Any object that you see in the world forms a pattern
e.g.
Image of Car.
Image of cow. Bus

When we talk about speech signal.
if ~~we~~ I Say yes. This sound will capture with the Microphone. then if see ~~you~~
o O/P of Microphone on oscillascope
then it oscillating wave forms oscillascope ~~forms~~.

*. A pattern is something which describes what we see in real world & what we hear, what we sense so an electrical signal. Compartion of what we sense is what is we means in this case as a pattern.

* Job of pattern Recognition is.

A M/C ~~who~~ should be able to understand what we are seeing around us ~~aroun~~ ~~and~~ or ~~hear~~ ~~around~~ ~~us~~. and what we are speaking. that is what we mean by a pattern recognition. & we want a M/C to enable to do this.

Something something or speak
by something called this has to be repres
descriptive or features.

When these descriptive or featur
are addence in particular fashi
in a form of a vector which
is known as feature vector.

So for a set of pattern if I
have / compute say D no. of
such features & form addence
them in a form of vector it
becomes ~~form~~ ~~three~~ D dimension vector. That
means a pattern now represented
by ~~three~~ ~~dimension~~ ~~vector~~, a point
in D - Dimension space.

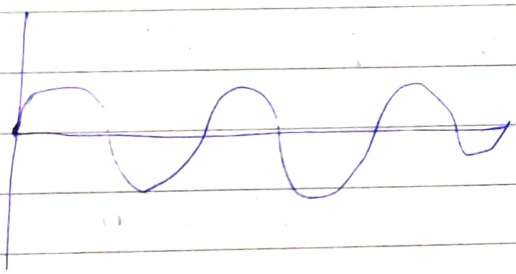
When I go for recognition

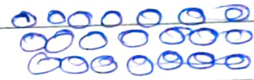
* When I say yes & the M/C
understands yes this yes is
represented by a ~~point~~ point in D- Dimensional
space and next time any word
is being represented by a similar such vector / point
in D- dimensional space. whether
that is yes or not I go in
D- dimensional space, try to compar

Visual pattern / Images
Spoken Pattern / words / Sentences.

The distance b/w Yes & New point if these ~~distance~~ are Same / identical then distance is Zero & Not identical then we find distance. if distance is very small then word appears to be yes otherwise it is Not.

Now, talk about pattern.





pattern

* The problem of pattern recognition is given some signal whether it is one dimensional or a two dimensional ~~or~~ ~~3~~ three dimensional signal we want a M/C to recognize ~~that~~ the P signal that is a problem of pattern recognition.

~~Now, look at a~~

The ~~past~~ problem of Pattern recognition is stated earlier with the efforts of ~~in~~ understanding Intelligence.

*. What is meant by Intelligence?

~~An~~ Intelligence ~~as~~ ~~an~~ the ability to comprehend or to understand and profit from experience.

~~exo~~ example of Sharp Obj pen. tip of pen or sharp will hurt us. it is previous experience & we learn we will not touch that tip

fire example. etc. {explain in your way}

* It is capability to acquire & applying knowledge.

example of picture :-
Can we recognize two diff. picture

i) Painting of ajanta caves.
ii) picture of Parliament house new delhi.

{explain in your lang. how ~~you~~ you recognize (previous experiences)}

* Plato -(427 to 347 BC) (2500 yr ago

The Concept of abstract ideas are known to us aapriore, through a Mystic Connection with world.

जो Knowledge पहले से हो apriori

e.

~~The~~

*. example of Taj Mahal & Bibi Ka Makbara (Aurangabad)

both looks almost same.

So, the concept which plato give is not sufficient, so we should adopt our knowledge.

The knowledge which we acquire should be incremental.

*. Aristotle (plato students)
(not fully satisfied with plato.

Data Analysis method that uses machine learning Algorithms to automatically recognize patterns & regularities in data.

Data can be anything from text & Images to Sound or other definable qualities.

*. Pattern recognition Systems can recognize familiar patterns quickly and accurately.

