ITA - Last Min Notes

Owner	N nithu
Tags	

UNIT 1

Domain,Range, Union,Intersection and Disjoint set

Dualization Law: (A+B) com = A(com)B(com)

Probability: measure of uncertainity

P=1(sure)

P=0(sure of not happening)

Mutually exclusive : P(A int B)=0

Independent: P(A int B)=P(A)P(B)

R.V → a function/a mapping

Real valued function defined over the space of a random experiment

ω	$X(\omega)$	$P\{\omega\}$
HHH	3	p^3
HHT	1	$p^2(1-p)$
HTH	1	$p^2(1-p)$
THH	1	$p^2(1-p)$
HTT	-1	$p(1-p)^2$
THT	-1	$p(1-p)^2$
TTH	-1	$p(1-p)^2$
TTT	-3	$(1-p)^3$

Frequency Distribution - Applicable for both quantitative and qualitative data

Rel. Freq: Counts the fractional part of the data that belong to a category

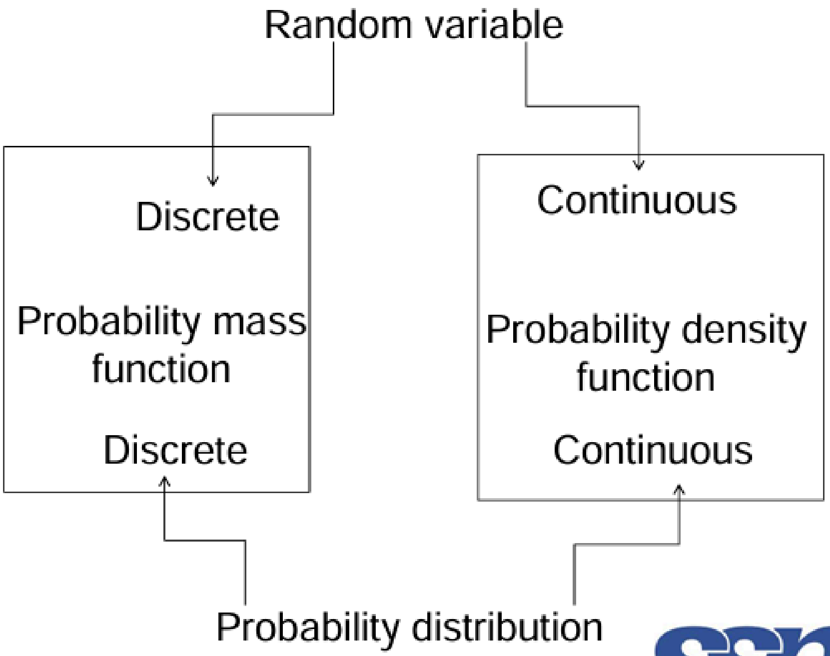
Continuous data: A smooth curve → continuous data

Average: x1+x2...xn/n

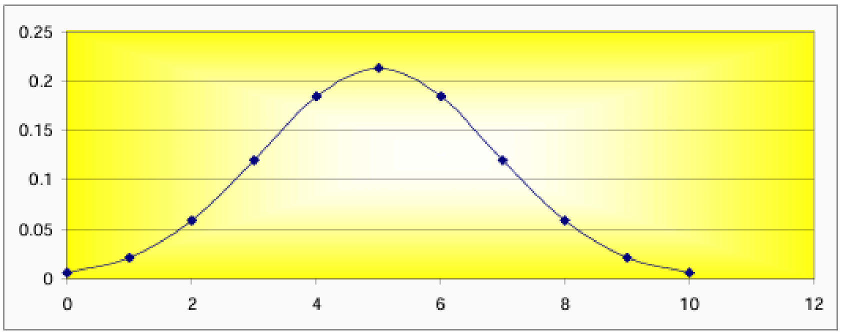
1/n is called weighing function

Expectation : sum(xp(x))

