EECS 545

HW-1

Name:- Sai Satya Charan. Malladi

UMID: 44166852

unique name: charanm

Honor Code

Q1) (a) True

(b) True

(c) False

Q2) PSD Matrices

a) Let, $M = A^TA$, and $A \in \mathbb{R}^{m \times n}$

Consider $u \in \mathbb{R}^{n \times 1}$ and u be any arbitrary vector

m, n are some positive integers.

$u^TMu \Rightarrow u^TA^TAu$

$\Rightarrow (Au)^T(Au)$ → inner product of $\langle Au, Au\rangle$ → $(AB)^T = B^TA^T$

$\Rightarrow \|Au\|^2$

$u^TMu = \|Au\|^2 \geq 0$

Hence matrix M is PSD, where $M = A^TA$

b Let, $\mathbb{E}[(x-\mathbb{E}[x])(x-\mathbb{E}[x])^T] = M$

Given, $X \in \mathbb{R}^{n \times 1}$ is a random coloumn vector

Let, $u \in \mathbb{R}^{n \times 1}$ be a arbitrary Column vector

$u^TMu = u^T\mathbb{E}[(x-\mathbb{E}[x])(x-\mathbb{E}[x])^T]u$ — (#)

From the property,

$$a\mathbb{E}(Y) = \mathbb{E}(aY)$$

we can re-write the equation (#) as

$$u^T \mathbb{E}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right]u = \mathbb{E}\left[u^T(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T u\right]$$

$$* \quad (AB)^T = B^T A^T \quad \leftarrow \quad = \mathbb{E}\left[\left((X-\mathbb{E}[X])^T u\right)^T \left((X-\mathbb{E}[X])^T u\right)\right]$$

$$= \mathbb{E}\left[\| (X-\mathbb{E}[X])^T u \|^2\right]$$

$$\therefore \; u^T M u = \mathbb{E}\left[\underbrace{\| (X-\mathbb{E}[X])^T u \|^2}_{\geq 0}\right]$$

$$= \mathbb{E}\left[\text{positive quantity}\right]$$

$$\Rightarrow \; u^T M u \geq 0$$

Hence $M = \mathbb{E}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right]$ is PSD Matrix

c) Let M be any Gram Matrix, $M \in \mathbb{R}^{d\times d}$

$$M = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \cdots & \langle x_d, x_1 \rangle \\ \langle x_1, x_2 \rangle & \langle x_2, x_2 \rangle & & \vdots \\ \langle x_1, x_3 \rangle & \vdots & \ddots & \vdots \\ \vdots & & & \\ \langle x_1, x_d \rangle & \cdots & \cdots & \langle x_d, x_d \rangle \end{bmatrix}$$

and let x_i's $\in \mathbb{R}^{n\times 1}$, where n is any random integer

We can split the Matrix M as follows.

$$M = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}^T \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix} \rightarrow (*)$$

where $\begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$ is of dimension $(n \times d)$

The equation $(*)$ resembles $M = A^TA$ where $A = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$

We have proved in 2 a) that matrices of the form A^TA are PSD

and Hence, $x^TMx \geq 0$, and M is PSD, where M is a gram Matrix.

3) Probability

Given discrete random variables X and Y with pmf $p(x,y)$

a) $\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}[X|Y]\right]$

$\mathbb{E}[X] = \sum_x x\, p_X(x)$ } where $p_X(x)$ is the marginal pmf of X

From the property,

$$\sum_y p_{X|Y}(x|y)\, p_Y(y) = p_X(x)$$

$p_X(x)$ can be substituted,

$$\mathbb{E}[X] = \sum_x x \sum_y p_{X|Y}(x|y)\, p_Y(y)$$

I have brought in $\sum_x x$ into the $\sum_y (\ )$

$$= \sum_y \sum_x \underbrace{x\, p_{X|Y}(x|y)\, p_Y(y)}$$

$$= \sum_y \left[\sum_x x\, p_{X|Y}(x|y)\right] p_Y(y) \rightarrow \sum_y \mathbb{E}[X|Y]\, p_Y(y)$$

$$= \underbrace{\sum_y \mathbb{E}[X|Y]\, P_Y(y)}$$ $\}$ property $\to$ $\mathbb{E}(g(X)) = \sum g(x)\, P_X(x)$