*. Applications :-
Image processing
Speech & fingerprint recognition
aerial photo interpretation, ~~opticals~~
optical Character recognition in scanned documents such as contracts & photographs, & even medical imaging & diagnosis.

*. Pattern recognition is also the technology behind data Analytics.

e.g.

The technique can be used to predict stock market.

Topic

* Bayes Decision Theory :- Min err. rate classification, Classifiers, Decision surfaces,

*. Supervised Learning :- given.

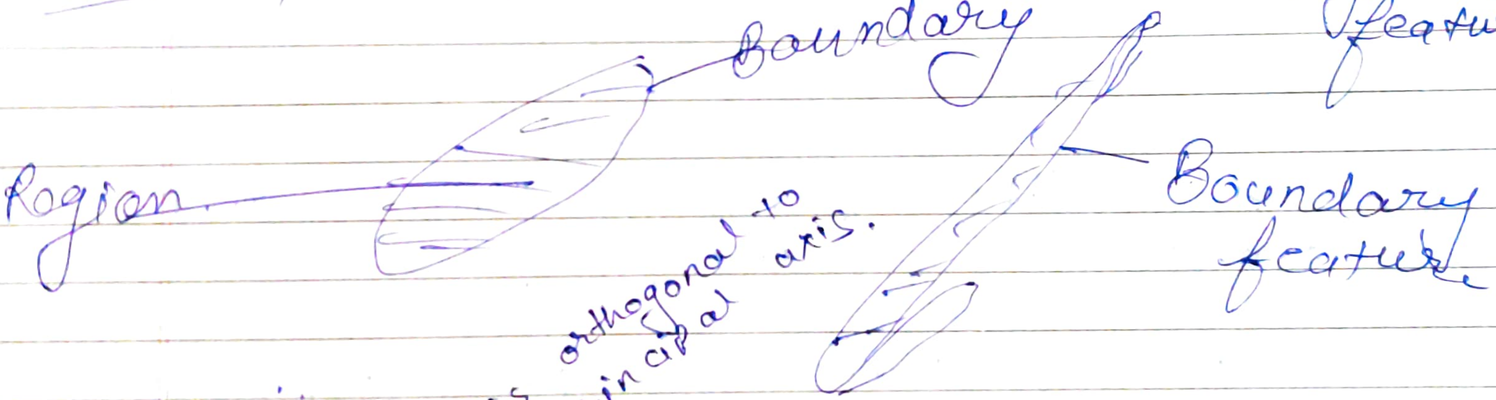
*. Unsupervised Learning.

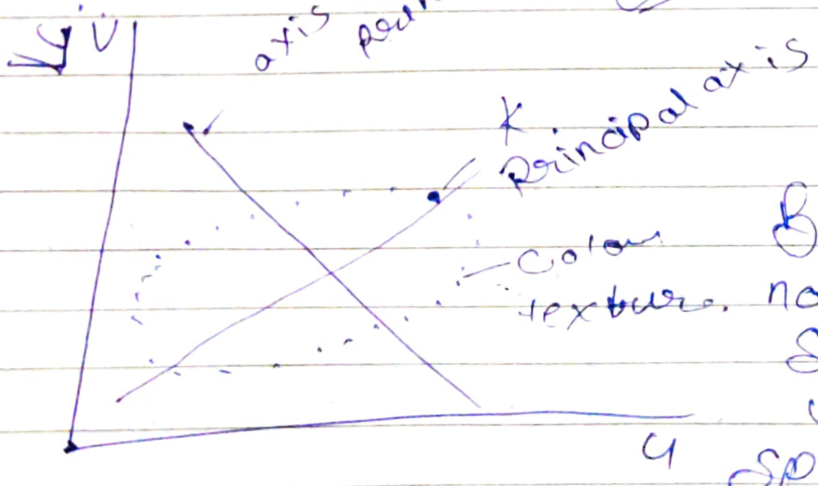
1) ~~Fea~~ we can identify on different types of feature.

1) feature Extraction

boundary features.

Region features.





