# Bayes Decision Problem

↓

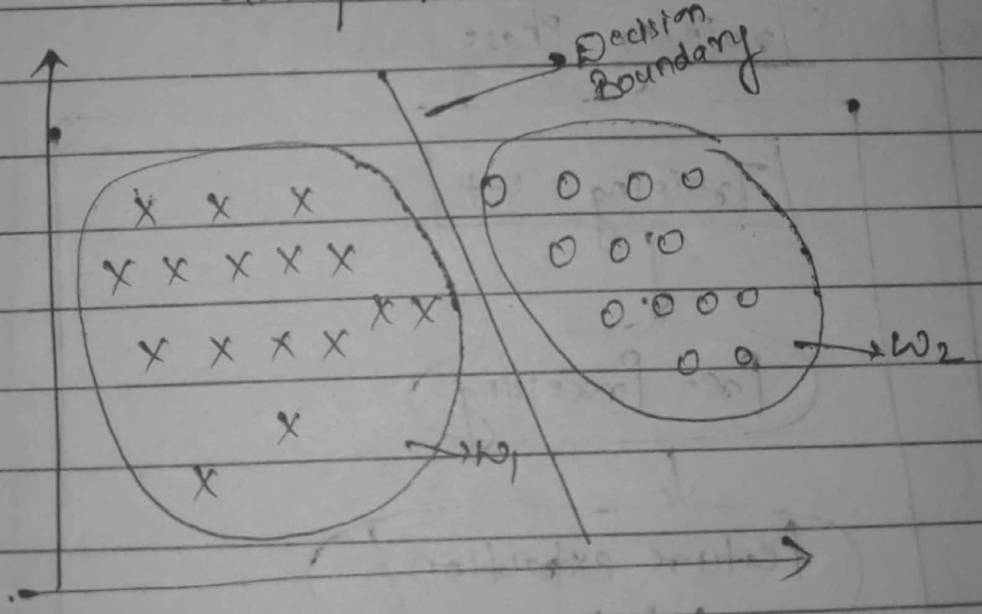
Statistical Pattern Recognition

→ Many of the features taken together, they can describe an object to some degree of accuracy. These features are called feature vectors.

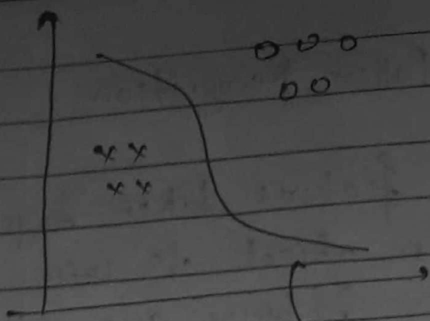
→ They had to be concatenated in a particular order & whichever order we concatenate them throughout our model is modelling problem as well as of the pattern as well as recognition of pattern.

→ When, I put all these diffⁿ features in a particular order, they are called feature vector.

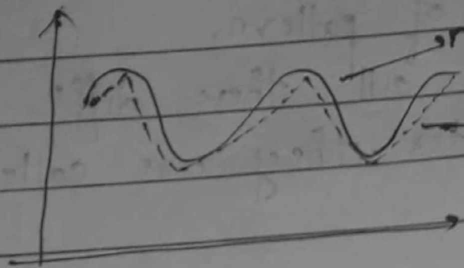
# Dimension → no. of features considered for classification.




→ Clear, which object belongs to which class if decision boundary is obtained.