std . dev = sum(pi(x-mu)^2



Gaussian distribution – mean and standard deviation



$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

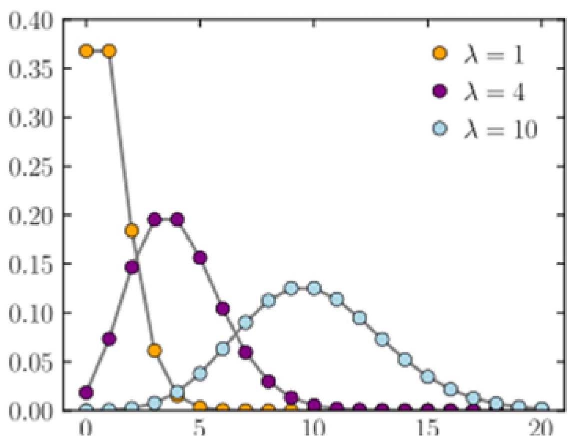
Defined from
-∞ to +∞

$$E(X) = \bar{X} = \text{first-order moment of } X = \sum_{k=1}^{n} p_k x_k$$

$$E(X^2) = \overline{X^2} = \text{second-order moment of } X = \sum_{k=1}^{n} p_k {x_k}^2$$

$$E(X^3) = \overline{X^3} = \text{third-order moment of } X = \sum_{k=1}^{n} p_k {x_k}^3$$

Poisson's distribution



$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

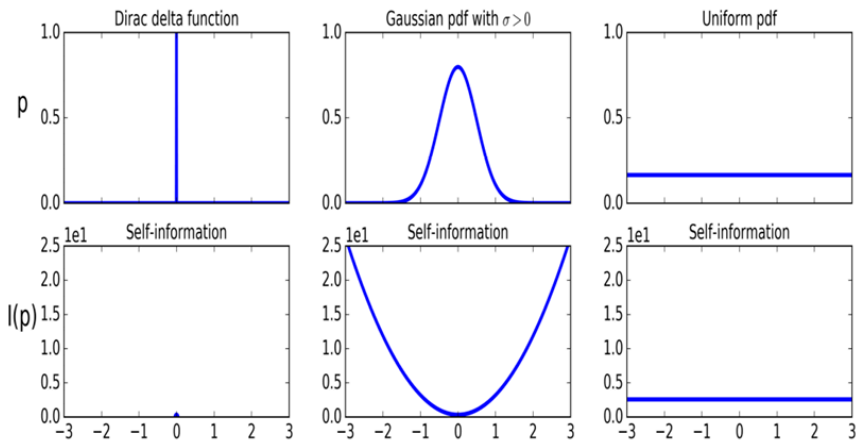
Kernel estimator

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right)$$

UNIT 2 - INTRO to info theory

Information Content - The information content is simply 0s and 1s it takes to transmit it. It can be considered as a measurable physical quantity.

Self-information(I) = -log(p) → Shannon's measure



More entropy , more is the disorder(higher entropy means random,disorganized, disordered with more uncertainity)

Shannon's Entropy:

Entropy is always positive

$$\sum_{i=1}^{r} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{r} p_i \log_2 p_i$$

Properties: Adding an impossible event does not change the entropy.

Entropy becomes 0, when one event is certain to happen

Two independent probabilities:
H=H (PX) + H (QY)

Discrete : sum(pi)=1

continuous: int(px)dx=1

Broader the distribution higher the entropy
Gaussian becomes broader with increasing σ (std dev.)
Higher the σ, more the entropy and viceversa
In the limit, $\sigma \to \infty$, Gaussian $\to$ uniform distribution
In the limit, $\sigma \to 0$, Gaussian $\to$ delta distribution

Joint Entropy: H(X,Y) = P(x,y) log(1/P(x,y))

Conditional Entropy: H(X|y=b k)= P(x|y=b k) log(1/P(x,y=b k))

P(y).P(x|y)=P(x,y)

log(1/P(x,y))=log(1/P(y))+log(1/P(x|y))

