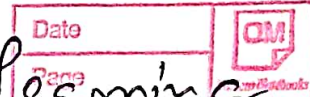


Introduction to Machine Learning

Problem Set #0

1. $f: \mathbb{R}^n \rightarrow \mathbb{R}$

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

Gradient

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

Hessian

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

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

~~Let~~

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

$$\nabla b^T x = \left[\frac{\partial b^T x}{\partial x_1} \quad \dots \quad \frac{\partial b^T x}{\partial x_n} \right]^T$$

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

$$= \underline{\underline{b}}$$

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

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

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

$$= A x \quad \left\{ \text{as } A = A^T \right\}$$

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

$$= \underline{\underline{A x + b}}$$

⑥ $f(x) = g(h(x))$

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

⑦ $f(x) = \frac{1}{2} x^T A x + b^T x$

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

$$\nabla^2 f(x) = \underline{\underline{A}}$$

① $f(x) = g(a^T x)$

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

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

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

2. ② $A = Z Z^T$

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

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

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

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

③ $A = Z Z^T$

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

$$Z Z^T x = 0$$

→ is zero if $x \perp Z$

⇒ Null space of A contain every vector orthogonal to Z .

↳ So rank = ~~1~~ 1

© $BA B^T$ $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} N^T N \times 10 \times 10$ $P = (x)^T + 11$

$$X^T B A B^T X = X^T B N^T N B^T X = (X^T B N^T) (N B^T X) = (X^T) \Delta$$

$$= (N^T X)^T (N^T X)^{-1} (X^T y)^T \beta =$$

$$= \|NB^T x\|^2 \geq \|0\|^2 = 0 = f(x) \quad \nabla$$

3. (a) $A = T \Lambda T^{-1}$ $T = [t^{(1)} \ t^{(2)} \ \dots \ t^{(n)}]$ $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

$$A = \text{diag}(\lambda_1, \dots, \lambda_n)$$



$$A^T = T \Lambda$$

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

$$\Rightarrow [A t^{(1)} \quad A t^{(2)} \quad \dots \quad A t^{(n)}] = [\lambda_1 t^{(1)} \quad \lambda_2 t^{(2)} \quad \dots \quad \lambda_n t^{(n)}]$$

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

$$A t^{(2)} = \lambda_1 t^{(2)}$$

$$At^{(n)} \in \lambda_{nt}^{(n)}$$

b k c

$$\lambda_{\min} \leq \|A\| \leq \lambda_{\max}$$