Q1. (1) $AB = BA$ (False)

Counterexample:

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$AB = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \quad BA = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$$

$\Rightarrow AB \neq BA$.

(2) $tr(AB) = tr(BA)$ (True)

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

$$(BA)_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj}$$

$$trace(AB) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} b_{ki} \quad (i=j)$$

$$trace(BA) = \sum_{i=1}^{n} \sum_{k=1}^{n} b_{ik} a_{ki} \quad (i=j)$$

Hence $tr(AB) = tr(BA)$

(3) $(AB)^T = A^T B^T$ (False)

Counterexample: (Taking the same matrix A & B in (1) as example).

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad AB = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \quad (AB)^T = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad B^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad A^T B^T = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \Rightarrow A^T B^T \neq (AB)^T$$

(4) $\det(AB) = \det(A)\det(B)$ (True).

① A is invertible

Then A can be expressed as the product of series of elementary matrices.

$$A = E_1 E_2 \cdots E_n$$

$$AB = E_1 E_2 \cdots E_n B$$

By mathematical induction:

$$\begin{aligned}\det(AB) &= \det(E_n E_{n-1} \cdots E_2 E_1 B) \\ &= \det(E_n)\det(E_{n-1} \cdots E_1 B) \\ &= \det(E_n)\det(E_{n-1}) \cdots \det(E_1)\cancel{B}\det(B) \\ &= \det(E_n E_{n-1} \cdots E_1)\det(B) \\ &= \det(A)\det(B)\end{aligned}$$

② If A is non-invertible, then AB is also non-invertible,

$\det(A) = 0$, $\det(AB) = 0$

$$\Rightarrow \det(AB) = \det(A)\det(B) = 0.$$

Q2: $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 1 & 5 & 5 \end{bmatrix}$ $\Rightarrow$ Reduced Row echelon form:

$$\begin{bmatrix} 1 & 0 & \frac{5}{3} \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 \end{bmatrix}$$

a) Rank of A:

$\text{rank}(A) = 2$

b) minimal basis for row span is $\left\{ \begin{bmatrix} 1 \\ 0 \\ \frac{5}{3} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \frac{2}{3} \end{bmatrix} \right\}$.

Q3: (1) $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix}$

$$(A-\lambda I) = \begin{bmatrix} 1-\lambda & 2 & 3 \\ 3 & 2-\lambda & 1 \\ 2 & 1 & 3-\lambda \end{bmatrix}$$

$$\det(A-\lambda I) = (1-\lambda)\begin{vmatrix} 2-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} - 2\begin{vmatrix} 3 & 1 \\ 2 & 3-\lambda \end{vmatrix} + 3\begin{vmatrix} 3 & 2-\lambda \\ 2 & 1 \end{vmatrix}$$

$$= (1-\lambda)\left[(2-\lambda)(3-\lambda)-1\right] - 2\left[(3-\lambda)3-2\right] + 3\left[3-2(2-\lambda)\right]$$

$$= (1-\lambda)(6-2\lambda-3\lambda+\lambda^2-1) - 2(9-3\lambda-2) + 9-6(2-\lambda)$$

$$= (1-\lambda)(6-5\lambda+\lambda^2-1) - 2(7-3\lambda) + 9-12+6\lambda$$

$$= (1-\lambda)(\lambda^2-5\lambda+5) - 14+6\lambda+6\lambda-3$$

$$= \lambda^2-5\lambda+5-\lambda^3+5\lambda^2-5\lambda-17+12\lambda$$

$$= -\lambda^3+6\lambda^2+2\lambda-12$$

$\lambda_1 = 6 \quad \lambda_2 = -\sqrt{2} \quad \lambda_3 = \sqrt{2}$

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{-13-9\sqrt{2}}{7} \\ \frac{5+11\sqrt{2}}{7} \\ 1 \end{bmatrix} \quad v_3 = \begin{bmatrix} \frac{-13+9\sqrt{2}}{7} \\ \frac{5-11\sqrt{2}}{7} \\ 1 \end{bmatrix}$$

(2) $\lambda_1 = 6 \quad \lambda_2 = -\sqrt{2} \quad \lambda_3 = \sqrt{2}$

We can see that $\lambda_2 = -\sqrt{2} < 0$

Hence A is not posite semi-definite.

