441/541 Math & Conceptual Preliminaries for AI/ML

Linear Algebra

$$\vec{x} \in \mathbb{R}^d \rightarrow \vec{x} = \langle x_1, x_2, \ldots, x_d \rangle$$

$$\underbrace{A}_{\text{Matrix}} \in \mathbb{R}^{m \times n} \rightarrow A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \longleftrightarrow A_{ij}$$

$$\underbrace{A}_{\text{Tensor}} \in \mathbb{R}^{m \times n \times p} \rightarrow A = \overset{p}{\begin{bmatrix} A_{ij} \end{bmatrix} \cdots \begin{bmatrix} A_{ij} \end{bmatrix}}$$

Operations

$$x, y \in \mathbb{R}^n$$

$$x \cdot y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R} \quad \text{(a scalar)}$$

x, y <u>orthogonal if</u> : $x \cdot y = 0$

$$\underset{m\times n}{\underset{\sim}{A}} \cdot \underset{n\times p}{\underset{\sim}{B}} = \underset{m\times p}{\underset{\sim}{C}}$$

$$A \cdot B = \underset{A}{\begin{bmatrix} \\ \text{row}_i \\ \\ \end{bmatrix}} \underset{B}{\begin{bmatrix} \\ \text{col}_j \\ \\ \end{bmatrix}} \to A_{ij} = A_{\text{row}_i} \cdot B_{\text{col}_j}$$

$\boxed{\text{Scalar multiplication}}$

$$c\vec{x} = c\langle x_1, \ldots, x_n \rangle = \langle cx_1, \ldots, cx_n \rangle$$

$$cA = c\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} ca_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ ca_{n1} & & ca_{nn} \end{bmatrix}$$

$A(BC) = (AB)C$ Matrix Multiplication is Associative

But Not commutative : $AB \neq BA$

Transpose : $$A^T = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \\ a_{n1} & & a_{nn} \end{bmatrix}^T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & & a_{nn} \end{bmatrix}$$

i.e. $A^T_{ij} = A_{ji}$

Norms

Note: $\|\vec{x}\|_2^2 = \vec{x}\cdot\vec{x}$

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \qquad \text{e.g. } \|\langle 1,2\rangle\|_2 = \sqrt{1^2+2^2} = \sqrt{5}$$

$$\|\vec{x}\|_\infty = \max_i |x_i| \qquad \|\langle 1,2\rangle\|_\infty = |2| = 2$$

$$\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|\langle 1,2\rangle\|_1 = |1| + |2| = 3$$

Can also define Matrix Norms; The above norms are examples of p-norms.

Special Matrices

$$\underline{I_n} = \begin{bmatrix} 1 & & \emptyset \\ & \ddots & \\ \emptyset & & 1 \end{bmatrix}_{n\times n} \qquad AI = IA = A$$

Identity Matrix

$$\underline{A^T = A} \qquad \text{e.g.} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

Symmetric Matrix

Note: $(AB)^T = B^T A^T \qquad (A+B)^T = B^T + A^T$

Diagonal Matrix: $D = \begin{bmatrix} a_1 & & \phi \\ & \ddots & \\ \phi & & a_n \end{bmatrix}$ (e.g. I_n)

upper-Triangular: $U = \begin{bmatrix} x & & * \\ & x & \\ \phi & x & \\ & & x \;\; x \end{bmatrix}$ e.g. $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$

Pots

Lower-Triangular: $L = \begin{bmatrix} x & & \phi \\ & x & \\ * & x & \\ & & x \end{bmatrix}$ e.g. $\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$

Orthogonal: $\boxed{A^T A = I = A A^T}$

(semi-) **Positive-Definite:** if for all $x \in \mathbb{R}^n$, $x^T A x \underbrace{\geq}_{PSD} 0$

PD: $x^T A x \underbrace{>}_{\text{strict}} 0$ e.g. Covariance Matrix

$A_{n \times n}$ is invertible (i.e. non-singular) if there exists $A^{-1}_{n \times n}$ such that: $A A^{-1} = A^{-1} A = I_n$