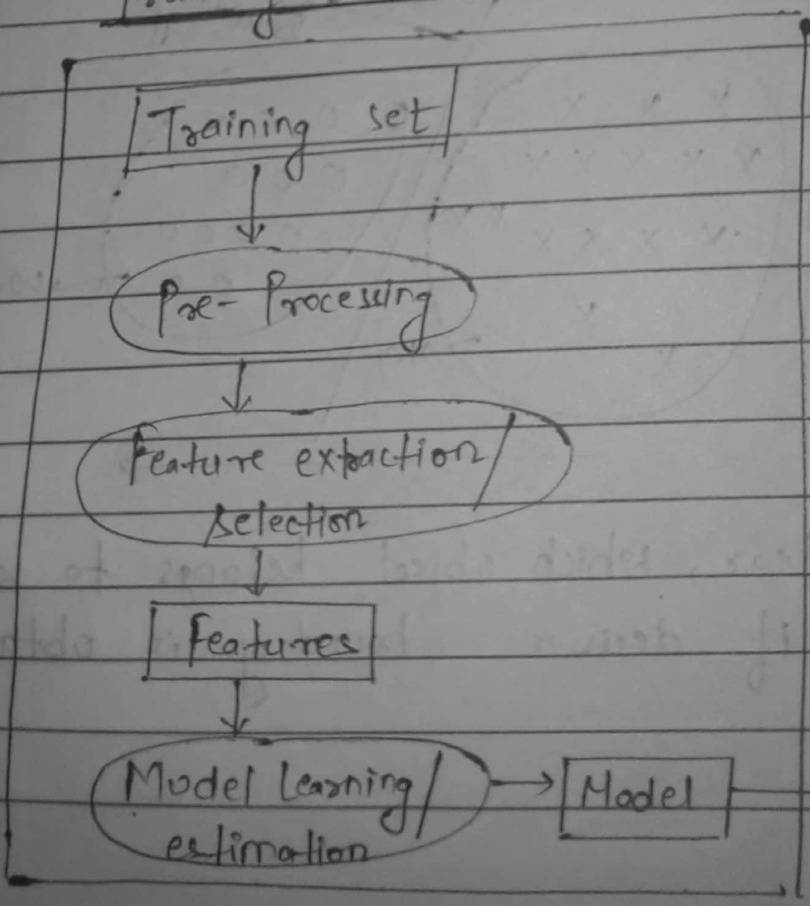
non-linear boundary
(may be cubic or quadratic)
↓
above this degree,
we won't be able to perform
well.



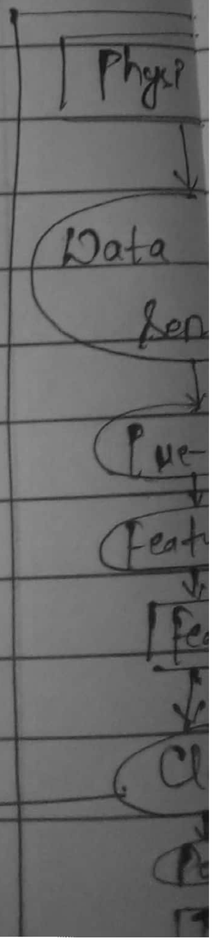
non-linear decision boundary

approximate, to
ease calculation &
make it reliable.

