Pattern & Anomaly Detection Theory

Assignment -1

Submitted By:
Dhruv Singhal
500075346
R177219074
AIML B3

Submitted To:
Dr. Gopal Phartiyal
Asst. Professor
SOCS
UPES

Q.1

[Dhruv Singhal
500075346
R177219074]

Ans Mean

$E(X+Y) = \iint (x+y)\, p(x,y)\, dx\, dy$

As x & y are independent

$P(x,y) = P(x) \cdot P(y)$

$= \iint (x+y)\, p(x)\, p(y)\, dx\, dy$

$= \iint x\, p(x)\, p(y)\, dx\, dy + \iint y\, p(x)\, p(y)\, dx\, dy$

$= \iint p(y)\, dy\, x\, p(x)\, dx + \iint p(x)\, dx\, y\, p(y)\, dy$

As $\int p(y)\, dy = 1$, $\int p(x)\, dx = 1$

So, $= \int x\, p(x)\, dx + \int y\, p(y)\, dy$

$= E(X) + E(Y)$

$\therefore \boxed{E(X+Y) = E(X) + E(Y)}$

Variance

$Var(X+Y) = \iint ((x+y) - E(x+y))^2\, p(x,y)\, dx\, dy$

$= \iint \{(x+y)^2 - 2(x+y)E(x+y) + E^2(x+y)\}\, p(x,y)\, dx\, dy$

$= \iint (x+y)^2\, p(x,y)\, dx\, dy - 2E(x+y) \iint (x+y)\, p(x,y)\, dx\, dy + E^2(x+y)$

$= \iint (x+y)^2\, p(x,y)\, dx\, dy - E^2(x+y)$

$= \iint (x^2+y^2+2xy)\, p(x)\, p(y)\, dx\, dy - E^2(x+y)$

$= \iint x^2\, p(x)\, p(y)\, dx\, dy + \iint y^2\, p(x)\, p(y)\, dx\, dy + \iint 2xy\, p(x)\, p(y)\, dy\, dx - E^2(x+y)$

$= \iint x^2\, p(x)\, p(y)\, dx\, dy + \iint p(x)\, p(y)\, y^2\, dx\, dy + 2\iint xy\, p(x)\, p(y)\, dx\, dy - E^2(x+y)$

$= E(x^2) + E(y^2) - E^2(x+y) + 2\iint xy\, p(x)\, p(y)\, dx\, dy$

$= E(x^2) + E(y^2) - (E(x) + E(y))^2 + 2\iint xy\, p(x)\, p(y)\, dx\, dy$

$$= E(x^2) - E^2(x) + E(y^2) - E^2(y) - 2E(x)E(y) + 2\iint xy\, p(x)\, p(y)\, dx\, dy$$

$$= Var(X) + Var(Y) - 2E(X)E(Y) + 2\left(\int x\, p(x)\, dx\right)\left(\int y\, p(y)\, dy\right)$$

$$= Var(X) + Var(Y) - \cancel{2E(X)E(Y)} + \cancel{2E(X)E(Y)}$$

$$\boxed{\therefore Var(X+Y) = Var(X) + Var(Y)}$$

Q-2

Dhruv Singhal
500075346
R177219074

$P(a) = 1/3$
$P(b) = 1/3$
$P(c) = 1/3$

$P_{xa} = P_{ya} = P_a$
$P_{xb} = P_{yb} = P_b$
$P_{xc} = P_{yc} = P_c$

$\therefore H(X) = -[P(a)\log_2 P(a) + P(b)\log_2 P(b) + P(c)\log_2 P(c)]$

a) $\boxed{H(X) = +1.584}$

b) $\boxed{H(Y) = H(X) = 1.584}$

$H(X,Y) = -[P_{aa}\log_2 P_{(aa)} + P_{ab}\log_2 P_{(ab)} + \dots + P_{cc}\log_2 P_{(cc)}]$