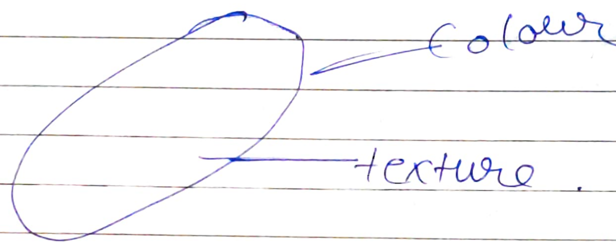
Boundary is nothing but a set of points in 2-dimension space.

Every point in boundary is represented by complex no.

$$S(K) = u(K) + jv(K)$$

K have different value on different points.

$S(K)$: $K = 0 \ldots\ldots N-1$



All diff. feature gives diff. Numerical values.

each values give some information

All values in particular order feature vector. M.

~~obje~~ct. ~~reco~~gnize c

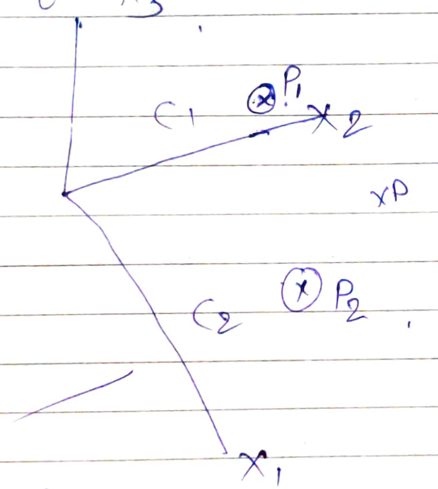
feature vector is statical vector M. then M dimensional vector.

~~every obj to~~

Transformming a object from its space or time domain to feature ~~vector~~ space.

M dimensional feature space.

If M = 3 — 3 dimensional feature vector.



feature space

$P_1 \approx \langle 3, 5, 1 \rangle$

$P_2 \approx \langle f_1, f_2, f_3 \rangle$

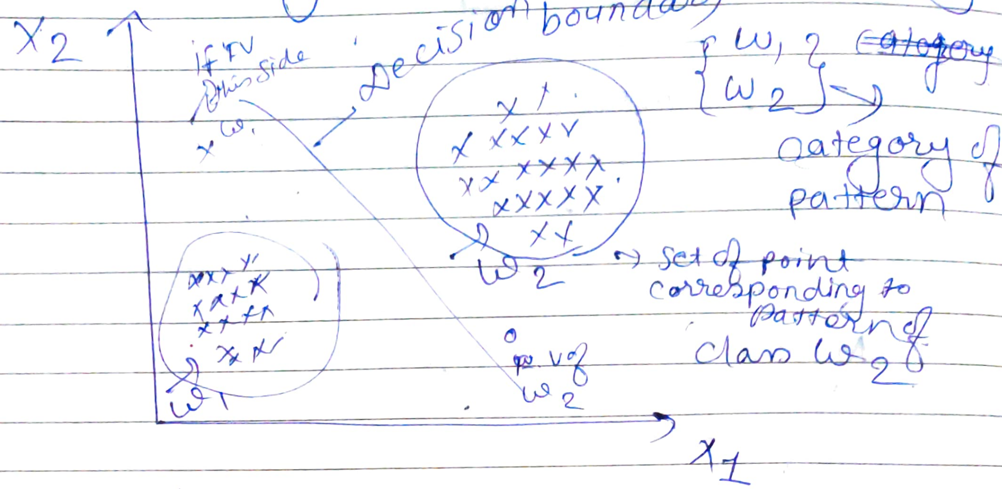
$P \approx \langle x_1, x_2, x_3 \rangle$

$d(P_1, P) > d(P_2, P)$.

P_1 is unknown to P.

More similar to P_2.

Bayes Decision Theory



Feature vector: $(x_1$ & $x_2)$
Diff. pattern on diff. point.

Class ω_1 → points will be very close to each other.

