$-\frac{1}{5}\cancel{(\log(\frac{1}{4}))}$

$P(\text{positive} \mid \text{sick}) = P(\text{positive}, \text{sick}) / P(\text{sick}) = 0.009 / 0.01 = 0.9$ ~~0.09~~

$P(- \mid \text{healthy}) = 0.891 / 0.99 = 0.9$

$P(\text{sick} \mid +) = 0.009 / 0.108 = 0.0833$

$P(\text{Healthy} \mid -) = 0.891 /$

Entropy

- Entropy is measure of randomness of random variable z.
- Represented by $H(z)$, for single event / random variable

$$H(z) = -\sum_{r \in z} P_z(z) \log_2 P_z(z)$$

- Entropy is maximum if $P_z(z) = \frac{1}{M}$ ie equiprobable for all values of z

Eg: Sun rises can rise from any direction / Toss of coin it can be H or T

- It is minimum : $P(z) = 1$ for a value of z (others are null)

Eg: Sun rises in east, there is no uncertainity

- If there several random variables it is called as ensemble of entropy

$$\sum_i p_i = 1$$

$$H = -1 = -\sum_i p_i \log_2 p_i \rightarrow \text{Marginal probability}$$

- Entropy of an ensemble is simply average of all elements in it. It can be computed by their average entropy by weighing each of $\log p_i$ contributions by its probability p_i occuring.

- In case of several events entropy is multiply by weight/probability in ensemble but not in case of entropy with 1 event since there is only 1 event & average would be 1

Example:

let z be an alphabet from A to Z but all prob of all alphabets is equiprobable

$$H = -\left[\frac{1}{26} \times 26 \log_2 \frac{1}{26}\right]$$

$$= 4.7$$

$$H = -\left[0.105 \times \log_2 0.105 + 0.072 \log_2 0.072 + 0.066 \log_2 0.066 + 0.063 \log_2 0.063 + 0.059 \log_2 0.059 + 0.055 \log_2 0.055 + 0.054 \log_2 0.054 + 0.052 \log_2 0.052 + 0.047 \log_2 0.047 + 0.035 \log_2 0.035 + 0.029 \log_2 0.029 + 0.023 \log_2 0.023\right]$$

$$= -\left[-0.105 \times 3.2515 - 0.072(3.8) - 0.066(3.92) - 0.063(3.9) - 0.059(4.08) - 0.055(4.18) - 0.054(4.21) - 0.052(4.26) - 0.047(4.4) - 0.035(4.8) - 0.029(5.1) - 0.023(5.42)\right]$$

$$= \underline{\underline{2.69}}$$

~~Conditional Entropy:~~

Joint Entropy:

- $$H = -\sum p(x,y) \log_2 p(x,y)$$

$$H = \sum_{x,y} P(x,y) \log_2 \frac{1}{P(x,y)}$$

- Joint entropy is additive

$$H(X,Y) = H(X) + H(Y) \qquad P(X,Y) = P(X) \cdot P(Y)$$

Given: $$H(X,Y) = \sum_{x,y} P(X) \cdot P(Y) \log_2\left(\frac{1}{P(X) \cdot P(Y)}\right)$$

$$= -\sum_{x,y} P(X) \cdot P(Y) \log_2 (P(X) \cdot P(Y))$$

$$= -\sum_{x,y} P(X) \cdot P(Y) \left[\log_2 P(X) + \log_2 P(Y)\right]$$

$$= -\sum_{x,y} P(X) \cdot P(Y) \log_2 P(X) + P(X) \cdot P(Y) \log_2 P(Y)$$

$$H(X) = -\sum_{X} P(X) \log_2 P(X) \qquad H(Y) = -\sum_{Y} P(Y) \log_2 P(Y)$$

$$H(X) + H(Y) = -\left[\sum_{X} P(X) \log_2 P(X) + \sum P(Y) \log_2 P(Y)\right]$$

$$= -\sum_{X} P(X)\log_2 P(X) \sum_{Y} P(Y) - \sum_{Y} P(Y)\log_2 P(Y) \sum_{X} P(X)$$