(3) $$\mu A + \gamma I = \begin{bmatrix} \mu+\gamma & 2\mu+\gamma & 3\mu+\gamma \\ 3\mu+\gamma & 2\mu+\gamma & \mu+\gamma \\ 2\mu+\gamma & \mu+\gamma & 3\mu+\gamma \end{bmatrix}.$$

$$(\mu A + \gamma I - I\lambda) = \begin{bmatrix} \mu+\gamma-\lambda & \cdots & \cdots \\ \cdots & 2\mu+\gamma-\lambda & \cdots \\ \cdots & \cdots & 3\mu+\gamma-\lambda \end{bmatrix}.$$

$\Rightarrow$ eigenvalues of $2A$ $\Rightarrow$ $2\cdot$ eigenvalue of A.

$\Rightarrow$ $(\mu A + \gamma I)\vec{x} = \nu\vec{x}$, where ν is the eigenvalue of $(\mu A + \gamma I)$

$\mu A\vec{x} + \gamma\vec{x} = \nu\vec{x}$

We know that $A\vec{x} = \lambda\vec{x}$, where λ is the eigenvalue of A.

$\mu\lambda\vec{x} + \gamma\vec{x} = \nu\vec{x}$

$(\mu\lambda+\gamma)\vec{x} = \nu\vec{x}$

$\Rightarrow \nu = \mu\lambda + \gamma$.

Hence $\nu_1 = \mu\cdot\lambda_1 + \gamma$; $\nu_2 = \mu\lambda_2 + \gamma$, $\nu_3 = \mu\lambda_3 + \gamma$

$= \mu + \gamma$ $(\lambda_1 = 1)$; $= -\sqrt{2}\mu + \gamma$ $(\lambda_2 = -\sqrt{2})$; $= \sqrt{2}\mu + \gamma$ $(\lambda_3 = \sqrt{2})$.

(4) eigenvalues of A^2: $A^2 = A^T A$

$A^2 \vec{x} = w\vec{x}$

$A^T A\vec{x} = w\vec{x}$

$A^T \cdot \lambda\vec{x} = w\vec{x}$

$\lambda A^T \vec{x} = w\vec{x}$

$\lambda \cdot \lambda \vec{x} = w\vec{x}$

$w = \lambda^2$.

Hence eigenvalues of A^2 are: $w_1 = \lambda_1^2 = 1$; $w_2 = \lambda_2^2 = 2$; $w_3 = \lambda_3^2 = 2$

Q4. $\vec{x} \in \mathbb{R}^{n\times 1}$ ☆ [Vector Calculus].

(1). $f(\vec{x}) = \vec{w}^T\vec{x}$; $\vec{w} \in \mathbb{R}^{n\times 1}$ $\Rightarrow$ $f(\vec{x}) = \underset{1\times n}{\vec{w}^T}\underset{n\times 1}{\vec{x}} = w_1x_1 + w_2x_2 + \cdots + w_nx_n$.

$\Rightarrow 1\times 1$

$\nabla_{\vec{x}} f(\vec{x}) = \frac{df}{d\vec{x}} = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \cdots \frac{\partial f(x)}{\partial x_n}\right]$;

$= [w_1 \quad w_2 \quad \cdots \quad w_n]$

$= w^T$.

(2) $f(\vec{x}) = \vec{x}^T\vec{x} \Rightarrow f(\vec{x}) = x_1^2 + x_2^2 + \cdots + x_n^2$.

$\nabla_{\vec{x}} f(\vec{x}) = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \cdots \frac{\partial f(x)}{\partial x_n}\right]$.

$= [2x_1 \quad 2x_2 \quad \cdots \quad 2x_n]$

$= 2\vec{x}^T$

(3) $f(\vec{x}) = (y - \vec{w}^T\vec{x})^2$; $\vec{w} \in \mathbb{R}^{n\times 1}$, $y \in \mathbb{R}$,

$\nabla_{\vec{x}} f(\vec{x}) = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \cdots \frac{\partial f(x)}{\partial x_n}\right]$.

$$f(\vec{x}) = (y-\vec{w}^T\vec{x})^2 = (y-\vec{w}^T\vec{x})^T(y-\vec{w}^T\vec{x})$$

$$= y^2 - y\vec{w}^T\vec{x} - (\vec{w}^T\vec{x})^T y + (\vec{w}^T\vec{x})^T(\vec{w}^T\vec{x})$$

$$= y^2 - yw^Tx - x^Twy + \cancel{wx^T}\, x^Tww^Tx$$

$$= y^2 - 2yw^Tx + x^Tww^Tx.$$

$$\nabla_x f(x) = \left[\frac{\partial f(x)}{\partial x_1} \; --- \; \frac{\partial f(x)}{\partial x_n}\right]^T$$

