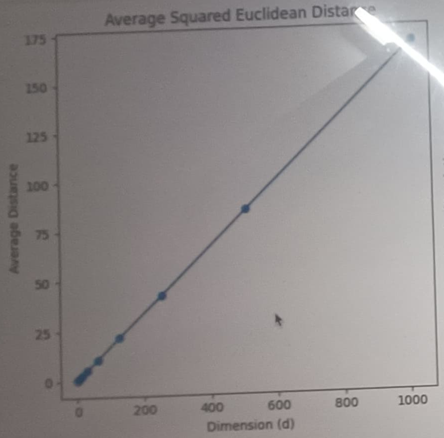
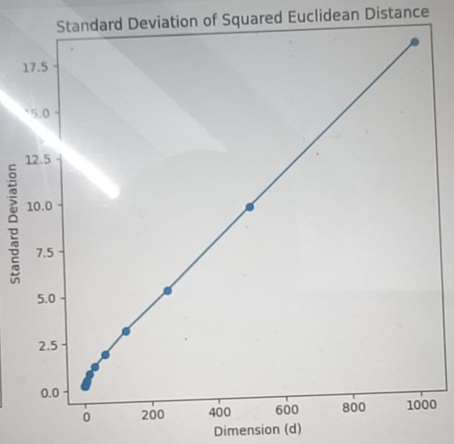
File Edit View Run Kernel Settings Help





1(b) $E(x) = \frac{1}{2}$

$E(Y) = \frac{1}{2}$

$E((X-Y)^2) = E(X^2+Y^2-2XY)$

$= E(X^2)+E(Y^2)-2E(X\cdot Y)$

$X, Y \to$ Independent

$\therefore E(XY) = E(X)\cdot E(Y)$

$\therefore E((X-Y)^2) = \frac{1}{3}+\frac{1}{3}-2\times\frac{1}{2}\times\frac{1}{2}$

$= \frac{1}{6}$

Interval $\to [0,1]$

Variance $(\sigma^2) = E(X^2) - (E(X))^2$

$= \frac{1}{3} - \frac{1}{4}$

$= \frac{1}{12}$

$Var(X) = Var(Y) = \frac{1}{12}$

$\therefore Var((X-Y)^2)$

$= E((X-Y)^4) + -E((X-Y))$

$= E(X^4+Y^4+6X^2Y^2-4XY^3-4X^3Y) - (E)^2$

$= \frac{1}{5} + \frac{1}{5} + \frac{6}{9} - \frac{1}{2} - \frac{1}{2} - \frac{1}{36}$