$\sum_{X} P(X) = 1$ & $\sum_{Y} P(Y) = 1$

$$= -\left[\sum_{X} P(X)\log_2 P(X) + \sum_{Y} P(Y)\log_2 P(Y)\right]$$

$$= H(X) + H(Y)$$

Conditional Entropy

Measures uncertainty about random variable X when R.V Y has taken some random value

$$H(X \mid Y=b_j) = \sum_{X} P(X \mid Y=b_j)\log_2 \frac{1}{P(X \mid Y=b_j)}$$

↳ This is for single event

$$H(X \mid Y) = \sum_{Y} P(Y)\left[\sum_{X} P(X \mid Y)\log_2 \frac{1}{P(X \mid Y)}\right]$$

$$= \sum_{X,Y} P(Y)\cdot P(X \mid Y)\cdot \log_2 \frac{1}{P(X \mid Y)}$$

$$= \sum_{X,Y} P(X,Y)\log_2 \frac{1}{P(X \mid Y)}$$

$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$ or $P(Y \mid X)\cdot P(X) = P(X,Y)$ $\Rightarrow P(X) = \frac{P(X,Y)}{P(Y \mid X)}$

Chain Rule for entropy

Prove: $H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$

$$H(X) + H(Y \mid X) = -\left[\sum_{X} P(X)\log_2 P(X) + \sum_{X,Y} P(X,Y)\log_2 P(Y \mid X)\right]$$

$$= -\left[\cancel{\sum_{X} P(X)\cdot \sum_{X,Y} P(X,Y)\log_2 P(X)}\right.$$

$$= -\left[\sum_{X}\right.$$

Krusha Doshi

$$H(X,Y) = -\sum_{X,Y} P(X,Y) \log_2 P(X,Y)$$

$$= -\sum P(Y|X)\, P(X) \log_2 \left[P(Y|X)\, P(X) \right]$$

$$= -\sum P(Y|X)\, P(X) \left[\log_2 P(Y|X) + \log_2 P(X) \right]$$

$$= -\left[\sum_{X,Y} P(Y|X) \cdot P(X) \log_2 P(Y|X) + \sum_{X,Y} P(Y|X)\, P(X) \cdot \log_2 P(X) \right]$$

$$= \left[-\sum_{X,Y} P(Y|X) \cdot P(X) \log_2 P(Y|X) - \sum_{X} P(X) \log_2 P(X) \sum_{Y} P(Y|X) \right]$$

$$= \left[-\sum_{X,Y} P(X,Y) \log_2 P(Y|X) + H(X) \underbrace{\left[\sum_{Y} P(Y|X) \right]}_{1} \right]$$

1 since its summation of $P(Y|X)$ & X is alre[...] given

$$= \left[-\sum_{X,Y} P(X,Y) \log_2 P(Y|X) + H(X) \right]$$

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

~~Assignment~~

Prove $H(X_1, X_2, \cdots X_n) \leq \sum_{i=1}^{n} H(X_i)$

$H(X) = -\sum_{x} P(x) \log P(x)$

$\sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} -\sum_{x} P(X_i) \log P(X_i)$

$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \cdots + H(X_n|X_1, X_2, \cdots, X_{n-1})$ $\Rightarrow$ By chain rule for entropy

By formula, $H(X, Y) = -\sum_{x,y} P(X,Y) \log P(X,Y)$

$\therefore H(X_1, X_2, \cdots, X_n) = -\sum_{x} P(X_1, \cdots, X_n) \log P(X_1, \cdots, X_n)$

Each conditional entropy $H(X_i|X_1, \cdots X_{i-1})$ is non negative, indicating

$\therefore$ ~~$H(X_1, X_2, \cdots, X_n) \leq \sum_{i=1} H\, H(X_i)$~~ that joint entropy will not exceed ~~marginal~~ total entropy of individual components. This follows the rule that $H(X|Y) \leq H(X)$ as existence of known value for Y tend to reduce random in X

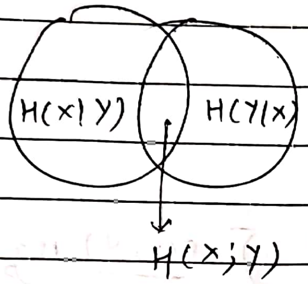
Hence,

$$\boxed{H(X_1, X_2, \cdots, X_n) \leq \sum_{i=1}^{n} H(X_i)}$$

Mutual Information between X & Y

Mutual info between 2 R.V measures amount of information that one conveys about other. Equivalently measures avg reduction in uncertainty about X that results from learning Y

$$H(X;Y) = \sum_{x,y} P(x,y) \log_2 \frac{P(x,y)}{P(x)\cdot P(y)}$$



- When X & Y are independent, mutual entropy is 0

$$I(X;Y) = \sum_{x,y} P(x,y) \log_2 \underbrace{\frac{P(X)\cdot P(Y)}{P(X)\cdot P(Y)}}_{1}$$

$$= \underline{\underline{0}} \qquad \{\log_2 1 = 0\}$$

Relative Entropy Mutual Info:

- It is always non negative or ≥ 0

$$P(X,X) = P(X)$$

self info

$$I(X;X) = \sum_{x} P(X) \log_2 \frac{P(X)}{P(X)\cdot P(X)}$$

$$= \sum_{x} P(x) \log_2 \frac{1}{P(x)}$$

$$= -\sum P(X) \log_2 P(X)$$

$$= H(X)$$

$\rightarrow$ 1

$I(X;Y) = H(X) - H(X|Y)$

$= I(Y;X)$

$= H(X) + H(Y) - H(X,Y) \rightarrow$ 2

Prove:

1 $I(X;Y) = H(X) - H(X|Y)$

$$= -\sum_{x} P(X) \log_2 P(X) - \left[-\sum_{x,y} P(X|Y) \log_2 P(X|Y) \right]$$

By Bayes theorem,

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$$= +\sum_{x} P(X) \log_2 \frac{1}{P(X)} + \sum_{x,y} \frac{P(X,Y)}{\cancel{P(X)}P(Y)} \log_2 \frac{P(X|Y)}{\cancel{X,Y}}$$

$$= \sum_{x} P(X) \cdot \log_2 \frac{1}{P(X)} + \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(Y)} \right)$$

$$= \cancel{\sum_{x} P(Y)} \sum_{x} P(X) \log_2 \frac{1}{P(X)} + \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(Y)} \right)$$

$$= \sum_{x} P(X) \log_2 \frac{1}{P(X)} + \sum_{x,y} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(Y)} \right)$$

$$I(X;Y) = \sum_{x,y} P(X,Y) \log_2 \frac{P(X,Y)}{P(X) \cdot P(Y)}$$

$$= \sum P(X|Y) \cdot P(Y) \log_2 \frac{P(X|Y) \cdot \cancel{P(Y)}}{P(X) \cdot \cancel{P(Y)}}$$

$$= \sum P(X|Y) \cdot P(Y) \log_2 \frac{P(X|Y)}{P(X)}$$

$$= \sum P(X|Y) \cdot P(Y) \log_2 P(X|Y) - \sum P(X|Y) P(Y) \log_2 P(X)$$

$$= -\left[\sum P(X,Y) \log_2 P(X|Y) \right] + \left[\sum P(X|Y) P(Y) \log_2 P(X) \right]$$

$$= \left[-H(X|Y) + H(X) \right]$$

$$= H(X) - H(X|Y)$$

$$= H(X) - H(X|Y)$$

$$= H(X) - \left[-H(Y) + H(X,Y) \right]$$

$$= H(X) + H(Y) - H(X,Y)$$

Cross Entropy

If there are 2 distributions $P(X)$ & $q(X)$ over same set of outcomes

$$H(P,q) = -\sum_{x} P(x) \log q(x)$$

If $P(x) = q(x)$

$$H(P) = \sum_{x} P(X) \log P(X)$$

- Cross entropy is asymmetric

$$H(P,q) \neq H(q,P)$$

Distance between X & Y (R.V)

$$D(X,Y) = H(X,Y) - I(X;Y)$$

$D(X,Y) \geq 0$ ∵ distances are non-negative

$D(X,X) = 0$

$D(X,Y) = D(Y,X) \Rightarrow$ symmetric

$D(X,Z) \leq D(X,Y) + D(Y,Z)$ (Triangle Inequality rule)

KL divergence

$$D_{KL}(p\|q) = \sum_{X} P(X) \log \frac{P(X)}{q(X)}$$

Here $p(x)$ & $q(x)$ are diff probability distributions defined over x

Application :

~~A model~~ say $x \rightarrow y_x$ is a training data has a model, this model cannot be applied on $z \rightarrow y_z$ if $p(x)$ & $q(z)$ are very different. They can be determined as different using ~~KL~~ distance between them & we can use KL divergence

$$D_{KL}(p\|q) = H(P,q) - H(P)$$

Scanned with OKEN Scanner

If variables are independent, they are always uncorrelated

If variables are uncorrelated, not necessary are independent under condition variables follows Gaussian joint distribution

