EECS 545

Homework One

1) Linear Algebra

(i) True. Let $A$ denotes the symmetric matrix mentioned.

$$\therefore A^T = A$$

$$\Rightarrow (A^{-1})^T = (A^T)^{-1} = A^{-1}$$

$A^{-1}$ is a symmetric matrix.

(ii) True. Let $A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ denotes the $2\times2$ orthogonal matrix

$$\Rightarrow A^T A = AA^T = I$$

$$\therefore \begin{bmatrix} a & c \\ b & d \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\therefore \begin{cases} a^2+b^2 = 1 \\ ac+bd = 0 \\ c^2+d^2 = 1 \end{cases}$$

And there are two solutions of the equation above:

① $\begin{cases} a = \cos\alpha \\ b = \sin\alpha \\ c = \sin\alpha \\ d = -\cos\alpha \end{cases}$ ② $\begin{cases} a = \cos\alpha \\ b = \sin\alpha \\ c = -\sin\alpha \\ d = \cos\alpha \end{cases}$

i.e. $\begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}$ or $\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$

(iii) False.

$$C = \begin{bmatrix} \cdots & c_1^T & \cdots \\ \cdots & c_2^T & \cdots \\ \cdots & c_3^T & \cdots \end{bmatrix} \qquad C^T = \begin{bmatrix} \vdots & \vdots & \vdots \\ c_1 & c_2 & c_3 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$CC^T = \begin{pmatrix} c_1^T c_1 & & \\ & c_2^T c_2 & \\ & & c_3^T c_3 \end{pmatrix}.$$

$$c_i^T c_i = c_i^2 \geq 0.$$

But in $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -1 & -2 \end{bmatrix}$

which diagonal can not be all non-negative.

So. A can't be written as $A = CC^T$.

(iii)

2) Probability

(i) $E_Y[E_X(X|Y)] = \underline{E_Y}[\sum_x x \cdot P(X=x|Y)]$

$$= \sum_y \left[\sum_x x \cdot P(X=x|Y=y)\right] P(Y=y)$$

$$= \sum_x x \sum_y P(X=x, Y=y)$$

$$= \sum_x x \cdot P(X=x)$$

$$= E[X].$$

---

(ii) $I(X\in C) = \begin{cases} 1, & \text{if } X\in C \\ 0, & \text{otherwise} \end{cases}$

$$E[I(X\in C)] = \sum_{i=0}^{1} i \cdot P(I(X\in C)=i)$$

$$= 0 \cdot P(I(X\in C)=0) + 1 \cdot P(I(X\in C)=1)$$

$$= P(I(X\in C)=1)$$

$$= P(X\in C)$$

---

(iii) $Var(X) = E(X^2) - (E(X))^2$

$Var(X|Y) = E(X^2|Y) - (E(X|Y))^2$

$E_Y[Var_X(X|Y)] + Var_Y[E_X(X|Y)]$

$$= \left[E_Y(E_X(X^2|Y)) - E_Y(E_X(X|Y) \cdot E_X(X|Y))\right] + \left[E((E(X|Y))^2) - (E(E(X|Y)))^2\right]$$

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

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

(iv) $X$ and $Y$ are independent. i.e. $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$$E(XY) = \iint f_{XY}(x,y)\,dx\,dy$$

$$= \iint f_X(x)\, f_Y(y)\, dx\,dy$$

$$= \int f_X(x)\,dx \cdot \int f_Y(y)\,dy$$

$$= E(X)\cdot E(Y).$$

Scanned by CamScanner

(IV) $E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = P(X=1)$

Same, $E(Y) = P(Y=1)$

$\cancel{E(X,Y)}$ $E(XY) = P(X=1, Y=1) = P(X=1) \cdot P(Y=1)$

$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) = P(X=1, Y=0) + P(X=1) \cdot P(Y=1)$.

