ITD of

Pattern Recognition

Page No.
Date

Short note 1Q with example

Human Perception: Taking action & decision based on environment

Pattern: A pattern is an object, process (series of action) or event that can be given a name.

Independent existence

An image is not a pattern but when a name is given to an image like ~~bik~~ car, etc. is a pattern.

Pattern Recognition: It is the study of how machines can
* observe the environment.
* Learn to distinguish pattern of interest
* Make sound and reasonable decisions about the categories of the pattern.

Eg. Face recognition

Class (Pattern or class): A pattern class (or category) is a set of patterns sharing common attributes and usually originating from the same source. ~~Each a~~

Feature extraction and classification:

Extraction: To ~~categorise~~ characterize an object, to be recognised by measurements whose values are very similar for objects in the same category (or class) and very different for objects in different category.

Classification: To assign a category or class to the object based on feature vector provided during feature extraction.

noise :- any property of the sensed pattern which is not due to the true underlying model but instead to randomness in the world or sensors

Page No.
Date

Basic structure of ~~Patter~~ Pattern Recognition System :

1) **Data Acquisition & Sensing :**

eg Recognise the address written on envelope.

* Some measurement or observation about the object to be classified.
* Image contains multiple information (there are several objects).
* Instruments are involved in sensing ~~like~~
  Parameters: Resolution, latency, sensitivity, data format, distortion.

2) **Preprocessing :**

* Remove the noise from data

3) **Isolation of patterns of interests from the background (Segmentation) :**

3) **Feature Extraction :**

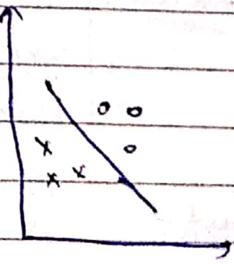
* Discriminating features (distinguishing features) invariant to irrelevant information
* Find a new representation in terms of features
* Features of a class should be close to each other.

4) **Model learning and estimation :**

* Learning a mapping mechanism between features and pattern groups and categories

5) **Classification :**

* Test Samples :- Assigning unknown object (datapoint) into one of the predefined class label.



6) **Postprocessing (optional) :**

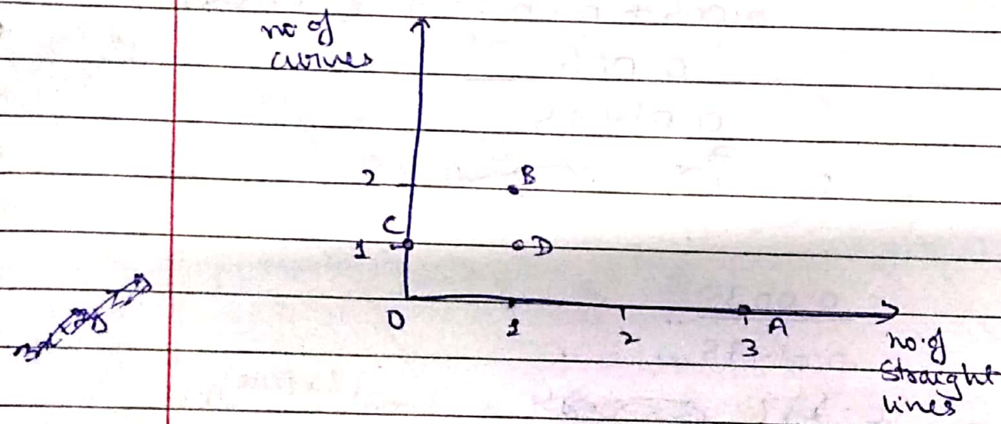
* ~~Confidence of our~~ Evaluation of confidence in decisions
* Exploitation of content to improve performance
* Combination of experts view.

Handwriting Recognition System:

Develop a model to recognise characters (A, B, C, D) from a handwritten text.

Feature extraction

No. of straight lines	No. of curves	Class
3	0	A
1	2	B
0	1	C
1	1	D



Testing Data:

	line	curve
A	2	1

Bayes Theorem:

Ques Desk Lab produced by the luminar company was found to ~~de~~ be ~~defected~~ defective. There are 3 factories A, B, C where such desk labs are manufactured. A Quality control manager (QCM) is responsible for investigating the source of found defects. The QCM would like to answer the following question If a randomly selected lamp is defective, (i) what is the probability that the lamp was manufactured in factory C, (ii) factory A and (iii) factor B

Conditional, A Priori → based on observation
Apostoriori → based on history

Conditional probability

A Priori probability	% of total production	Probability of defective products
A	0.35 = P(A)	0.015 = P(D/A)
B	0.35 = P(B)	0.010 = P(D/B)
C	0.30 = P(C)	0.020 = P(D/C)

A Postoriori

(i) $P(C/D) = \dfrac{P(D/C) * P(C)}{P(D/C) * P(C) + P(D/B) * P(B) + P(D/A) * P(A)}$

