

## Norms of vectors :-

Notions of size and distance in a vector space are described by norms (eg:- approximations, convergence, similarity of graphs, images etc are measured by norms)

\* A norm is real-valued function defined on vector space  $C(\mathbb{R}^m \rightarrow \mathbb{R})$  that is commonly denoted as  $x \rightarrow \|x\|$  and satisfies the following properties

- ① Non negative i.e  $\|x\| \geq 0$  ✓
- ② Positive on non-zero vectors i.e  $\|x\| = 0 \iff x = 0$  ✓
- ③ For every vector  $x$  and scalar  $\alpha$ , one has  $\|\alpha x\| = |\alpha| \|x\|$
- ④ Triangle inequality  $\|x+y\| \leq \|x\| + \|y\|$

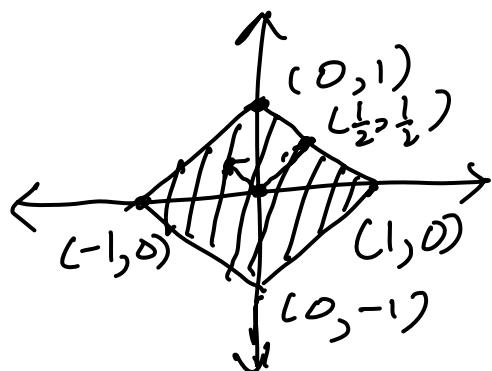
Most important vector norms are

p-norms :-

$$\textcircled{1} \quad \|\underline{x}\|_1 = \sum_{i=1}^m |x_i|$$

$$\underline{x} \in \mathbb{R}^m \quad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\mathbb{R}^2$$

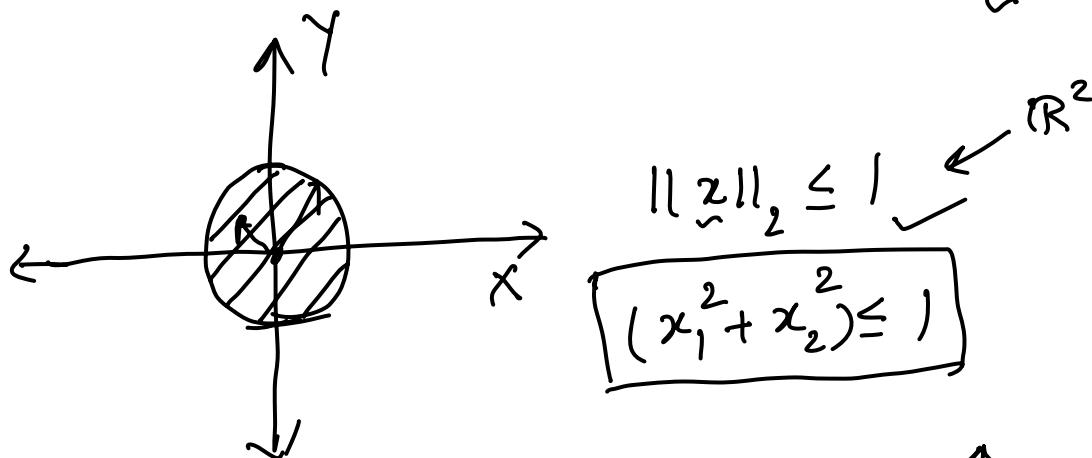


$$\|\underline{x}\|_1 \leq 1$$

$$(|x_1| + |x_2|) \leq 1$$

$$\textcircled{2} \quad \|\underline{x}\|_2 = \left( \sum_{i=1}^m |x_i|^2 \right)^{1/2} = \sqrt{\underline{x}^T \underline{x}}$$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