Correlation & covariance Matrix & Wiener Filter

1\. Hermitian Matrix

$$C^H = C$$

{ what is conjugate ? $y = a + jb \Rightarrow y^* = a - jb$ }

H → Hermitian operator

→ conjugate + transpose

C → conjugate Matrix

2\. covariance Matrix

Auto covariance Matrix

$$C_{p\times p} = \left[\begin{bmatrix} (x_1 - \mu x_1) \\ (x_2 - \mu x_2) \\ \vdots \\ (x_p - \mu x_p) \end{bmatrix}_{1\times p} \left[(x_1 - \mu x_1), \cdots (x_p - \mu x_p) \right]^*_{p\times 1} \right]$$

$$= \begin{bmatrix} C_{x_1 x_1} & C_{x_1 x_2} & \cdots & C_{x_1 x_p} \\ \vdots & & & \\ \vdots & - & - & C_{x_p x_p} \end{bmatrix}_{p\times p}$$

Prove : covariance matrix is always symmetric

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$C = E\left[(\underline{x} - \mu \underline{x})(\underline{x} - \mu \underline{x})^H \right]$$

Given $C^H = C \Rightarrow$ To prove symmetric

$$C = \begin{bmatrix} C_{x_1 x_1} & C_{x_1 x_2} & \cdots & C_{x_1 x_p} \\ \vdots & & & \\ - & - & - & C_{x_p x_p} \end{bmatrix}_{p\times p}$$

$$C^H = \begin{bmatrix} C_{x_1 x_1} & - & - & C_{x_p x_1} \\ C_{x_1 x_2} & & & \\ \vdots & & & \\ C_{x_1 x_p} & - & & C_{x_p x_p} \end{bmatrix}$$

$$(AB)^H = \left[(AB)^T \right]^* = \left[B^T A^T \right]^* = B^H A^H$$

$$C^H = E\left[(\underbrace{(\underline{x} - \mu \underline{x})}_{A} \underbrace{(\underline{x} - \mu \underline{x})^H}_{B})^H \right]$$

$$= E\left[(\underline{x} - \mu \underline{x})^{H} (\underline{x} - \mu \underline{x})^H \right]$$

$$= \underline{\underline{C}}$$

→ covariance matrix is always hermitian matrix

— If mean is 0 then

↳ $E[(x)(x)^H]$

$\Downarrow$

correlation / variance / auto-covariance

$\mu_x = E[x$

$\sigma_x^2 = E[|(x-\mu_x)|^2]$

1. $C_{\underline{x},\underline{y}} = E[(\underline{x}-\underline{\mu_x})(\underline{y}-\underline{\mu_y})^H]$

Here $\underline{x}$ & $\underline{y}$ are vectors

2. $C_{x,y} = E[(x-\mu x)(y-\mu y)^H]$

Here x & y are scalars ~~vector~~

$= E[(x-\mu x)(y-\mu y)^*]$

$= E[xy^*] - E[\mu x\, y^*] - E[x\,\mu y^*] +$
$E[\mu x\, \mu y^*]$

∵ $\mu x = E[x]$ Expectation & expectation on expectation is always same ($\mu x = E[x]$)

$= E[xy^*] - \mu x\, E[y^*] - \mu y^* E[x] + \mu x\, \mu y^*$

$= E[xy^*] - \mu x\, \mu y^* - \cancel{\mu y^* \mu x} + \cancel{\mu x\, \mu y^*}$

$= E[xy^*] - \mu x\, \mu y^*$

if $C_{x,y} = 0$ ie

$E[xy^*] = \mu x\, \mu y^* \Rightarrow$ uncorrelated

[correlation between 2 random variables equals to product of means]

Random process ⇒ Random variable becomes f^n of time

Stationarity

- The data is now wrt time.

First order stationary

$$E[x(n)] = \mu_x = \mu$$

$x(n) \rightarrow$ Discrete time random process

~~No~~ Power = 1 ⇒ First order

Eg: So if there are 500 samples are taken & it results into mean μ, now again another 500 samples are taken & mean is taken again which comes out to μ & so on

Then the series is stationary.

Second order stationary / wide sense stationarity (WSS).