$= \dfrac{0.30 * 0.20}{0.30 * 0.020 + 0.35 * 0.010 + 0.35 * 0.0}$

$= \dfrac{0.006}{0.006 + 0.0035 + 0.00525}$

$= \dfrac{0.006}{0.01475}$

$= 0.407 \approx$

(ii) $P(B/D) = \dfrac{0.0035}{0.01475}$

$= 0.237$

(iii) $P(A/D) = \dfrac{0.00525}{0.01475}$

$= 0.356$

X ← variable like polish of desk

↓ quantize

$\underbrace{\text{good, very good}}_{\text{accept}}$, $\underbrace{\text{bad, poor}}_{\text{reject}}$

Conditional probability in pattern recognition is class conditional probability.

Date

variable	ex1	ex2	ex3	ex4	ex5	ex6	ex7	ex8	ex9	ex10
value	good	verygood	good	verygood	good	bad	poor	poor	bad	bad
category	accept	accept	reject	accept	~~reject~~ accept	reject	reject	accept	accept	reject

$P(accept) = \frac{6}{10}$, $P(reject) = \frac{4}{10}$

$P(good/accept) = \frac{2}{6}$, $P(verygood/accept) = 1$

$P(bad/accept) = \frac{1}{6}$, $P(poor/accept) = \frac{1}{6}$

$P(good/reject) = \frac{1}{4}$, $P(verygood/reject) = 0$

$P(bad/reject) = \frac{2}{4}$, $P(poor/reject) = \frac{1}{4}$

Testing variable : ext → bad ~~poor~~

$$P(accept/bad) = \frac{\frac{6}{10}\times\frac{1}{6}}{\frac{6}{10}\times\frac{1}{6} + \frac{4}{10}\times\frac{2}{4}} = \frac{1}{1+2} = \frac{1}{3} \; 0.5$$

$$P(reject/bad) = \frac{\frac{4}{10}\times\frac{2}{4}}{\frac{6}{10}\times\frac{2}{4} + \frac{4}{10}\times\frac{2}{4}} = \frac{2}{3}$$

$P(reject/bad) > P(accept/bad)$

∴ Rejected

Bayes Decision :

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

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

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

∴ Posteriori Prob, $P(\omega_i|x) = \dfrac{P(x|\omega_i) \cdot P(\omega_i)}{P(x)}$

$\omega \rightarrow$ omega

Page No.
Date

Decision rule: $P(x|\omega_1) \cdot P(\omega_1) > P(x|\omega_2)\, P(\omega_2) \Rightarrow \omega_1$

	ex_1	ex_2	ex_3	ex_4	ex_5	ex_6	ex_7	ex_8
Polish	thick	thick	thin	thin	thin	thick	thick	thick
Class	accept	accept	accept	accept	reject	reject	reject	reject

Polish(x)
↳ quantize — thick / thin

$P(\text{accept}) = 0.5$, $P(\text{reject}) = 0.5$

$P(\text{thick}/\text{accept}) = \frac{2/4}{1/2}$, $P(\text{thick}/\text{reject}) = \frac{3/4}{1/2}$

$P(\text{thick}/\text{accept}) = \frac{1/2}{1/2}$, $P(\text{thick}/\text{reject}) =$

$P(\text{thick}/\text{accept}) = 1$, $P(\text{thick}/\text{reject})$

$P(\text{thin}/\text{accept}) = \frac{P(\text{thin} \cap \text{accept})}{P(\text{accept})}$

Apriori $P(\text{accept}) = 0.5$, $P(\text{reject}) = 0.5$

class probability:

$P(\text{thick}/\text{accept}) = \frac{2}{4}$, $P(\text{thick}/\text{reject}) = \frac{3}{4}$

$P(\text{thick}/\text{accept}) = \frac{1}{2}$, $P(\text{thick}/\text{reject}) = 0.75$

$P(\text{thin}/\text{accept}) = \frac{2}{4}$, $P(\text{thin}/\text{reject}) = \frac{1}{4}$

$P(\text{thin}/\text{accept}) = \frac{1}{2}$, $P(\text{thin}/\text{reject}) = 0.25$

$P(\text{accept}/\text{thick}) = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2}} = \frac{\frac{1}{4}}{\frac{2}{8} + \frac{3}{8}} = \frac{2/8}{5/8} = \frac{2}{5}$

$= 0.4$

Scanned by CamScanner

$P(\text{reject} / \text{thick}) = 0.6$

$$P(\text{accept} / \text{thin}) = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2}} = \frac{\frac{2}{8}}{\frac{2}{8} + \frac{1}{8}}$$

$$P(\text{accept} / \text{thin}) = \frac{2}{3}$$

$$P(\text{reject} / \text{thin}) = \frac{1}{3}$$

If x = thick,
decision = reject ($\because P(\text{accept}/\text{thick}) < P(\text{reject}/\text{thick})$)

