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 ($\mathbb{R}^m \to \mathbb{R}$) that is commonly denoted as $x \to \|x\|$ and satisfies the following properties

① Nonnegative 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 α, one has $\|\alpha x\| = |\alpha| \, \|x\|$

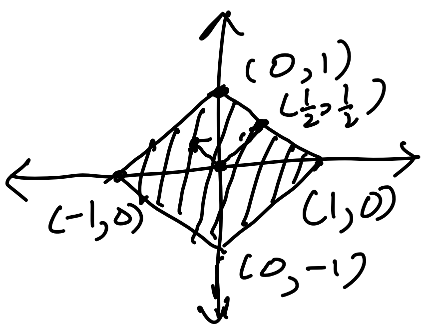
④ Triangle inequality $\|x + y\| \leq \|x\| + \|y\|$ ✓

Most important vector norms are p-norms :-

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

① $\|\underset{\sim}{x}\|_1 = \sum_{i=1}^{m} |x_i|$

$\mathbb{R}^2$

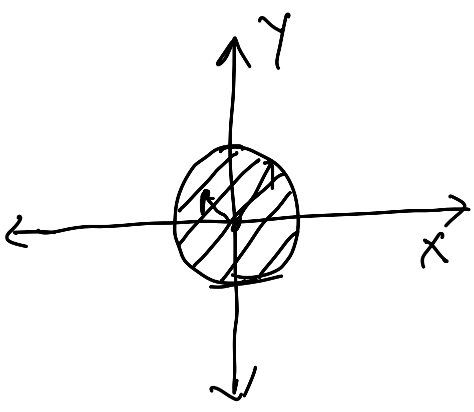


$\boxed{\|\underset{\sim}{x}\|_1 \le 1}$

$\boxed{(|x_1|+|x_2|) \le 1}$ ✓

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

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

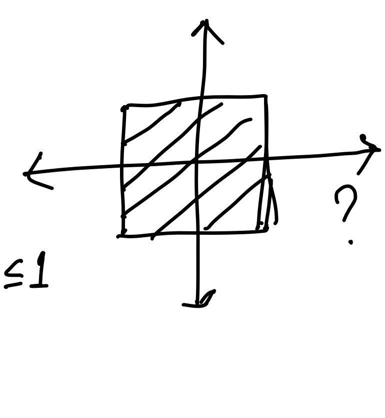


$\mathbb{R}^2$

$\|\underset{\sim}{x}\|_2 \le 1$ ✓

$\boxed{(x_1^2 + x_2^2) \le 1}$

③ $\|\underset{\sim}{x}\|_\infty = \max_{1 \le i \le m} |x_i|$



?

$\|x\|_\infty \le 1$

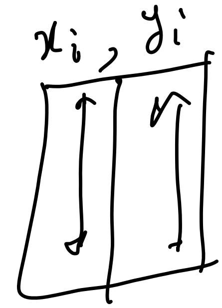
④ $\underline{\|\underset{\sim}{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p}}$ $\quad 1 \le p < \infty$

<u>Relevance</u> :-

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 model parameter vector

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

<u>Model</u> :- $\boxed{y_m = \underline{\underline{w_0}} + \underline{w_1}\, x}$ ✓



<u>Model parameters</u> :- w_0, w_1

Least Square minimization is as follows :-

i.e $$E = \sum_{i=1}^{m} \left(\underline{y_m(x_i)} - \underline{y_i} \right)^2 \quad —(1)$$

% ub	GDP
x_i	y_i

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad \underset{\sim}{X} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{m \times 2}$$

$$\underset{\sim}{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\downarrow$$

$$\underset{m \times 1}{\underset{\sim}{\varepsilon}} = \underset{m \times 1}{\underset{\sim}{y}} - \underbrace{\underset{m \times 2}{\underset{\sim}{X}} \; \underset{2 \times 1}{\underset{\sim}{w}}}$$

$\|\underset{\sim}{\varepsilon}\|_2^2$ is minimized

Hence $E = \|\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{w}\|_2^2$

Now in order to prevent overfitting the modified objective function is

$$E = \boxed{\|\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{w}\|_2^2} + \boxed{\lambda \|\underset{\sim}{w}\|_1}$$

(LASSO Regression)

or

$$E = \|\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{w}\|_2^2 + \boxed{\lambda \|\underset{\sim}{w}\|_2}$$

(Ridge 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 $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$

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

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

* The induced matrix norm is defined as the maximum factor by which $\underset{\sim}{A}$ can stretch any vector $\underset{\sim}{x}$

