ML CS229

#PS0

NOTE P24

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

$(--)\begin{pmatrix} \vdots \end{pmatrix}$ $(--)\begin{pmatrix} \vdots \end{pmatrix}$

$1\times n \quad n\times n \quad n\times 1 \qquad 1\times n \quad n\times 1$

$\nabla(b^T x)$

$b^T x = (b_1, \cdots b_n) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= \sum_{i=1}^{n} b_i x_i$

$\frac{\partial}{\partial x_k} \sum_{i=1}^{n} b_i x_i$

$= b_k$

$\begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$

$\therefore \nabla(\vec{b}^T x) = \vec{b}$

$\nabla(\frac{1}{2}\vec{x}^T A \vec{x})$

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

$= (x_1, \cdots x_n) \begin{pmatrix} A_{11} & A_{12} & \cdots \\ & & A \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= \begin{pmatrix} x_1 A_{11} & x_1 A_{12} & \cdots \\ & & \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= (x_1 A_{11} \cdot x_1 + x_1 A_{12} \cdot x_2 + \cdots) + \cdots$

$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$

$\frac{\partial}{\partial x_k} \sum\sum x_i A_{ij} x_j$

$= \frac{\partial}{\partial x_k} \left[\sum_{i\neq k}\sum_{j\neq k} x_i A_{ij} x_j + \sum_{i\neq k} x_i A_{ik} \cancel{x_k} + \sum_{j\neq k} \underline{x_k} \cancel{A_{kj}} x_j + \underline{x_k^2} \cdot A_{kk} \right]$

$\begin{pmatrix} \cdots & A_{1k} & \cdots \\ & A_{2k} & \\ & \vdots & \\ & A_{nk} & \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= \sum_{i\neq k} x_i A_{ik} + \sum_{j\neq k} A_{kj} \cdot x_j + 2x_k \cdot A_{kk}$

As A is symmetric, supplement $i=k$, $j=k$

so $\Rightarrow A_{ik} = A_{ki}$

$= 2\sum_{i=1} x_i A_{ik}$ *

$n\times 1 \quad n\times n$

$\therefore \nabla(\frac{1}{2}\vec{x}^T A \vec{x}) = \cancel{2}\vec{x} A$

$\frac{1}{2}\times 2 A\vec{x}$

Therefore

$\nabla f(x) = A\vec{x} + \vec{b}$

(b) $\nabla f(x)$

$f(x) = g(h(x))$

let intermediate variable $u = h(x)$

$\therefore \frac{\partial g(u)}{\partial x_i} = \frac{d g(u)}{d u} \frac{\partial u}{\partial x_i}$ $\quad x \in R^n$

$u \in R$ $\quad = g'(u) \cdot \frac{\partial h(x)}{\partial x_i}$

$\therefore \nabla f(x) = g'(h(x)) \cdot \nabla h(x)$

(c) $= \nabla (A\vec{x} + \vec{b})^T$

$\frac{\partial \sum A x_i + b}{\partial x_k}$

$\overset{n\times n}{A} \cdot \overset{n\times 1}{\vec{x}}$

$A\vec{x} + \vec{b} = \begin{pmatrix} A_{11}x_1 + \cdots A_{1n}x_n \\ \vdots \\ A_{n1}x_1 + \cdots A_{nn}x_n \end{pmatrix}$

$(A_{11}x_1 + \cdots A_{1n}\cdot x_n, \; \cdots x_1 A_{n1} + \cdots A_{nn}x_n)$

i-column

$\nabla_x \frac{\partial f(x)}{\partial x_i} = \frac{\partial \sum_{j=1} A_{ij} \cdot x_j}{\partial x_k}$ ← same as *

$\uparrow x_l$

$= \frac{\partial A_{ik} x_k}{\partial x_k} = A_{ik}$

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

II

$\nabla^2 f(x) = [\nabla(A\vec{x}+\vec{b})_1, \cdots \nabla(A\vec{x}+\vec{b})_n]$