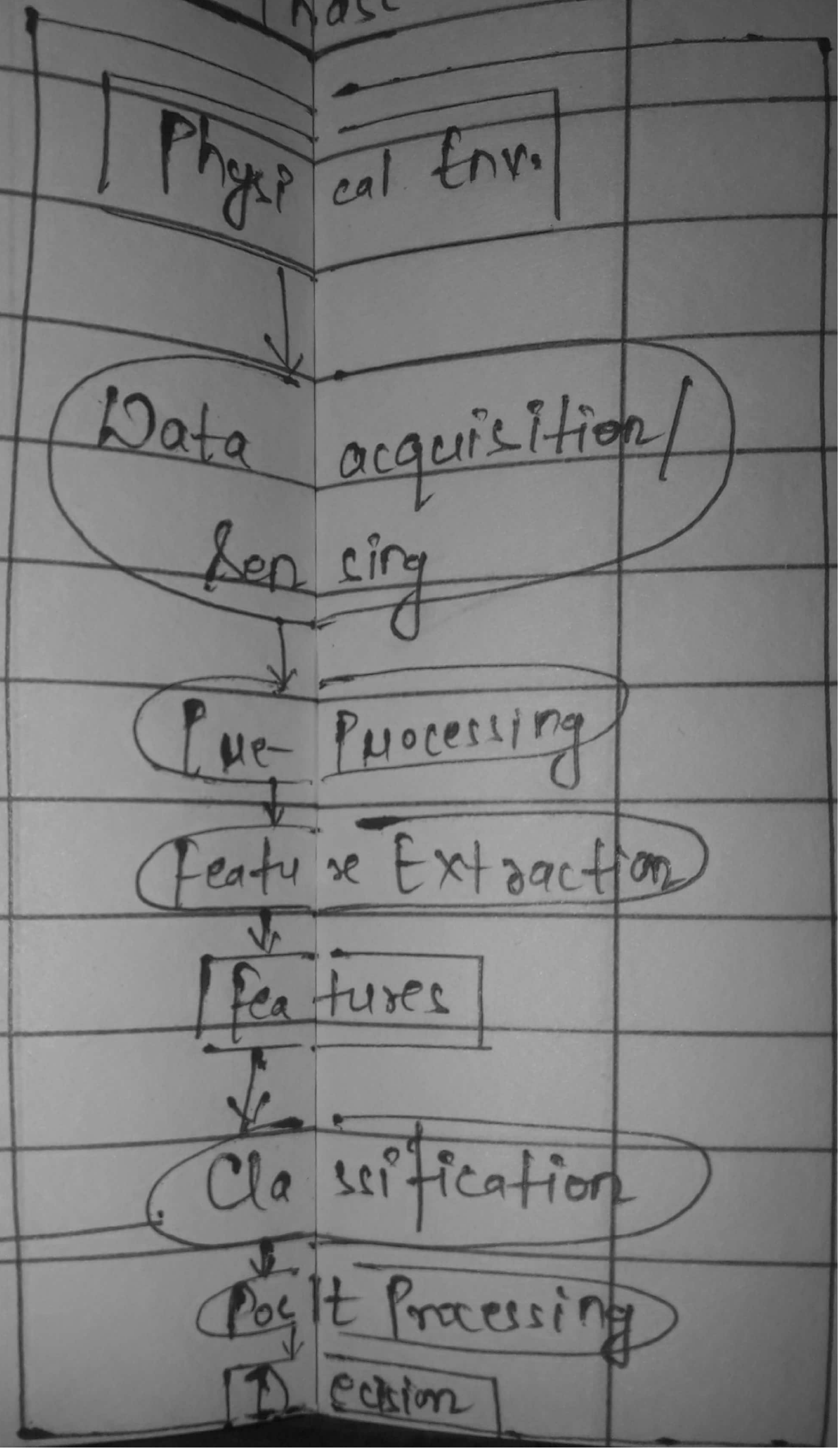
## ## Training Phase



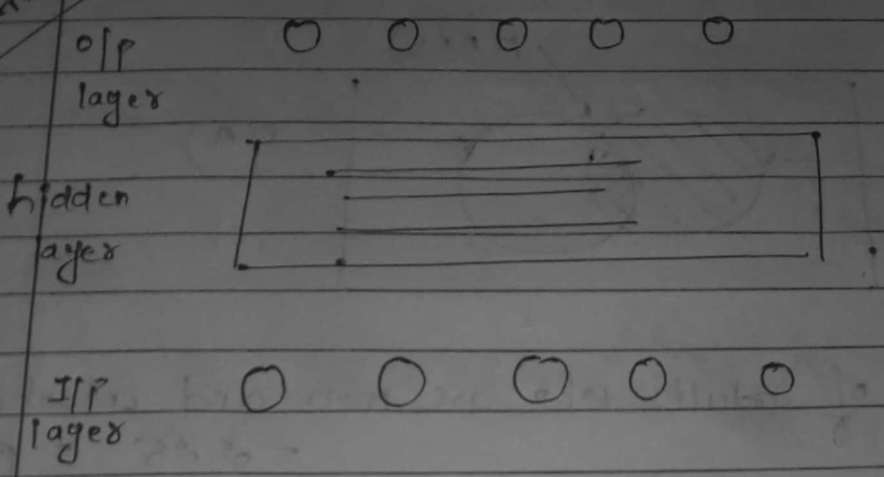

Test Ph



## Test Phase


Physical Env.
Data acquisition/ Sensing
Pre-Processing
Feature Extraction
Features
Classification
Post Processing
Decision

Structure of neural network.




$$\boxed{O_i = f\left(\sum w_{ij} x_i\right)}$$

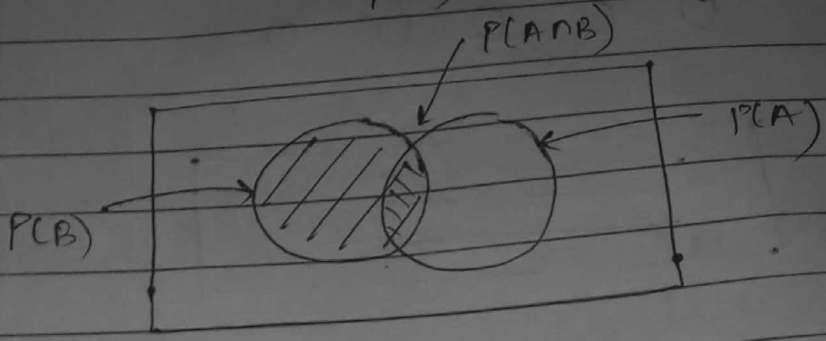
→ Overlapping boundaries ~~are~~ makes it tough to be distinguished for patterns.