$= -[6 \times \frac{1}{6}\log_2(\frac{1}{6})]$

d) $\boxed{H(X,Y) = 2.584}$

$\because H(X,Y) = H(X|Y) + H(Y)$ or
$= H(Y|X) + H(X)$

e) $\therefore$ $\underline{H(X|Y) = H(X,Y) - H(Y) = 1}$

c) $\underline{H(Y|X) = H(X,Y) - H(X) = 1}$

f) Mutual Information:

$I(X,Y) = H(X) - H[X|Y) = H(Y) - H(Y|X)$
$= 1.584 - 1$

$\boxed{I(X,Y) = 0.584}$

[Dhruv Singhal
500075346
R177219074]

Q-3

x	$P(x)$	$P(x) \cdot x$
0	1/5	0
1	1/5	1/5
2	1/5	2/5
3	1/5	3/5
4	1/5	4/5

$\therefore$ Mean $E(x) = 0 + 1/5 + 2/5 + 3/5 + 4/5$

$= 10/5$

$\boxed{\text{Mean } E(x) = 2}$

Variance $Var(x) = \left[0^2\left(\frac{1}{5}\right) + 1^2\left(\frac{1}{5}\right) + 2^2\left(\frac{1}{5}\right) + 3^2\left(\frac{1}{5}\right) + 4^2\left(\frac{1}{5}\right)\right] - (Ex)^2$

$= \left[\frac{1}{5} + \frac{4}{5} + \frac{9}{5} + \frac{16}{5}\right] - (2)^2$

$= \frac{30}{5} - 4$

$= 6 - 4$

$\boxed{Var(x) = 2}$

$\therefore$ Standard Deviation $(\sigma) = \sqrt{Var(x)}$

$= \sqrt{2}$

$\boxed{\text{Standard Deviation} = 1.414}$

Dhruv Singhal 500075346 R177219074

Q-4

$x = \{0,1,2,3,4\}$, $N = 500$

$x = x_k$	0	1	2	3	4
Number of Prisoners	80	265	100	40	15
$f(x) = P(X=x)$	80/500	265/500	100/500	40/500	15/500
$f(x) = P(X=x)$	0.16	0.53	0.2	0.08	0.02

Mean, Variance?

Soln

$\because$ Mean $E(X) = \sum_{i=0}^{4} f(x_i)(x_i)$

$\therefore$ $E(X) = 0.16(0) + 0.53(1) + 0.2(2) + 0.08(3) + 0.02(4)$

$E(x) = 1.29$

$\boxed{\text{Mean} = 1.29}$

Variance $Var(X) = \left[\sum_{i=0}^{4} f(x_i)(x_i)^2\right] - (Ex)^2$

$Var(x) = \left[0^2(0.16) + 1^2(0.53) + 2^2(0.2) + 3^2(0.08) + 4^2(0.02)\right] - (1.29)^2$

$= 2.53 - 1.66$

$\boxed{Var(x) = 0.87}$

Q-5 Dhruv Singhal R177219074 500075346

Ans Let us Determine the number of trials, success & the failure. The trial is the drawing a Card 5 times. Thus,

$n = 5$, $x = 3$ (exactly 3 heart)

success (p) (draw a heart card) $= \frac{13}{52} = \frac{1}{4} = 0.25$

failure (q) (not a heart card) $= 1 - p = 0.75$

$\because$ binomial distribution is ${}^{n}C_{x}\, p^{x} q^{n-x}$

$$\therefore = {}^{5}C_{3}(0.25)^{3}(0.75)^{2}$$

$$= \frac{5!}{3!\,2!} \times \left(\frac{1}{4}\right)^{3}\left(\frac{3}{4}\right)^{2}$$

$$= \frac{5 \times 4 \times \cancel{3!}}{\cancel{3!} \times 2!} \times \left(\frac{1}{4}\right)^{5} \times 9$$

$$= \frac{\cancel{10}^{5} \times 9}{4 \times 4 \times 4 \times 4 \times \cancel{4}_{2}} = \frac{45}{128 \times 4}$$