Commonly we solve linear systems of equations:

$$\begin{cases} x_1 + x_2 = 1 \\ 2x_1 + 3x_2 = 5 \end{cases} \to \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}}_{\text{"}A\text{"}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\text{"}\vec{x}\text{"}} = \underbrace{\begin{bmatrix} 1 \\ 5 \end{bmatrix}}_{\text{"}\vec{b}\text{"}}$$

Solution, for A non-singular is uniquely given:

$$A\vec{x} = \vec{b} \to \boxed{\vec{x} = A^{-1}\vec{b}}$$

inversion takes $\mathcal{O}(n^3)$ time

Properties: $(A^T)^{-1} = A^{-T}$ $\quad (AB)^{-1} = B^{-1}A^{-1}$, if A, B non-singular

(some) Matrix Factorizations

$\boxed{A = LU}$ → Result of Gaussian Elimination (not all matrices have one)

L: lower-triangular matrix of "multipliers" used in G.E.

U: upper-triangular (Echelon form) after G.E.

$\boxed{PA = LU}$ (all matrices have this)

Like "LU", except we perform pivoting/row exchange via permutation matrix P.

$\boxed{A = QR}$ Q: orthogonal; R: upper-triangular

$\boxed{A = V\Lambda V^T}$ <u>Eigendecomposition</u>; Λ: diagonal matrix of eigenvalues; V: cols. are eigenvectors (see below)

$\boxed{A = LL^T}$ <u>Cholesky Decomposition</u>; A is positive definite; L: lower triangular (gives better numerical stability).

$$\boxed{\underset{n\times p}{A} = \underset{n\times n}{U}\,\underset{n\times p}{\Sigma}\,\underset{p\times p}{V^T}}$$

<u>SVD Decomposition</u>

V: eigenvectors of A^TA (orthogonal matrix)

U: eigenvectors of A^TA (orthogonal matrix)

Σ: Diagonal Matrix of "singular values" of A; elements are square roots of eigenvalues of AA^T.

How is this useful?

(singular values)

<u>Data Compression</u>: $A \approx \sum_{i=1}^{M} \sigma_i u_i v_i^T$

<u>Determinants</u> : $|A_{n\times n}|$: increase in "(unit) volume" for linear transformation defined by A.

$$|A| = \sum_{i=1}^{k} a_{ij}(-1)^{i+j} M_{ij}$$

(ij Minor of A)

<u>Note:</u> $|A| = 0$ if & only if A is <u>singular</u>.

<u>Some properties:</u> $|AB| = |A|\cdot|B|$ $\quad |A^T| = |A|$

Eigenvalues λ an eigenvalue of $A_{n\times n}$ if there exists $\vec{v} \neq \vec{0}$ such that

$$\boxed{A\vec{v} = \lambda\vec{v}}$$

(Note $\lambda = 0$ is OK)

$\vec{v}$: The eigenvector associated with λ.

How to find eigenvalues/eigenvectors?

<u>Solve:</u> $\underbrace{|A - \lambda I|}_{\text{polynomial in } \lambda} = 0 \longrightarrow$ ① get eigenvalues

② get eigenvectors: $A\vec{v} = \lambda\vec{v}$

$\boxed{\text{Liner Independence}}$

Set of vectors: $\{\vec{v}_1, .., \vec{v}_n\}$ is $\boxed{\text{linearly independent}}$ if

$\alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \vec{0}$ implies <u>all $\alpha_i = 0$</u>.

(otherwise vectors are <u>linearly dependent</u>).

$\boxed{\text{Span}}$

<u>Span</u> of a set of vectors is the set of all

liner combinations of the vectors:

$$\text{span}\{\vec{v}_1, .., \vec{v}_n\} = \underbrace{\alpha_1 \vec{v}_1 + \cdots \alpha_n \vec{v}_n}_{\text{over all } \{\alpha_i \in \mathbb{R}\}}$$

$\boxed{\text{Basis}}$

Set of vectors is a $\boxed{\text{basis}}$ (for a vector space) if

set is both: ① linearly independent

&

② "spans" space.

e.g. $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is a basis for $\mathbb{R}^2$.

