CS229

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

$$\nabla_x f(x) = \nabla_x\left(\frac{1}{2} x^T A x\right) + \nabla_x (b^T x).$$

$$= \frac{1}{2} \cdot 2Ax + b.$$

$$= Ax + b.$$

notes P10.

(b). $f(x) = g(h(x))$ g and h are both differentiable.

$$\nabla_x f(x) = g'(h(x)) \nabla h(x).$$ chain rule.

(c). $\nabla^2 f(x) = \nabla (Ax+b)$

$$= A.$$

(d). $f(x) = g(a^T x)$.

$$\nabla f(x) = \nabla\left(g(a^T x)\right) = g'(a^T x) \nabla (a^T x)$$

$$= g'(a^T x) a = \begin{bmatrix} g'(\Sigma a_i x_i) a_1 \\ \vdots \\ g'(\Sigma a_i x_i) a_n \end{bmatrix}.$$

$$\nabla^2 f(x) = \begin{bmatrix} g''(\Sigma a_i x_i) a_1 a_1 & g''(\Sigma a_i x_i) a_2 a_1 & \\ g''(\Sigma a_i x_i) a_1 a_2 & & \vdots \\ & \ddots & \\ g''(\Sigma a_i x_i) a_1 a_n & g''(\Sigma a_i x_i) a_2 a_n & \end{bmatrix}.$$

$$= g''(a^T x) a a^T$$

2. A is PSD. $A \succeq 0$ if all vectors $x^T A x \geq 0$, $A = A^T$

A PD. $A \succ 0$ if all vectors $x^T A x > 0$, $A = A^T$.

I. satisfies $x^T I x = \|x\|_2^2 = \sum_{i=1}^{n} x_i^2$.

(a). $z \in R^n$ is an n-vector.

$A = zz^T$

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

$= x^T z z^T x$ associative

$= x^T z x^T z$

$= (x^T z)^2 \geq 0$.

(b) $z \in R^n$ nonzero. n-vector

$A = zz^T$.

$N(A) = \{x \in R^n : Ax = 0\}$

$zz^T x = 0$.

Since z is nonzero, $z^T x = 0$.

$N(A) = \{x \in R^n : z^T x = 0\}$. x is orthognal to z^T.

$z^T \cdot z = z \cdot z^T$ is nonzero.

$\text{rank}(A) = n - \text{nul}(A)$

$= n - (n-1)$. z.

$= 1$.

(c) A PSD. $x^T A x \geq 0$, $A = A^T$.
B. arbitrary.

$$x^T (BAB^T) x = x^T BAB^T x$$
$$= (x^T B) A (B^T x)$$
$$= (B x^T) A (B^T x)$$
$$= (B^T x)^T A (B^T x)$$

$\because$ A is PSD.

$\therefore$ $(B^T x)^T A (B^T x) \geq 0$.

$\therefore$ BAB^T is PSD.

3. $A \in R^{n \times n}$ $\qquad A x = \lambda x$

$\uparrow$ eigenvalue

(a). $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$

$T = [t^{(1)}\ t^{(2)} \cdots t^{(n)}]$

show $A t^{(i)} = \lambda_i t^{(i)}$

$$A = T \Lambda T^{-1}$$
$$AT = T\Lambda$$

$$A \begin{bmatrix} | & | & & | \\ t^{(1)} & t^{(2)} & \cdots & t^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ t^{(1)} & t^{(2)} & \cdots & t^{(n)} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

$$\begin{bmatrix} | & | & & | \\ At^{(1)} & At^{(2)} & \cdots & At^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 t^{(1)} & \lambda_2 t^{(2)} & \cdots & \lambda_n t^{(n)} \\ | & | & & | \end{bmatrix}$$

$\therefore$ $A t^{(i)} = \lambda_i t^{(i)}$

(b). U is orthorgonal if $U^T U = I$.

A is symmetric if $A = A^T$.

$U^T A U = \Lambda$

$A = U \Lambda U^T$

$\lambda_i = \lambda_i(A)$ ith eigenvalue of A.

$AU = U\Lambda$.

From (a) we know $A u^{(i)} = \lambda_i u^{(i)}$

$u^{(i)}$ is an eigenvector of A.

Show that

(c). If A is PSD, that is $x^T A x \geq 0$,

then $\lambda_i(A) \geq 0$ for each i.

$A t^{(i)} = \lambda_i t^{(i)}$

$t^{(i)T} A t^{(i)} = t^{(i)T} \lambda_i t^{(i)} \geq 0$. A is PSD.

$= \lambda_i t^{(i)T} t^{(i)}$ $t^{(i)T} t^{(i)} \geq 0$.

$= \lambda_i \geq 0$.

4. $X = (X_1 \cdots X_n)$

$X \sim N(\mu, \Sigma)$.

(a) $Y = X_1 + X_2 + \cdots X_n$

$E[Y] = E[X_1 + X_2 \cdots X_n]$

$= n\mu = a^T\mu$

$a = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ $a \in \mathbb{R}^n$

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

$= a^T E(XX^T) a - a^T\mu\mu^T a$

$= a^T Cov(\vec{X}) a$ Y is Gaussian distribution.

$= \sum_{i=1}^{n} \Sigma_{ii}$

(b). $E[X^T\Sigma^{-1}X]$

$= \cancel{E[Y^T A^T (AA^T)^{-1} A Y]}$

$\cancel{= E[Y^T A^T (A^T A)^{-1} A Y]}$

?

$\cancel{= E[Y^T Y]}$

$\cancel{= E[\sum_{i=1}^{n} Y_i^2]}$

$\cancel{\Sigma = AA^T}$

$\cancel{\text{Let } Y = A^{-1}X}$

$\cancel{X = AY}$

$E[X^T\Sigma^{-1}X]$

$= E[tr(XX^T\Sigma^{-1})]$

$= tr(E[XX^T\Sigma^{-1}])$

$= tr(E[XX^T]\Sigma^{-1})$

$= tr((\mu\mu^T + \Sigma)\Sigma^{-1})$

$= tr(\mu\mu^T\Sigma^{-1}) + n.$

$= tr(\mu^T\Sigma^{-1}\mu) + n = \mu^T\Sigma^{-1}\mu + n.$