P(exactly 3 heart in 5 throws) $= \underline{0.088}$

$\because$ Mean $\mu = np = 5 \times \frac{1}{4} = \frac{5}{4} = \underline{1.25}$

$\because$ Variance, $\sigma^{2} = npq = 5 \times \frac{1}{4} \times \frac{3}{4} = \frac{15}{16} = \underline{0.9375}$

Q-6 Dhruv Singhal 500075346 R177219074

Ans) It is based on Baye's Theorem

Hence it is given that head will occur.

But we don't know if it will be due to the regular or fake coins.

Regular coin=2

fake coin=1

Total coins=3

$\therefore$ P(regular coin) = 2/3

P(fake coin) = 1/3

$\therefore$ P(head|regular coin) = 1/2 [As normal coin]

P(head|fake coin) = 1 [$\because$ It is double headed]

a) $\therefore$ P(head) = P(fake) x P(head|fake) + P(regular) x P(head|regular)

$= \frac{1}{3} \times 1 + \frac{\cancel{2}}{3} \times \frac{1}{\cancel{2}}$

$= \frac{1}{3} + \frac{1}{3}$

$P(head) = \frac{2}{3}$

b) $P(\text{fake coin}|\text{head}) = \frac{P(\text{fake coin}) * P(\text{head}|\text{fake coin})}{P(\text{head})}$

$= \frac{\frac{1}{3} \times 1}{\frac{2}{3}} = \frac{1 \times \cancel{3}}{\cancel{3} \times 2}$

P(fake coin|head) = 1/2

Q-7 Dhruv Singhal 500075346 R177219071

$n = 8,$

States (x)	a	b	c	d	e	f	g	h
Probabilities	1/2	1/4	1/8	1/16	1/64	1/64	1/64	1/64
Code String	0	10	110	1110	111100	111101	111110	111111

a) entropy $= -\sum_{i=1}^{n} P(x_i)\log_2 x_i$

$\therefore$ entropy

$$H(X) = -\left[\left(\tfrac{1}{2}\right)\log_2\left(\tfrac{1}{2}\right) + \tfrac{1}{4}\log_2\left(\tfrac{1}{4}\right) + \tfrac{1}{8}\log_2\left(\tfrac{1}{8}\right) + \tfrac{1}{16}\log_2\left(\tfrac{1}{16}\right) + 4\times\tfrac{1}{64}\log_2\left(\tfrac{1}{64}\right)\right]$$

$$= -[-0.5 - 0.5 - 0.375 - 0.25 - 0.375]$$

$\boxed{\text{entropy } H(X) = 2}$ bits/symbol

b) Average Code length $= \sum_{i=1}^{n} P(x_i)$ (Code string length of x_i)

$$= \left[\tfrac{1}{2}(1) + \tfrac{1}{4}(2) + \tfrac{1}{8}(3) + \tfrac{1}{16}(4) + \tfrac{1}{64}(6)\times 4\right]$$

$$= \left[\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{3}{8} + \tfrac{1}{4} + \tfrac{3}{8}\right]$$

$$= \left[\tfrac{1}{2} + \tfrac{3}{4} + \tfrac{6}{8}\right]$$

$$= \left[\tfrac{4+6+6}{8}\right] = \tfrac{16}{8} = 2.00 \text{ bits/symbol}$$

Average Code length $L(X) = 2.00$ bits/symbol

c) Coding Efficiency $\eta = \dfrac{H(X)}{L(X)} = \dfrac{2}{2} = 1$

Dhruv Singhal 500075346 R177219074

Q-8

Ans PCA [Principal Component Analysis] is used for Dimensionality Reduction. It includes 4 steps.