H(X,Y)=H(Y)+H(X|Y)=H(X)+H(Y|X)

I(X,Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)

KLD

KLD → Distance between two PDFs

KLD(relative entropy) can be used

Measures on how one prob.dist diverges from another dist

$$D_{KL}(P \| Q) = -\sum_x P(x) \log \frac{Q(x)}{P(x)}$$
$$= \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

Assymetrical : Two dist p and q

p(0)=1-r; p(1)=r

q(0)=1-s; q(1)=s

$$D(p||q) = (1-r)\log\frac{1-r}{1-s} + r\log\frac{r}{s}$$

$$D(q||p) = (1-s)\log\frac{1-s}{1-r} + s\log\frac{s}{r}.$$

If r=s then D(p||q)=D(q||p)=0

J-divergence:

$$\sqrt{\frac{1}{2}\left[D_{KL}(P \parallel Q)\right]^2 + \frac{1}{2}\left[D_{KL}(Q \parallel P)\right]^2}$$

Entropy, Relative entropy & cross entropy

Relative entropy = Cross entropy – Shannon's entropy

$D_{KL}(P||Q) = H(P,Q) - H(P)$

= sigma P(x) log(P(x)) - sigma P(x) log(Q(x))

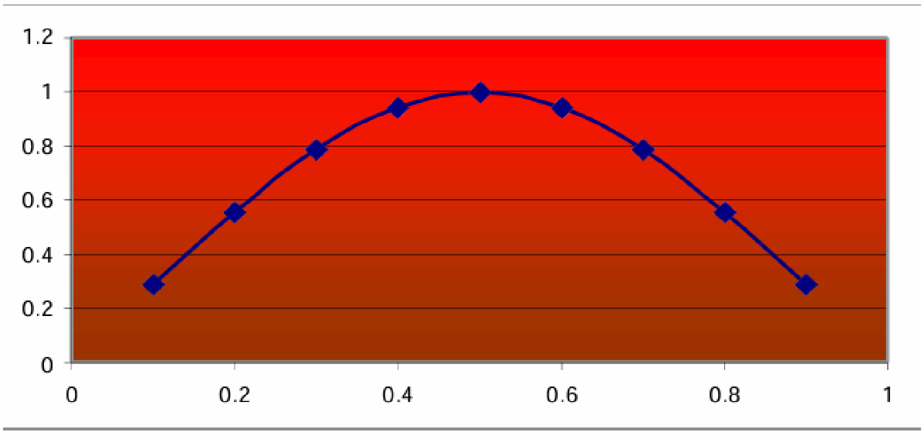
Renyi's Entropy

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p_i^\alpha\right)$$

when alpha =2

where the sigma term → Information Potential

H(x)=-1log(sigma(p i ^2)



Renyi entropy

$$H = \frac{1}{\alpha - 1} log\left(E\left(p^{(\alpha-1)}\right)\right)$$

Tutorials are the main!

UNIT 3 - CHANNEL CODING

Efficient and reliable data transmission →by removal of redundancy and correction of errors.

DATA COMPRESSION : Source Coding → compress from the source

ERROR CORRECTION : Channel Coding → Extra bits to make the transmission robust to disturbances

SOURCE CODES

→ Make the source data smaller

→ Fewer bits carry the entire info.

ENCODER: A device or an element which involved source coding

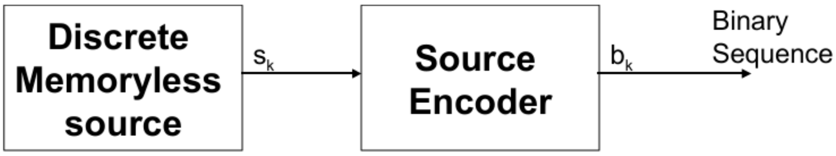
Fixed Length Codign(FLC)
Variable Length Coding(VLC)

If each character of the code is represented by the same no of characters → FLC
Eg: ASCII

Variable Length Coding(VLC) → Maps source symbol to variable no of bits.