$g(x) \to$ discrete RV

$$\boxed{\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}[X|Y]\right]}$$

Hence proved

b) $\mathbb{E}\left[I[x \in C]\right] = P(X \in C)$

$$I = \begin{cases} 1 & x \in C \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}\left[I[x \in C]\right] = \sum_x I \cdot P_X(x)$$

Property
$\mathbb{E}[g(x)] = \sum g(x)\, P_X(x)$
if $g(x)$ is discrete RV

$$\Rightarrow \sum_{x \in C} I \cdot P_X(x) + \sum_{x \notin C} I \cdot \cancel{P_X(x)}$$

0

$\{x \notin C, I = 0\}$

$$\Rightarrow \sum_{x \in C} 1 \cdot P_X(x)$$

$$\mathbb{E}\left[I(x \in C)\right] = \underline{P(x \in C)}$$

Hence proved

Property.
$$\left[\begin{array}{l} P_{XY}(x,y) = P_X(x)\, P_Y(y) \\ \text{if } X, Y \text{ are independent} \end{array}\right]$$

c) $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

$\mathbb{E}[XY] = \sum_x \sum_y xy\, P_{XY}(x,y)$

$\mathbb{E}[XY] = \sum_x \sum_y xy\, P_X(x)\, P_Y(y)$ — ★

when X and Y are independent, we can de-couple the equation ★

$$\mathbb{E}[XY] = \underbrace{\sum_x x\, p_X(x)}\ \underbrace{\sum_y y\, p_Y(y)}$$

$$= \mathbb{E}[X]\ \mathbb{E}[Y]$$

Hence proved.

d) X, Y takes values in $\{0,1\}$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$,

then X, and Y are independent

$$\mathbb{E}[XY] = \sum_x \sum_y xy\, p_{XY}(x,y)$$

$$= p_{1,1}(1,1)$$

$$\mathbb{E}[X]\mathbb{E}[Y] = \left[\sum_x x\, p_X(x)\right]\left[\sum_y y\, p_Y(y)\right]$$

$$= \left[0 + 1 \cdot p_1(1)\right]\left[0 + p_1(1)\right]$$

$$= p_X(1)\ p_Y(1)$$

From the given condition $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

we get $\Rightarrow$ $\boxed{p_{X,Y}(1,1) = p_X(1)\ p_Y(1)}$ — (#)

From the law of total probability

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

$$p_X(x) = p_{XY}(x,0) + p_{XY}(x,1) \quad — ①$$

Similarly

$$p_Y(y) = p_{XY}(0,y) + p_{XY}(1,y) \quad — ②$$

Also, from the law of propbability

$$\sum_x P_x(x) = 1$$

$$\Rightarrow P_x(1) + P_x(0) = 1 \quad - (3)$$

Similarly

$$\sum_y P_y(y) = 1$$

$$\Rightarrow P_y(1) + P_y(0) = 1 \quad - (4)$$

From eqn (#) and (3)

$$P_{xy}(1,1) = P_x(1)\, P_y(1)$$

$$= [1 - P_x(0)]\, P_y(1)$$

From eqn (2)

$$P_y(1) - P_{xy}(1,1) = P_{xy}(0,1)$$

$$P_{xy}(1,1) \Rightarrow P_y(1) - P_x(0)\, P_y(1)$$

$$P_x(0)\, P_y(1) = \underbrace{P_y(1) - P_{xy}(1,1)}$$

$$\boxed{P_x(0)\, P_y(1) = P_{xy}(0,1)} \quad - (\#2)$$

From eqn (#) and (4)

$$P_{xy}(1,1) = P_x(1)\, P_y(1)$$

$$= P_x(1)\, [1 - P_y(0)]$$

$$P_{xy}(1,1) \Rightarrow P_x(1) - P_x(1)\, P_y(0)$$

From eqn (1)

$$P_x(1) - P_{xy}(1,1) = P_{xy}(1,0)$$

$$\underbrace{P_x(1) - P_{xy}(1,1)} = P_x(1)\, P_y(0)$$

$$\boxed{P_{xy}(1,0) = P_x(1)\, P_y(0)} \quad - (\#3)$$

From eqn (#3) and (3)