ω_2 →

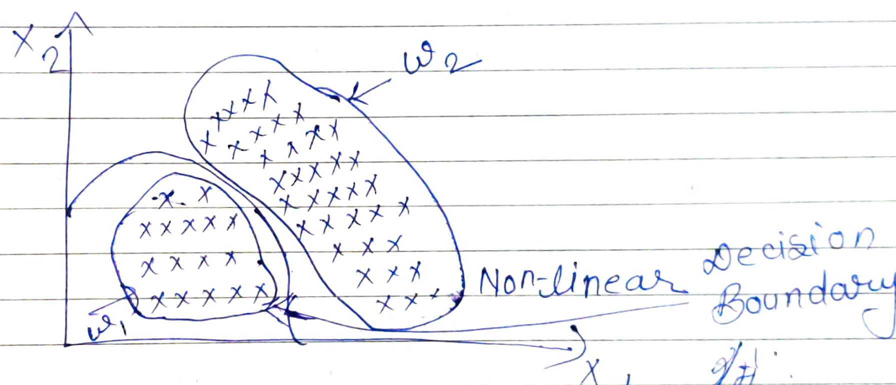
points of class ω_1 & ω_2 are likely to be apart from each other.

** Designing of this Classifier / Training of classifier.

means to find the equation of this decision boundary from the training samples.

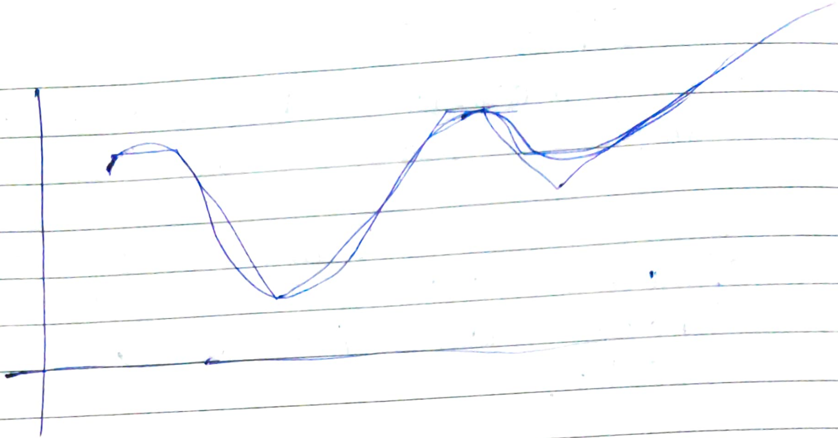
for Linear Decision Boundary we take training Samples from class of ω_1 & ω_2 Seperated these ω_1 & ω_2 are Classify by Linear Boundary this is also known as linearly Seperable class.

*. Now, suppose.



It most popular is Quadratic Classifier or by cubic classifies

Quadratic Classifier :-

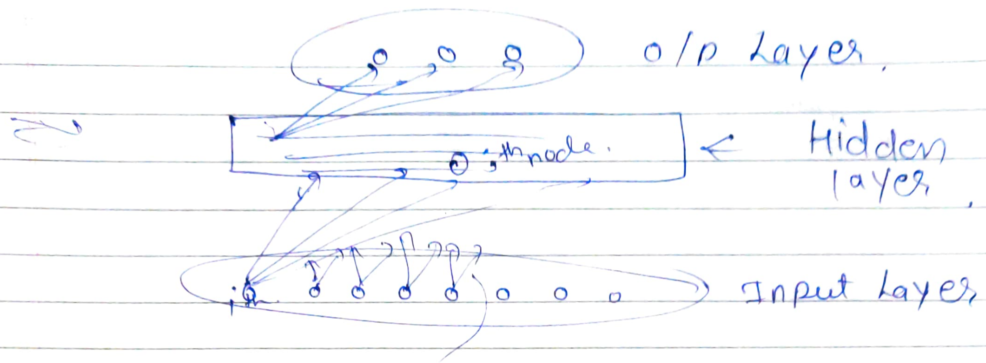


What to do if no. of classes increasing & boundary is non-linear having Analytic expression will become more Complicated.