Eg: Morse Code

n→ length of the code



Avg Codeword Length:

$$\bar{L} = \sum_{k=0}^{K} p_k l_k$$

Efficiency(n) = L min/ L bar or H(S)/L bar

Source Coding Theorem:

Givn a dms of entropy H(S), the avg code word length for any source encoding is bounded as

L(bar) ≥ H(S)

From coding eff.

L(bar)≥ L(min)

From Shannon's Source Coding Theorem

L(bar)≥H(S)
From 1 and 2

L(min)=H(S)

DATA COMPACTION:

→ To remove redundancy.

→ Lossless communication

CODING SCHEMES:

Prefix coding: Initial part of the codeword . A codeword should not be the prefix of other code words. They are uniquely decodable(converse not true) and instantaneous codes.

The number of bits in the prefix of the n-bit codeword ≤n

KMI Inequality(Kraft McMillan Inequality)

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

The average codeword length $(\overline{L})$ can be bounded as follows:

$$H(S) \leq \overline{L} < H(S) + 1$$

$$H(S) = \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}}$$

If a prefix code satisfies an efficient KMI

L(bar) = H(S)

Huffmann Coding: Bottom up approach

Disadvantages: Changing ensemble and does not consider blocks of symbols

Shannon-Fano Coding:

A more frequent message has to be encoded by a shorter encoding vector (word) and a less frequent message has to be encoded by a longer encoding vector (word)".

$$\sigma^2 = \sum_{k=0}^{K-1} \left(l_k - \overline{L}\right)^2 p_k$$

Not unique

Top-down approach

LZ Coding

CHANNEL CODING

Mapping → data seq into a channel input

Inverse-mapping → channel output into a output sequence

To increase resistance to channel noise via error control coding

Errors:

Single-bit error: One bit will be changed

Least likely

Burst errors: More than one bit is changed

Most likely to happen in serial transmission.

Redundancy check:

Vertical redundancy check: For single bit errors and detect burst errors if the no.of errors is odd

Longitudinal redundancy check: Burst errors

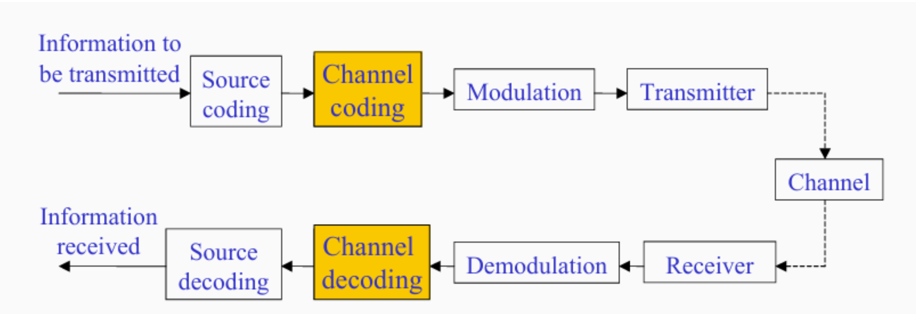
Cyclic redundancy check: (k+n) bits

Sender: Divides into k sections, n bits each, all add using one's complement to get the sum , then complement that,checksum is sent

Receiver: Divides into k sections, n bits each, all add using one's complement to get the sum , then complement that, if they add up to zero, it is accepted else rejected

m= data bits, r=redundancy bits, m+r = total msg

2^r≥m+r+1



Forward Error Correction: To transmit enough redundant data to allow receiver to recover from errors all by itself.

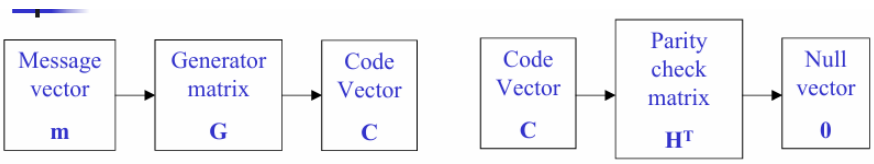
No sender retransmission is required.