1. $E[x(n)] = \mu$
2. $E[\underset{A}{x(n)}\ \underset{B}{x^*(n-k)}] = r(k)$ k : lag/gap

$$r^*(-k) = E[x(n+k)\, x^*(n)] \qquad n+k = m$$
$$= E[x(m)\, x^*(m-k)] \qquad n = m-k$$
$$= r(k)$$

Hence, it is Hermitian/conjugate symmetry

If $k = 0$,

$$r(0) = E[|x(n)|^2] \geq 0 \Rightarrow \text{Average power of signal}$$

Auto correlation :

- Correlation with same random variable
- $\mu = 0$

$$R_{p \times p} = E[\underline{x}\ \underline{x}^H]$$

$$= \begin{bmatrix} E[|x(n)|^2] & E[x(n)\, x^*(n-1)] & -- & E[x(n)\, x^*(n-p)] \\ | & & & \\ | & - & & E[|x(n-p)|^2] \end{bmatrix}$$

To get this value $\Rightarrow n-(n-k)$

$$= \begin{bmatrix} r(0) & r(1) & \cdots & r(p) \\ \vdots & & & \\ r(-p) & r(1-p) & \cdots & r(0) \end{bmatrix}$$

$\because r^*(-k) = r(k)$

$$= \begin{bmatrix} r(0) & r(1) & \cdots & r(p) \\ \vdots & & & \\ r^*(p) & r^*(p-1) & \cdots & r(0) \end{bmatrix}$$

Toeplitz Matrix $\Rightarrow$ If ~~above~~ upper diagonal values are known then we can compute lower diagonal values.

DATE 03 11 24

~~Weiner Filter~~

⇒ If $T_{p\times p}$ is a hermition matrix or a toeplitz matrix $\{ T_{p\times p} = T^H_{p\times p} \}$

$x \neq 0$

$\underline{T}\,\underline{x} = \lambda \underline{x}$ → Eigen value decomposition

Matrix Eigen vector

$(\underline{T}\,\underline{x})^H = \lambda^* \underline{x}^H$ { order doesn't matter in λx since λ is a scalar }

$\underline{x}^H \underline{T}^H = \lambda^* \underline{x}^H$

$\underline{x}^H \underline{T} = \lambda^* \underline{x}^H$

$\underline{x}^H \underline{T}\,\underline{x} = \lambda^* \underline{x}^H x$ (Multiplying x both the sides)

$\underline{x}^H \lambda \underline{x} = \lambda^* \underline{x}^H \underline{x}$ We multiply after $\underline{x}^H \underline{T}$ since or multiplication might gets messed up)

$\lambda \underline{x}^H \underline{x} = \lambda^* \underline{x}^H x$

$(\lambda - \lambda^*)\, \underline{x}^H \underline{x} = 0$

{ By property, $\underline{x}^H \underline{x} = \sum_{i=1}^{p} |x_i|^2$ ∴ $\underline{x}^H \underline{x} \neq 0$ }

∴ $(\lambda - \lambda^*) = 0$

∴ $\lambda = \lambda^*$

Hence, eigen values are always real

⇒ Prove x_1^H & x_2 are orthogonal if $\lambda_1 \neq \lambda_2$

$\underline{T}\,\underline{x_1} = \lambda_1 \underline{x_1}$ — ①

$\underline{T}\,\underline{x_2} = \lambda_2 \underline{x_2}$ — ②

Apply Hermition on eqn ①

$(\underline{T}\,\underline{x_1})^H = (\lambda_1 x_1)^H$

$\underline{x_1}^H \underline{T}^H = \lambda_1^* \underline{x_1}^H$ — ③

Multiply $\underline{x_2}$ on eqn ③

$\underline{x_1}^H \underline{T}^H \underline{x_2} = \lambda_1^* \underline{x_1}^H \underline{x_2}$

$\underline{x_1}^H \underline{T}\,\underline{x_2} = \lambda_1^* \underline{x_1}^H \underline{x_2}$

$\underline{x_1}^H \lambda_2 \underline{x_2} = \lambda_1^* \underline{x_1}^H \underline{x_2}$

$\lambda_2 \underline{x_1}^H \underline{x_2} = \lambda_1^* \underline{x_1}^H \underline{x_2}$

$(\lambda_2 - \lambda_1^*)\, \underline{x_1}^H \underline{x_2} = 0$ { ∵ $\lambda_1 = \lambda_1^*$ }

since $\lambda_1 \neq \lambda_2$ ∴ $\underline{x_1}^H \underline{x_2} = 0$ ⇒ orthogonal