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

$$\boxed{\|A\|_p^{(m,n)} = \max_{\substack{\underset{\sim}{x} \in \mathbb{R}^n \\ \underset{\sim}{x} \neq \underset{\sim}{0}}} \frac{\|\underset{\sim}{A}\underset{\sim}{x}\|_p^{(m)}}{\|\underset{\sim}{x}\|_p^{(n)}}} \quad \text{— (1)}$$

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

then $\boxed{\dfrac{\|\underset{\sim}{A}\underset{\sim}{x}\|_p^{(m)}}{\|x\|_p^{(n)}} \leq C}$

i.e $\|\underset{\sim}{A}\underset{\sim}{x}\|_p^{(m)} \leq C\|x\|_p^{(n)}$

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

In a more convenient way

$$\boxed{\|\underset{\sim}{A}\|^{(m,n)} = \max_{\substack{\bar{\underset{\sim}{x}} \in \mathbb{R}^n \\ \|\bar{\underset{\sim}{x}}\| = 1}} \|\underset{\sim}{A}\bar{\underset{\sim}{x}}\|^{(m)}} \quad \text{— (2)}$$

Why? ↗ For any vector $\underset{\sim}{x} \in \mathbb{R}^n$, define a unit vector $\boxed{\bar{\underset{\sim}{x}} = \dfrac{\underset{\sim}{x}}{\|x\|^{(n)}}} \longrightarrow \underset{\sim}{x} = \bar{\underset{\sim}{x}}\,\|x\|^{(n)}$

then $\|A\|^{(m,n)} = \max_{\substack{\underset{\sim}{x} \in \mathbb{R}^n \\ \underset{\sim}{x} \neq \underset{\sim}{0}}} \dfrac{\|\underset{\sim}{A}\underset{\sim}{x}\|^{(m)}}{\|\underset{\sim}{x}\|^{(n)}}$

$$= \max_{\substack{\underset{\sim}{x} \in \mathbb{R}^n \\ \underset{\sim}{x} \neq \underset{\sim}{0}}} \frac{\left\| \underset{\sim}{A} \frac{\underset{\sim}{x}}{\|\underset{\sim}{x}\|^{(n)}} \right\|^{(m)}}{\left\| \frac{\underset{\sim}{x}}{\|\underset{\sim}{x}\|^{(n)}} \right\|^{(n)}}$$

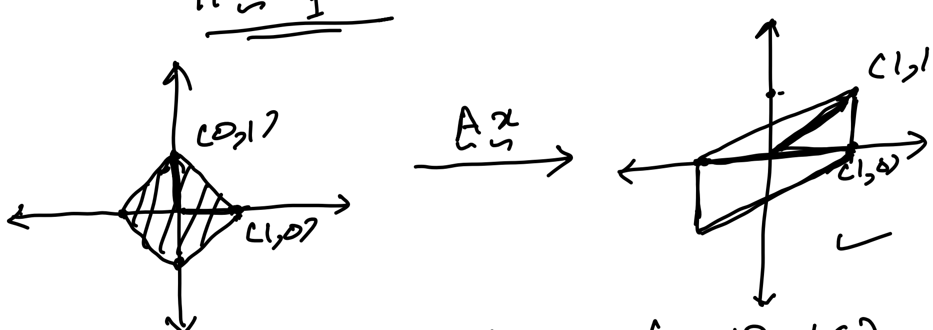
$$= \max_{\substack{\underset{\sim}{x} \in \mathbb{R}^n \\ \underset{\sim}{x} \neq 0}} \frac{\|\underset{\sim}{x}\|^{(n)} \, \|\underset{\sim}{A} \bar{x}\|^{(m)}}{\|\underset{\sim}{x}\|^{(n)} \, \|\bar{x}\|^{(n)}}$$

$$= \max_{\bar{x} \in \mathbb{R}^n} \|\underset{\sim}{A} \bar{x}\|^{(m)}$$

Eg:- Matrix $\underset{\sim}{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ maps $\mathbb{R}^2 \to \mathbb{R}^2$

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

1-norm $\quad \|\underset{\sim}{x}\|_1 \leq 1 \qquad \underset{\sim}{A}\underset{\sim}{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$