$$= -2yw^T + 2w^Tw^Tx$$

(4). $f(x) = \log(1+e^{-yw^Tx})$; $w \in \mathbb{R}^{n\times 1}$

$$\nabla_x f(x) = \frac{-yw^Te^{-yw^Tx}}{1+e^{-yw^Tx}}$$

(5) $f(x) = x^TAx$; $A \in \mathbb{R}^{n\times n}$ $x \in \mathbb{R}^{n\times 1}$ $\Rightarrow f(x) = x^TAx = \sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j$

$$\nabla_x f(x) = \left[\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; --- \frac{\partial f}{\partial x_n}\right]$$

$$= \left[\sum_{j=1}^n (a_{1j}+a_{j1})x_j \;\; \sum_{j=1}^n (a_{2j}+a_{j2})x_j \; --- \; \sum_{j=1}^n (a_{nj}+a_{jn})x_j\right]$$

$$= (A+A^T)x$$

If A is symmetric $\Rightarrow A^T = A$

$$\Rightarrow \nabla_x f(x) = 2Ax.$$

Q5: Hyperplane $\mathcal{H}: w^T x + b = 0$, $w \in \mathbb{R}^n$ & $b \in \mathbb{R}$.

(1) when $b = 0$. $\mathcal{H}$ passes through origin.

(2) $$\frac{|(x_0 - x)\cdot w|}{\|w\|} = \frac{|x_0 \cdot w - x \cdot w|}{\|w\|} = \frac{|x_0 \cdot w - (-b)|}{\|w\|} = \frac{|w^T x_0 + b|}{\|w\|}.$$

(We assume the coordinate of x_0 is $\vec{x_0}$).

Q6: [Vector Norms].

$x \in \mathbb{R}^n$. $\|x\|_p \to$ p-norm of x $\quad \|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$

① $\|x\|_2 \le \|x\|_1$ (True)

$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} \qquad \|x\|_2^2 = \sum_{i=1}^n |x_i|^2$

$\|x\|_1 = \sum_{i=1}^n |x_i| \qquad \|x\|_1^2 = \left(\sum_{i=1}^n |x_i|\right)^2$

$\Rightarrow \|x\|_2^2 = \sum_{i=1}^n |x_i|^2 \le \left(\sum_{i=1}^n |x_i|\right)^2 = \|x\|_1^2$

$\Rightarrow \|x\|_2 \le \|x\|_1$

② $\|x\|_1 \le \sqrt{n}\|x\|_2$ (True)

$\|x\|_1^2 = \left(\sum_{i=1}^n |x_i|\right)^2$

$(\sqrt{n}\|x\|_2)^2 = n\|x\|_2^2 = n\sum_{i=1}^n |x_i|^2$

Cauchy-Schwarz: $|\langle u, v\rangle|^2 \le \langle u, u\rangle \cdot \langle v, v\rangle$

$\left(\sum_{i=1}^n u_i v_i\right)^2 \le \left(\sum_{i=1}^n u_i^2\right)\left(\sum_{i=1}^n v_i^2\right)$

$\|x\|_1^2 = \left(\sum_{i=1}^n |x_i| \cdot 1\right)^2 \le \sum_{i=1}^n |x_i|^2 \cdot \sum_{i=1}^n 1^2 = n\sum_{i=1}^n |x_i|^2 = n \cdot \|x\|_2^2$

$\Rightarrow \|x\|_1 \le \sqrt{n}\|x\|_2$

③ $n\|X\|_\infty \le \|X\|_1$ (False).

$\|X\|_\infty = \max_{i=1\cdots n} |X_i|$

$\|X\|_1 = \sum_{i=1}^{n} |X_i| \underline{\le} n \cdot \max_{i=1\cdots n} |X_i| = n\|X\|_\infty$.

④ $\|X\|_0 \le \|X\|_1$ (False)

$\|X\|_0 \Rightarrow$ # of non-zero elements in vector.

$\|X\|_1^2 = (\sum_{i=1}^{n} |X_i|)^2 \le \sum_{i=1}^{n} |X_i|^2 \cdot \sum_{i=1}^{n} 1^2 = n \cdot \sum_{i=1}^{n} |X_i|^2$

Counterexample:

$\vec{x} = [0.5 \ \ 0.1 \ \ 0.2]$

$\|\vec{x}\|_0 = 3$.

$\|\vec{x}\|_1 = \sum_{i=1}^{3} |X_i| = 0.5 + 0.1 + 0.2 = 0.8$.

$\|\vec{x}\|_0 \ge \|\vec{x}\|_1$ here.