1. Centre the data: By subtracting mean from each data point so that data is centered around 0 & therefore has zero mean.

$$\hat{x} = x - \bar{x}$$

2. Compute Covariance Matrix:

$$S = \frac{1}{N} \hat{x}^T x$$

3. Compute Eigen values & vectors using Eigen value Decomposition:

$$Cov(W) = \begin{pmatrix} Var(w_1) & 0 & - - & 0 \\ 0 & Var(w_2) & - - & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & Var(w_m) \end{pmatrix}$$

So we perform Eigen value Decomposition of the covariance matrix S

$$S = \Gamma \Lambda \Gamma^T \qquad [\Gamma = \text{Eigenvectors}$$

4. Dimensionality Reduction:

$$Z = \hat{x} \Gamma'$$

we truncate Γ by keeping only the eigen vectors corresponding to the largest M eigen values. we call the truncated Γ matrix as Γ'.

PCA can be formulated in two ways:

a) Maximum Variance Formulation

b) Minimum Variance formulation

In this question, we are going to address Minimum Variance formulation.

Dhruv Singhal 500075348 R177219074

Minimum Error Formulation:

Let $\{x_n\}, n=1,2,\dots,N$ & $x_n \in R^D$ be the set of data observations, then our goal is to find transformation that minimises the reconstruction error.

$$J(x, \tilde{x}) = \frac{1}{N}\sum_{n=1}^{N} ||x_n - \tilde{x}_n||^2 \quad -①$$

[$\tilde{x}_n$ = reconstruction generated from lower dimensional latent variable]

Here, we have a complete D dimensional orthonormal (orthogonal & unit length) basis w_i, where $w_i = 1,2,\dots,D$ Then

$$w_i^T w_j = \begin{cases} 1, & i=j \\ 0, & \text{else} \end{cases}$$

Since basis is complete, we can represent any vector as a linear combination of the basis vectors.

$$X_n = \sum_{i=1}^{D} \alpha_{ni} w_i$$

$\alpha_{ni} = X_n^T w_i$ (property of orthonormal bases)

$$\therefore X_n = \sum_{i=1}^{D} (X_n^T w_i) w_i$$

We can represent M dimensional space using the first M basis vectors

$$\tilde{X}_n = \sum_{i=1}^{M} Z_{ni} W_i + \sum_{i=M+1}^{D} b_i w_i$$

$[Z_{ni} = X_n^T w_i]$

$$\therefore X_n - \tilde{X}_n = \sum_{i=M+1}^{D} Z_{ni} W_i - b_i w_i \quad -②$$

$\therefore$ putting ② in ①

$$J = \frac{1}{N}\sum_{n=1}^{N} || \sum_{i=M+1}^{D} Z_{ni} W_i - b_i w_i ||^2$$

Taking derivative w.r.t z_{nj} & setting to zero, we get

$$Z_{nj} = X_n^T w_j$$

Taking derivative w.r.t b_j & setting to zero, we get

$$b_j = \tilde{X}^T w_j$$

$\therefore J(w) = \sum_{i=M+1}^{D} w_i^T S w_i$

[Dhruv Singhal
R177219074
500075346]

$[S = (x_n - \tilde{x})(x_n - \tilde{x})^T]$

Our goal is to minimise $J(w)$, but there exist a trivial solution when $w=0$. To overcome this, we again use orthonormal basis property i.e.,

$$\|w\|^2 = 1$$

Now, using example $D=2$ & $M=1$, choose direction w_2 so that with to minimise

$$J = w_2^T S w_2 + \lambda_2 (1 - w_2^T w_2)$$

taking derivative w.r.t w_2 & setting to 0

$$S w_2 = \lambda_2 w_2$$

[w_2 is eigen vector corresponding to eigen value λ_2]

$$\therefore S w_i = \lambda_i w_i$$

$\therefore$ The minimum reconstruction error (distortion measure) is given by

$$J = \sum_{I=M+1}^{D} \lambda_i$$