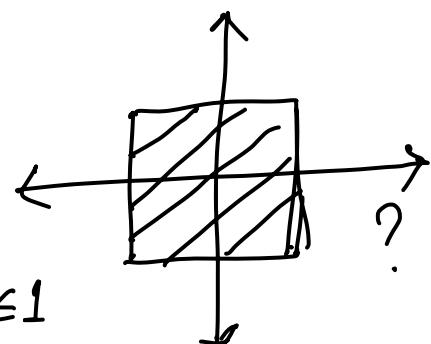


$$\|\underline{x}\|_2 \leq 1$$

$$(x_1^2 + x_2^2) \leq 1$$

$$\textcircled{3} \quad \|\underline{x}\|_\infty = \max_{1 \leq i \leq m} |x_i|$$

$$\|\underline{x}\|_\infty \leq 1$$



$$(4) \quad \|x\|_p = \left( \sum_{i=1}^m |x_i|^p \right)^{1/p} \quad 1 \leq p < \infty$$

Note: As  $p$  tends to infinity, one can show  $p$ -norm definition in (4) tends to the infinity-norm definition in (3)

Relevance :-

① 1-norm / 2-norm :- Often in regression to prevent overfitting, the model parameters are bounded by certain values. Eg:- LASSO regression, constraint is on the  $L_1$  norm or  $L_2$  norm of the model parameter vector

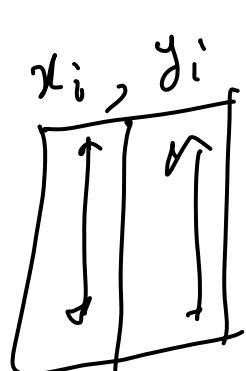
Eg:-  $(x_i, y_i) \quad i=1 \text{ to } m$  data pts

Model:- 
$$y_m = \omega_0 + \omega_1 x$$

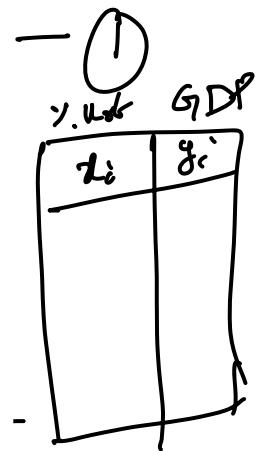
Model parameters :-  $\omega_0, \omega_1$

Least Square minimization

is as follows:-



$$\text{i.e } E = \sum_{i=1}^m (y_m(x_i) - \hat{y}_i)^2$$



$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\underline{x} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

$$\underline{\omega} = \begin{bmatrix} \omega_0 \\ \omega_1 \end{bmatrix}$$

$$\underline{\varepsilon} = \underline{y} - \underline{x} \underline{\omega} \quad (\text{Define error vector})$$

$m \times 1 \quad m \times 1 \quad m \times 2 \quad 2 \times 1$

$\|\underline{\varepsilon}\|_2^2$  is minimized

$$\|\underline{\varepsilon}\|_2^2$$

Hence  $E = \|\underline{y} - \underline{x} \underline{\omega}\|_2^2$   
Now in order to prevent overfitting

the modified objective function

$$\text{if } E = \|\underline{y} - \underline{x} \underline{\omega}\|_2^2 + \lambda \|\underline{\omega}\|_1$$

Puts constraint  
on model  
parameters

$$E = \|\underline{y} - \underline{x} \underline{\omega}\|_2^2 + \lambda \|\underline{\omega}\|_2$$

or

CLASS  
Regression

CRidge  
Regression)

# Matrix Norms induced by

## vector norms:-

Induced matrix norms are defined in terms of the behaviour of a matrix as an operator between its normed domain and range space.

i.e  $A \in \mathbb{R}^{m \times n}$

→ Now a norm on the domain of  $A$  can be represented as  $\|\cdot\|^{(n)}$  (since  $A$  acts on vectors in  $\mathbb{R}^n$ )

→ A norm on the range of  $A$  would be  $\|\cdot\|^{(m)}$  (since  $Ax = y \in \mathbb{R}^m$ )

\* The induced matrix norm is defined as the maximum factor by which  $A$  can stretch any vector  $x$  (relative stretch)