Let $p$ denotes $P(X=1)$, $q$ denotes $P(Y=1)$

$$P(X=1, Y=0) = P(X=1) - P(X=1)P(Y=1)$$
$$= p - pq = p(1-q) = P(X=1) \cdot P(Y=0)$$

$$P(X=0, Y=0) = 1 - pq - p(1-q) - q(1-p)$$
$$= 1 - pq - p + pq - q + pq$$
$$= (1-p)(1-q)$$
$$= P(X=0)P(Y=0)$$

$$P(X=0, Y=1) = P(X=0)P(Y=1)$$

(b) (i) $\leq$. $P(H=h, D=d) = \cancel{P(H=h \wedge D=d)}\ P(H=h \wedge D=d) \leq P(H=h)$

(ii) ? depend. $P(H=h \mid D=d) = \dfrac{P(H=h, D=d)}{P(D=d)}$

(iii) $\geq$. $LHS = \dfrac{P(H=h, D=d)}{P(D=d)}$

$RHS = \dfrac{P(D=d, H=h)}{P(H=h)} \cdot P(H=h) = P(H=h, D=d)$

$P(D=d) \leq 1.$

$P(H=h \mid D=d) \geq P(D=d \mid H=h)\, P(H=h)$

(3) (a) ① $A$ is PSD $\Rightarrow \lambda_i \geq 0$.

$A = \sum_{i=1}^{d} \lambda_i u_i u_i^T$ $\qquad x^T A x \geq 0$.

$x^T A x = \sum_{i=1}^{d} \lambda_i (x^T u_i)(u_i^T x)$

$= \sum_{i=1}^{d} \lambda_i (x^T u_i)^2$.

$\sum_{i=1}^{d} \lambda_i (x^T u_i)^2 \geq 0$. $\qquad (x^T u_i)^2 \geq 0$

$\lambda_i \geq 0 \quad \forall i, \ i = 1, 2, \ldots d$

② $\lambda_i \geq 0 \Rightarrow A$ is PSD.

since $\lambda_i \geq 0$.

then $x^T A x = \sum_{i=1}^{d} \lambda_i (x^T u_i)^2 \geq 0$.

$A$ is PSD.

(b) ① $A$ is PD $\Rightarrow \lambda_i > 0$.

$A = \sum_{i=1}^{d} \lambda_i u_i u_i^T$, $A$ is PD $\Rightarrow x^T A x > 0$

$x^T A x = \sum_{i=1}^{d} \lambda_i (x^T u_i)(u_i^T x) = \sum_{i=1}^{d} \lambda_i (x^T u_i)^2 > 0$

Since $(x^T u_i)^2 \geq 0$, $\lambda_i > 0$.

② $\lambda_i > 0 \Rightarrow A$ is PD

Since $\lambda_i > 0$.

$x^T A x = \sum_{i=1}^{d} \lambda_i (x^T u_i)^2 > 0$

$A$ is PD.

(iii) False.

$$C = \begin{bmatrix} \cdots c_1^T \cdots \\ \cdots c_2^T \cdots \\ \cdots c_3^T \cdots \end{bmatrix} \qquad C^T = \begin{bmatrix} \vdots & \vdots & \vdots \\ c_1 & c_2 & c_3 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$CC^T = \begin{pmatrix} c_1^T c_1 & \vdots & \vdots \\ \vdots & c_2^T c_2 & \vdots \\ \vdots & \vdots & c_3^T c_3 \end{pmatrix}.$$

$$c_i^T c_i = c_i^2 \geq 0.$$

But in $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -1 & -2 \end{bmatrix}$

which diagonal can not be all non-negative.

So. A can't be written as $A = CC^T$.

(iii) False.

If A ~~could~~ can be written as $A = CC^T$ ~~is a symmetric matrix~~.

Then $(\mathrm{eig}(C))^2 = \mathrm{eig}(CC^T) = \mathrm{eig}(C^TC) = \mathrm{eig}(A)$.

if A have ~~can~~ eigen decomposition $\Rightarrow A = U\Lambda U^T$

$$= (U\sqrt{\Lambda})(\sqrt{\Lambda}U^T)$$

$$= (U\sqrt{\Lambda})(U^T\sqrt{\Lambda}^T)^T$$

$$= (U\sqrt{\Lambda})(U\sqrt{\Lambda})^T \quad \text{since } \Lambda \text{ is diagonal.}$$

$$= CC^T$$

But $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -9 & -2 \end{bmatrix}$ apparently not a symmetric

If A could be written as $CC^T$,

then A could be written as $(U\sqrt{\Lambda})(U\sqrt{\Lambda})^T$ ie. $U\Lambda U^T$

which means A needs to be symmetric.

So, the assumption is not right.

A can not be written as $A = CC^T$.