$= \frac{7}{180}$

1(c) $E(S) = E(Z_1 + Z_2 + Z_3 + \dots + Z_d)$

$$= \frac{1}{6} + \frac{1}{6} + \dots + \frac{1}{6}$$

$$= \frac{d}{6}$$

$Var(S) = E(S^2) - (E(S))^2$

$E(S^2) = d(E(Z_i)^2) + (d^2+d)\left(\frac{1}{36}\right)$

$$= \frac{d}{15} + \frac{d^2}{36} + \frac{d}{36}$$

$$= \frac{d^2}{36} + \frac{7d}{180}$$

$$(E(S))^2 = \frac{d^2}{36}$$

$Var(S) = \frac{7d}{180}$

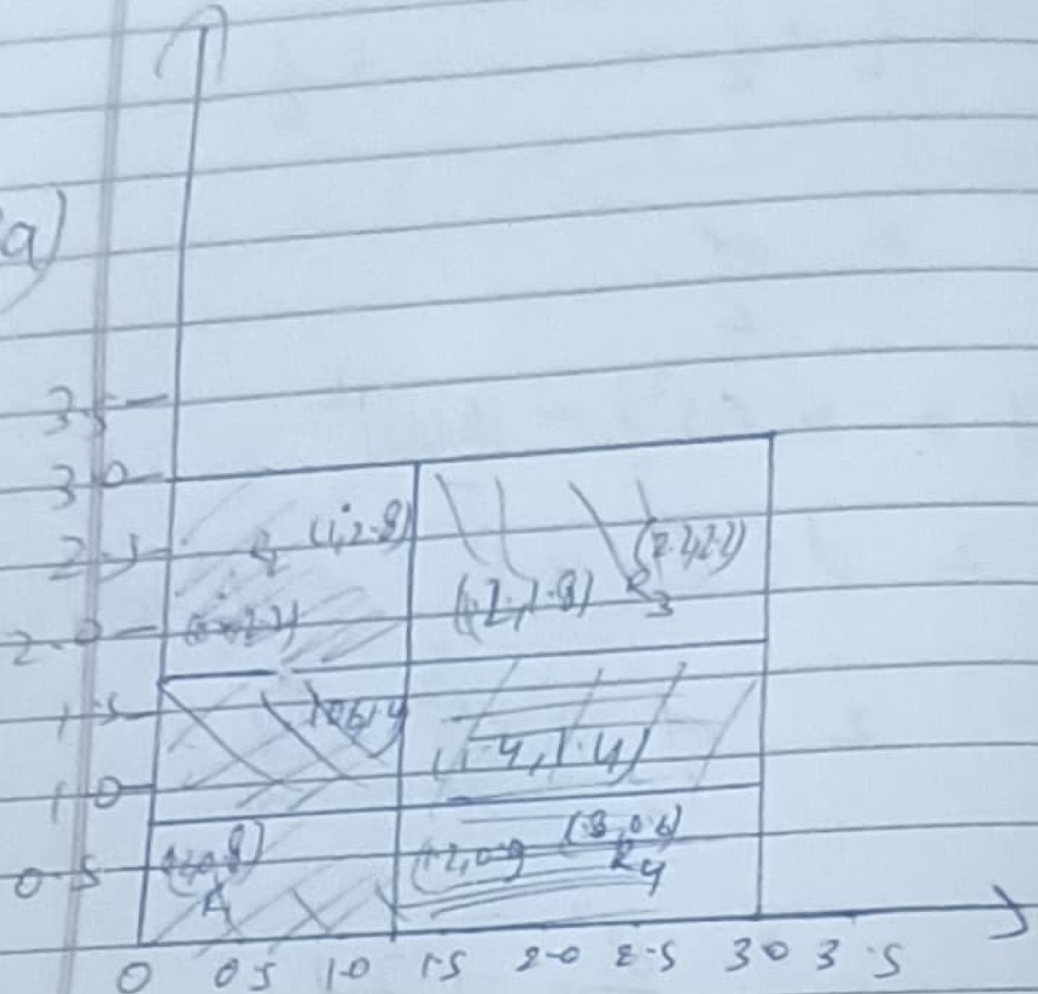
1d) Low dimensional. Easy to figure out the btw pts

High dimensional. hard to figure out rln. btw pts.

Since avg distance btw pts in high D sp. is more or less same, there is no such thing as 'near' or 'far' btw pts.

If $a \uparrow$, RHS $\downarrow$ & hence, probability of $z - E(z) \downarrow$

3a)