BLOCK CODES: Info divided into k-sized blocks

Add r error bits

n= k+r

Efficiency : k/n = k/k+r



Code	Generator polynomial g(x)	Parity check bits
CRC-12	$1+x+x^2+x^3+x^{11}+x^{12}$	12
CRC-16	$1+x^2+x^{15}+x^{16}$	16
CRC-CCITT	$1+x^5+x^{15}+x^{16}$	16

Convolutional Codes:

Encode info. streams rather than info.blocks

Channel Capacity:

$$C = B\log_2\left(1+\frac{P}{N_0 B}\right) \quad bits/second$$

Channel Coding Theorem - II

DMS, H(S) bps

Avg Info rate → H(S)/Ts

Channel Coding theorem : DMS - with alphabet S and entropy H(S), producing symbols every Ts sec

DMC → capacity C , be used once in every Tc sec

H(S)/T s ≤ C/Tc(the source output can be transmitted)

C/Tc → critical rate

else :

H(S)/T s > C/Tc(it is not possible)

Message Word	Parity bits	Code words
0000	000	0000000
0001	101	1010001
0010	111	1110010
0011	010	0100011
0100	011	0110100
0101	110	1100101
0110	100	1000110
0111	001	0010111
1000	110	1101000
1001	011	0111001
1010	001	0011010
1011	100	1001011
1100	101	1011100
1101	000	0001101
1110	010	0101110
1111	111	1111111

Repetition Codes: repeat every bit some prearranged no of times

s	0	0	1	0	1	1	0
t	000	000	111	000	111	111	000
n	000	001	000	000	101	000	000
r	000	001	111	000	010	111	000

Received sequence **r**	Decoded sequence $\hat{s}$
000	0
001	0
010	0
100	0
101	1
110	1
011	1
111	1

Run length Coding - count the number of adjacent pixels with the same grey-value→ this is called run length

UNIT 4 - DIGITAL IMAGE PROCESSING

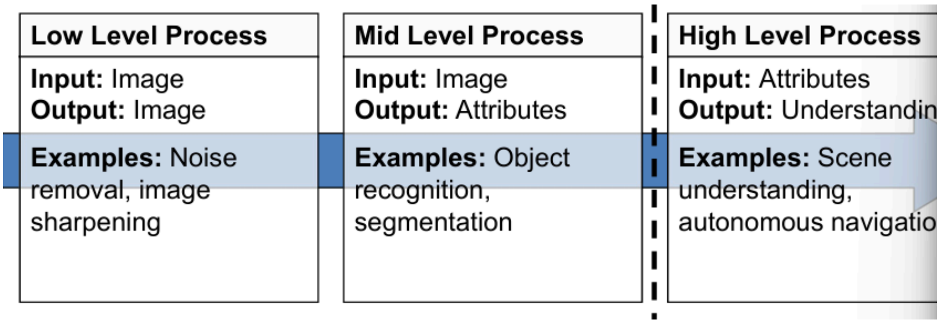
BASICS

Digital Image - representation of 2D image as pixels(finite set of digital values) → approximation

Pixel values → grey levels,colours,heights,opacities..

Digital Image Processing: Improvement of pictorial information (Image enhancement)

Processing of image data for storage, transmission and representation



Applications: Medicine(CAT< MRI< Gamma Ray Scans< X Ray< Tomography)