$P_{xy}(1,0) = P_x(1)\, P_y(0)$

$= [1 - P_x(0)]\, P_y(0)$

$P_{xy}(1,0) = P_y(0) - P_x(0)\, P_y(0)$

From eqn (2)

$P_y(0) - P_{xy}(1,0) = P_{xy}(0,0)$

$\underbrace{-P_{xy}(1,0) + P_y(0)} = P_x(0)\, P_y(0)$

$$\boxed{P_{xy}(0,0) = P_x(0)\, P_y(0)}$$

We have proved that

$P_{xy}(x,y) = P_x(x)\, P_y(y) \quad \forall x, y \in \{0,1\}$

Hence the random variable X, Y are independent

4) Guassian Level Sets.

$\Sigma = U \Lambda U^T \qquad U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \qquad \theta = -\pi/4$

$x \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

a) Boundary of $C : \{x \mid (x-\mu)^T \Sigma^{-1} (x-\mu) \le r^2\}$ $\quad r = 1$

$(x-\mu)^T \Sigma^{-1} (x-\mu) = 1$ is the standard form of the ellipse

$$\boxed{\text{Center} = \mu = \begin{bmatrix} -2 \\ 3 \end{bmatrix}}$$

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = (x-\mu)^T [U^T \Lambda^{-1} U] (x-\mu)$$

$$= [U(x-\mu)]^T \Lambda^{-1} [U(x-\mu)]$$

Since Λ is diagonal $(U_{i,:}) \rightarrow$ row of U

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \sum \frac{(U_{i,:}(x-\mu))^2}{\lambda_i}$$

} equation of ellipse with axis given by $U_{i,:}(x-\mu)$

Hence,

length of semi-major axis $= \sqrt{\lambda_{max}} = \sqrt{5}$

$$\boxed{\begin{array}{l} \text{Length of major axis} = 2\sqrt{5} \\ \text{Length of minor axis} = 2 \times \sqrt{\lambda_{min}} = 2\sqrt{1} = 2 \end{array}}$$

~~Equation $U_{i,:}(x-\mu) = \left[\frac{1}{\sqrt{2}} \; \frac{1}{\sqrt{2}}\right](x-\mu)$~~

~~eigen vector corresponding to~~

from $U^T \Lambda^{-1} U$

$$\Lambda^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1 \end{bmatrix} \quad U^T = \begin{bmatrix} 1/\sqrt{2} & +1/\sqrt{2} \\ -\frac{1}{\sqrt{2}} & 1/\sqrt{2} \end{bmatrix}$$

eigen vector corresponding to $\frac{1}{5} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \rightarrow$ major axis