Example:

$$A = \begin{bmatrix} 1+2j & 2 \\ 3 & 4+2j \end{bmatrix} \qquad A^H = \begin{bmatrix} (1+2j)^* & 3^* \\ 2^* & (4+2j)^* \end{bmatrix}$$

$$A^H A = \begin{bmatrix} (1+2j)^* & 3 \\ 2 & (4+2j)^* \end{bmatrix} \begin{bmatrix} 1+2j & 2 \\ 3 & 4+2j \end{bmatrix}$$

$$= \cancel{\left[(1+2j)^* (1+2j) + 9 + 4 + (4+2j)^* (4+2j) \right]}$$

$$= \cancel{(1-2j)(1+2j) + 13 + (4-2j)(4+2j)}$$

$$= \cancel{(1-(4j^2)) + 13 + 16 - 4j^2}$$

$$= \cancel{30 - 8j^2}$$

$$= \cancel{38}$$

$$= \begin{bmatrix} 1+4+9 & 2-2j+12+6j \\ 2+4j+12-6j & 4+16+4 \end{bmatrix}$$

$$= \begin{bmatrix} 14 & 14+4j \\ 14-2j & 24 \end{bmatrix}$$

$\Rightarrow$ If all eigen values & vectors are combined into matrix

$$T \left[e_1 \; e_2 \; e_3 \; \cdots \; e_p \right] = \left[\lambda_1 e_1 \; \lambda_2 e_2 \; \cdots \; \lambda_p e_p \right]$$

$\Downarrow$

lets call as E

$$TE = \left[e_1 \; e_2 \; \cdots \; e_p \right] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & & 0 \\ \vdots & & & & \\ 0 & \cdots & & & \lambda_p \end{bmatrix}$$

$$TE = ED$$

$$E^H E = \begin{bmatrix} e_1^H \\ e_2^H \\ \vdots \\ e_p^H \end{bmatrix} \left[e_1 \; e_2 \; \cdots \; e_p \right] \quad \text{—④}$$

Apply E^H both the sides on eqn (1)

$$I \, E \, E^H = E \, D \, E^H$$

As we proved in previous theorem, x_i^H & x_i are orthogonal

$$I \, I = E \, D \, E^H$$

$$\boxed{I = E \, D \, E^H}$$ decomposition of Toeplitz matrix

* { $E E^H = I$ or $E^{-1} = E^H$ } We use to hermition since it avoids inverse matrix computation
* computation of inverse matrix is difficult for non square matrix
* If $|A| = 0$ then it might be problematic

Weiner Filters

- Important for optimization in image processing → considers both degradation fn & noise
- Uses supervised learning ~~tech~~
- Also called Minimum Mean sq error or least square error filter

⇒ It removes additive noise & inverts bluring simultaneously

Minimizing error

$$e(n) = y(n) - d(n)$$

$e(n)$ ↓ error, $y(n)$ ↓ o/p, $d(n)$ ↓ actual value

$$E^2 = E[e^2(n)]$$ ⇒ We apply statistical operation since there will be randomness in signals

Mean squared error

$$= E[e(n) \, e^H(n)]$$

$$= E[e(n) \, e^T(n)] \quad \{\because e^H(n) \text{ is } e^T(n) \text{ as } e(n) \text{ has real values}\}$$

$$= E[(d(n) - y(n)) \, e^T(n)]$$

$$= E[(d(n) - W^T x(n)) \, e^T(n)] \quad \{\because y(n) = W^T x(n)\}$$

⇓ Weiner filter is linear estimation of original image

$$= E[(d(n) - W^T x(n))(d(n) - W^T x(n))^T]$$

$$\begin{bmatrix} w_0 & w_1 \end{bmatrix} \begin{bmatrix} x(n) \\ x(n-1) \end{bmatrix} = w_0 x(n) + w_1 x(n-1)$$

$$x^T(w) = \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$= E\left[d^2(n) - w^T x(n) d(n) - w^T x(n) d(n) + w^T x^T(n)\right]$$

{ ∵ $d(n)$ is a scalar ∴ $d^T(n) = d(n)$ }

$$= E\left[d^2(n) - \underline{w}^T \underline{x}(n) d(n) - d(n)(\underline{w}^T \underline{x}(n))^T + \underline{w}^T \underline{x}(n) (\underline{w}^T \underline{x}(n))^T\right]$$

$$= E\left[d^2(n) - \underline{w}^T \underline{x}(n) d(n) - d(n) \underline{x}^T(n) \underline{w} + \underline{w}^T \underline{x}(n)(\underline{x}^T(n) \underline{w})\right]$$