Denoting  $\|A\|$  as the matrix norm induced by  $\|\cdot\|^{(n)}$  and  $\|\cdot\|^{(m)}$

$$\|A\|_p^{(m,n)} = \max_{\substack{\underline{x} \in \mathbb{R}^n \\ \underline{x} \neq 0}} \frac{\|\underline{Ax}\|_p^{(m)}}{\|\underline{x}\|_p^{(n)}} \quad - \textcircled{1}$$

Denote  $\|A\|_p^{(m,n)}$  by some +ve scalar  $C$ ,

then

$$\frac{\|\underline{Ax}\|_p^{(m)}}{\|\underline{x}\|_p^{(n)}} \leq C$$

$$\text{i.e. } \|\underline{Ax}\|_p^{(m)} \leq C \|\underline{x}\|_p^{(n)}$$

' $C$ ' is smallest number which satisfies above inequality and there exists  $\underline{x}$  such that this inequality becomes equality.

In a more convenient way

$$\|A\|^{(m,n)} = \max_{\substack{\bar{x} \in \mathbb{R}^n \\ \|\bar{x}\|=1}} \|\underline{A}\bar{x}\|^{(m)} \quad - \textcircled{2}$$

Why? → For any vector  $\underline{x} \in \mathbb{R}^n$ , define a

unit vector

$$\bar{x} = \frac{\underline{x}}{\|\underline{x}\|^{(n)}}$$

$$\rightarrow \underline{x} = \bar{x} \|\underline{x}\|^{(n)}$$

$$\text{then } \|A\|^{(m,n)} = \max_{\substack{\underline{x} \in \mathbb{R}^n \\ \underline{x} \neq 0}} \frac{\|\underline{Ax}\|^{(m)}}{\|\underline{x}\|^{(n)}}$$

$$= \max_{\substack{\underline{x} \in \mathbb{R}^n \\ \underline{x} \neq 0}} \frac{\|\underline{A}\underline{x}\|^{C^n} \|\underline{x}\|^{C^m}}{\|\underline{x}\|^{C^n} \|\underline{x}\|^{C^m}}$$

$$= \max_{\substack{\underline{x} \in \mathbb{R}^n \\ \underline{x} \neq 0}} \frac{\|\underline{x}\|^{C^n} \|\underline{A}\underline{x}\|^{C^m}}{\|\underline{x}\|^{C^n} \|\underline{x}\|^{C^m}}$$

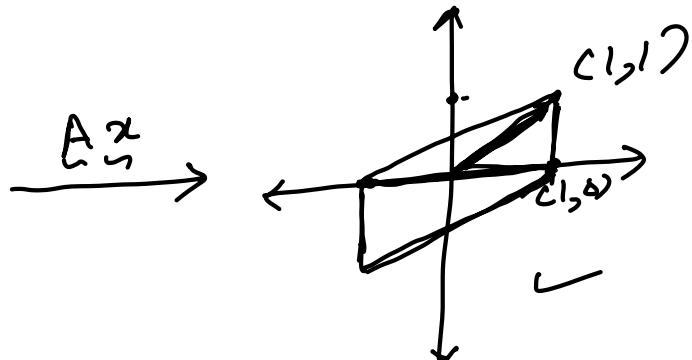
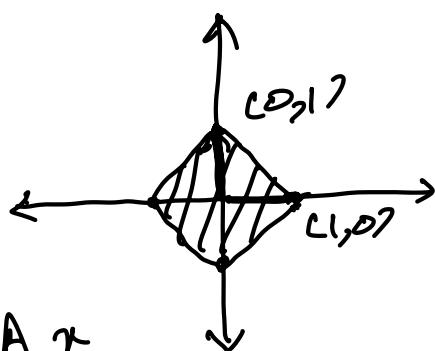
$$= \max_{\underline{x} \in \mathbb{R}^n} \|\underline{A}\underline{x}\|^{C^m}$$

Eq.: Matrix  $\underline{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  maps  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$

$$\|\underline{A}\|^{(2,2)}$$

$$\text{1-norm} \quad \underline{\|\underline{x}\|_1 \leq 1}$$

$$\underline{A}\underline{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$



$$\underline{A}\underline{x} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\underline{A}\underline{x} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