If x = thin,
decision = accept ($\because P(\text{accept}/\text{thin}) > P(\text{reject}/\text{thin})$)

$P(\text{error} \mid x) = \begin{cases} P(\omega_1 \mid x) & \text{if we decide in favour of } \omega_2 \\ P(\omega_2 \mid x) & \text{if we decide in favour of } \omega_1 \end{cases}$

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

$\cancel{P(\text{error} \mid x) = \sum_{i=1} P(\omega_i \mid x)}$

Multiple Class :

	ex1	ex2	ex3	ex4	ex5	ex6	ex7	ex8	ex9	ex10
Polish	good	verygood	good	verygood	bad	good	bad	good	bad	verygood
Class	mid	high	mid	high	low	low	low	high	mid	mid

$P(\text{high}) = \frac{3}{10}$, $P(\text{mid}) = \frac{4}{10}$, $P(\text{low}) = \frac{3}{10}$

Class probability

$P(\text{good}/\text{high}) = \frac{1}{3}$, $P(\text{good}/\text{mid}) = \frac{2}{4}$, $P(\text{good}/\text{low}) = \frac{1}{3}$

$P(\text{vgood}/\text{high}) = \frac{2}{3}$, $P(\text{vgood}/\text{mid}) = \frac{1}{4}$, $P(\text{vgood}/\text{low}) = 0$

~~P(error)~~

P(error / x) = summation of all other probabilities

If two probabilities are same, then post processing

Page No.

Date

class probability $\begin{cases} P(\text{bad}/\text{high}) = 0, \quad P(\text{bad}/\text{mid}) = \frac{1}{4}, \quad P(\text{bad}/\text{low}) = \frac{2}{3} \end{cases}$

$$P(\text{high}/\text{good}) = \frac{\frac{1}{3} \times \frac{3}{10}}{\frac{1}{3}\times\frac{3}{10} + \frac{2}{4}\times\frac{4}{10} + \frac{1}{3}\times\frac{3}{10}} = \frac{1}{1+2+1} = \frac{1}{4}$$

$$P(\text{mid}/\text{good}) = \frac{2}{4} = \frac{1}{2}$$

$$P(\text{low}/\text{good}) = \frac{1}{4}$$

$$P(\text{high}/\text{vgood}) = \frac{\frac{2}{3} \times \frac{3}{10}}{\frac{2}{3}\times\frac{3}{10} + \frac{1}{4}\times\frac{4}{10} + 0} = \frac{2}{2+1} = \frac{2}{3}$$

$$P(\text{mid}/\text{vgood}) = \frac{1}{3}, \quad P(\text{low}/\text{vgood}) = 0$$

$$P(\text{high}/\text{bad}) = 0, \quad P(\text{mid}/\text{bad}) = \frac{\frac{1}{4}\times\frac{4}{10}}{0 + \frac{1}{4}\times\frac{4}{10} + \frac{2}{3}\times\frac{3}{10}} = \frac{1}{1+2} = \frac{1}{3}$$

$$P(\text{low}/\text{bad}) = \frac{2}{3}$$

If x = good, decision = mid

If x = vgood, decision = high

If x = bad, decision = low

P(Class / fea)

Page No.
Date

Feature Vector:

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

Mutually Independent attributes (Naive Bayes):

$$x = (x_1, x_2, \ldots x_n)$$

Class conditional probability, for feature vector $P(x/c_j) = \prod_{i=1}^{n} P(x_i/c_j)$

$$\cancel{P(c_j)} \quad \boxed{P(c_j) \prod_{i=1}^{n} P(x_i/c_j)}$$

class maximizes

Eg

Chilly	runny nose	headache	Fever	Flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

