Northeastern University
College of Engineering
Group member name(s): Yanming Liu
Group member UID(s): 002199402

June 16, 2022
Homework 2

IE 7374: Machine Learning

Answer to Question 1

For $A \in R^{m \cdot n}$, considering that $||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$, show that $||A||_F^2 = Trace(A^T A)$

$$Trace(A^T A) = tr \begin{bmatrix} a_1^T a_1 & & \\ & \ddots & \\ & & a_n^T a_n \end{bmatrix} = \sum_{i=1}^{n} a_i^T a_i = \sum_{i=1}^{n} [a_{i1} \dots a_{in}] \begin{bmatrix} a_{i1} \\ \vdots \\ a_{n1} \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$$

Then, $Trace(A^T A) = ||A||_F^2$

Answer to Question 2

A. Prove that L_2 norm is a norm.

To prove it is a norm, we need confirm the following four properties:

1) **Positivity** $||x||_2 >= 0, x \in \mathbb{R}^n$

$$L_2 = ||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} >= 0 \qquad (1)$$

2) $||x||_2 = 0$, iff $x = 0$
From the formula (1), we can easily tell that if $||x||_2 = 0$, iff $x_i^2 = 0$, where $i = 1, \dots, n \rightarrow \mathrm{x}_i = 0$. That is $x = 0$.

3) **Homogeneity**, $||cx||_2 = |c|||x||_2$

$$||cx||_2 = \sqrt{\sum_{i=1}^{n} c^2 x_i^2} = \sqrt{c^2 \sum_{i=1}^{n} x_i^2} = |c| \sqrt{\sum_{i=1}^{n} x_i^2} = |c|||x||_2$$

4) **Triangle inequality**, $||x + y||_2 <= ||x||_2 + ||y||_2$

$$\begin{aligned} ||x+y||_2^2 &= \sum_{i=1}^{n} (x_i + y_i)^2 \\ &= \sum_{i=1}^{n} (x_i^2 + y_i^2 + 2x_i y_i) \\ &= \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} 2x_i y_i \\ &= ||x||_2^2 + ||y||_2^2 + 2xy \\ &<= ||x||_2^2 + ||y||_2^2 + 2|x||y| \\ &= (||x||_2 + ||y||_2)^2 \end{aligned}$$

That means $||x + y||_2 <= ||x||_2 + ||y||_2$

B. Prove that L_∞ is a norm.

To prove it is a norm, we need confirm the following four properties:

1) **Positivity** $||x||_\infty >= 0, x \in \mathbb{R}^n$

$$L_\infty = ||x||_\infty = \max_i |x_i| >= 0$$

2) $||x||_\infty = 0$, iff $x = 0$
From the formula (1), we can easily tell that if $||x||_\infty = 0$, iff $|x_i| = 0$, where $i = 1, \ldots, n \rightarrow$x$_i = 0$. That is $x = 0$.

3) **Homogeneity**, $||cx||_\infty = |c|||x||_\infty$

$$||cx||_\infty = \max_i |cx_i| = |c| \max_i |x_i| = |c|||x||_\infty$$

4) **Triangle inequality**, $||x + y||_\infty <= ||x||_\infty + ||y||_\infty$

$$\begin{aligned} ||x + y||_\infty &= \max_i |x_i + y_i| \\ &<= \max_i (|x_i| + |y_i|) \\ &= \max_i |x_i| + \max_i |y_i| \\ &= ||x||_\infty + ||y||_\infty \end{aligned}$$

Answer to Question 3

Prove that for $A \in \mathbb{R}^{m \cdot n}$ and $B \in \mathbb{R}^{n \cdot m}$; $Trace(AB) = Trace(BA)$

$$\begin{aligned} Trace(AB) &= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji} \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} B_{ji} A_{ij} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} \\ &= \sum_{j=1}^{n} (BA)_{jj} \\ &= Trace(BA) \end{aligned}$$