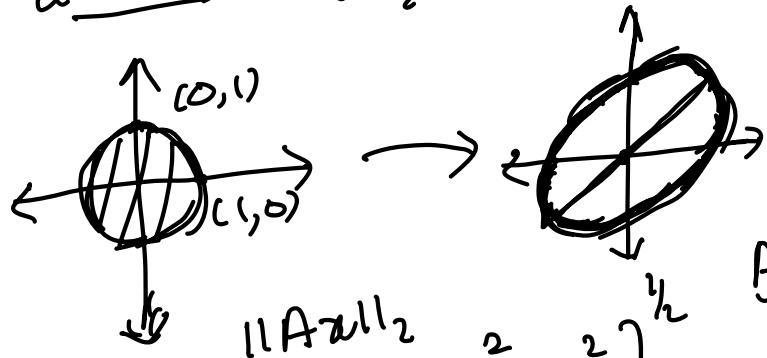
$$\|\underline{A}\underline{x}\|_1 = 2$$

$$\Rightarrow \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\|\underline{A}\|_1^{(2,2)} = ?$$

$$\underline{A}\underline{x} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

2-norm  $\|\underline{x}\|_2 \leq 1$



$$\begin{aligned} \|\underline{Ax}\|_2 &= \sqrt{(\sin\theta + \cos\theta)^2 + (\cos\theta)^2} \\ &= \sqrt{(\sin^2\theta + 2\sin\theta\cos\theta + \cos^2\theta) + \cos^2\theta} \\ &= \sqrt{\sin^2\theta + 2\sin\theta\cos\theta + \cos^2\theta} \end{aligned}$$

Let  $\underline{x} = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}$

$$\|\underline{A}\underline{x}\|_2 = \max \|\underline{A}\underline{x}\|_2$$

$$\underline{A}\underline{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix} = \begin{bmatrix} \sin\theta + \cos\theta \\ \cos\theta \end{bmatrix}$$

$$\max_{\theta} (\underline{f}(\theta)) \text{ same as } \max_{\theta} f(\theta)$$

$$\text{i.e. } \max_{\theta} (\sin^2\theta + \cos^2\theta + 2\sin\theta\cos\theta + \cos^2\theta)$$

$$\Rightarrow \max_{\theta} (1 + 2\sin\theta\cos\theta + \cos^2\theta) \Rightarrow \theta^* = \frac{1}{2} \tan^{-1}(2)$$

$$\Rightarrow \|\underline{A}\underline{x}\|_2 = \sqrt{1 + 2\sin^2\theta\cos^2\theta + \cos^2\theta} = 1.618$$

Eg:- Diagonal matrix p-norm!

$$\underline{D} = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_m \end{bmatrix}$$

$$\underline{p}=1$$

$$\|\bar{\underline{x}}\|_1 = 1$$

$$\bar{\underline{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\|\underline{D}\underline{x}\|_1 = \max_{\bar{\underline{x}} \in \mathbb{R}^m} \|\underline{D}\bar{\underline{x}}\|_1$$

$$\|\bar{\underline{x}}\|_1 = 1$$

$$\sum_i |x_i| = 1$$

$$\max \sum_{i=1}^m |d_i x_i|$$

$$\|\bar{\underline{x}}\|_1 = 1$$

Above quantity has a max value

$$\|\underline{D}\underline{x}\|_1 = \max_{1 \leq i \leq m} |d_i|$$

$$\|D\|_2 = \max_{\|\bar{x}\|_2=1} \|D\bar{x}\|_2$$

$$\max (\dfrac{d_1^2 x_1^2 + d_2^2 x_2^2 + \dots + d_n^2 x_n^2}{x_1^2 + x_2^2 + \dots + x_n^2})^{1/2}$$

$$\|D\bar{x}\|_2 = 1$$

$$\|\bar{x}\|_2 = 1$$

Again above quantity has a maximum value  $\max_{1 \leq i \leq m} |d_i|$

$$\boxed{\|D\|_p = \max_{1 \leq i \leq m} |d_i|}$$

Result (a) :-

1-norm of a matrix

If  $A \in \mathbb{R}^{m \times n}$ , then  $\|A\|_1$  is equal to maximum column sum of  $A$ .