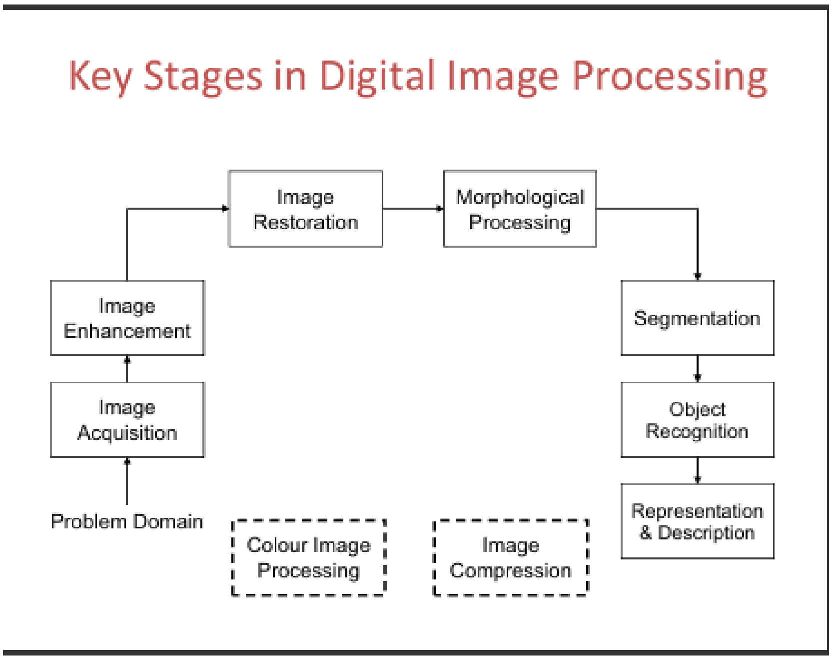
Astronomical images(Ranger 7< Hubble)

GIS(Satellites<Meterology)

Industrial Inspection(PCB)

Law Enforcement(CCTV<Fingerprints<Number plate)

HCI



Resolution: Details contained in a pixel

Representing color: RGB Values

(255,255,255)→ white

(0,0,0)→ Black

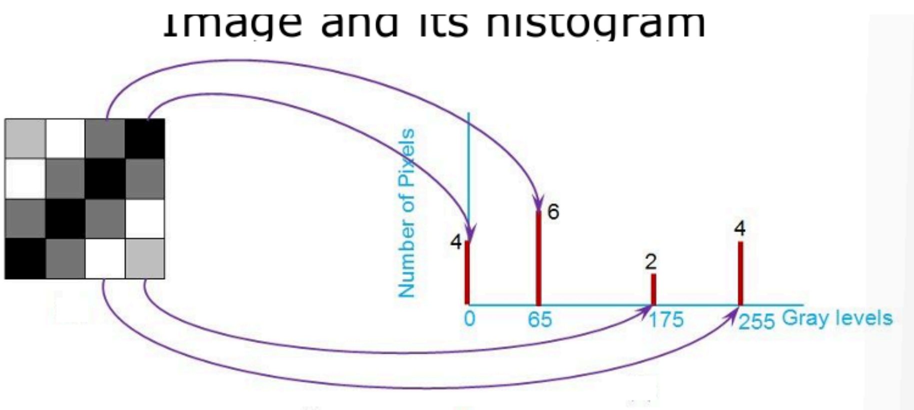
BInary Images → interpreting in form of 0s and 1s (light or dark,black or white)

Raster vs Vector:

Raster: representing as pixels(Eg: BMP,GIF[gaphics interchange format],JPEG[joint photographic experts group])

Vector : SVG(Scalable Vector Graphics) → text-based scripts

HISTOGRAMS



For levels h(1)...h(255)

p(1)=h(1)/N

p(255)=h(255)/N

(i*j) has total pixes = 3(i*j)

