac{6}{10}\right)$$

$$= \cancel{0.8} - (1.32) - 0.44$$

$$= 1.7$$

$$\Delta(i_N) = 1.7 - \left(\frac{1}{4}(0.8) + \frac{6}{10}(0)\right)$$

$$= 1.7 - 0.2$$

$$= 1.5$$