Lecture 1 : Linear Algebra :

DS 4400
Instructor : Ehsan Elhamifar

We will work with vectors, matrices and tensors with real numbers.

$$x \in \mathbb{R}^n \longrightarrow x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{where } x_i \in \mathbb{R}$$

$$A \in \mathbb{R}^{m \times n} \longrightarrow A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{where } a_{ij} \in \mathbb{R}$$

$$(A)_{ij} = a_{ij} \quad . \quad (A^T)_{ij} = a_{ji} \quad \text{transpose.} \qquad \text{Ex)} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T = [x_1 \ x_2]$$

$$A = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \quad \text{where } a_i \in \mathbb{R}^m \quad \text{or} \quad A = \begin{pmatrix} - \bar{a}_1^T - \\ - \bar{a}_2^T - \\ \vdots \\ - \bar{a}_m^T - \end{pmatrix} \quad \text{where } \bar{a}_i \in \mathbb{R}^n$$

Inner Products : $x \in \mathbb{R}^n, \ y \in \mathbb{R}^n \Longrightarrow x^T y = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i \ \in \mathbb{R}$

Matrix Multiplication : $A \in \mathbb{R}^{m \times n}, \ B \in \mathbb{R}^{n \times p} \longrightarrow C = AB \in \mathbb{R}^{m \times p}$

$$C = \underbrace{\begin{pmatrix} - \bar{a}_1^T - \\ - \bar{a}_2^T - \\ \vdots \\ - \bar{a}_m^T - \end{pmatrix}}_{m \times n} \underbrace{\begin{pmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{pmatrix}}_{n \times p} = \begin{pmatrix} \bar{a}_1^T b_1 & \cdots & \bar{a}_1^T b_p \\ \vdots & & \vdots \\ \bar{a}_m^T b_1 & \cdots & \bar{a}_m^T b_p \end{pmatrix} \Rightarrow (C)_{ij} = \underbrace{\bar{a}_i^T}_{\text{row } i \text{ of } A} \underbrace{b_j}_{\text{col } j \text{ of } B}$$

$$C = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} - \bar{b}_1^T - \\ - \bar{b}_2^T - \\ \vdots \\ - \bar{b}_n^T - \end{pmatrix} = a_1 \bar{b}_1^T + a_2 \bar{b}_2^T + \cdots + a_n \bar{b}_n^T = \sum_{i=1}^{n} \underbrace{a_i \bar{b}_i^T}_{\text{outer product}}$$

Outer product : $x \in \mathbb{R}^m, \ y \in \mathbb{R}^p$

$$x y^T = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} (y_1 \cdots y_p) = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_p \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_p \end{pmatrix}$$

Ex) System of linear equations :

$$\begin{cases} y_1 = 3x_1 - 5x_2 + x_3 \\ y_2 = -x_1 + 2x_2 + 4x_3 \end{cases}$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} 3 & -5 & 1 \\ -1 & 2 & 4 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{x} \Longrightarrow y = Ax$$

$$y = \begin{pmatrix} \bar{a}_1^T \\ \bar{a}_2^T \end{pmatrix} x = \begin{pmatrix} \bar{a}_1^T x \\ \bar{a}_2^T x \end{pmatrix} \quad \text{or} \quad y = \begin{pmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} | \\ a_1 \\ | \end{pmatrix} x_1 + \begin{pmatrix} | \\ a_2 \\ | \end{pmatrix} x_2 + \begin{pmatrix} | \\ a_3 \\ | \end{pmatrix} x_3$$

* Identity Matrix : $I \in \mathbb{R}^{n\times n}$: $I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$: $AI = IA = A$
* Diagonal Matrix : $D = \text{diag}(d_1, \dots, d_n) \overset{n\times n}{=} \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \in \mathbb{R}^{n\times n}$ $\rightarrow I = \text{diag}(\underbrace{1,1,\dots,1}_{n \text{ times}})$

* For a square matrix $A \in \mathbb{R}^{n\times n}$:

- A is symmetric iff $A^T = A$ and anti-symmetric if $A^T = -A$
- $\text{tr}(A) = \sum_{i=1}^{n} A_{ii}$ = sum of diagonal elements of A

$\rightarrow$ $\text{tr}(A^T) = \text{tr}(A)$, $\text{tr}(\alpha A) = \alpha\, \text{tr}(A)$, $\text{tr}(\underset{}{\overset{n\times n}{A}}\overset{n\times n}{B}) = \text{tr}(BA)$

