2/13/2021 Homework - 2 Ronit Mankad, 001062358

Q1: For $A \in R^{m \cdot n}$,

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Show that trace $\|A\|_F^2 = \text{Trace}(A^TA)$

⇒ Since A is a $R^{m \cdot n}$ matrix, A^TA will be an $R^{n \cdot n}$ matrix like,

$$A^TA = \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{m1} \\ a_{12} & & \vdots \\ \vdots & & \vdots \\ a_{1n} & & a_{mn} \end{bmatrix}}_{m} \times \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix}}_{n}$$

$$\therefore A^TA = \begin{bmatrix} \sum_{i=1}^{m} |a_{i1}|^2 & \cdots & \sum_{i=1}^{m} a_{i1} a_{in} \\ \vdots & & \\ \sum_{i=1}^{m} a_{in} \cdot a_{i1} & \cdots & \sum_{i=1}^{m} |a_{in}|^2 \end{bmatrix}$$

Scanned with CamScanner

$$\therefore \text{Trace}(A^TA) = \sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2$$

$$\therefore \text{Trace}(A^TA) = \left(\sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2}\right)^2$$

$$\therefore \text{Trace}(A^TA) = \|A\|_F^2$$

Hence Proved.

Q2:

(a) Prove that L_2 norm is a norm

$\Rightarrow$ For a function $f(x) : R^n \to R$ to be a norm, it should satisfy 4 conditions-

Let us prove all of them for the L-2 norm.

$$\text{L-2 Norm} = \|x\|_2 \triangleq \sqrt{\sum_{i=1}^{n} x_i^2},$$

i) $\forall x \in R^n, f(x) \geq 0$.

Now since L-2 norm is a square root, $\sqrt{\sum_{i=1}^{n} x_i^2}$,

it is always greater than zero. So this condition is met.

ii) $f(x) = 0$ iff $x = 0$ [Non-Negativity].

Now, L-2 norm $= \sqrt{\sum_{i=1}^{n} x_i^2}$

This value can be zero if and only if all values of the n-dimensional vector X are zero. i.e X is a zero vector.

$\therefore$ This condition is also met by L-2 Norm.

iii) Homogenity; $\forall x \in R^n, t \in R \Rightarrow f(tx) = |t| f(x)$.

Let us multiply the vector $X \in R^n$ with some ~~real number t~~ Scalar t.

Then $tX = \begin{bmatrix} tx_1 \\ tx_2 \\ \vdots \end{bmatrix}$

and, $\|tx\|_2 = \sqrt{\sum_{i=1}^{n} (tx_i)^2}$

$= \sqrt{\sum_{i=1}^{n} t^2 (x_i)^2}$

$= t\sqrt{\sum_{i=1}^{n} (x_i)^2}$

$\therefore \|tx\|_2 = t\|x\|_2$

Hence the homogenity constraint is also met.

iv) Triangular Inequality,

$\forall x, y \in R^n, \; f(x+y) \leq f(x) + f(y)$

now, $\|x+y\|_2 = \sqrt{\sum_{i=1}^{n} (x_i + y_i)^2}$

Squaring both sides,

$\|x+y\|_2^2 = \sum_{i=1}^{n} (x_i + y_i)^2$

$\therefore \|x+y\|_2^2 = \sum_{i=1}^{n} (x_i)^2 + \sum_{i=1}^{n} (y_i)^2 + 2\sum_{i=1}^{n} x_i y_i$

$\therefore \|X+Y\|_2^2 = \|X\|_2^2 + \|Y\|_2^2 + 2\sum_{i=1}^{n} x_i y_i$

$= \|X\|_2^2 + \|Y\|_2^2 + 2X\cdot Y$

↳ vector dot product.

Now, using the Cauchy-Shwarz inequality,

$$|X\cdot Y| \leq \|X\|_2 \|Y\|_2$$

$\therefore \cancel{\|X\|_2^2 + X}$

$$\|X+Y\|_2^2 \leq \|X\|_2^2 + \|Y\|_2^2 + 2\|X\|_2\|Y\|_2$$

$$\therefore \|X+Y\|_2^2 \leq \left(\|X\|_2 + \|Y\|_2\right)^2$$

Taking square root both side,

$$\|X+Y\|_2 \leq \|X\|_2 + \|Y\|_2$$

Hence L-2 Norm follows the triangular inequality.

$\therefore$ We can say that L-2 Norm is a norm.

(b) L_∞ norm is a norm.

L_∞ norm is defined as $\triangleq \max_{i=1 \ldots N} |x_i|$

Let us prove that it is a norm by showing that is displays all four properties of a norm:

(i) $\forall x \in R^n, f(x) \geq 0$

Now $L^\infty = \max_{i=1 \ldots N} |x_i|$

and $|x_i|$ will always be greater than or equal to zero regradless if x_i is negative or not.

$\therefore L^\infty$ satisfies this property.

(ii) $f(x) = 0$ iff $x = 0$ [Non-Negativity]

Now, $L^\infty = \max_{i=1 \ldots N} |x_i|$ can only be zero in the case where all ~~ele~~ elements of the vector X are zero. Otherwise it will be > 0.