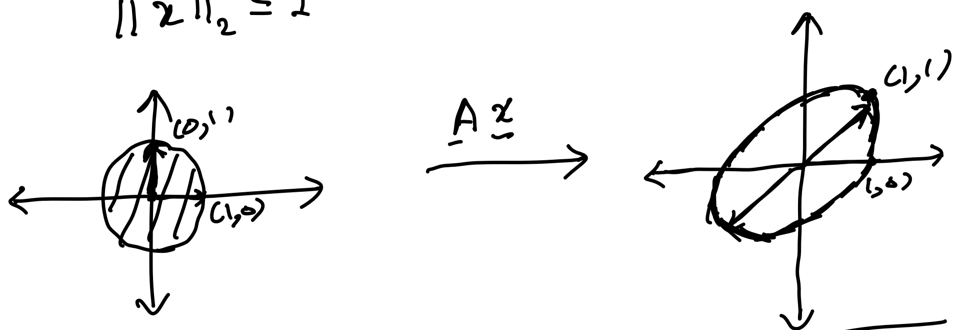
$$\underset{\sim}{A}\underset{\sim}{x} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_1 = 2$$

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

2- norm

$$\|x\|_2 \le 1$$



$$\|A\|_2 = \sqrt{1^2+1^2} = \sqrt{2} = 1.414$$

∞ - norm

Eg:- Diagonal matrix p- norm:-

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

$\underline{p=1}$ $\quad \boxed{\|\bar{x}\|_1 = 1}$ $\quad \bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$

$$\|D\|_1 = \max_{\substack{\bar{x} \in \mathbb{R}^m \\ \|\bar{x}\|_1 = 1}} \|D\bar{x}\|_1 \qquad \boxed{\sum_i |x_i| = 1}$$

$$\max_{\|\bar{x}\|_1 = 1} \sum_{i=1}^{m} |d_i x_i|$$

Above quantity has a max value

$$\boxed{\|D\|_1 = \max_{1 \le i \le m} |d_i|}$$

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

$$\max_{\|\bar{x}\|_2=1} (d_1^2 x_1^2 + d_2^2 x_2^2 + \ldots d_n^2 x_n^2)^{1/2} \quad \|\bar{x}\|_2 = 1$$

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

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

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:- $A = \left[a_1 \,\middle|\, a_2 \,\middle|\, \cdots \,\middle|\, a_n \right]$

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

$$\|Ax\|_1 = \left\|\sum_{j=1}^{n} x_j a_j\right\|_1 \le \sum_{j=1}^{n} \|x_j a_j\|_1$$

(Triangle inequality)

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_1 \le \sum_{j=1}^{n} \|x_j \underset{\sim}{a_j}\|_1 = \sum_{j=1}^{n} |x_j| \, \|\underset{\sim}{a_j}\|_1$$

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

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_1 \le \sum_{j=1}^{n} |x_j| \, \|a_j\|_1 \le \sum_{j=1}^{n} |x_j| \, \|a_{j_{max}}\|_1$$

$$= \|a_{j_{max}}\|_1 \boxed{\underbrace{\sum_{j=1}^{n} |x_j|}_{1}}$$

We know $\|\underset{\sim}{x}\|_1 = 1$

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_1 \le \|a_{j_{max}}\|_1$$

By choosing $\underset{\sim}{x} = \underset{\sim}{e}_{j_{max}}$, we achieve this bound i.e

$$\|\underset{\sim}{A}\,\underset{\sim}{e}_{j_{max}}\|_1 = \|a_{j_{max}}\|_1$$

Hence $\|A\|_1^{(m,n)} = \max\limits_{1 \le j \le n} \|a_j\|_1$

Result (b):- The ∞-norm of a matrix is maximum row sum

Pf:- Exercise

Eg:- $\underset{\sim}{A} = \begin{pmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{pmatrix}$

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

Result (c):-

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

Holder inequality:-

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

$$|\underset{\sim}{x}^T \underset{\sim}{y}| \le \|\underset{\sim}{x}\|_p \|\underset{\sim}{y}\|_q$$

If $p=q=2$, we get Cauchy-Schwartz inequality $|\underset{\sim}{x}^T \underset{\sim}{y}| \le \|\underset{\sim}{x}\|_2 \|\underset{\sim}{y}\|_2$ — (*)

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

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_2 = \|\underbrace{\underset{\sim}{u}}_{m\times 1}\underbrace{\underset{\sim}{v}^T}_{1\times n}\underbrace{\underset{\sim}{x}}_{n\times 1}\| = |\underset{\sim}{v}^T\underset{\sim}{x}|\,\|u\|_2$$