* People take use of Neural N/W in such case.

But the decision boundary / Info of decision boundary is encoded in weight vector or weight Matrix in NN

* NN actually tries to form Straight line boundaries.



one or More hidden Layer is Multi Layer perceptron.

evernode in input Layer has Connection with everynode in hidden Layer. with some connection weight & every node in hidden Layer have Connection with O/p Layer.

$$f\left(\sum_i^{total} W_{ij}\, x_i\right) = O_j$$

total input — output of j^{th} node.

Bayes Decision Theory

Take Industry example.

ω_1 Accept — ω_2 Reject

from previous experience of supervised learning how many are accepted & how many are rejected based on that generate probability.

$P(\omega_1)$ — $P(\omega_2)$

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

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

Apriori probabilities
Combine some feature x.

x. feature x.

probability density function of feature x is used.

$p(x/\omega_1)$ — $p(x/\omega_2)$

Class Conditional PDF

PDF: Probability Density function.

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

then $\Rightarrow \omega_1$ in favour of ω_1

$P(\omega_1 / x) < P(\omega_2 / x) \Rightarrow \omega_2$ in favour ω_2 decision

Combining these 2 have more logical Decision Rule.

How to Combine (joint probability)

An obj belongs class (ω_i) & at the same time have the feature x

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

or

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

P of x given ω_i

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

likelyhood

Apriori

hypothesis

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

a postiriori probability

$P(x) \rightarrow$ marginal.

Decision Rule. Baye's theory.

$P(\omega_i)$ = apriori probility.

$P(x / \omega_i)$: Class condition probability density function of x given ω_i by

~~$P(x)$~~ :

which are accepted which are rejected.

$P(x)$: for an unknown obj measure the feature x.

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

* Bayes Decision Rule.

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

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

if expand.

$p(x/\omega_1) \cdot p(\omega_1) > p(x/\omega_2)\, p(\omega_2)$
$\Rightarrow \omega_1$

if $p(\omega_1/x) = p(\omega_2/x)$ object to be likely accepted & rejected then decision based on $P(x/\omega_1)$ & $P(x/\omega_2)$ but if $P(x/\omega_1)$ & $p(x/\omega_2)$ is same then decision based on apriori algo beyond $p(\omega_1)$ & $p(\omega_2)$

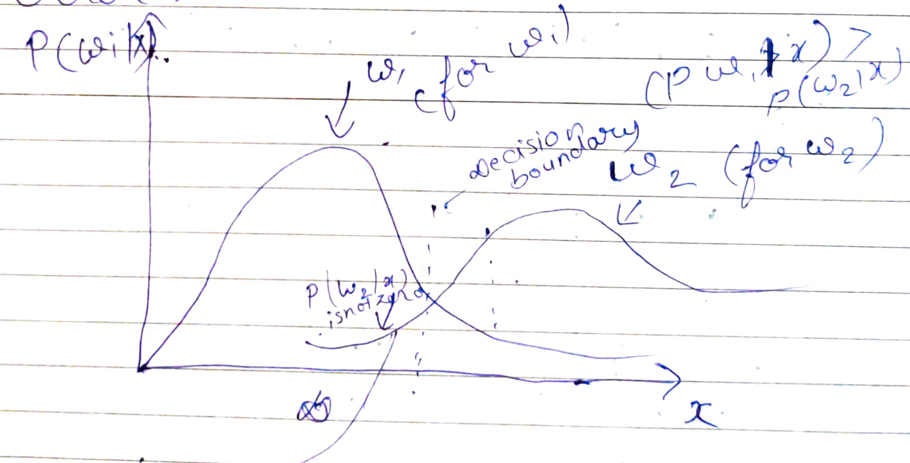
*. If Cannot able to make decision based on Observation then we use apriori

- If does not able to make decision on apriori we use Observation.

- In other Cases Consider both to take decisions.

* This include both apriori & class Conditional probability to take decision.