Proof :-  $A = \left[ \begin{array}{c|c|c|c} a_1 & a_2 & \dots & a_n \end{array} \right]$

where  $a_j \in \mathbb{R}^m$ . Consider the action of  $A$  on some unit vector  $\underline{x}$  i.e  $\|\underline{x}\|_1 = 1$

$$\|A\underline{x}\|_1 = \left\| \sum_{j=1}^n x_j a_j \right\|_1 \leq \sum_{j=1}^n \|x_j a_j\|_1$$

(Triangle inequality)

$$\|A\bar{x}\|_1 \leq \sum_{j=1}^n |x_j| \|a_j\|_1 = \sum_{j=1}^n |x_j| \|a_j\|_1$$

Let  $j_{\max}$  be the column for which  $\|a_j\|_1$  is maximum for  $j = 1, 2, \dots, n$

$$\begin{aligned} \|A\bar{x}\|_1 &\leq \sum_{j=1}^n |x_j| \|a_j\|_1 \leq \sum_{j=1}^n |x_j| \|a_{j_{\max}}\|_1 \\ &= \|a_{j_{\max}}\|_1 \left[ \sum_{j=1}^n |x_j| \right] \end{aligned}$$

We know  $\|\bar{x}\|_1 = 1$

$$\|A\bar{x}\|_1 \leq \|a_{j_{\max}}\|_1$$

By choosing  $\bar{x} = e_{j_{\max}}$ , we achieve this bound i.e

$$\|Ae_{j_{\max}}\|_1 = \|a_{j_{\max}}\|_1$$

$$\text{Hence } \|A\|_1 = \max_{1 \leq j \leq n} \|a_j\|_1$$

Result (b) :- The  $\infty$ -norm of a matrix is maximum row sum

Pf:- exercise

$$\text{eg:- } A = \begin{pmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{pmatrix}$$

$$\|A\|_1 = 9 \quad \|A\|_\infty = 15$$

Result :-

$p \neq 1, \infty$ , we need other methods to evaluate matrix-norms.

Hölder inequality :-

If  $p$  and  $q$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1$  with  $1 \leq p < \infty$ , Hölder inequality states that for any vectors  $x$  and  $y$

$$|x^T y| \leq \|x\|_p \|y\|_q$$

If  $p = q = 2$ , we get Cauchy-Schwarz inequality  $|x^T y| \leq \|x\|_2 \|y\|_2$  —  $\checkmark$

Eg:- Let  $u \in \mathbb{R}^m$ ,  $v \in \mathbb{R}^n$  and  $A = u v^T$ , evaluate  $\|A\|_2$

$$\|A x\|_2 = \|u v^T x\|_2 = \|v^T x\|_2 \|u\|_2$$

Use Cauchy-Schwarz

$$\begin{aligned} \|A x\|_2 &= \|v^T x\|_2 \|u\|_2 \leq \|u\|_2 \|v\|_2 \|x\|_2 \\ &= \|u\|_2 \|v\|_2 \end{aligned}$$

$$\|A x\|_2 \leq \|u\|_2 \|v\|_2$$

This is a tight bound because the

above inequality becomes equality when

$$\underline{x} = \frac{\underline{v}}{\|\underline{v}\|}$$

Hence  $\boxed{\|\underline{A}\underline{U}\|_2^{(m,n)} = \|\underline{U}\|_2 \|\underline{V}\|_2}$

Result (d)

$$\underline{A} \in \mathbb{R}^{L \times m}, \quad \underline{B} \in \mathbb{R}^{m \times n}$$

$$\|\underline{A}\underline{B}\| \leq \|\underline{A}\| \|\underline{B}\|$$

Pf:-  $\|\underline{A}\underline{B}\underline{x}\| \stackrel{(1)}{\leq} \|\underline{A}\| \|\underline{B}\underline{x}\| \stackrel{(2)}{\leq} \|\underline{A}\| \|\underline{B}\| \|\underline{x}\| \rightarrow 1$

Since we are considering all possible unit-vectors

$\|\underline{x}\|$ , we have  $\|\underline{A}\underline{B}\underline{x}\| \leq \|\underline{A}\| \|\underline{B}\| + \underline{x} \in \mathbb{R}$  ①

In general we cannot find  $\underline{x}$  which makes the above inequality an equality.