Binning: Bin 0 - 0 - 63

Bin 1 64-127

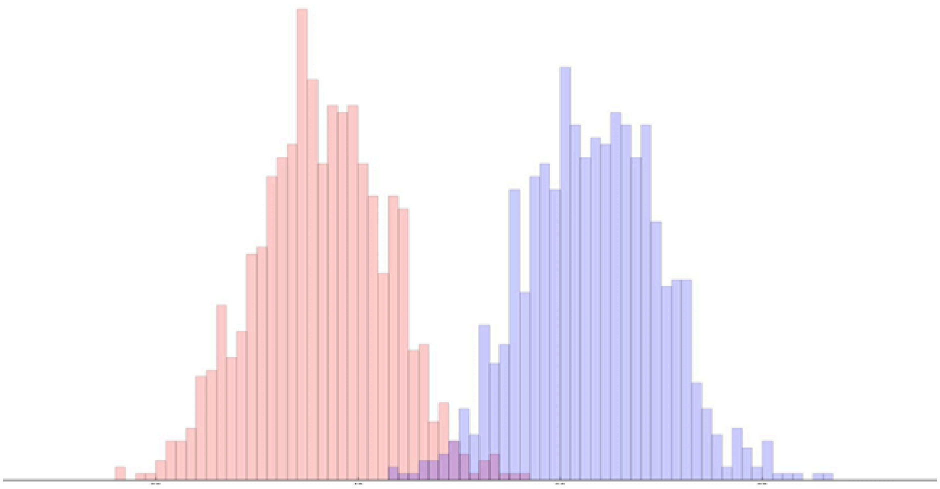
Bin 2 128-191

Bin 3 192-255

Combinations: 4*4*4*4=256

Histogram intersection algorithm

$$intersection = \sum_{j=1}^{n} \min\left(I_j, M_j\right)$$



Distance → Intersection captures similarity between two histograms

Distance = (1- similarity)

Normalized histogram= PDF
Comparing histograms= Comparing PDFs(use KLD)

UNIT 5

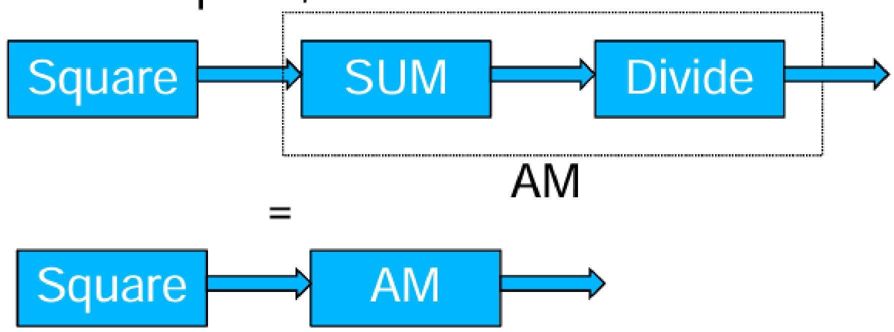
Sum of Squared errors: d1^2... dk^2 → a minimum

Root mean square error

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\left(y_i - f\left(x_i\right)\right)^2}$$

Squared Mean

SM: Square, sum and divide



then take root

Limitation: errors are squared and summed

Characteristics: Squaring a big no, becomes a bigger no

Weighted AM:

Weights of MSE

- Bigger numbers $\Rightarrow$ bigger weights
- If outliers happen to be larger number then big weight is allotted

$$weighted\ AM = \sum normal\_weight\ X\ normal\_data + reduced\_weight\ X\ outlier$$

- **Probability = normalized frequency distribution**

$$\sum_{i=1}^{r} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{r} p_i \log_2 p_i$$

MEE(Minimum error entropy)

- Compared to mean square error, entropy can handle the outliers effectively

Adaptive system with cost function as MSE

- Least mean square error

Adaptive system with cost function as entropy

- Minimum error entropy

entropy=–E(log(P(error))

Graph Theoretic Clustering

Parzen Window and PDF

$$\hat{F}(i) = \frac{-1}{2N\sigma^2} \sum_{j=1}^{N} \hat{V}_{i,j} d_{i,j}$$

$$\hat{V}(X) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{V}_{i,j}$$

where
$d_{i,j} = x_i - x_j$
$V_{ij} = G(d_{ij})$

BAYES CLASSIFICATION

Posterior=Prior*likelihood

$$P(flu|sneezing, builder) = \frac{P(sneezing, builder|flu).P(flu)}{P(sneezing, builder)}$$