$P(Y) = \frac{5}{8}$, $P(N) = \frac{3}{8}$

$\cancel{P(1st/y) = P(Y/N) * P(N/N) * P(Mild/N) * P(Y/N)}$

=

$P(Chills = Y \mid Flu = Y) = \frac{3}{5}$

$P(Chills = Y \mid Flu = N) = \frac{1}{\cancel{5}\,3}$

$P(Chills = N \mid Flu = Y) = \frac{2}{5}$

$P(Chills = N \mid Flu = N) = \frac{2}{3}$

$P(runny\ nose = Y \mid Flu = Y) = \frac{4}{5}$

$P(runny\ nose = \overset{Y}{\cancel{N}} \mid Flu = N) = \frac{1}{3}$

$P(runny\ nose = N \mid Flu = Y) = \frac{1}{5}$

$P(runny\ nose = N \mid Flu = N) = \frac{2}{3}$

$P(headache = No \mid Flu = Y) = \frac{1}{5}$

$P(headache = \overset{No}{\cancel{Mild}} \mid Flu = N) = \frac{1}{3}$

$P(headache = Mild \mid Flu = Y) = \frac{2}{5}$

$P(headache = Mild \mid Flu = N) = \frac{1}{3}$

$P(headache = Strong \mid Flu = Y) = \frac{2}{5}$

$P(headache = Strong \mid Flu = N) = \frac{1}{3}$

$P(\text{Fever} = Y \mid \text{Flu} = Y) = \frac{4}{5}$

$P(\text{fever} = Y \mid \text{Flu} = N) = \frac{1}{3}$

$P(\text{Fever} = N \mid \text{Flu} = Y) = \frac{1}{5}$

$P(\text{Fever} = N \mid \text{Flu} = N) = \frac{2}{3}$

Testing data :-

x - Chills = Y, runny nose = N, headache = Mild, Fever = N

$P(x / Y) = P(Y) \cdot [P(\text{Chills} = Y / \text{Flu} = Y)$ & $P(\text{runny nose} = N / \text{Flu} = Y)$ & $P(\text{headache} = \text{Mild} / \text{Flu} = Y)$ & $P(\text{Fever} = N / \text{Flu} = Y)]$

$P(x/Y) = \frac{\cancel{5}}{8} * \frac{3}{\cancel{5}} * \frac{1}{5} * \frac{2}{5} * \frac{1}{5} = \frac{6}{1000} = 0.006$

$P(x/N) = P \frac{\cancel{3}}{\cancel{8}_{2}} * \frac{1}{\cancel{3}} * \frac{\cancel{2}}{3} * \frac{1}{3} * \frac{\cancel{2}}{3} = \frac{\cancel{4}}{\cancel{216}} \frac{1}{54} = 0.01$

Prediction = N

Text Classification :

$$P(w_k / +) = \frac{n_k + 1}{n + |\text{vocabulary}|}$$

$$P(v_j / W) = P(v_j) \prod_{w \in \text{words}} P(w / v_j)$$

Scanned by CamScanner

Date

Doc	Text	Class
1	"I loved the movie"	+ve
2	"I hated the movie"	-ve
3	"A great movie, good movie"	+ve
4.	"Poor, acting"	-ve
5	"Great acting. a good movie"	+ve

	I	loved	the	movie	hated	a	great	good	poor	acting	class
1.	1	1	1	1							+ve
2.	1		1	1	1						-ve
3				2		1	1	1			+ve
4									1	1	-ve
5				1		1	1	1		1	+ve

For positive class:

I	loved	the	movie	hated	a	great	good	poor	acting
1	1	1	1						
			2		1	1	1		
			1		1	1	1		1

P

For negative class:

I	loved	the	movie	hated	a	great	good	poor	acting
1	–	1	1	1					
–	–	–	–	–	–	–	–	1	1

$P(+ve) = \frac{3}{5}$, $P(-ve) = \frac{2}{5}$

~~P(I|+ve)~~

$P(I|+ve) = \frac{1+1}{14+10} = \frac{1}{12}$, $P(I|-ve) = \frac{1+1}{6+10} = \frac{1}{8}$

$P(\text{loved}|+ve) = \frac{2}{24} = \frac{1}{12}$ $P(\text{loved}|-ve) = \frac{1}{16}$

$P(\text{the}|+ve) = \frac{2}{24}$ $P(\text{the}|-ve) = \frac{2}{16}$

$P(\text{movie}|+ve) = \frac{\cancel{2}5}{24}$ $P(\text{movie}|-ve) = \frac{2}{16}$

$P(\text{hated}|+ve) = \frac{\cancel{2}1}{24}$ $P(\text{hated}|-ve) = \frac{2}{16}$

$P(\text{a}|+ve) = \frac{3}{24}$ $P(\text{a}|-ve) = \frac{1}{16}$

$P(\text{great}|+ve) = \frac{3}{24}$ $P(\text{great}|-ve) = \frac{1}{16}$

$P(\text{poor}|+ve) = \frac{1}{24}$ $P(\text{poor}|-ve) = \frac{2}{16}$

$P(\text{Acting}|+ve) = \frac{2}{24}$ $P(\text{Activing}|-ve) = \frac{2}{16}$

$P(\text{good}|+ve) = \frac{3}{24}$ $P(\text{good}|-ve) = \frac{1}{16}$

Testing Document: I hated the poor acting

$P(+ve|\text{test}) = \cancel{P(I|+ve) \times P(}$

$= \frac{2}{24} \times \frac{1}{24} \times \frac{2}{24} \times \frac{1}{24} \times \frac{2}{24} \times \frac{3}{5} = \frac{24}{39{,}813{,}120}$

$= 6.02 \times 10^{-7}$

$P(-ve|\text{test}) = \frac{1}{8} \times \frac{2}{16} \times \frac{2}{16} \times \frac{2}{16} \times \frac{2}{16} \times \frac{2}{5} = \frac{64}{5{,}242{,}880}$

$= 1.2 \times 10^{-5}$

$\Rightarrow$ ~~Reject~~ −ve review.