(neither linear nor non-linear)

⇒ (Multilayer Perceptron Problem)

# Conditional Probablity :-

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$




Eg. The % of adult who are man and alcoholic $= 2.25 = P(A \cap M)$

Men and women are assumed to be equal.

$$P(A/M) = \frac{P(A \cap M)}{P(M)} = \frac{2.25}{100 \times 0.5} = \frac{0.0225}{0.5}$$

$$= \frac{\cancel{225}^{45} \times 10}{\cancel{5} \times 1000}$$

$$= \underline{0.045}$$

# Baye's Theorem

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Now, $\boxed{P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}}$

Page No.
Date : / /

Q. What is the probablity of 2 girls in a couple. when given atleast, one is girl.

$P(2G / \text{atleast one is } G) = \dfrac{P(\text{atleast one is } G / 2G) \times P(2G)}{P(\text{atleast one is } G)}$

2G

$\rightarrow \frac{1}{3}$

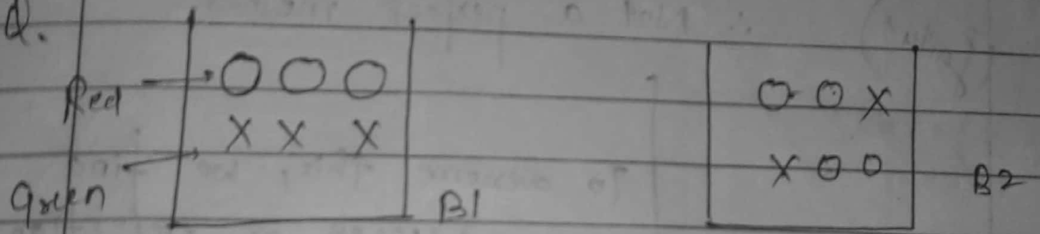
$= \dfrac{1 \times \frac{1}{4}}{3/4}$

GB
BG
GG
BB

possible combination.

$= \frac{1}{3}$

Q.




If, I randomly draw a green ball, What is the probablity of having it in B1.

$P(B1 / G) = \dfrac{P(G / B1) \times P(B1)}{P(G)}$

$= \dfrac{\frac{1}{2} \times \frac{1}{2}}{\left(\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{2}{6}\right)} = \dfrac{\frac{1}{4}}{\left(\frac{1}{4} + \frac{1}{6}\right)}$

$= \dfrac{\frac{1}{4}}{\frac{6+4}{24}} = \frac{1}{4} \times \frac{24}{10} = \frac{3}{5}$

Page No.
Date : / /

# Bayes Decision Theory

$\omega_1$ — Accept — $P(\omega_1)$

$\omega_2$ — Reject — $P(\omega_2)$

$P(\omega_1) > P(\omega_2) \Rightarrow \omega_1$

$P(\omega_1) < P(\omega_2) \Rightarrow \omega_2$

(a priority probablity)

# if (.90 is success probablity, .10 is rejection ")

$\Rightarrow$ machine will always accept in this case. $\because P(\omega_1) > P(\omega_2)$

(failure) $\therefore$ Not a perfect method.

To overcome this, we find success or rejection based on some feature using feature vector.

Given class:—

feature is known & we need to find class.

$P(x|\omega_1)$ $P(x|\omega_2)$

Class Cond^n PDF (Probablity Distribution function)

$P(\omega_i, x) = P(\omega_i|x).P(x)$

$= P(x/\omega_i).P(\omega_i)$

$= P(\omega_i \cap x)$

Page No.
Date: / /

$$P(\omega_i/x) \cdot P(x) = P(x/\omega_i) \cdot P(\omega_i)$$

Bayes form.

$$\Rightarrow \boxed{P(\omega_i/x) = \frac{P(x/\omega_i) \cdot P(\omega_i)}{P(x)}}$$

if, $x$ i.e. a feature is given then we need to find its class.

