1.

let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$, then $A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$

let $b_{rs} = a_{sr}$, for $1 \le r, s \le n$

$$\therefore \det(A^T) = \det([b]_n) = \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} b_{i,\sigma_i})$$
$$= \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{\sigma_i, i})$$
$$= \det([a]_n)$$
$$= \det(A)$$

2. let $I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & 1 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

$$\therefore \det(I) = \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{i,\sigma_i})$$

~~Obvios~~ Obviously, except $\prod_{i=1}^{n} a_{i,i} = 1 \cdot 1 \cdot 1 \cdots 1 = 1$, ~~there it~~ all items must contains 0

$$\therefore \det(I) = \cancel{\sum_{\sigma \in S_n}} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{i,i}) = 1$$

3. let A is a $n \times n$ upper/lower ~~tran~~ triangular ~~matrice~~ matrix.

$\therefore$ $A - \lambda I$ is also a upper/lower triangular matrix.

let $\det(A - \lambda I) = 0$

$$\Rightarrow \prod_{i=1}^{n} (a_{ii} - \lambda) = 0$$

$$\Rightarrow \lambda = a_{ii} \quad (i = 1, 2, \dots n).$$

$\therefore$ the eigen-values are equal to the diagonal elements.

4. support A is a 2×2 non-negative matrix, so

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad a, b, c, d \geq 0.$$

$$\therefore A - \lambda I = \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}$$

$\therefore \det(A - \lambda I) = 0$

$\Rightarrow (a-\lambda)(d-\lambda) - bc = 0$

$\Rightarrow \lambda^2 - (a+d)\lambda + ad - bc = 0$

$\Rightarrow (\lambda - \frac{a+d}{2})^2 = (\frac{a+d}{2})^2 + bc - ad$

$\Rightarrow \lambda = \frac{a+d}{2} \pm \sqrt{(\frac{a+d}{2})^2 + bc - ad}$

$\therefore$ when $bc - ad > 0$, $\sqrt{(\frac{a+d}{2})^2 + bc - ad} > \frac{a+d}{2} \Rightarrow \frac{a+d}{2} - \sqrt{(\frac{a+d}{2})^2 + bc - ad} < 0$.

for example, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\lambda_1 = 1$, $\lambda_2 = -1$.

5.

a) $\frac{\partial f(x)}{\partial x} = a$

b) $\frac{\partial f(x)}{\partial x} = \frac{1}{2}(P + P^T)x + q$

c) $\frac{\partial f(x)}{\partial x} = \frac{1}{2}(P + P^T)x = Px$

d) let $g(x) = e^{a^Tx+b}$, then $\frac{\partial g(x)}{\partial x} = a \cdot e^{a^Tx+b}$

$\therefore \frac{\partial f(x)}{\partial x} = \frac{\partial f(g)}{\partial g} \cdot \frac{\partial g(x)}{\partial x}$

$\therefore f(x) = \frac{g(x)}{1+g(x)}$

$\Rightarrow \frac{\partial f}{\partial g} = \frac{1}{(1+g(x))^2}$

$\therefore \frac{\partial f(x)}{\partial x} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x} = \frac{g'(x) \cdot (1+g(x)) - g(x) \cdot g'(x)}{(1+g(x))^2} = \frac{a \cdot e^{a^Tx+b}}{(1+e^{a^Tx+b})^2}$

e) $\nabla f = \frac{e^{-g(x)} \cdot \nabla_x g(x)}{(1+e^{-g(x)})^2}$

$\because e^{-g(x)} = \frac{1}{f(x)} - 1$

$\therefore \nabla f = (\frac{1}{f(x)} - 1) f^2(x) \cdot \nabla_x g(x)$

6. Let $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$, $Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{12} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}$

$\therefore tr(X^T Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} y_{ij}$

Proof.

1) $\langle X, Y \rangle = \langle Y, X \rangle$

$\because \langle X, Y \rangle = tr(X^T Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} y_{ij} = \sum_{j=1}^{m}\sum_{j=1}^{n} y_{ij} x_{ij} = tr(Y^T X) = \langle Y, X \rangle$

2) $\langle X+Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$

$\because \langle X+Y, Z \rangle = tr((X+Y)^T Z)$

$= \sum_{i=1}^{m}\sum_{j=1}^{n} (x_{ij} + y_{ij}) \cdot z_{ij}$

$= \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} \cdot z_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij} \cdot z_{ij}$

$= tr(X^T Z) + tr(Y^T Z)$

$= \langle X, Z \rangle + \langle Y, Z \rangle$

3) $\langle cX, Y \rangle = c \langle X, Y \rangle$

$\langle cX, Y \rangle = tr((cX)^T Y)$

$= \sum_{i=1}^{m}\sum_{j=1}^{n} c x_{ij} y_{ij}$

$= c \cdot \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} y_{ij}$

$= c\, tr(X^T Y)$

$= c \langle X, Y \rangle$

4) $\langle X, X \rangle \geq 0$, $\langle X, X \rangle = 0 \Leftrightarrow X = 0$

$\langle X, X \rangle = tr(X^T X) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2 \geq 0$

if $\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 = 0$

$\Rightarrow x_{ij} = 0$, $(i \in 1, \dots, m,\ j \in 1, \dots, n)$

$\Rightarrow X = 0$

7)

$$x^T A x = x^T(-A^T)x = \cancel{0} - x^T A^T x = -(x^T A x)^T$$

$\therefore x^T A x = 0$

8) $\because$ X and Y be two-mean random variables.

$\therefore E[X] = 0$, $E[Y] = 0$

$\therefore Cov(X, Y) = E[XY] - E[X]E[Y] = E[XY]$

$var(X) = E[X^2] - E^2[X] = E[X^2]$

$var(Y) = E[Y^2] - E^2[Y] = E[Y^2]$

as Cauchy-Schwarz inequality, $E[XY] \le \sqrt{E[X^2]E[Y^2]}$

$\Rightarrow (Cov(X, Y))^2 \le var(X)\, var(Y)$