CS229 (Fall 2018)

Introduction to Machine Learning

Problem Set #0

1. $f: R^n \to R$

$$\nabla f(x) = \begin{bmatrix} \frac{\delta f(x)}{\delta x_1} \\ \vdots \\ \frac{\delta f(x)}{\delta x_n} \end{bmatrix} \quad \text{where } X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Gradient

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\delta^2 f(x)}{\delta x_1^2} & \frac{\delta^2 f(x)}{\delta x_1 \delta x_2} & \cdots & \frac{\delta^2 f(x)}{\delta x_1 \delta x_n} \\ \frac{\delta^2 f(x)}{\delta x_2 \delta x_1} & \frac{\delta^2 f(x)}{\delta x_2^2} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\delta^2 f(x)}{\delta x_n \delta x_1} & \cdots & \cdots & \frac{\delta^2 f(x)}{\delta x_n^2} \end{bmatrix}$$

Hessian

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

$\Rightarrow$ Let $b = [b_1\ b_2 \cdots b_n]$ & $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$b^T x = b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$$

$$\nabla b^T x = \left[\frac{\delta b^T x}{\delta x_1} \cdots \frac{\delta b^T x}{\delta x_n}\right]^T$$

$$= [b_1\ b_2\ \cdots\ b_n]^T$$

$$= b$$

$$\nabla\left(\frac{1}{2}x^T A x\right) = \nabla\left(\frac{1}{2}x^T(Ax)\right) + \nabla\left(\frac{1}{2}(x^T A)x\right)$$

$$= \frac{1}{2}Ax + \frac{1}{2}A^T x$$

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

$$= Ax \quad \{as\ A = A^T\}$$

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

$$= Ax + b$$

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

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

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

$$\nabla f(x) = Ax + b$$

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

Date
Page

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

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

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

$\nabla^2 f(x) = a g''(a^T x) a^T$

2. (a) $A = zz^T$

$\Rightarrow x^T A x \geq 0$

$\Rightarrow x^T z z^T x \geq 0$

$\Rightarrow (z^T x)^T (z^T x) \geq 0$

$\Rightarrow \|z^T x\|^2 \geq 0$

(b) $A = zz^T$

$Ax = 0$ if $x \in N(A)$

$\underbrace{z z^T x}_{\text{is zero if } x \perp z} = 0$

$\Rightarrow$ Null space of A contain every vector orthogonal to z.

$\Rightarrow$ So rank $=$ ~~[illegible]~~ 1

Date
Page

(c) $BAB^T \qquad A = N^T N$

$x^T B A B^T x = x^T B N^T N B^T x$

$= (NB^T x)^T (NB^T x)$

$= \|NB^T x\|^2 \geqslant 0$

3. (a) $A = T \Lambda T^{-1} \qquad T = [t^{(1)}\ t^{(2)} \cdots t^{(n)}]$

$\Lambda = \mathrm{diag}(\lambda_1 \cdots \lambda_n)$

$AT = T\Lambda$

$$A\,[t^{(1)} \cdots t^{(n)}] = [t^{(1)} \cdots t^{(n)}] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

$$\Rightarrow [At^{(1)}\ At^{(2)} \cdots At^{(n)}] = [\lambda_1 t^{(1)}\ \lambda_2 t^{(2)} \cdots \lambda_n t^{(n)}]$$

$At^{(1)} = \lambda_1 t^{(1)}$

$At^{(2)} = \lambda_2 t^{(2)}$

$\vdots$

$At^{(n)} = \lambda_n t^{(n)}$

(b) & (c) $\longrightarrow \lambda_{min} \leqslant \|A\| \leqslant \lambda_{max}$