But we know

$$\|\underline{A}\underline{B}\underline{x}\| \stackrel{(1)}{\leq} \|\underline{A}\underline{B}\| \stackrel{(2)}{\leq} \|\underline{A}\| \|\underline{B}\|$$

- ②

and  $\|\underline{A}\underline{B}\|$  is a tight bound i.e. one can find  $\underline{x}$  where this bound becomes equality

Let that  $\underline{z}$  be  $\underline{z}$

$$\|\underline{AB}\hat{\underline{z}}\|^{(l)} = \|\underline{AB}\|^{(l,m)} \quad - \textcircled{3}$$

From  $\textcircled{3}$  and  $\textcircled{1}$ , we can say that

$$\|\underline{AB}\|^{(l,m)} \leq \|\underline{A}\|^{(l,m)} \|\underline{B}\|^{(m,n)}$$

Hilbert-Schmidt Norm (Frobenius norm)

$$\underline{A} \in \mathbb{R}^{m \times n}; \|\underline{A}\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}$$

This is the 2-norm of matrix when viewed as  $mn$ -dimensional vector

$$\|\underline{A}\|_F = \left( \sum_{j=1}^n \|\underline{a}_j\|_2^2 \right)^{1/2}$$

$$\|\underline{A}\|_F = \sqrt{\text{tr}(\underline{A}^T \underline{A})} = \sqrt{\text{tr}(\underline{A} \underline{A}^T)}$$

Recall  $\text{tr}(X)$  is a sum of diagonal entries of matrix  $X$

Result e.i:-

$$\|\underline{AB}\|_F^2 \leq \|\underline{A}\|_F^2 \|\underline{B}\|_F^2$$

$$\underline{A} \in \mathbb{R}^{n \times k}_{k \times m}$$

$$\text{Pf:-} \quad \text{Let } \underline{C} = \underline{A} \underline{B} \quad \underline{B} \in \mathbb{R}^{k \times m}$$

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

$$= \underline{a}_i^T \underline{b}_j \quad \text{--- (4)}$$

$\underline{a}_i^T$  is row vector of  $\underline{A}$  and  $\underline{b}_j$  is the column vector of  $\underline{B}$

$$\begin{aligned} \|C\|_F^2 &= \|AB\|_F^2 \\ &= \sum_{i=1}^n \sum_{j=1}^m |C_{ij}|^2 \\ &= \sum_{i=1}^n \sum_{j=1}^m (\underline{a}_i^T \underline{b}_j)^2 \end{aligned}$$

Using Cauchy-Schwarz inequality

$$\begin{aligned} \|C\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^m |\underline{a}_i^T \underline{b}_j|^2 \leq \sum_{i=1}^n \sum_{j=1}^m \|\underline{a}_i\|_2^2 \|\underline{b}_j\|_2^2 \\ &= \underbrace{\sum_{i=1}^n \|\underline{a}_i\|_2^2}_{\|A\|_F^2} \underbrace{\sum_{j=1}^m \|\underline{b}_j\|_2^2}_{\|B\|_F^2} \end{aligned}$$

$$= \|A\|_F^2 \|B\|_F^2$$

$$\boxed{\|AB\|_F^2 \leq \|A\|_F^2 \|B\|_F^2}$$

Result :-

For any  $A \in \mathbb{R}^{m \times n}$  and orthogonal matrix  $Q \in \mathbb{R}^{m \times m}$ , we have  $\|QA\|_2 = \|A\|_2$  and  $\|QA\|_F = \|A\|_F$

i.e. 2-norm and Frobenius norm is  
invariant under multiplication by  
orthogonal matrix.

Proof:- Exercise!