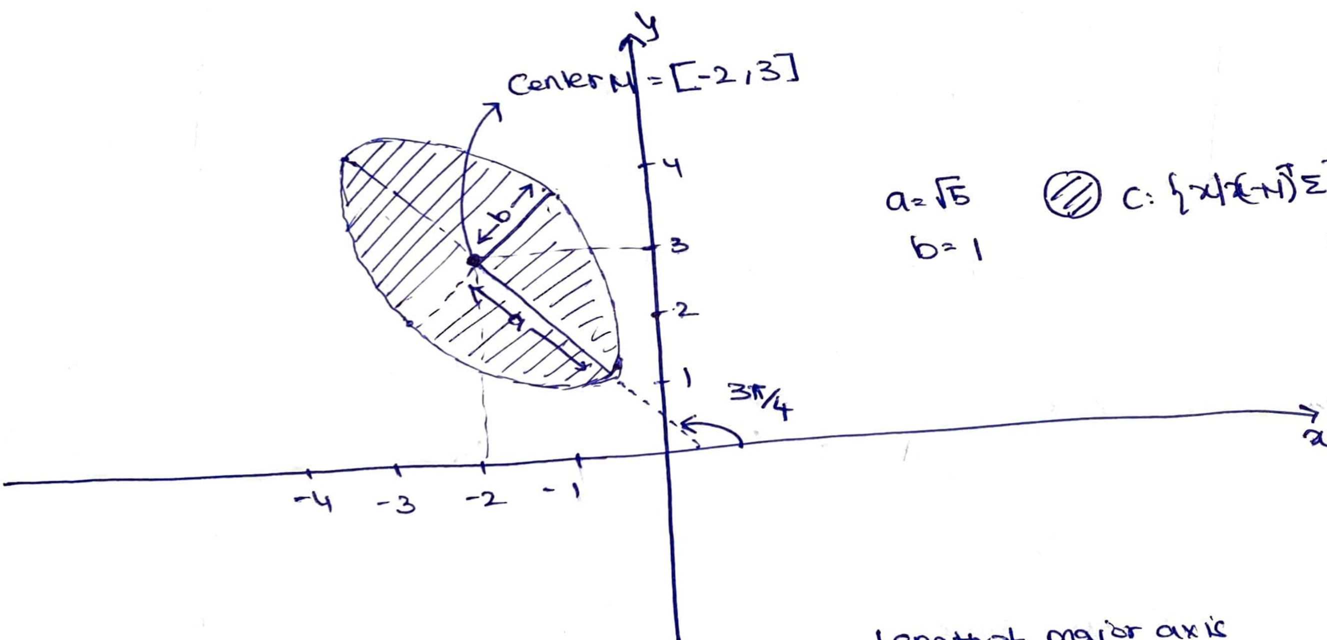
$1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \rightarrow$ minor axis

Angle made with x-axis $= 3\pi/4$ (anti-clockwise sense) end points on major axis

$$M = \pm\sqrt{5}\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -3.581 \\ 4.581 \end{bmatrix}, \begin{bmatrix} -0.419 \\ 1.419 \end{bmatrix}$$

end points of minor axis

$$M = \pm\sqrt{1}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -2.707 \\ 2.293 \end{bmatrix}, \begin{bmatrix} -1.293 \\ 3.707 \end{bmatrix}$$



$a = \sqrt{5}$

$b = 1$

$C: \{x \mid (x-M)^T \Sigma^{-1} (x-M) \le 1\}$

length of major axis

$= 2a$

$= \underline{\underline{2\sqrt{5}}}$

length of minor axis $= 2b$

$= \underline{\underline{2}}$

b) $(x-\mu)^T \Sigma^{-1} (x-\mu) = (x-\mu)^T [U \Lambda^{-1} U^T] (x-\mu)$

$$\Rightarrow \left[(x-\mu)^T U \Lambda^{-1/2}\right]\left[\Lambda^{-1/2} U^T (x-\mu)\right] \quad \text{— (\#)}$$

Since we know Λ^{-1} is pd and diagonal

$$\Lambda^{-1} = \Lambda^{-1/2} \Lambda^{-1/2}$$

also $(\Lambda^{-1/2})^T = \Lambda^{-1/2}$ due to diagonal nature of matrix.

Equation (#) can be re written as.

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \left[(U \Lambda^{-1/2})^T (x-\mu)\right]^T \left[\Lambda^{-1/2} U^T (x-\mu)\right]$$

$$= \underbrace{\left[\Lambda^{-1/2} U^T (x-\mu)\right]}_{Z}^T \left[\Lambda^{-1/2} U^T (x-\mu)\right]$$

$x \sim (\mu, \Sigma)$

From the property, If $X \sim N(\mu, \Sigma)$
$AX \sim N(A\mu, A\Sigma A^T)$

$(x-\mu) \sim N(0, \Sigma)$

$$\Lambda^{-1/2} U^T (x-\mu) \sim N\left(0, \Lambda^{-1/2} U^T \Sigma (\Lambda^{-1/2} U^T)^T\right)$$

$$\Lambda^{-1/2} U^T (x-\mu) \sim N\left(0, \Lambda^{-1/2} U^T U \Sigma U^T U \Lambda^{-1/2}\right)$$

$$\sim N\left(0, \Lambda^{-1/2} \Lambda \Lambda^{-1/2}\right)$$

$\left[U^T U = I \right.$ as U is orthogonal $\left.\right]$

$$\underbrace{\Lambda^{-1/2} U^T (x-\mu)}_{Z} \sim N(0, I)$$

$(x-\mu)^T \Sigma^{-1} (x-\mu) = z^T z$, where $z \sim N(0, I)$

$= \underbrace{z_1^2 + z_2^2}\sim \chi_2^2$ $\quad z = \Lambda^{-1/2} U^T (x-\mu)$