Answer to Question 4

For any $A \in R^{n \cdot n}$
1) Prove $x \in R^n$, $\frac{\partial x^T A x}{\partial x} = (A + A^T)x$

Assume that $y = x^T A x$, then

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$

Next, we write $x^T A x$ explicitly

$$
\begin{aligned}
x^T A x &= \sum_{i=1}^{n} x_i (Ax)_i \\
&= \sum_{i=1}^{n} x_i (\sum_{j=1}^{n} A_{ij} x_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \\
&= \sum_{i=2}^{n} \sum_{j=2}^{n} A_{ij} x_i x_j + \\
&\quad \sum_{j=1}^{n} A_{1j} x_1 x_j + \sum_{i=1}^{n} A_{i1} x_i x_1 \\
&= \sum_{i=2}^{n} \sum_{j=2}^{n} A_{ij} x_i x_j + 2 \sum_{j=1}^{n} A_{1j} x_1 x_j
\end{aligned}
$$

The reason for the countdown from the first step to the last step is $A_{ij} = A_{ji}$. We calculate the derivative of y with respect to x_1,

$$\frac{\partial y}{\partial x_1} = 2 \sum_{j=1}^{n} A_{1j} x_j$$

Similarly,

$$\frac{\partial y}{\partial x} = \begin{bmatrix} 2\sum_{j=1}^{n} A_{1j} x_j \\ \vdots \\ 2\sum_{j=1}^{n} A_{nj} x_j \end{bmatrix} = 2 \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 2Ax$$

We can easily tell that $2A = A + A^T$, for A is symmetric, $A_{ij} = A_{ji}$

$$2 \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix} = \begin{bmatrix} A_{11} + A_{11} & \dots & A_{1n} + A_{1n} \\ \vdots & & \vdots \\ A_{n1} + A_{n1} & \dots & A_{nn} + A_{nn} \end{bmatrix} = \begin{bmatrix} A_{11} + A_{11} & \dots & A_{1n} + A_{n1} \\ \vdots & & \vdots \\ A_{n1} + A_{1n} & \dots & A_{nn} + A_{nn} \end{bmatrix}$$

2) if $X \in R^{n \cdot n}$ then $\frac{\partial Trace(A^T X)}{\partial X} = A$

$$Trace(A^T X) = tr \begin{bmatrix} a_1^T x_1 & & \\ & \ddots & \\ & & a_n^T x_n \end{bmatrix} = \sum_{i=1}^{n} a_i^T x_i$$

$a_i, x_i \in \mathbb{R}^n$, so the deviate of above function with respect to x_{ij}

$$\frac{\partial Trace(A^T X)}{\partial x_{ij}} = a_{ij},\ i.e.\ \frac{\partial Trace(A^T X)}{\partial X} = A$$

Answer to Question 5

For each equation prove if it is convex: (either using linear algebra or prove by points)

A. $f(x) = exp(ax)$

Prove: $f''(x) = a^2 exp(ax) >= 0$, thus $f(x)$ is a convex.

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

Prove: $f''(x) = \frac{1}{x^2} > 0$, also a convex.

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

Prove: $f''(x) = e^{g(x)}[g''(x)+(g'(x))^2] > 0$, as $g''(x) >= 0$ for $g(x)$ is a convex, and $e^{g(x)}, (g'(x))^2 > 0$

D. L_∞

Prove: By the triangle inequality and homogeneity of the norm, for any $x_i, y_i \in \mathbb{R}$ and any $\theta \in (0, 1)$,

$$L_\infty = \max_i |\theta x_i + (1-\theta)y_i| <= \max_i |\theta x_i| + \max_i |(1-\theta)y_i| = \theta \max_i |x_i| + (1-\theta) \max_i |y_i|$$

Therefore, L_∞ is convex.

E. $f(x) = x^T Ax, A \in \mathbb{R}^n$ and is PSD

Prove: $f''(x) = 2A >= 0$, for A is PSD.Theroefore, $f(x)$ is convex.