Minimum Risk Classifier :

$c \rightarrow$ state of nature $\{\omega_1, \omega_2, \ldots \omega_c\}$
$a \rightarrow$ action $\{\alpha_1, \alpha_2, \ldots \alpha_a\}$
$\lambda \rightarrow$ d-dimensional feature vector

~~Loss~~ Loss function :- $\lambda(\alpha_i|\omega_j) \rightarrow$ Loss occured for taking action α when the true state of nature is ω_j

~~Ques~~ ~~Important mail ω_1~~

~~Ques~~ Problem : Incoming email is either normal (important mail), ω_1 or junk mail, ω_2. We have two actions
$\alpha_1 \Rightarrow$ keep the mail.
$\alpha_2 \rightarrow$ move the mail to trash.

$\lambda(\alpha_1|\omega_1) = 0$, $\lambda(\alpha_1|\omega_2) = 1$

(Imp. mail in trash, extra wt.) $\lambda(\alpha_2|\omega_1) = 3$, $\lambda(\alpha_2|\omega_2) = 0$

Risk function
Avg loss
Avg Risk
Conditioned Risk
Expected loss

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

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

~~R~~

$R(\alpha_1|x) = \lambda_{11} P(\omega_1|x) + \lambda_{12} P(\omega_2|x)$
$R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)$

if $\alpha_1|x < \alpha_2|x \Rightarrow$ take the action α_1,
then this condition holds true
$(\lambda_{21} - \lambda_{11}) P(\omega_1|x) > (\lambda_{12} - \lambda_{22}) P(\omega_2|x)$

Date

Minimum Error rate classifier :

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

$$R(\alpha_i|x) = \sum_{i\neq j} P(\omega_j|x)$$

Eg

~~$R(\alpha_i|x) = \sum$~~

$$R(\alpha_1|x) = P(\omega_2|x)$$
$$R(\alpha_2|x) = P(\omega_1|x)$$

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

$$R(\alpha_1|x) = 1 - P(\omega_1|x)$$
$$R(\alpha_2|x) = 1 - P(\omega_2|x)$$

minimize   maximize

maximize   minimize

Ques: $P(\omega_1) = 0.4$ & $P(\omega_2) = 0.6$

$P(x|\omega_1) = 0.35$

$P(x|\omega_2) = 0.65$

$$P(\omega_1|x) = \frac{0.35 * 0.4}{0.35*0.4 + 0.65*0.6} = \frac{0.14}{0.14+0.39} = \frac{0.14}{0.53} = 0.264$$

$$P(\omega_2|x) = \frac{0.39}{0.53} = 0.736$$

$\lambda(\alpha_1|\omega_1) = 0$, $\lambda(\alpha_1|\omega_2) = 1$

$\lambda(\alpha_2|\omega_1) = 3$, $\lambda(\alpha_2|\omega_2) = 0$

Page No.
Date

$R(\alpha_1|x) = 0 + 0.736 = 0.736$
$R(\alpha_2|x) = 3 * 0.264 + 0 = 0.792$

Predicted action α_2 i.e. move the mail to trash.

M-estimate:

$$P_{tail} = \frac{N_{tail} + m\pi_{tail}}{N_{all} + m}$$

← prior expectation

↑ confidence

if $N_{all} = N_{tail} = 0$ ⇒ degenerates to a prior expectation

If N_{all} & N_{tail} → large ⇒ relative frequency

Toss no.	1	2	3	4	5
Outcome	tail	tail	head	tail	head
Relative frequency	1.0	1.0	2/3 = 0.67	3/4 = 0.75	3/5 = 0.60
m-estimate	2/3	3/4	3/5	4/6 = 2/3	4/7

Application ↓ 2+ Chance for head is not 0 till 2nd observation

$\pi_{tail} = 0.5$, $m = 2$

* Prior expectation help us to improve probability estimates in domains with insufficient no of observations.