[In current Eg. classes = 2
feature = 1]

$$P(x) = \sum_{i=1}^{2} P(x/\omega_i)\, P(\omega_i)$$

** {in general, $i \in [1 \text{ to } n]$ } — no. of classes
Here, ($n=2$)

1. $P(\omega_1/x) > P(\omega_2/x) \longrightarrow \omega_1$ (accept)
2. $P(\omega_1/x) < P(\omega_2/x) \longrightarrow \omega_2$ (reject)

Now,

1. $$\frac{P(x/\omega_1)\, P(\omega_1)}{P(x)} > \frac{P(x/\omega_2)\, P(\omega_2)}{P(x)}$$

$$\Rightarrow \boxed{P(x/\omega_1)\, P(\omega_1) > P(x/\omega_2)\, P(\omega_2)}$$

presence of 'x' in '$\omega_1$' bucket

presence of 'x' in '$\omega_2$'

already learned at the time of training

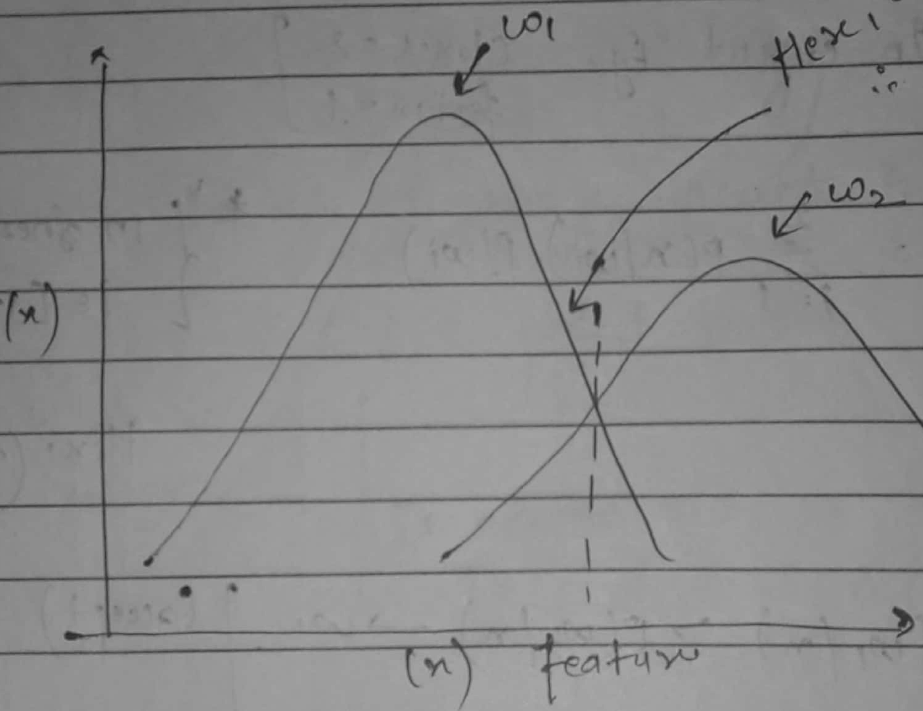
Now, assume, $\boxed{P(\omega_1) = P(\omega_2)}$
in ideal case.

$Eq^n$ reduces to,

$$\boxed{P(x|\omega_1) > P(x|\omega_2)} \rightarrow \omega_1 \text{ (accept)}$$

* completely depends on feature 'x' i.e. learning of training set.

## Error




Here, $\omega_1 > \omega_2$ ∴ we put it in '$\omega_1$' but '$\omega_2$' is also present ⇒ There is some error

✓ Error when :- x is placed in '$\omega_1$' $\boxed{P(\omega_2/x)}$

✓ Error when :- x is placed in '$\omega_2$' $\boxed{P(\omega_1/x)}$

Page No.
Date : / /

(overall, error)

$$P(error) = \int_{-\infty}^{\infty} P(error, x)\, dx$$

$$\boxed{P(error) = \int_{-\infty}^{\infty} P(error, x) \cdot P(x)\, dx.}$$

$$P(error|x) = \min\{P(\omega_1/x),\ P(\omega_2/x)\}$$

| $\omega_1$ | $\omega_2$ |
|---|---|
| $\downarrow$ | $\downarrow$ |
| accept | Reject |
| $P(\omega_1)$ | $P(\omega_2)$ |

$$P(\omega_1) > P(\omega_2) \Rightarrow \omega_1$$
$$\Rightarrow \omega_2$$

$$\underbrace{P(x/\omega_1) \qquad\qquad P(x/\omega_2)}$$

$\downarrow$ Class Conditional PDF (Probablity Density Function).