↑ Sum of 2 standard Normal = Chi^2 distribution

So, $P_r(x \in C) = P\left((x-\mu)^T \Sigma^{-1} (x-\mu) \leq r^2\right)$ $\quad r=1$

$= P(z^T z \leq 1)$

$= P(z_1^2 + z_2^2 \leq 1)$

$= P(\chi_2^2 \leq 1)$

CDF of $\chi_2^2 = \frac{1}{\Gamma(k/2)} \gamma\left(k/2, x/2\right)$

Here $k=2 \quad x=1$

$P(\chi_2^2 \leq 1) = \frac{1}{\Gamma(1)} \gamma\left(1, \frac{1}{2}\right)$

$= 0.3935$

$P(x \in C) = 0.3935$

5) a) Let $F(x)$ and $g(x)$ be two convex functions.

and $H(x) = F(x) + g(x)$ $\quad x \in \mathbb{R}^{n \times 1}$

Since $F(\alpha w_1 + (1-\alpha) w_2) \leq \alpha F(w_1) + (1-\alpha) F(w_2)$ $\forall \alpha \in [0,1]$

$g(\alpha w_1 + (1-\alpha) w_2) \leq \alpha g(w_1) + (1-\alpha) g(w_2)$ $\quad$ & $\forall w_i \in$ domain.

for any random $w_1, w_2 \in$ Domain.

$$H(\alpha w_1 + (1-\alpha) w_2) = F(\alpha w_1 + (1-\alpha) w_2) + g(\alpha w_1 + (1-\alpha) w_2)$$

$$\leq \alpha F(w_1) + (1-\alpha) F(w_2) + \alpha g(w_1) + (1-\alpha) g(w_2)$$

$$\leq \alpha [F(w_1) + g(w_1)] + (1-\alpha) [F(w_2) + g(w_2)]$$

$$H(\alpha w_1 + (1-\alpha) w_2) \leq \alpha H(w_1) + (1-\alpha) H(w_2)$$

Hence proved that $H(X)$ is also convex as it satisfies the convexity criterion.

b) Let $F(x)$, $g(x)$ be a convex functions

$\Rightarrow \nabla^2 F(w)$ and $\nabla^2 g(w)$ are PSD

$$\Rightarrow \left.\begin{array}{l} \nabla^2 F(w) \geq 0 \\ \nabla^2 g(w) \geq 0 \end{array}\right\} \; w \in \mathbb{R}^d$$

Let $H(x) = F(x) + g(x)$

$$\nabla^2 H(w) = \nabla^2 F(w) + \nabla^2 g(w) \geq 0$$

Since $\nabla^2 H(w) \succeq 0 \quad \forall w \in \mathbb{R}^d$

From the property that a function is convex iff $\nabla^2(\ ) \succeq 0$

$H(x)$ which is the sum of two convex functions is also

Convex.

c) $f(x) = \frac{1}{2} x^T A x + b^T x + C$, $A \in \mathbb{R}^{d \times d}$, and A = symmetric.

$\Rightarrow A^T = A$

~~$\nabla f(x) = \frac{1}{2} A x + \frac{1}{2} x^T A + b$~~

$$\frac{df(x,y)}{dx} = \frac{\partial f(x,y)}{\partial x} + \frac{\partial(y^T(x))}{dx} \frac{\partial(f(x,y))}{\partial y}$$

$$\nabla f(x) = \frac{1}{2} A x + \frac{1}{2} A^T x + b$$

We know that $A^T = A$ (as A is symmetric)

$$\nabla f(x) = Ax + b$$

$$\nabla^2 f(x) = A$$

f is convex, when A is PSD $\Rightarrow x^T A x \geq 0 \ \forall x \in \mathbb{R}^{d \times 1}$

f is strictly convex, when A is PD $\Rightarrow x^T A x > 0 \ \forall x \in \mathbb{R}^{d \times 1}$