Lastly, Re 4 Fundamental Subspaces of a matrix $A_{m\times n}$.

1. Col(A): column space of A - set of all linear combinations of column vectors of A.

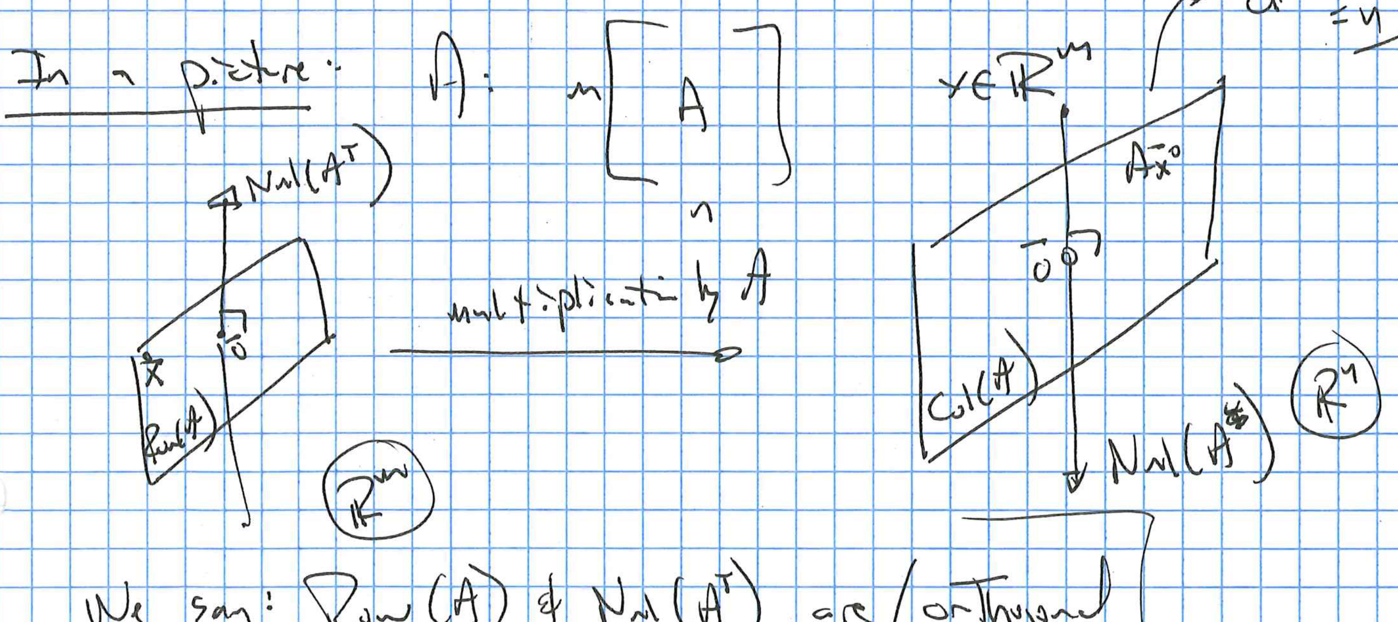
2. Nul(A): Nullspace of A - set of all vectors $\vec{x}$, where: $A\vec{x} = \vec{0}$.

3. Row(A): Rowspace of A - set of all linear combinations of rows of A.

4. $Nul(A^T)$: Nullspace of A^T - i.e. set of all vectors $\vec{x}$, where $A^T\vec{x} = \vec{0}$.

$\boxed{FTLA}$: $\dim(Col(A)) + \dim(Nul(A^T)) = m$

In a picture:



We say: Row(A) & $Nul(A^T)$ are $\boxed{\text{orthogonal}}$

Col(A) & Nul(A) are orthogonal