$P(\omega_1/x)$ $\qquad$ $P(\omega_2/x)$

$$P(\omega_1, x) = \frac{P(x/\omega_1) \cdot P(\omega_1)}{P(x)}$$

Page No.
Date: / /

$$P(\omega_2/x) = \frac{P(x/\omega_2)\,P(\omega_1)}{P(x)}$$

$$P(x) = \sum_{i=1}^{2} P(x/\omega_i) \cdot P(\omega_i)$$

$$\boxed{P(\omega_i/x) = \frac{P(x/\omega_i) \cdot P(\omega_i)}{P(x)}}$$

a posteriori probability.

$$\frac{P(x/\omega_i)\,P(\omega_i)}{\cancel{P(x)}} > \frac{P(x/\omega_j)\,P(\omega_2)}{\cancel{P(x)}} \Rightarrow \omega_1$$

# Priori Prob.

Que

| | | | | |
|---|---|---|---|---|
| 1. | Quality of Course | good | fair | bad |
| 2. | Prob (Priori) | 0.2 | 0.4 | 0.4 |
| 3. | Class ~~Class~~ Conditionals $P(x/\omega_i)$ | good | fair | bad |
| 4. | Interesting | 0.8 | 0.5 | 0.1 |
| 5. | Boring | 0.2 | 0.5 | 0.9 |

Page No.
Date : / /

Loss function :-

| | $P[\lambda(a_i/w_i)]$ | good | fair | bad |
|---|---|---|---|---|
| 1. | | | | |
| 2. | Taking | 0 | 5 | 10 |
| 3. | Not taking | 20 | 5 | 0 |

# Generalisation :-

* Use more than 2 states of nature
  ↳ classes

* Use more than 1 features
  ↳ feature vector

* Allow other action other than merely deciding states of nature

* Introduce a loss function more general than Probablity of error.

$C_1 \rightarrow$ states of nature

$\omega_1, \omega_2, \omega_3 \cdots\cdots \omega_c$

$a \rightarrow$ Actions

$\{\alpha_1, \alpha_2, \cdots, \alpha_a\}$

loss function $\rightarrow$

$\lambda(\alpha_i / \omega_j) \rightarrow$ loss incurred error taking action $\alpha_i$ when true state of nature is $\omega_j$.

Page No.
Date : / /

$X \Rightarrow d$ - dimensional feature vector

Now,

$X$

Action $\alpha_i$

$$R(\alpha_i/x) = \sum_{j=1}^{c} \lambda(\alpha_i/\omega_j) \, P(\omega_j/x)$$

function

NOTE:-

* action with least risk factor will be the desired one.

Risk Function / Conditional Risk / Expected loss

* (It is calculated for all values of $\alpha_i$)

$\Rightarrow$ Min Risk Classifier

Eg. Class :- $\omega_1$ $\omega_2$

Action:- $\alpha_1$ $\alpha_2$

$$\lambda(\alpha_i/\omega_j) = \lambda_{ij}$$

$$\boxed{R(\alpha_i/x) = \sum_{j=1}^{c} \lambda(\alpha_i/\omega_j) \, P(\omega_j/x)}$$

$$\rightarrow R(\alpha_1/x) = \sum_{j=1} \lambda_{11}(\alpha_1/\omega_1) P(\omega_1/x) + \lambda_{12}(\alpha_1/\omega_2) . P(\omega_2/x)$$

$$\Rightarrow \boxed{R(\alpha_1/x) = \lambda_{11} P(\omega_1/x) + \lambda_{12} P(\omega_2/x)}$$

Page No.
Date : / /

$\longrightarrow R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) + \lambda_{22} (P(\omega_2|x))$

* We assumed that $R(\alpha_2|x) > R(\alpha_1|x) \longrightarrow$ we should select '$\omega_1$'.