$\therefore L^\infty$ satisfies this property.

(iii) $\forall x \in R^n, t \in R$, $f(tx) = |t| f(x)$ [Homogenity]

Now $tX = \begin{bmatrix} tx_1 \\ tx_2 \\ \vdots \\ \vdots \end{bmatrix}$

$$\therefore L^\infty(tX) = \max_{i=1..N} |tx_i| = |t| \max_{i=1..N} |x_i|$$

$$\therefore L^\infty(tx) = |t| L^\infty(x)$$

$\therefore L^\infty$ is homogenous.

(iv) $\forall x, y \in R^n \Rightarrow f(x+y) \leq f(x) + f(y)$

~~Here $L^\infty(x+y) = \max_{i=1..N} |x_i + y_i|$~~

$$\therefore L^\infty(x+y) \leq L^\infty(x) + L^\infty(y)$$

Looking at LHS,

[Property of modulus
$|A+B| \leq |A| + |B|$]

$$\max_{i=1..N} |x_i + y_i| \leq \max_{i=1..N} |x_i| + \max_{i=1..N} |y|$$

$$\leq L^\infty(x) + L^\infty(y)$$

Hence Proved

∴ We can say that L-∞ Norm is a norm since it satisfies all the properties of a norm.

Q3: Prove that for $A \in R^{m.n}$ and $B \in R^{n.m}$,

$$\text{Trace}(AB) = \text{Trace}(BA)$$

$$\Rightarrow \quad A = m\underbrace{\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1n} \\ \vdots & & & \\ \vdots & & & \\ a_{m1} & \dots & \dots & a_{mn} \end{bmatrix}}_{n} \qquad B = \underbrace{\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ \vdots & & & \\ \vdots & & & \\ b_{n1} & \dots & \dots & b_{nm} \end{bmatrix}}_{m}$$

$$A \cdot B = \underbrace{\begin{bmatrix} \sum_{i=1}^{n} a_{1i} \cdot b_{i1} & \dots & \sum_{i=1}^{n} a_{1i} b_{im} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} a_{mi} b_{i1} & \dots & \sum_{i=1}^{n} a_{mi} b_{im} \end{bmatrix}}_{m} m$$

$$\text{Trace}(AB) = \cancel{\sum_{i=1}^{n} a_{1i}}\ \sum_{i=1}^{n} a_{1i} \cdot b_{i1} + \sum_{i=1}^{n} a_{2i} \cdot b_{i2} + \dots \sum_{i=1}^{n} a_{mi} b_{in}$$

$$\therefore \text{Trace}(AB) = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{ji} \cdot b_{ij}$$

Now,

$$B \cdot A = \begin{bmatrix} \sum_{i=1}^{m} b_{1i} a_{i1} & \cdots & \sum_{i=1}^{m} b_{1i} a_{in} \\ \vdots & & \\ \sum_{i=1}^{m} b_{ni} a_{i1} & \cdots & \sum_{i=1}^{m} b_{ni} a_{im} \end{bmatrix}$$

$$\text{Trace}(B \cdot A) = \sum_{i=1}^{m} b_{1i} a_{i1} + \sum_{i=1}^{m} b_{2i} a_{i2} + \dots \sum_{i=1}^{m} b_{ni} a_{im}$$

$$\therefore \text{Trace}(BA) = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} \cdot a_{ij}$$

now, in Trace(BA) let us switch i and j so that the notation is the same as Trace(AB)

$$\therefore \text{Trace}(BA) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} a_{ji}$$

as we can see,

$$\text{Trace}(AB) = \text{Trace}(BA)$$

Hence Proved.

P4: For $A \in R^{n \cdot n}$, prove that,

A. If $x \in R^{n \cdot 1}$ then $\frac{\partial x^T A x}{\partial x} = (A + A^T)x$

Let us use the definition of derivative to calculate the derivative here,

$$\frac{\partial f(x)}{\partial x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \text{ where } h \to 0$$

$$\frac{\partial}{\partial x}(x^T A x) = \lim_{h \to 0} \left[\frac{(x+h)^T A (x+h) - x^T A x}{h} \right]$$

$$= \lim_{h \to 0} \left[\frac{x^T A x + x^T A h + h^T A x + h^T A h - x^T A}{h} \right.$$

Here, $(h^T A x)^T = h A^T x^T$

$$= \lim_{h \to 0} \left(\frac{x^T A h + x^T A^T h + h^T A h}{h} \right)$$

$$= \lim_{h \to 0} \left(\left(\frac{x^T (A + A^T) + h^T A}{h} \right) h \right)$$

$$= \lim_{h \to 0} \left(x^T (A + A^T) + h^T A \right)$$

Now since, $h \approx 0$ (h is really small), $h^T A = 0$.

$$\therefore \frac{\partial x^T A x}{\partial x} = x^T (A + A^T)$$

Hence Proved.