* Norms : $f: \mathbb{R}^n \to \mathbb{R}$ is a function that satisfies 4 properties :
(1) $\forall x \in \mathbb{R}^n$: $f(x) \geq 0$ (non-negativity)
(2) $f(x) = 0$ if and only if $x = 0$ (definiteness)
(3) $\forall \alpha \in \mathbb{R}$, $\forall x \in \mathbb{R}^n$: $f(\alpha x) = |\alpha| f(x)$ (homogeneity)
(4) $\forall x, y \in \mathbb{R}^n$: $f(x+y) \leq f(x) + f(y)$ (triangle - inequality)

Common norms :
ℓ_2- norm : $\|x\|_2 \triangleq \sqrt{\sum_{i=1}^{n} x_i^2} \longrightarrow \|x\|_2^2 = \sum_{i=1}^{n} x_i^2 = x^T x \rightarrow \|x\|_2 = \sqrt{x^T x}$
ℓ_1 - norm : $\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$
ℓ_∞ -norm : $\|x\|_\infty \triangleq \max_{i=1,\dots,n} |x_i|$

We can also define norm for matrices : $A \in \mathbb{R}^{m\times n}$

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\text{tr}(A^T A)}$$

* Linear independence and Rank :

A set of vectors $\{x_1, \dots, x_n\}$ are linearly independent if $\sum_{i=1}^{n} \alpha_i x_i = 0 \longrightarrow$ only solution is $\alpha_1 = \dots = \alpha_n = 0$.
If $\exists$ a solution $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \neq 0 \longrightarrow$ linearly dependent !
$2x_1 + 3x_2 + 5x_3 = 0 \longrightarrow x_1 = -\frac{1}{2}(3x_1 + 5x_3) \rightarrow$ we can write one as a linear comb. of others.

* The rank of a matrix is the largest number of linearly independent columns (or rows) of A.

$\forall A \in \mathbb{R}^{m \times n}, \forall B \in \mathbb{R}^{n \times p}$:

+ $\text{rank}(A) \leqslant \min(m, n)$
+ $\text{rank}(A) = \text{rank}(A^T)$
+ $\text{rank}(AB) \leqslant \min\{\text{rank}(A), \text{rank}(B)\}$
+ $\text{rank}(A+B) \leqslant \text{rank}(A) + \text{rank}(B)$

* Inverse of A : $A \in \mathbb{R}^{n \times n}$: $A^{-1} \in \mathbb{R}^{n \times n}$ is inverse of A iff $A^{-1}A = AA^{-1} = I_n$

$\longrightarrow$ A is invertible iff A is full rank, $rk(A) = n$ (all columns of A are LI)

A : non-invertible or singular, otherwise.

* $A, B \in \mathbb{R}^{n \times n}$:

$(A^{-1})^{-1} = A$

$(AB)^{-1} = B^{-1}A^{-1}$

$(A^T)^{-1} = (A^{-1})^T$

* Orthogonal and Orthonormal Matrices:

$U \in \mathbb{R}^{n \times n}$ orthogonal iff $U^TU = \begin{pmatrix} * & & 0 \\ & * & \\ 0 & & * \end{pmatrix}$ = diagonal $\begin{pmatrix} -u_1^T- \\ \vdots \\ -u_n^T- \end{pmatrix} \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} * & & 0 \\ & * & \\ 0 & & \ddots * \end{pmatrix}$

$\longrightarrow$ $u_i^Tu_j = 0 \ \forall i \neq j \Rightarrow$ distinct columns are orthogonal to each other.

$U^TU = I_n \longrightarrow U$: orthonormal $\longrightarrow$ $u_i^Tu_i = 1 \longrightarrow \|u_i\|_2 = 1 \longrightarrow$ length of each col is 1.

$\longrightarrow$ $u_i^Tu_j = 0 \quad \forall i \neq j$

$\Longrightarrow$ For orthonormal U : $\|Ux\|_2 = \|x\|_2$ preserves the norm of x

* Determinant : For $A \in \mathbb{R}^{n \times n}$: $\det(A) \in \mathbb{R}$ satisfies 3 properties [Also denote $\det(A) = |A|$]

* $\det(I) = 1$ * $\det([\alpha a_1, a_2 \cdots a_n]) = \alpha \det(A)$ * $\det([a_2, a_1, a_3 \cdots a_n]) = -\det(A)$

$\left| \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right| = a_{11}a_{22} - a_{12}a_{21}$ Measures area (or volume) spanned by columns or rows of A.