Now,

$$\boxed{(\lambda_{21} - \lambda_{11}) \cdot P(\omega_1|x) > (\lambda_{12} - \lambda_{22}) P(\omega_2|x)}$$

belongs to wrong class $\alpha_2$ but $\omega_1$ $\Rightarrow$ +ve.

belong to same $\Rightarrow 0$

$\simeq 1y$

$\Rightarrow (\lambda_{21} - \lambda_{11}) > 0$

$$P(\omega_1/x) > P(\omega_2/x) \Rightarrow \omega_1$$

$$(\lambda_{21} - \lambda_{11}) P(\omega_1|x) > (\lambda_{12} - \lambda_{22}) P(\omega_2|x)$$

$$\xrightarrow{\text{LR}} \omega_1$$

# Min Error Rate Classification :-

$\alpha_i \rightarrow$ True state of nature is $\omega_i$

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & ; i=j \\ 1 & ; i \neq j \end{cases} \quad \forall\, i,j = 1,2,\ldots,c$$

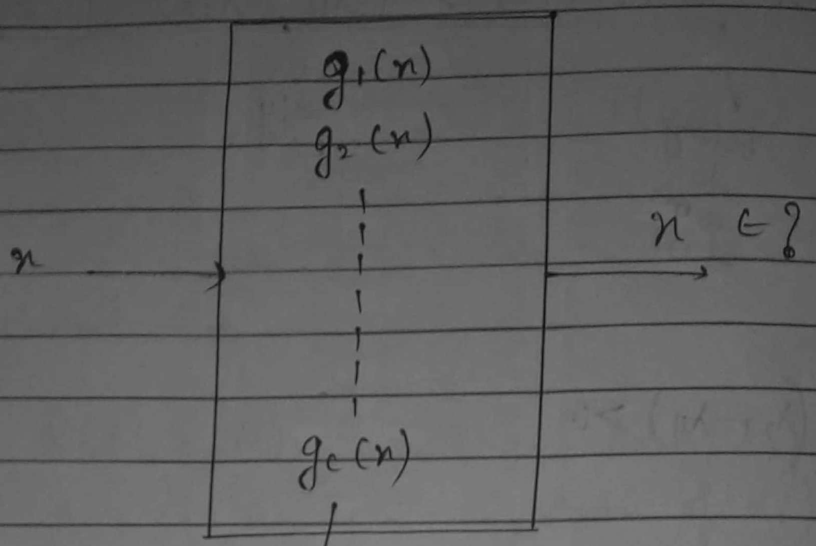
$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x)$$

$$= \boxed{\sum_{i \neq j} P(\omega_j|x)} \Longrightarrow$$ for this case, $\lambda(\alpha_i|\omega_j) = 1$
for, $i=j$, $\lambda(\alpha_i|\omega_j) = 0$.

Page No.

Date : / /

$= \boxed{1 - P(\omega_i/x)}$ → easier to calculated probability of one.




c is the no. of classes

(CLASSIFIER)

$g(x) \rightarrow$ Discriminant $F^n$

$\omega_1, \omega_2, \omega_3, \text{----}, \omega_c \rightarrow$ No. of classes

$g_i(x) : \quad i = 1 \text{ ---- } c$

$\boxed{g_i(x) > g_j(x)} \quad \forall j \neq i$

$\Rightarrow x \in \omega_i$

(min risk will always be -ve)

* Min Risk Classifier

$\boxed{g_i(x) = -R(\alpha_i/x)}$

∵ we take risk to be min. ∴ -ve of that will be max

Page No.
Date : / /

## * Min Error Rate

$$\sum_{i\neq j} P(\omega_i/x) \longrightarrow 1-P(\omega_i/x)$$

$$g_i(x) = \sum_{i\neq j} P(\omega_i|x) = 1-P(\omega_i|x)$$

$$g_i(x) = -R(\alpha_i|x)$$
$$= -(1-P(\omega_i|x))$$
$$= -1 + P(\omega_i|x)$$

$$\boxed{g_i(x) = P(\omega_i|x)}$$

$$g_i(x) = P(\omega_i|x)$$
$$= \frac{P(x/\omega_i)\cdot P(\omega_i)}{\sum_{j=1}^{c} P(x/\omega_j)P(\omega_j)} \longrightarrow P(x)$$

* Imp
we have removed $P(x)$ i.e. denominator bcz. our main concern is comparison that's why we have removed denominator bcz. it is same for all.

$$\boxed{g_i(x) = P(x/\omega_i)P(\omega_i)}$$

- We have taken log bcz. product is a costly operation.

$$\Rightarrow \ln(g_i(x)) = \ln(P(x/\omega_i)) + \ln(P(\omega_i))$$

$$\text{if } (g_1(x) > g_2(x)). \Rightarrow \omega_1$$

$g_1(x) = g_2(x) \longrightarrow$ decision boundary.

$$g(x) = P(\omega_1/x) - P(\omega_2/x)$$
$$= P(x/\omega_1)\cdot P(\omega_1) - P(x/\omega_2)\cdot P(\omega_2)$$

-ve sign bcz. if we get +ve then $\omega_1$ else $\omega_2$.