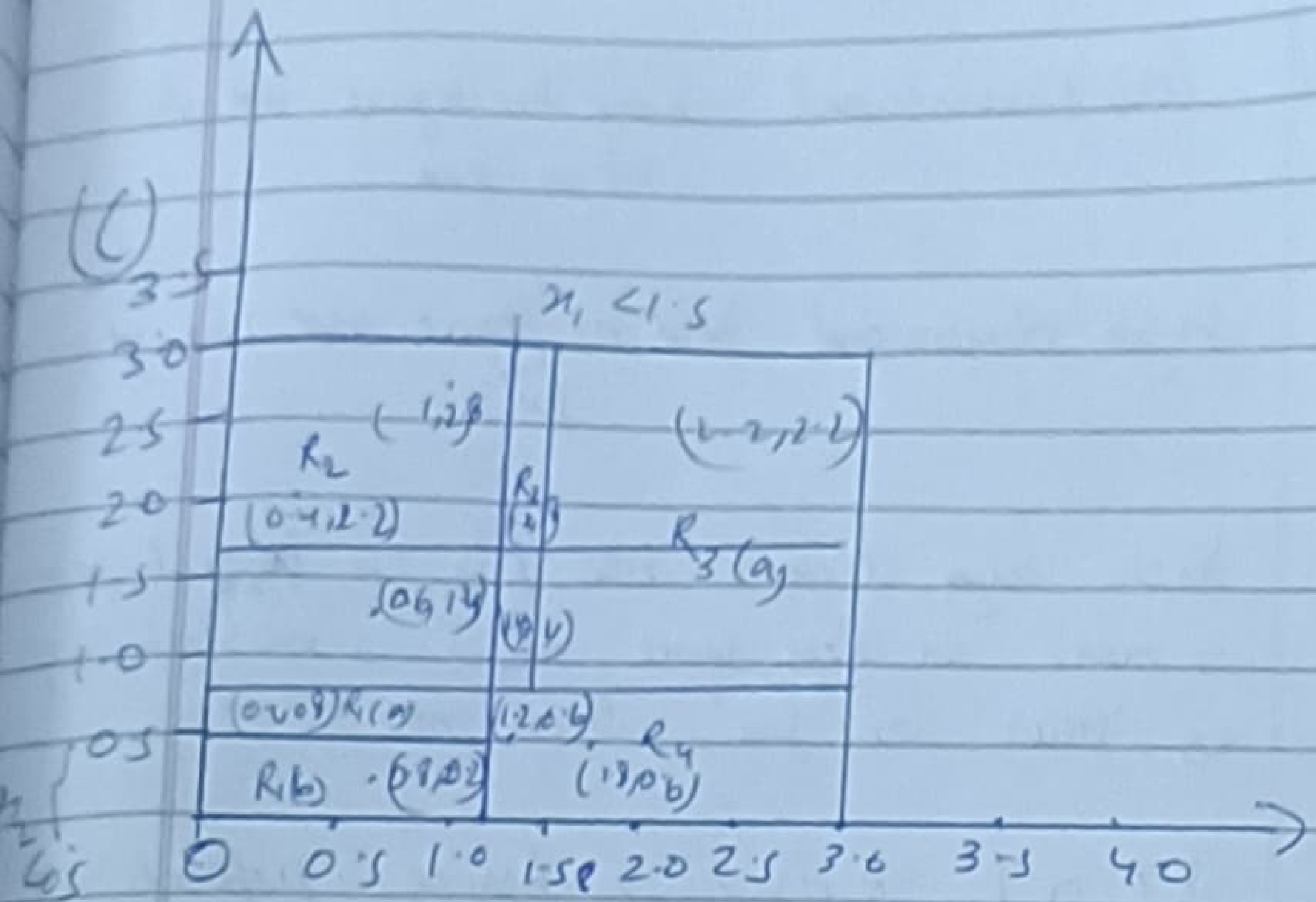


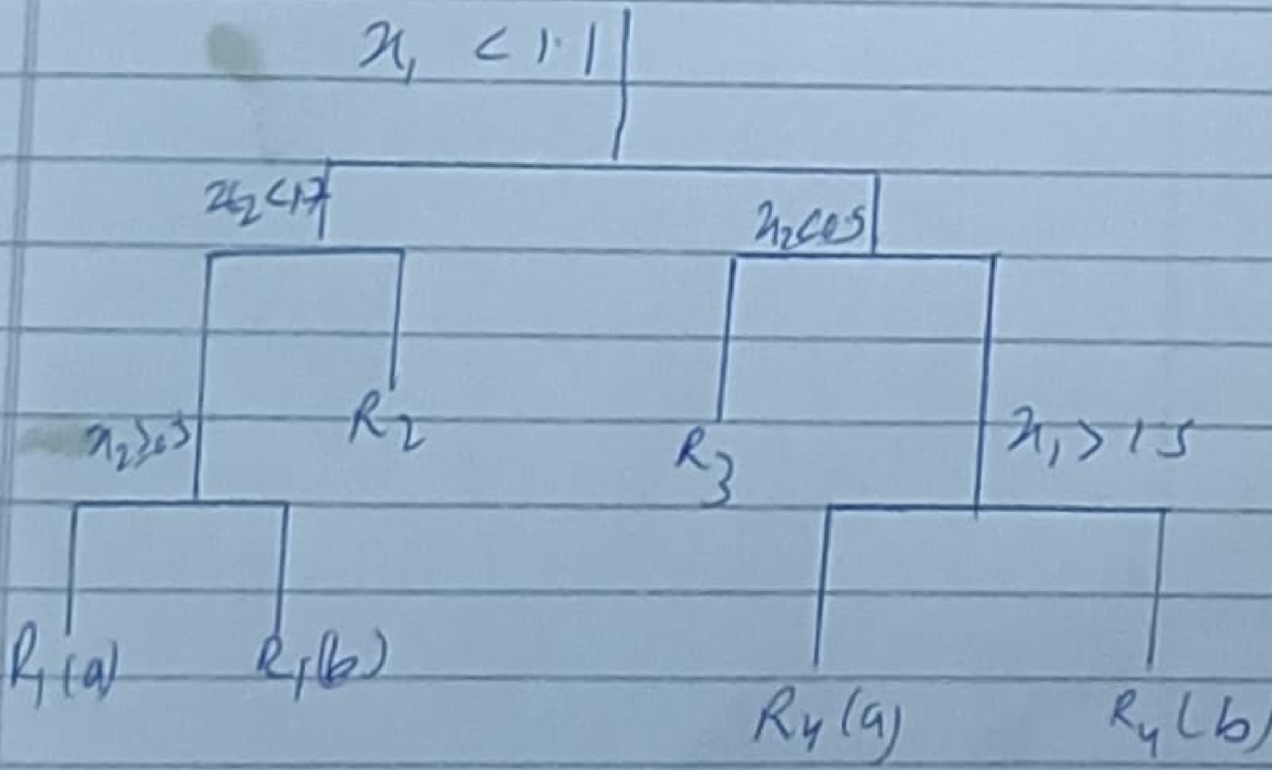
b) $\text{mean} = \dfrac{0.1 + 0.6 - 0.1}{3}$

$= \dfrac{1}{15}$

≈ 0.0066

(c)


$x_1 < 1.5$
R_2
R_3
R_4
$R_1(a)$
$R_1(b)$
(2.2, 2.2)
(0.4, 2.2)
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


$x_1 < 1.1$
$x_2 < 1.7$
$x_2 < 0.5$
$x_2 \geq 0.5$
R_2
R_3
$x_1 > 1.5$
$R_1(a)$
$R_1(b)$
$R_4(a)$
$R_4(b)$

5a) Given,

$$\hat{\theta} = \arg\max \ln p(y/X;\theta)$$

$$(X^T X)\hat{\theta} = X^T y.$$

$$P(y/X;\theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(\frac{-1}{2\sigma_\varepsilon^2}(y^{(i)} - x^{(i)T}\theta\right.$$

Applying 'ln' on LHS,

$$\ln(p(y/X;\theta)) = -\frac{N}{2}\ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^{i}$$

$$\arg\max \ln(P(y/X;\theta)) = \arg\max_{(\theta)}\left(-\sum_i (y^i - x^{iT}\theta)\right)$$

$$= \arg\min_{(\theta)}\left(\sum_i (y^i - x^{iT}\theta)^2\right)$$

$$\hat{\theta} = \arg\min_{(\theta)} (Y - X^T\theta)^T(Y - X^T\theta)$$

$$= \arg\min_{\theta}(Y^T Y - \theta^T X Y - Y^T X^T \theta + \theta^T X^T X \theta$$