Error.



if $p(\omega_2|x)$ is zero then there is no error,
but if there is a value then there may be a probability that the object belongs to ω_2

If I decide

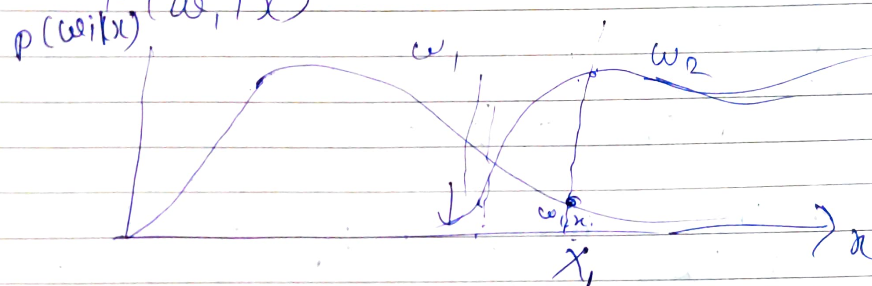
$\omega_1 \rightarrow P(\omega_2 / x)$

if I decide ω_1 then the probability of error (will be in ω_2) $P(\omega_2 | x)$

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

if decide ω_2 then the probability of error (will be in ~~ω_1~~ ω_1)

$P(\omega_1 | x)$



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

if I decide the obj ~~beto~~ to belong to ω_1 the probabity of error $P(\omega_2 | x)$ is high.

* The bayes decision rule ensure the probability of error is minimize.

Total error.

$$p(\text{error}) = \int_{-\infty}^{\infty} p(\text{error}, x)\, dx.$$

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

$$P(\text{error} \mid x) = \min \{ p(\omega_1 \mid x), P(\omega_2 \mid x) \}$$

Generalization :-

* Use more than 2 States of nature $\longrightarrow$ Classes. (~~we allow more~~

*. Use more than one features $\rightarrow$ feature vector

* Other actions other merely deciding States of Nature.

*. Introduce a loss function: more general than probability of error.

Sol

$C \rightarrow$ States of Nature $(\omega_1, \omega_2, \ldots\ldots \omega_c)$

$a \rightarrow$ Actions
$\{\alpha_1, \alpha_2, \ldots \ldots \alpha_a\}$

Loss Function

$\lambda(\alpha_i / \omega_j) \rightarrow$ Loss incurred for taking action α_i when true state of nature is ω_j

$X \Rightarrow$ d - dimensional feature vector

Now, Suppose.
x. is a feature vector.
Action α_1
but ~~if~~ we don't know true state of nature.

~~true state of nature~~

$\cancel{\lambda(\alpha_i / \omega_j)} =$

for that we ensure a Loss.

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

(expected Loss)
also called
Risk function / conditional Risk (depends on x)
/ expected Loss

also called Minimum Risk classifier.

Suppose.

State of nature, ω_1 ω_2

Action → α_1 α_2

$$\lambda(\alpha_i | \omega_j) \Rightarrow \lambda_{ij}$$

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

In two class problem.

$$R(\alpha_1 | x) = \lambda_{11} \cdot P(\omega_1 | x) + \lambda_{12} \cdot P(\omega_2 | x)$$

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

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

$\Rightarrow \omega_1$

condition for taking decision in favour of class ω_1

$\lambda_{11}, \lambda_{22}$ = Loss ~~incurred~~ occured during taking correct decision.

$\lambda_{21}, \lambda_{12}$: Loss occured during taking right decision.