$P(tail) = 0.75$ → 4 times toss, 3 times appear tail

$$P_{tail} = \frac{3 + 2 \times 0.5}{4 + 2} = \frac{4}{6} = \frac{2}{3}$$

If m = 100

Toss no.	1	2	3
Outcome	tail	tail	head
M-estimate	51/101	52/102 = 26/51	52/103

closer to prior expectation

If m = 1

Toss	1	2	3
Outcome	tail	tail	head
M-estimate	1.5/2 = 0.75	2.5/3 = 0.84	2.5/4 = 0.625

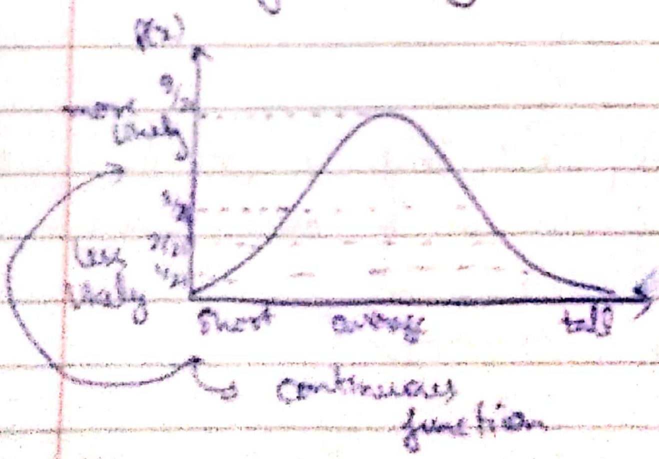
Continuous Attribute - Age, wt → (continuously changing)
Discrete Attribute → have fixed value

bins [[0,10), ... [90, 100)]

$\frac{N_i}{N}$ & $\sum \frac{N_i}{N} = 1$

To smoothen out graph, decrease interval size

Probability Density Function (Pdf):



			P(x)
0-10	-	1	1/21
11-20	-	2	2/21
21-30	-	3	3/21
31-40	-	9	9/21
41-50	-	3	3/21
51-60	-	2	2/21
61-70	-	1	1/21

Standard PDF: (Gaussian Distribution)

$P(x) = k \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Scanned by CamScanner

$(x-\mu)^2 \Rightarrow$ value slopes down with same angle
Slope steep — depends on σ^2
Greater variance → flatter bell

$$K = \frac{1}{\sqrt{2\pi\sigma^2}}$$

<u>Approximating PDFs:</u>

Eg ⇒ Distribution of age in a group that mixes — drama school children with their parents.

Diverse data → not better representation of data
Split the data, represent it & then combine

<u>Combine Gaussian function:</u>

$$P(x) = K \sum_{i=1}^{m} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}$$