We know that $\frac{d\hat{\theta}}{d\theta} = 0$

$\therefore -2X^T Y + 2X^T X\theta = 0$

5 b) $|X^T X| = 0$

$\downarrow$

Multiple solution

$|X^T X| \neq 0 \rightarrow$ Solution would be uniq

(7a) $h(x) = \frac{e^x}{1+e^x}$ (logistic function)

Taking ln on both sides

$$\ln(h(x)) = x - \ln(1+e^x)$$

$$\frac{d}{dx}\ln(h(x)) = 1 - \frac{d}{dx}\left(\ln(1+\ldots\right.$$

$$\frac{1}{h(x)}\frac{dh(x)}{dx} = 1 - \frac{1}{(1+e^x)}(e^x)$$

$$\left(\frac{d(h(x))}{dx}\right) = h(x)\left[1 - h(x)\right]$$

7(b) Class labels :- [0, 1]

Aim: To find out the negative log likelihood function.

$$P(Y=y; X=x) = [\phi(\underline{x}; \underline{\theta})]^{y} [1-\phi(\underline{x}; \underline{\theta})]^{1-y}$$

If $y=0$

$$P(Y=0; X=x) = (1-\phi(\underline{x}; \underline{\theta}))$$

If $y=1$,

$$P(Y=1; X=x) = \phi(\underline{x}; \underline{\theta})$$

$$P(\bar{y} \mid \underline{x}; \underline{\theta}) = \prod_{i=1}^{n} (\phi(\underline{x}; \underline{\theta}))^{y(i)} \cdot (1-\phi(\underline{x}; \underline{\theta}))^{y(i)}$$

Finally in order to obtain our aim, we take the "ln" & multiply by 1

$$L(\theta) = -\ln(P(\bar{y} \mid \underline{x}; \underline{\theta}))$$

$$= \sum_{i=1}^{n} (-y^{(i)} \ln(\phi(\underline{x}; \underline{\theta})) + (-1+y^{(i)}) + \ln(1-\phi(\underline{x}; \underline{\theta}))$$

Date / / Page no.

7c) We now look forward towards derivating the previous part.

$$\therefore \frac{\partial L(\theta)}{\partial \theta_i} = \frac{-\partial}{\partial \theta_i} y \log(Q(x;\theta)) - \frac{\partial}{\partial \theta_j}((1-y)\log(1-Q(x,\theta)$$

Substituting $Q(x;\theta)$ with $\sigma(\underline{\theta}^T \underline{x})$

$$\sigma(\underline{\theta}^T \underline{x}) = \frac{e^{\underline{\theta}^T \underline{x}}}{1+e^{\underline{\theta}^T \underline{x}}}$$

$$\therefore \frac{\partial L(\theta)}{\partial \theta_i} = \left(\frac{-\partial}{\partial \theta_i} \sigma(\underline{\theta}^T \underline{x})\right)\left(\frac{y}{\sigma(\underline{\theta}^T \underline{x})} + \frac{y-1}{1-\sigma(\underline{\theta}^T \underline{x})}\right)$$

$$= -(\sigma(\underline{\theta}^T \underline{x})) \cdot (1-\sigma(\underline{\theta}^T \underline{x}))x_i$$

$$\left(\frac{y}{\sigma(\underline{\theta}^T \underline{x})} + \frac{y-1}{1-\sigma(\underline{\theta}^T \underline{x})}\right)$$

$$\frac{\partial L(\theta)}{\partial \theta_i} = (\sigma(\underline{\theta}^T \underline{x}) - y)x_i$$

d) Hessian Matrix $= \dfrac{\partial^2 L(\theta)}{\partial \theta_i \, \partial \theta_j}$

$\Rightarrow \left(\dfrac{\partial}{\partial \theta_n}\left(\sigma(\underline{\theta}^T \underline{x})\right) \cdot \left(1-\sigma(\underline{\theta}^T \underline{x}\right)(x_i)(x_j\right.$