(>0 in condition)

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

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

Bayes

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

Generalized

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

Minimum - error - Rate Classification :-

$\alpha_i \rightarrow$ True State of nature is ω_i

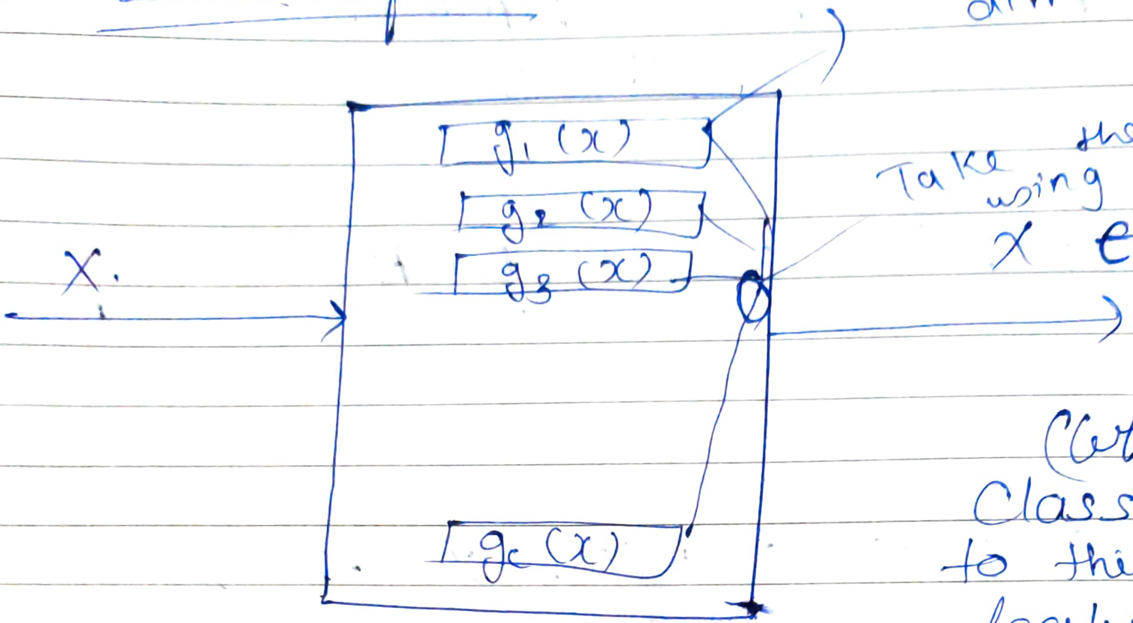
$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & \text{no loss whenever } i = j \\ 1 \text{ loss} & i \neq j \end{cases} \quad \text{for } i, j = 1, \dots c$$

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

$$= \sum_{i \neq j} P(\omega_j|x) = 1 - \frac{P(\omega_i|x)}{\lambda_{ij}}$$

Classifier :-

different Module / function unit / which computes different g(x)



Take the decision using either max / Min criteria.

$X \in ?$

if use generalization

(What is the class belonging to this obj having feature vector)

Classifier

X → feature vector

No. of function = Number of classes = No. of Actions.

$g(X) = g_1(x), g_2(x) \ldots g_c(x)$ are called discriminant function.

~~Nat~~

Nature of discriminant function.

$\omega_1 \ldots\ldots \omega_c \rightarrow$ c no. of classes

$g_i(x) \quad : \quad i = 1 \ldots\ldots c$

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

$\Rightarrow x \in \omega_i$

Min Risk classifier.

$R(\alpha_i / x)$

when related Risk function & discriminate

$$g_i(x) = -R(\alpha_i | x)$$

Min error rate :-

$$g_i(x) = P(\omega_i | x)$$

$g_i(x)$

$f(g_i(x))$

$f(\cdot) \rightarrow$ Monotonically increasing

$$g_i(x) = P(\omega_i | x)$$

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

for every value of g_i it is same so eliminate

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

$$g_i(x) = \ln P(x|\omega_i) + \ln P(\omega_i)$$ (concept of log)

Two categories

$\rightarrow$ ω_1 , ω_2

$g_1(x)$, $g_2(x)$

$g_1(x) > g_2(x) \longrightarrow \omega_1$

$g_1(x) < g_2(x) \longrightarrow \omega_2$

Decision boundary.

$$g_1(x) - g_2(x) = 0$$

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

$$g(x) = P(\omega_1|x) - P(\omega_2|x)$$

$$= \ln \frac{P(x|\omega_1)}{P(x|\omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$