$$K = \frac{1}{(2\pi)^{m/2}\sigma^m}$$

~~denote~~ denote μ_i, the value of x in the ith example

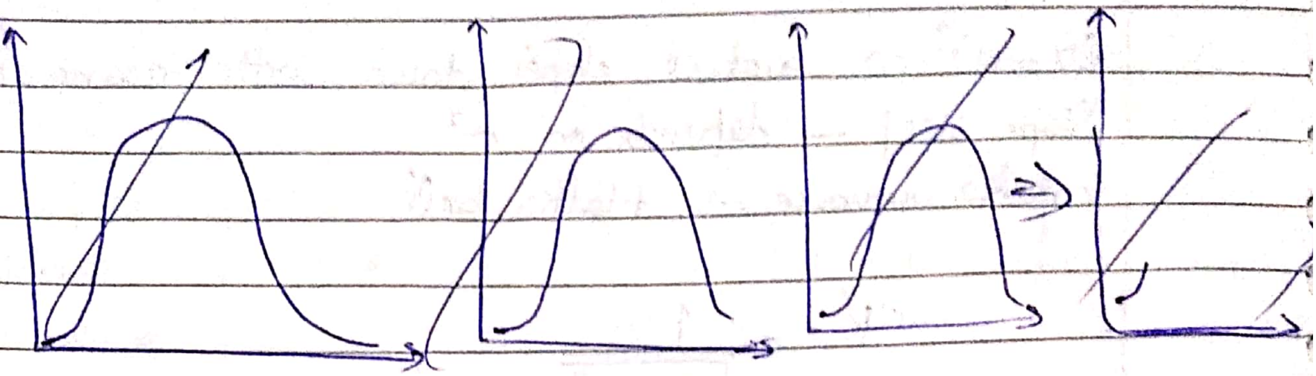
Eg

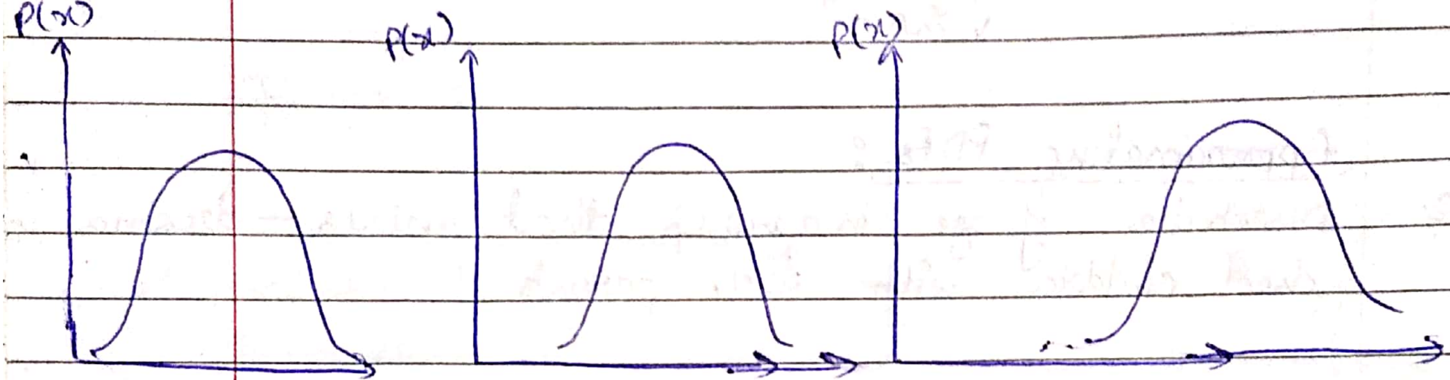
	ht	Age	Class
ex 1	0.4		+
ex 2	0.5		+
ex 3	0.7		+

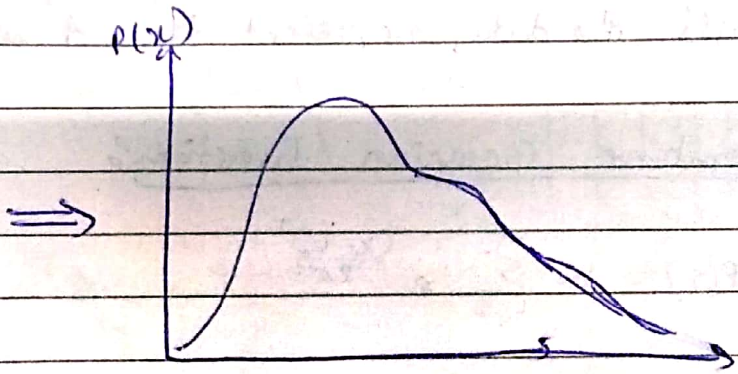
& $\sigma = 1$

$$P(x) = \frac{1}{(2\pi)^{3/2}(1)^3}\left[e^{-\frac{(x-0.4)^2}{2}} + e^{-\frac{(x-0.5)^2}{2}} + e^{-\frac{(x-0.7)^2}{2}}\right]$$

$$P(x) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x-0.4)^2}{2}} + e^{-\frac{(x-0.5)^2}{2}} + e^{-\frac{(x-0.7)^2}{2}}\right]$$







In general form,

eg

	Attr1	Attr2	Attr3	Class
ex1	3.1	2.1	2.3	+
ex2	4.2	6.2	7.6	+
ex3	7.8	1.3	0.5	+
ex4	2.3	5.2	2.4	−
ex5	6.4	3.2	4.3	−
ex6	1.3	5.8	3.3	−
ex7	?	?	?	?
ex8				

$$P(x_1/+) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_1-3.1)^2}{2}} + e^{-\frac{(x_1-4.2)^2}{2}} + e^{-\frac{(x_1-7.8)^2}{2}}\right] \cancel{e^{-\frac{(x_1-2.3)^2}{2}} + e^{-\frac{(x_1-6.4)^2}{2}} + e^{-\frac{(x_1-1.3)^2}{2}}\Big]}$$

$$P(x_2/+ve) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_2-2.1)^2}{2}} + e^{-\frac{(x_2-6.2)^2}{2}} + e^{-\frac{(x_2-1.3)^2}{2}}\right]$$

$$P(x_3/+ve) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_3-2.3)^2}{2}} + e^{-\frac{(x_3-7.6)^2}{2}} + e^{-\frac{(x_3-0.5)^2}{2}}\right]$$

$$P(x_1/-ve) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_1-2.3)^2}{2}} + e^{-\frac{(x_1-6.4)^2}{2}} + e^{-\frac{(x_1-1.3)^2}{2}}\right]$$

$$P(x_2/-ve) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_2-5.2)^2}{2}} + e^{-\frac{(x_2-3.2)^2}{2}} + e^{-\frac{(x_2-5.8)^2}{2}}\right]$$

$$P(x_3/-ve) = \frac{1}{(2\pi)^{3/2}}\left[e^{-\frac{(x_3-2.4)^2}{2}} + e^{-\frac{(x_3-4.3)^2}{2}} + e^{-\frac{(x_3-3.3)^2}{2}}\right]$$

$$P(+ve) = \frac{1}{2}, \qquad P(-ve) = \frac{1}{2}$$

Testing variable, x: $(9, 2.5, 3.2)$

$$P(+ve/x) = P(+ve) * P(x_1/+ve) * P(x_2/+ve) * P(x_3/+ve)$$

$$= \cancel{\frac{1}{2} * \frac{1}{(2\pi)^{3/2}}\Big[e}$$

$$= \frac{1}{2} * \frac{1}{(2\pi)^{3/2}} * \left[e^{-\frac{(9-3.1)^2}{2}} + e^{-\frac{(9-4.2)^2}{2}} + e^{-\frac{(9-7.8)^2}{2}}\right] *$$