$= \left[\frac{\partial(A\vec{x}+\vec{b})}{\partial x_1}, \cdots \frac{\partial(A\cdot\vec{x}+\vec{b})}{\partial x_n}\right]$

$= \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & & A_{nn} \end{bmatrix} = A$

III $\frac{\partial}{\partial x_l}\left[\frac{\partial f(x)}{\partial x_k}\right] = \frac{\partial}{\partial x_l}\frac{\partial}{\partial x_k}\left(\frac{1}{2}\vec{x}\cdots\right)$ ↳ symbols altogether

$= \frac{\partial}{\partial x_l} \cancel{\frac{\partial}{\partial x_k}} \left(\sum_{i=1}^{n} x_i A_{ik}\right)$

$= A_{lk}$ $\quad A$

(d) $\vec{x} \in \mathbb{R}^n$

$\vec{a}^T \cdot \vec{x} \to R$

$g(u)$ where $u = \vec{a}^T\vec{x} \in R$

$\nabla f(x) = \nabla g(u) = \frac{\partial g(u)}{du} \cdot \frac{\partial g(u)}{\partial x}$

$x \in R^n \quad u \in R$

$= g'(u) \cdot \vec{a}$

$= g'(\vec{a}^T\vec{x}) \cdot \vec{a}$

$\nabla^2 f(x) = \cancel{\frac{\partial}{\partial x}} [g'(\vec{a}^T \cdot \vec{x}) \cdot \vec{a}]$

$\left[\frac{\partial}{\partial x_1}[g'(\vec{a}^T\vec{x})\vec{a}_1], \cdots \frac{\partial}{\partial x_n}[g'(\vec{a}^T\vec{x})\vec{a}_n]\right]$

$\frac{\partial}{\partial x_j}[g'(\vec{a}^T\vec{x})\vec{a}_i]$

let $u = \vec{a}^T \cdot \vec{x}$ again

$= \frac{dg'(u)\cdot\vec{a}}{du} \; \frac{\partial u}{\partial x_j}$

$= g''(u)\vec{a}_i \, a_j$

$\therefore g''(\vec{a}^T\vec{x})\, \vec{a}\, \vec{a}^T$

$\vec{a} \cdot \vec{a}^T$

ML CS229

#PS0

II. $\dfrac{\partial^2 g(h(x))}{\partial x_i \, \partial x_j}$

let $U = h(x)$

$\dfrac{d^2 g(U)}{dU^2} \cdot \dfrac{\partial U}{\partial x_i} \cdot \dfrac{\partial U}{\partial x_j}$

$= g''(\vec{a}^T \cdot \vec{x}) \cdot a_i \cdot a_j$

$= g''(\vec{a}^T \vec{x}) \begin{bmatrix} a_1 & & \vdots \\ \vdots & \ddots & \vdots \\ & & a_{nn} \end{bmatrix} \sim \vec{a} \cdot \vec{a}^T$

2. (a) $z \in R^n$

1°: $A^T = (\vec{z}\vec{z}^T)^T = (\vec{z}^T)^T \cdot \vec{z}^T = \vec{z}\vec{z}^T = A$

Matrix A is symmetric.

2°: $\forall \vec{x} \in R^n$

$\vec{x}^T A \cdot \vec{x}$ $\qquad AQ = Q\Lambda$

Diagonize

$A = Q \cdot \Lambda \cdot Q^{-1}$

As A is symmetric $\Rightarrow$ Q orthonormal

$Q^{-1} = Q^T \qquad Q^T Q = I$

$\therefore A = Q \Lambda Q^T$

plug in $\vec{x}^T A \vec{x}$

we get $\vec{x}^T Q \Lambda Q^T \vec{x}$

let $\vec{y} = Q^T \cdot \vec{x}$

$= \vec{y}^T \cdot \Lambda \cdot \vec{y}$

$= (y_1 \cdots y_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$= \sum_{i=1}^{n} \lambda_i y_i^2 \geq 0$

where $\vec{y} = Q^T \cdot \vec{x}$

II. $\vec{x}^T A \cdot \vec{x}$

plug $A = \vec{z}\vec{z}^T$ into:

$\vec{x}^T \vec{z} \cdot \vec{z}^T \vec{x}$

$= (\underset{1\times n}{\vec{x}^T} \cdot \underset{n\times 1}{\vec{z}}) \cdot (\vec{z}^T \vec{x})$

$= (\vec{x}^T \vec{z}) \cdot (\vec{x}^T \cdot \vec{z})^T$

$= \underset{1\times 1}{(\vec{x}^T \cdot \vec{z})^2} \geq 0$

(b) $N(A) = \{\vec{x} \in R^n \mid A\vec{x} = \vec{0}\}$ set

$\vec{z} \neq 0$

$\vec{x}^T \cdot A \cdot \vec{x}$ this time

$= (\vec{x}^T \cdot \vec{z})^2 > 0$

when $\vec{x} \neq 0$ & $\vec{z} \neq 0$

positive definite.

$\therefore$ when $\vec{x} \neq 0$: $\vec{x}^T A \cdot \vec{x} \neq 0 \Rightarrow A\vec{x} \neq 0$

$\Rightarrow$ only if $\vec{x} = 0$: $A\vec{x} = 0$

number of $N(A) = 1$

$\therefore R(A^T) = n - N(A) = n - 1$

Span the space

$R(A) = R(\vec{z}\vec{z}^T) = 1$

$\vec{z} \cdot \vec{z}^T \neq 0$

$(\text{—})(|)$

(c) $A \in R^{n\times n} \geq 0$

$B \in R^{m\times n}$

$? \Rightarrow \quad BAB^T \geq 0$

Given that $A \geq 0$

$\Rightarrow \quad A^T = A$

$\forall x \in R^n \quad \vec{x}^T A \vec{x} \geq 0$

Proof. 1°. $(BAB^T)^T$

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

$= B \cdot A \cdot B^T$

Therefore BAB^T is symmetric.

2°: $\forall \vec{x} \in R^n$

$\vec{x}^T \cdot (\overbrace{B A B^T}^{m\times n \; n\times n \; n\times m}) \cdot \vec{x}$

$P \Lambda P^{-1}$ ✗ arbitrary

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

$= \underbrace{(\vec{x}^T B)}_{1\times m \;\; m\times n} A \underbrace{(\vec{x}^T B)^T}_{\vec{x}'} \geq 0$

$1\times n$

3. (a) proof: According to the def of eigenvector & eigenvalues $\Rightarrow \quad A\vec{t}^{(i)} = \lambda_i \vec{t}^{(i)}$

Given: $A = T\Lambda T^{-1}$

T is invertible

$AT = T\Lambda$

$A[\vec{t}^{(1)}, \ldots, \vec{t}^{(n)}] = \overbrace{[\vec{t}^{(1)} \ldots \vec{t}^{(n)}]}^{1\times n} \overbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}^{n\times n}$

$A[\vec{t}^{(1)} \ldots \vec{t}^{(n)}] = [\lambda_1 \vec{t}^{(1)}, \longrightarrow \lambda_n \vec{t}^{(n)}]$

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

(b) $U = [U^{(1)}, \ldots U^{(n)}]$,

where $U^{(i)} \in R^n$.

$A = U \cdot \Lambda U^T$

$\Lambda = diag(\lambda_1, \ldots \lambda_n)$

if U is orthogonal

$U \cdot U^T = I \qquad U^T = U^{-1}$

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

from (a). $\Rightarrow \quad AU^{(i)} = \lambda_i U^{(i)}$

(c) $A \geq 0$. PSD

$\Rightarrow A^T = A$

the same as 2 (a) proof

II eigenvector $\vec{t}^{(i)}$

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

$\vec{t}^{(i)T} A \cdot \vec{t}^{(i)} \geq 0$

$\vec{t}^{(i)T} \cancel{A} \lambda_i \vec{t}^{(i)} \geq 0$

$\Rightarrow \lambda_i \geq 0$