Q7: [optimization]

$f(x_1, x_2) = x_1^2 + x_2^4 - x_2^2$

1. $\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial (x_1^2 + x_2^4 - x_2^2)}{\partial x_1} \\ \frac{\partial (x_1^2 + x_2^4 - x_2^2)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 4x_2^3 - 2x_2 \end{bmatrix} = 0$

$\begin{cases} 2x_1 = 0 \\ 4x_2^3 - 2x_2 = 0 \end{cases}$ $\quad x_1 = 0$, $\quad 2x_2^3 = x_2$ $\Rightarrow x_2(2x_2^2 - 1) = 0$ $\Rightarrow x_2 = 0$ or $x_2 = \frac{\sqrt{2}}{2}$ or $x_2 = -\frac{\sqrt{2}}{2}$

$\Rightarrow$ critical points:

$\begin{cases} (0, 0) \\ \text{or } (0, \frac{\sqrt{2}}{2}) \\ \text{or } (0, -\frac{\sqrt{2}}{2}) \end{cases}$

2.

Hessian of $f(x_1, x_2)$:

$$H = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 - 2 \end{bmatrix}.$$

$$\det(H) = 2(12x_2^2 - 2) - 0$$
$$= 24x_2^2 - 4$$

~~when crit~~ critical points: $(0,0)$; $(0, \frac{\sqrt{2}}{2})$; $(0, -\frac{\sqrt{2}}{2})$

① $(0,0)$: $\det(H) = -4 < 0$, $\Rightarrow$ saddle point.

② $(0, \frac{\sqrt{2}}{2})$: $\det(H) = 24 \cdot \frac{1}{2} - 4 = 8 > 0$

$\Rightarrow H = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow \lambda_1 = 2, \lambda_2 = 4$

$\Rightarrow$ positive definite.

$\Rightarrow$ local ~~min~~ minima.

③ $(0, -\frac{\sqrt{2}}{2})$: $\det(H) = 8 > 0$

$H = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow$ local minima

3. $f(0, \pm\frac{\sqrt{2}}{2}) = 0^2 + (\frac{\sqrt{2}}{2})^4 - (\frac{\sqrt{2}}{2})^2$

$= \frac{1}{4} - \frac{1}{2}$

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

$\Rightarrow f(x_1, x_2)_{min} = -\frac{1}{4}$ on $\mathbb{R}^2$.

Q8: [Probability].

~~A: You go to beach on Saturday.~~

~~B: Sunny day.~~

~~P(B|A) = 10%~~

WTS: $P(A|B)$

A: Sunny Day.

B: Forecase Sunny.

$P(A) = 10\% = 0.1$; $P(B|A) = 1 - P(B'|A) = 1 - 0.15 = 0.85$; $P(B|A') = 0.05$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A')P(A')}$$

$$= \frac{0.85 \times 0.1}{0.85 \times 0.1 + 0.05 \times 0.9}$$

$$= \frac{0.085}{0.13}$$

$= \cancel{0.654}\ 0.6538$

Q9: $z \sim N(\mu, \sigma^2)$.

1、 $az+b \sim N(0,1)$

$$\begin{cases} a\mu + b = 0 \\ a^2\sigma^2 = 1 \end{cases} \Rightarrow \begin{matrix} a\mu = -b \\ a = \frac{1}{\sigma} \end{matrix} \Rightarrow \begin{cases} b = -\frac{1}{\sigma}\mu = -\frac{\mu}{\sigma} \\ a = \frac{1}{\sigma} \end{cases}$$

$\frac{z-\mu}{\sigma} \sim N(0,1)$

2、 $Var(z) = E[z^2] - (E[z])^2$

$\sigma^2 = E[z^2] - \mu^2$

$E[z^2] = \sigma^2 + \mu^2$

3、 $\bar{z} \sim N(\bar{\mu}, \bar{\sigma}^2)$ independent to z

~~z+z~~ $z+\bar{z}$ also normal

$\Rightarrow (z+\bar{z}) \sim N(\mu+\bar{\mu}, \sigma^2+\bar{\sigma}^2)$

Q10: ① ~~X~B(n,p)~~. $X \sim Geo(p)$

$P(X=k) = (1-p)^{k-1} \cdot p$

$E[X] = \frac{1}{p}$

② (1) first toss is head, second toss is tail

$p \times (1-p)$

(2) first toss is tail, second is head

$(1-p) \times p$

③ both heads:

$p \times p$

P(~~other~~ both heads | one of them are heads)

$$= \frac{P(\text{both heads})}{P(\text{one of heads})} = \frac{p^2}{p^2+2p(1-p)} = \frac{p^2}{p^2+2p-2p^2} = \frac{p^2}{2p-p^2} = \boxed{\frac{p}{2-p}}$$