B. If $X \in R^{n \cdot n}$, then, $\frac{\partial \, Trace(A^T X)}{\partial X} = A$

$\Rightarrow$

$$Tr(A^T X) = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_{ij}$$

Now, $\frac{\partial (Tr(A^T X))}{\partial X} =$

$$= \frac{\partial \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_{ij} \right)}{\partial X}$$

$$= \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} \left(\partial (a_{ij} x_{ij}) \right)}{\partial X}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}$$

$$\boxed{\frac{\partial (Tr(A^T X))}{\partial X} = A}$$

Q5. Prove if each of the equations are convex:

(A) $f(x) = e^{ax}$

Let's prove if $f(x)$ is convex by taking a double derivative and checking if the hessian is ~~convex~~ PSD.

i.e $$\frac{\partial^2 f(x)}{\partial x} \geq 0$$

Now, $$\frac{\partial^2 f(x)}{\partial x} = \frac{\partial (a.e^{ax})}{\partial x}$$

$$H = a^2 e^{ax} \geq 0$$

Now, we know that e^{ax} will always be greater than zero and $a^2 e^{ax}$ will also be greater than or equal to zero.

Therefore, ~~the~~ H is PSD.

Hence $f(x) = e^{ax}$ is convex

(B) $f(x) = -\log(x)$

Hessian $H = \frac{\partial^2 f(x)}{\partial x} = \frac{\partial}{\partial x}\left(-\frac{1}{x}\right)$

$\therefore H = \frac{1}{x^2} \geq 0$

Now $\frac{1}{x^2}$ will always be greater than zero. Hence H is PSD.

$\therefore -\log(x)$ is convex. Hence Proved.

(C) $f(x) = e^{g(x)}$ where $g(x)$ is convex.

Now, $H = \frac{\partial^2 f(x)}{\partial x}$

$= \cancel{\frac{\partial}{\partial x}\left(g(x)e^{g(x)} \cdot g'(x)\right)}$

$= \cancel{g^2(x)e^{g(x)} \cdot (g'(x))^2 + g(x)e^{g(x)} \cdot g''(x)}$

$$= \frac{\partial}{\partial x}\left(g'(x) e^{g(x)} \right)$$

$$= (g'(x))^2 e^{g(x)} + g''(x) e^{g(x)}$$

(using product rule of derivative)

$$\therefore H = e^{g(x)} \left[(g'(x))^2 + g''(x) \right]$$

Now, we know that $e^{g(x)}$ will always be ≥ 0 ~~since~~ because of the property of the exponent.

Moreover, $[(g'(x))^2 + g''(x)]$ will also be ~~great~~ ≥ 0 since the first term is a square and $g''(x)$ will be ≥ 0 since it is Hessian of $g(x)$ which is convex.

$\therefore H \geq 0$, $\Rightarrow$ H is PSD.

Hence, $e^{g(x)}$ is convex.

(D) L_∞ Norm

Now $f(x) = L(\infty) = \max_{i=1..N} |x_i|$

For a function to be convex,

$$F(\alpha\theta_1 + (1-\alpha)\theta_2) \leq \alpha f(\theta_1) + (1-\alpha) f(\theta_2)$$

for θ_1, θ_2 and $\alpha = [0,1]$

$$f(\alpha\theta_1 + (1-\alpha)\theta_2) = \max_{i=1..N} |\alpha\theta_1 + (1-\alpha)\theta_2|$$

$$\leq \max_{i=1..N} |\alpha\theta_1| + \max_{i=1..N} |(1-\alpha)\theta_2|$$

[Property of modulus]

$$= \alpha \max_{i=1..N} |\theta_1| + (1-\alpha) \max_{i=1..N} |\theta_2|$$

(~~sine~~ Homogeneity Property of Norm)

$$= \alpha L_\infty(\theta_1) + (1-\alpha) L_\infty(\theta_2)$$

$$L_\infty(\alpha\theta_1 + (1-\alpha)\theta_2) \leq \alpha L_\infty(\theta_1) + (1-\alpha) L_\infty(\theta_2)$$

Hence Proved. L_∞ is convex.

(e) $f(x) = x^T A x$, $A \in R^{m \cdot n}$ and is PSD.

$$x^T A x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \\ a_{n1} & & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\therefore x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Now $$\frac{\partial^2 f(x)}{\partial x} = \frac{\partial^2}{\partial x}\left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \right)$$

First calculating $\frac{\partial f(x)}{\partial x}$,

$$\frac{\partial f(x)}{\partial x} = a_{ij} \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \right)$$

$$= a_{ij}\left(\sum_{j=1}^{n} x_i + \sum_{j=1}^{n} x_j\right)$$

$$= \sum_{j=1}^{n} a_{ij} x_i + \sum_{j=1}^{n} a_{ij} x_j$$

$$= A^T X + A\cancel{X}$$

$$\frac{\partial f(x)}{\partial x} = (A + A^T)\cancel{X}$$

Now, $H = \frac{\partial}{\partial x}\left(\frac{\partial f(x)}{\partial x}\right) = \frac{\partial}{\partial x}\left((A + A^T)X\right)$

$$H = A + A^T$$

Now, A is PSD so $A \geq 0$. Also, the transpose of a PSD matrix will also be PSD $\therefore A^T \geq 0$ as well.

Hence, $H \geq 0$, $\therefore$ $f(x) = x^T A x$ is a convex function.