Use Cauchy-Schwartz

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_2 = |\underset{\sim}{v}^T\underset{\sim}{x}|\,\|\underset{\sim}{u}\|_2 \le \|\underset{\sim}{u}\|_2\|\underset{\sim}{v}\|_2 \cancelto{1}{\|\underset{\sim}{x}\|_2}$$

$$= \|\underset{\sim}{u}\|_2\|\underset{\sim}{v}\|_2$$

$\|\underset{\sim}{A}\underset{\sim}{x}\|_2 \le \|\underset{\sim}{u}\|_2\|\underset{\sim}{v}\|_2$

This is a tight bound because the

above inequality becomes equality when

$\underset{\sim}{x} = \dfrac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|}$

Hence $\boxed{\|\underset{\sim}{A}\|_2^{(m,n)} = \|u\|_2 \, \|v\|_2}$

<u>Result (d)</u>

$\underset{\sim}{A} \in \mathbb{R}^{l \times m}, \quad \underset{\sim}{B} \in \mathbb{R}^{m \times n}$

$\|\underset{\sim}{A}\underset{\sim}{B}\|^{(l,n)} \leq \|\underset{\sim}{A}\|^{(l,m)} \, \|\underset{\sim}{B}\|^{(m,n)}$

Pf:- $\|\underset{\sim}{A}\underset{\sim}{B}\underset{\sim}{x}\|^{(l)} \leq \|\underset{\sim}{A}\|^{(l,m)} \, \|\underset{\sim}{B x}\|^{(m)}$

$\leq \|\underset{\sim}{A}\|^{(l,m)} \, \|\underset{\sim}{B}\|^{(m,n)} \, \underbrace{\|x\|}^{(n)} \to 1$

Since we are considering all possible unit-vectors $\|x\|$, we have

$\|\underset{\sim}{A}\underset{\sim}{B}\underset{\sim}{x}\|^{(l)} \leq \|\underset{\sim}{A}\|^{(l,m)} \, \|\underset{\sim}{B}\|^{(m,n)} \quad \forall \underset{\sim}{x} \in \mathbb{R}^{l}$ — ①

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

But we know

$\|\underset{\sim}{A}\underset{\sim}{B}\underset{\sim}{x}\|^{(l)} \leq \|\underset{\sim}{A}\underset{\sim}{B}\|^{(l,n)}$ — ②

and $\|AB\|^{(l,n)}$ is a tight bound i.e one can find $\underset{\sim}{x}$ where this bound becomes equality

Let that $\underset{\sim}{x}$ be $\hat{\underset{\sim}{x}}$

$$\|\underset{\sim}{A}\underset{\sim}{B}\hat{\underset{\sim}{x}}\|^{(l)} = \|AB\|^{(l,n)} \quad -(3)$$

From (3) and (1), we can say that

$$\boxed{\|AB\|^{(l,n)} \le \|A\|^{(l,m)}\|B\|^{(m,n)}}$$

Hilbert - Schmidt Norm (Frobenius norm)

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

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

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

$$\|A\|_F = \sqrt{tr(\underset{\sim}{A}^T\underset{\sim}{A})} = \sqrt{tr(\underset{\sim}{A}\underset{\sim}{A}^T)}$$

Recall $tr(X)$ is a sum of diagonal entries of matrix X

Result e:-

$$\|\underset{\sim}{A}\underset{\sim}{B}\|_F^2 \le \|\underset{\sim}{A}\|_F^2 \|\underset{\sim}{B}\|_F^2 \qquad \underset{\sim}{A} \in \mathbb{R}^{n \times k}, \quad \underset{\sim}{B} \in \mathbb{R}^{k \times m}$$

Pf:- Let $\underset{\sim}{C} = \underset{\sim}{A}\underset{\sim}{B}$

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

$$= \underset{\sim}{a}_i^T \underset{\sim}{b}_j \quad \text{— (4)}$$

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

$$\|C\|_F^2 = \|AB\|_F^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} |c_{ij}|^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} |\underset{\sim}{a}_i^T \underset{\sim}{b}_j|^2$$

Using Cauchy-Schwarz inequality

$$\|C\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} |\underset{\sim}{a}_i^T \underset{\sim}{b}_j|^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{m} \|\underset{\sim}{a}_i\|_2^2 \|\underset{\sim}{b}_j\|_2^2$$

$$= \sum_{i=1}^{n} \underbrace{\|\underset{\sim}{a}_i\|_2^2} \sum_{j=1}^{m} \underbrace{\|\underset{\sim}{b}_j\|_2^2}$$

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

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

Result f:

For any $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ and orthogonal matrix $\underset{\sim}{Q} \in \mathbb{R}^{m \times m}$, 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!