$$\left[e^{-\frac{(2.5-2.1)^2}{2}} + e^{-\frac{(2.5-6.2)^2}{2}} + e^{-\frac{(2.5-1.3)^2}{2}}\right] *$$

$$\left[e^{-\frac{(3.2-2.3)^2}{2}} + e^{-\frac{(3.2-7.6)^2}{2}} + e^{-\frac{(3.2-\overset{0.5}{\cancel{7.6}})^2}{2}}\right]$$

$$P(+ve/x) = \frac{1}{2} \times \frac{1}{(2\pi)^{9/2}} \left[2.76\times10^{-8} + 992.9\times10^{-8} + 0.48\right]$$

$$\left[0.92 + 21.06\times10^{-3} + 0.48\right] \times$$

$$\left[0.66 + 6.2\times10^{-5} + 0.02\right]$$

$$P(+ve/x) = \frac{1}{2} \times \frac{1}{(2\pi)^{9/2}} \left[0.48\right) \times \left[1.4\right] \times 0.68)$$

$$P(+ve/x) = \cancel{\frac{1}{2} \times (2.55) \times 10^{-4}} = 5.84 \times 10^{-5}$$

$$P(-ve/x) = \frac{1}{2} \times \frac{1}{(2\pi)^{9/2}} \left[e^{-\frac{(9-2.3)^2}{2}} + e^{-\frac{(9-6.4)^2}{2}} + e^{-\frac{(9-1.3)^2}{2}}\right]$$

$$\times \left[e^{-\frac{(2.5-5.2)^2}{2}} + e^{-\frac{(2.5-3.2)^2}{2}} + e^{-\frac{(2.5-5.8)^2}{2}}\right]$$

$$\times \left[e^{-\frac{(3.2-2.4)^2}{2}} + e^{-\frac{(3.2-4.3)^2}{2}} + e^{-\frac{(3.2-3.3)^2}{2}}\right]$$

$$P(-ve/x) = \frac{1}{2} \times \frac{1}{\sqrt{(2\pi)^9}} \times \left[1.78\times10^{-10} + 0.034 + 1.133\times10^{-13}\right]$$

$$\times \left[0.026 + 0.782 + \overset{0.00431}{\cancel{4.31 \times 10^{-3}}}\right]$$

$$\times \left[0.726 + 0.546 + 0.995\right]$$

$$P(-ve/x) = \frac{1}{2} \times \frac{1}{\sqrt{(2\pi)^9}} (0.034) \times (0.81231) \times (2.267)$$

~~P(-ve/x) = 192.30~~

$$P(-ve/x) = 8.013 \times 10^{-6}$$

Prediction :- Positive [P(+ve/x) > P(-ve/

Maximum likelihood Estimation

YYYNNNNN

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

We denote probability by symbol π

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

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

YYY NN NNN

$\boxed{11100000}$

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

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

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

Assuming data independence, then joint probability

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

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

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

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

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

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

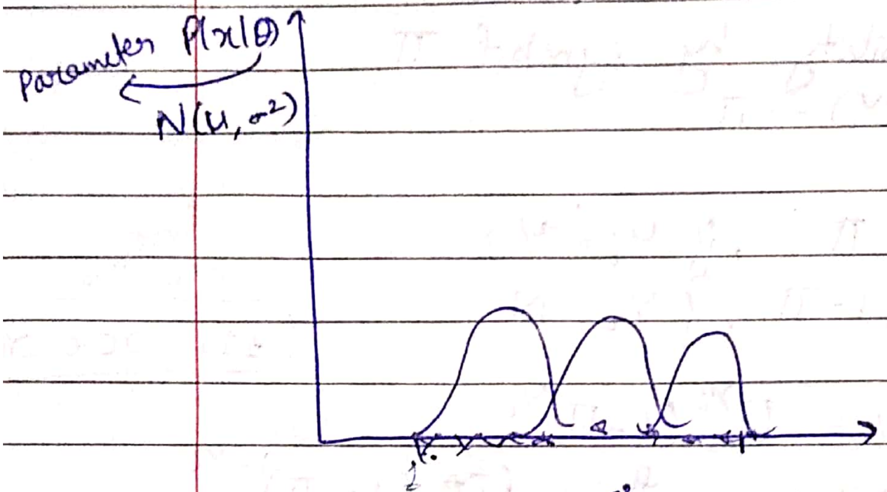
$\nabla \log \pi$

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

Scanned by CamScanner

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

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



Considering $\sigma^2 = 1$ (shape is ~~no~~ fixed)

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

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

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

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

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

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

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

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

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

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

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

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