{ $w^T x(n) = x^T(n) w$
$(w^T x(n))^T = x^T(n) w = w^T x(n)$ }

$$= E\left[d^2(n) - \underline{w}^T \underline{x}(n) d(n) - d(n) \underline{w}^T \underline{x}(n) + \underline{w}^T \underline{x}(n) \underline{x}^T(n) \underline{w}\right]$$

$$= E\left[d^2(n) - 2\underline{w}^T \underline{x}(n) d(n) + \underline{w}^T \underline{x}(n) \underline{x}^T(n) \underline{w}\right]$$

$$= E\left[d^2(n)\right] - 2E\left[\underline{w}^T \underline{x}(n) d(n)\right] + E\left[\underline{w}^T \underline{x}(n) \underline{x}^T(n) \underline{w}\right]$$

$$= \sigma_d^2 - 2\underline{w}^T E\left[\underline{x}(n) d(n)\right] + \underline{w}^T E\left[\underline{x}(n) \underline{x}^T(n)\right] \underline{w}$$

$$\boxed{E^2 = \sigma_d^2 - 2\underline{w}^T \underline{p} + \underline{w}^T \underline{R} \underline{w}} \quad \text{--- (1)}$$

$\underline{p}$ ⇓ cross correlation b/w $x(n)$ & $d(n)$

{ we don't apply expectation on w since its static in nature & doesn't change with time }

Gradient Descent optimization

objective : w should be optimized ie coefficients should be optimized

Take first order derivative of w for equ (1)

$\nabla_w E^2 = \underline{0}$ (∵ we want minima)

Since E^2 is a vector so we need to equate with 0 vector

$$\nabla_W E^2 = \frac{\partial}{\partial w_k}\left(\sigma_d^2 - 2\underline{w}^T \underline{p} + \underline{w}^T \underline{R} \underline{w}\right)$$

let $A = w^T p$

second order derivative → +ve → minima
second order derivative → -ve → Maxima

$\frac{\delta A}{\partial w_k} =$

$A = w_0 P(0) + w_1 P(1) + \dots + w_k P(k) + w_N P(N)$

$\frac{\partial A}{\partial w_k} = P(k)$

$$\nabla_w A = \begin{bmatrix} \frac{\partial A}{\partial w_0} \\ \frac{\partial A}{\partial w_1} \\ \vdots \\ \frac{\partial A}{\partial w_N} \end{bmatrix} = \begin{bmatrix} P(0) \\ P(1) \\ \vdots \\ P(k) \\ \vdots \\ P(N) \end{bmatrix} = \underline{P}$$

let $B = \underline{w}^T \underline{R} \underline{w}$ — w: $((N+1)\times 1)$, R: $(N+1) \times (N+1)$

$$B = \sum_{i=0}^{N} w_i \sum_{j=0}^{N} R_{ij} w_j$$

$$\nabla_w B = \begin{bmatrix} \frac{\partial B}{\partial w_0} \\ \frac{\partial B}{\partial w_1} \\ \vdots \\ \frac{\partial B}{\partial w_k} \\ \vdots \\ \frac{\partial B}{\partial w_N} \end{bmatrix} = \begin{bmatrix} \\ \end{bmatrix}$$

$$\cancel{\frac{\partial B}{\partial w_k}}\ B = \sum_{\substack{i=0 \\ i\neq k}}^{N} w_i \sum_{j=0}^{N} R_{ij} w_j + w_k \sum_{j=0}^{N} R_{kj} w_j$$

(j does contain k^{th} term)

$$\frac{\partial B}{\partial w_k} = \sum_{\substack{i=0 \\ i\neq k}}^{N} w_i R_{ik} + \frac{\partial}{\partial w_k}\left[w_k \sum_{\substack{j=0 \\ j\neq k}}^{N} R_{kj} w_j + w_k R_{kk} w_k \right]$$

$$= \sum_{\substack{i=0 \\ i\neq k}}^{N} w_i R_{ik} + \sum_{\substack{j=0 \\ j\neq k}}^{N} R_{kj} w_j + 2 w_k R_{kk}$$

$$= \sum_{\substack{i=0 \\ i\neq k}}^{N} w_i R_{ik} + \sum_{\substack{j=0 \\ j\neq k}}^{N} R_{kj} w_j + w_k R_{kk} + w_k R_{kk}$$