A : full_rank = invertible $\Longleftrightarrow |A| \neq 0$

* Positive Semidefinite:

why PSD is useful? It makes optimizations (as we will see) convex, which can be solved efficiently.

+ Symmetric $A \in \mathbb{R}^{n \times n}$ is positive semidefinite iff $\forall x \neq 0 \in \mathbb{R}^n$: $x^T A x \geq 0$
+ " " positive definite " " $x^T A x > 0$
+ " " negative semidefinite iff " $x^T A x \leq 0$
+ " " Indefinite iff for some $x \neq 0$ $x^T A x > 0$ and for some $x \neq 0$, $x^T A x < 0$.

* For any $A \in \mathbb{R}^{m \times n}$: $G = A^T A$ is PSD : $G \succeq 0$

Eigenvalues and Eigenvectors:

* For a square matrix $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^n$ is the corresponding eigenvector iff
  $Ax = \lambda x$, $x \neq 0$ (multiply A by an evector results only scaling of evector)

* All eigenvalues of A are given by $\det(\lambda I - A) = |\lambda I - A| = 0$.

* $A \in \mathbb{R}^{n \times n} \longrightarrow \lambda_1, \lambda_2, \ldots, \lambda_n$ ($x_1, x_2, \ldots, x_n$) $\longrightarrow$ $tr(A) = \sum_{i=1}^{n} \lambda_i$; $\det(A) = \prod_{i=1}^{n} \lambda_i \longrightarrow$ if $\exists \lambda_j = 0 \longrightarrow \det(A) = 0$ $\downarrow$ A is singular (colms are LD)

* $A \in \mathbb{R}^{n \times n}$ symmetric $\longrightarrow$ $\lambda_i \in \mathbb{R}$, $x_i \in \mathbb{R}^n$ real eval / evecs

Singular Value Decomposition:

Is a generalization of eigendecomposition for general non-square matrices $A \in \mathbb{R}^{m \times n}$.

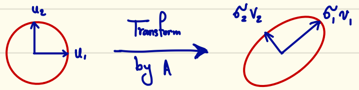
* A non-negative real number σ is a singular value of $A \in \mathbb{R}^{m \times n}$ if and only if there exist unit length vectors $u \in \mathbb{C}^m$, $v \in \mathbb{C}^n$ such that $Au = \sigma v$.

* A singular value decomposition (SVD) of Y corresponds to $A = U \Sigma V^T$ where
  - $U \in \mathbb{R}^{m \times m}$ is orthonormal : $U^T U = U U^T = I_m$ columns of U: left singular vectors
  - $V \in \mathbb{R}^{n \times n}$ is " : $V^T V = V V^T = I_n$ " " V: right " "
  - $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal ($\Sigma_{ij} = 0, i \neq j$) with non-negative real diagonal entries, $\Sigma_{11} \geq \Sigma_{22} \geq \cdots$

$\longrightarrow$ SVD appears in many problems: dimensionality reduction, clustering.

$\longrightarrow$ There exists efficient routines to get SVD.

Intuition behind SVD:



$$\begin{cases} A u_1 = \sigma_1 v_1 \\ A u_2 = \sigma_2 v_2 \end{cases}$$

If $\sigma_1 \simeq \sigma_2 \longrightarrow$ unit disk mapped to unit disk

If $\sigma_2 = 0 \longrightarrow$ unit disk mapped to a line

* rank(A) = rank(Σ) = # nonzero singular values of A.

We can also write : $A = U \Sigma V^T = [u_1 \cdots u_n] \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix} = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

Derivatives:

$f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$x \longmapsto f(x)$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{pmatrix} \in \mathbb{R}^n$$

$f: \mathbb{R}^{mn} \longrightarrow \mathbb{R}$

$X \longrightarrow f(X)$

$$\frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

* $\frac{\partial(\alpha^T x)}{\partial x} = \alpha$
* $\frac{\partial(x^T A x)}{\partial x} = (A^T + A) x$
* $\frac{\partial \operatorname{tr}(A^T X)}{\partial X} = A$