4. (a) $f(x) = a^T x + b$

$f(y) = a^T y + b$

$$f(tx + (1-t)y) = a^T(tx + (1-t)y) + b$$

$$= t a^T x + (1-t) y a^T + b$$

$$= t a^T x + (1-t) a^T y + (t + 1 - t) b$$

$$= t(a^T x + b) + (1-t)(a^T y + b)$$

$$= t f(x) + (1-t) f(y)$$

$$\le t f(x) + (1-t) f(y) \Longrightarrow \text{Convex}$$

$$-f(tx + (1-t)y) = -(t f(x) + (1-t) f(y))$$

$$\Downarrow$$

Same as, this is convex

$\therefore f(tx + (1-t)y)$ is concave.

Since $f(tx + (1-t)y) = t f(x) + (1-t) f(y)$

Not strictly inequality.

$f(x)$ is not strictly convex.

(b) $f$ is strictly convex.

then $f(tx_1+(1-t)x_2) < hf(x_1)+(1-h)f(x_2)$ , $t\in(0,1)$

Suppose $x_1$ is global minimizer, and $x_2$ is also a minimizer.

where $f(x_1) \le f(x_2)$

Since $t\in(0,1) \Rightarrow tf(x_1)+(1-t)f(x_2) \le tf(x_2)+(1-t)f(x_2)$

$$tf(x_1)+(2-t)f(x_2) \le (1-t)f(x_2)$$

$$f(tx_1+(1-t)x_2) < tf(x_1)+(1-t)f(x_2) \le f(x_2)$$

$$f(tx_1+(1-t)x_2) < f(x_2)$$

We assume $x_2$ is one of the global minimizer,

i.e. $f(x) \ge f(x_2)$ , which exactly the opposite of inequality shows ~~be~~ above.

So, ~~there is~~ $f$ at most has one global minimizer.

d ① $f$ convex $\longrightarrow \nabla^2 f(x) \geq 0$.

$$f(x+\lambda d) = f(x) + \nabla f(x)^T(\lambda d) + \frac{1}{2}(\lambda d)^T H(x)(\lambda d) + o(\|\lambda d\|^2)$$

$\because f(x+\lambda d) \geq f(x) + \langle \nabla f(x), \lambda d \rangle$

$\therefore \underline{\lambda^2}\left(\frac{1}{2} d^T H(x) d + \underline{\frac{o(\|\lambda d\|^2)}{\lambda^2}}\right) \geq 0$.

$\lambda^2 \geq 0$ and $\frac{o(\|\lambda d\|^2)}{\lambda^2} \geq 0$.

$\therefore d^T H(x) d \geq 0$. i.e. $\nabla^2 f(x) \geq 0$

② $\nabla^2 f(x) \geq 0 \longrightarrow f$ convex

$$f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T H(z)(y-x)$$

$\downarrow$

$t(y-x)+y$, where $t\in(0,1)$

$\because H(z) \geq 0$.

$\therefore f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$

$\therefore f$ is convex

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

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

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

$= Ax + b.$

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

if $A$ is a PSD matrix, then $f$ is convex.

if $A$ is a PD matrix, the $f$ is strictly convex.

Scanned by CamScanner

1C) $h$ denotes ~~to~~ an arbitrary vector. $\|h\|=1$.

use suggestion $f(x^*+th) = f(x^*) + \langle \nabla f(x^*), th \rangle + \frac{1}{2}\langle th, \nabla^2 f(x^*) ht \rangle + O(t^2h^2)$

$$\begin{cases} x = x^* + th \\ y = x^* \end{cases} \qquad x - y = th.$$

$$f(x^*+th) = f(x^*) + (th)^T \nabla f(x^*) + \frac{1}{2}(\nabla^2 f(x^*) ht)^T th + O(t^2h^2)$$

$$f(x^*+th) - f(x^*) = \cancel{th^T \nabla f(x^*)} + \frac{1}{2} t^2 h \nabla^2 f(x^*) h + O(t^2)$$

since $f(x^*)$ is local minizer.

the $f(x^*+th) - f(x^*) \geq 0$.

$\therefore \frac{1}{2} t^2 h \nabla^2 f(x^*) h + O(t^2) \geq 0$

$$h \nabla^2 f(x^*) h \geq 0.$$

$$\nabla^2 f(x^*) \geq 0$$

$\nabla^2 f(x^*)$ is ~~p~~ positive semidefinite.