$$= \sum_{i=0}^{N} R_{ik} W_i + \sum_{j=0}^{N} R_{kj} W_j$$

$\because$ R is hermitian, $R^T = R$ $\therefore R_{ji} = R_{ij}$

$$= \sum_{i=0}^{N} W_i R_{ki} + \sum_{j=0}^{N} R_{kj} W_j$$

$$= \left[W_0 R_{k0} + W_1 R_{k1} + \cdots W_N R_{kN} \right] + \left[R_{k0} W_0 + R_{k1} W_1 + \cdots + R_{kN} W_N \right]$$

$$= 2 \sum_{i=0}^{N} W_i R_{ki}$$

$$\frac{\partial B}{\partial w_0} = 2 \sum_{i=0}^{N} w_i R_{0i} = 2 \left[W_0 R_{00} + W_1 R_{01} + \cdots + W_N R_{0N} \right]$$

$$\frac{\partial B}{\partial w_1} = 2 \sum_{i=0}^{N} W_i R_{1i} = 2 \left[W_0 R_{10} + W_1 R_{11} + \cdots + W_N R_{1N} \right]$$

$$\therefore \nabla_{WB} = \begin{bmatrix} \frac{\partial B}{\partial w_0} \\ \frac{\partial B}{\partial w_1} \\ \vdots \\ \frac{\partial B}{\partial w_k} \\ \vdots \\ \partial B / \partial w_N \end{bmatrix} = 2 R \underline{W}$$

(R is matrix, P is vector)

$$\nabla_W E^2 = 0 - 2\underline{P} + 2R\underline{W} = 0$$

$$= -2\underline{P} + 2R\underline{W} = 0$$

$$\therefore \underline{P} = R\underline{W} \qquad \therefore \boxed{\underline{W}^* = \underline{R}^{-1} \underline{P}}$$

W is optimal weights / optimal filter

(When R is orthogonal then we can write $R^{-1} = R^{H}$)

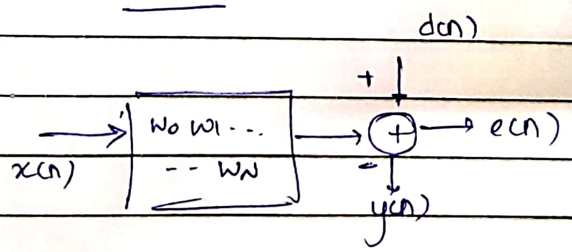
Ex: $R = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ $\quad P = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$

$W = ?$ using Gradient descent opti

$$|R| = ad - bc$$
$$= 1 - 0.81$$
$$= 0.19$$

$$R^{-1} = \frac{1}{0.19}\begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} = \begin{bmatrix} 5.26 & -4.73 \\ -4.73 & 5.26 \end{bmatrix}$$

$$W = R^{-1} P$$
$$= \frac{1}{0.19}\begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$$
$$= \begin{bmatrix} 5.26 & -4.73 \\ -4.73 & 5.26 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$$
$$= \begin{bmatrix} 0.579 \\ -0.42 \end{bmatrix}$$



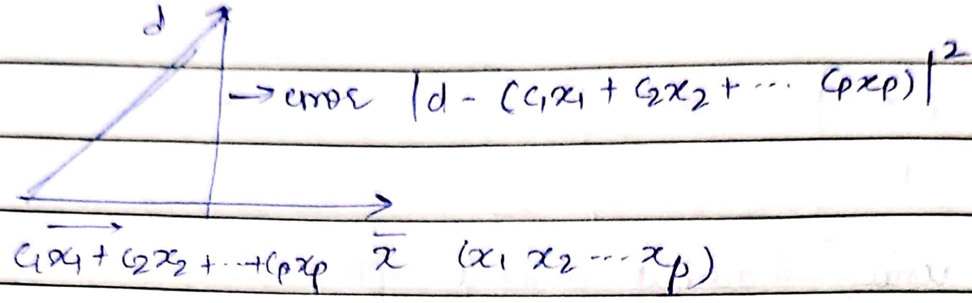
$$E^2 = E\left[e^2(n)\right]$$

Objective: Minimize mean square error.

Let us consider a random variable x whose values are $x_1, x_2 \dots x_p$

Suppose we want to estimate another random variable who is linear combination of x & it is called linear estimation

$$\hat{d} = c_1 x_1 + c_2 x_2 + \dots c_p x_p$$



if weights are optimal then error is minimum or is orthogonal to $\bar{x}$