Scanned by CamScanner

Page No.
Date : / /

again log both sides.

$$\ln(g(x)) = \ln\left(P(x/\omega_1)\cdot P(\omega_1) - P(x/\omega_2)\cdot P(\omega_2)\right)$$

$$= \ln\left(\frac{P(x/\omega_1)\cdot P(\omega_1)}{P(x/\omega_2)\cdot P(\omega_2)}\right)$$

$$= \ln\left(\frac{P(x/\omega_1)}{P(x/\omega_2)}\cdot\frac{P(\omega_1)}{P(\omega_2)}\right)$$

$$\boxed{\ln(g(x)) = \ln\left(\frac{P(x/\omega_1)}{P(x/\omega_2)}\right) + \ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right)}$$

Now,

Eg. Data set

| | | | | |
|---|---|---|---|---|
| 1 | R | Sports | Domestic | Yes |
| 2 | R | " | " | N |
| 3 | R | " | " | Yes |
| 4 | Y | " | " | N |
| 5 | Y | " | Imp | Yes |
| 6 | Y | SUV | " | N |
| 7 | Y | SUV | " | Yes |
| 8 | Y | SUV | Domestic | N |
| 9 | R | SUV | Imp | N |
| 10 | R | Sports | Imp | Yes. |

Page No.
Date: / /

Probablities :-

Color :

$P(R/Yes) = 3/5$ $P(R/No) = 2/5$

$P(Y/yes) = 2/5$ $P(Y/No) = 3/5$

Type :

$P(SUV/Yes) = 1/5$ $P(SUV/No) = 3/5$

$P(Sports/Yes) = 4/5$ $P(sports/No) = 2/5$

Origin :

$P(Dom/yes) = 2/5$ $P(Dom/No) = 3/5$

$P(Imp/yes) = 3/5$ $P(Imp/No) = 2/5$

$P(yes) = 5/10$

$P(No) = 5/10$

Q. (Red, SUV, Dom) → find, Yes Class or No Class.

$$P(yes/(Red, SUV, Dom)) = \frac{P(yes, Red, SUV, Dom)}{P(Red, SUV, Dom)}$$

$$\left(\frac{\cancel{3}}{5} / \frac{\cancel{2}}{8}\right)$$

$$P(No/(Red, SUV, Dom)) = \frac{P(No, Red, SUV, Dom)}{P(Red, SUV, Dom)}.$$

∵ Comparision is to be done, we remove denominator.

$$P(x/yes) \cdot P(yes) = P(Red/yes) \cdot P(SUV/yes) \cdot P(Dom/yes) \cdot P(yes)$$

$$= \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{2}{5}$$

Page No.
Date : / /

** ($\because P(yes) = P(No)$
$\therefore$ It has not been taken)

$P(x/No)\ P(No) = P(Red/No) \cdot P(SUV/No) \cdot P(Dom/No) \cdot P(No)$

$= \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot$

On comparing, ans will be No

**

Que Find optimal sol$^n$

$\Rightarrow$ Course should be (good, interesting)

$P((good, interesting)/taking) \cdot P(taking)$

$= P(good/taking) \cdot P(goo\ interesting/taking) \times P(taking)$

$= 0$

$P(good, interesting)/non\text{-}taking) \cdot P(non\text{-}taking)$

$= P(good/non\,taking) \cdot P(interesting/non\,taking) \cdot P(non\,taking)$

$= \cancel{0}$ x

Page No.
Date : / /

Basic structure of Pattern

↳ Data acq
→ Pre Processing
→ M/L estimation
→ Classification
→ Post Process

Baysian theory
→ Discrete (2 classes)
→ Continuous (more than 2 classes)

Two category classification
Cestimation of Posterior prob
Bayes Thm
conditional Risk
Risk Classifier
Discriminat fn