FIT3181 Deep Learning

Week 00: Extra Content – Linear Algebra

Dr. Trung Le
trunglm@monash.edu

Acknowledge Prof. Dinh Phung for this material.

Faculty of Information Technology, Monash University, Australia

Linear Algebra

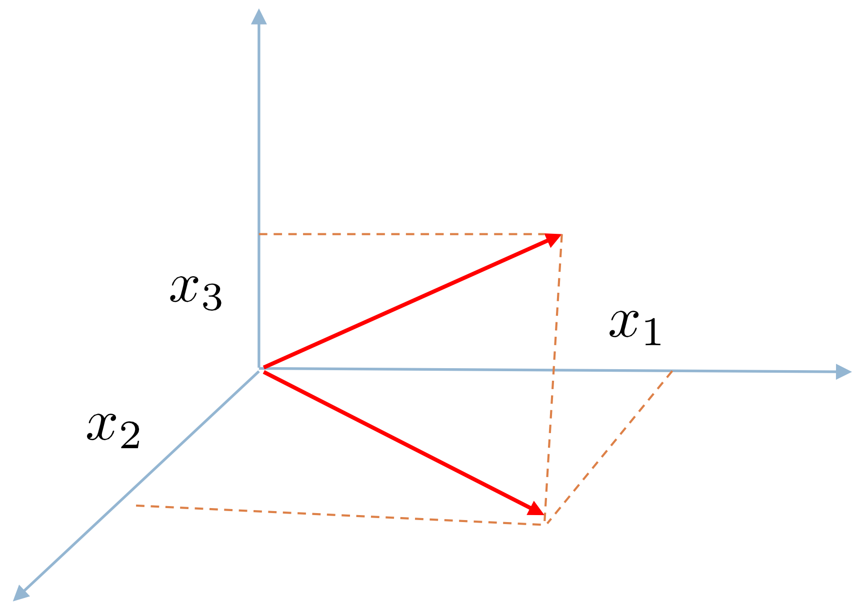
Vector

n-dimensional vector

dimension

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

x_i : i-th element



MONASH University

Vector

Operations on vector

transpose: column vector to row vector

$$\mathbf{x}^{\mathsf{T}} = [x_1 \; x_2 \; \ldots \; x_n]$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

addition

$$\mathbf{x}^{\mathsf{T}}\mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

inner product

Vector

p-norm

$$||\mathbf{x}||_p = (|x_1|^p + \ldots + |x_D|^p)^{\frac{1}{p}}$$

1. $||\mathbf{x}|| > 0$ when $\mathbf{x} \neq \mathbf{0}$ and $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. $||k\mathbf{x}|| = ||k||||\mathbf{x}||$ for any scalar k.
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

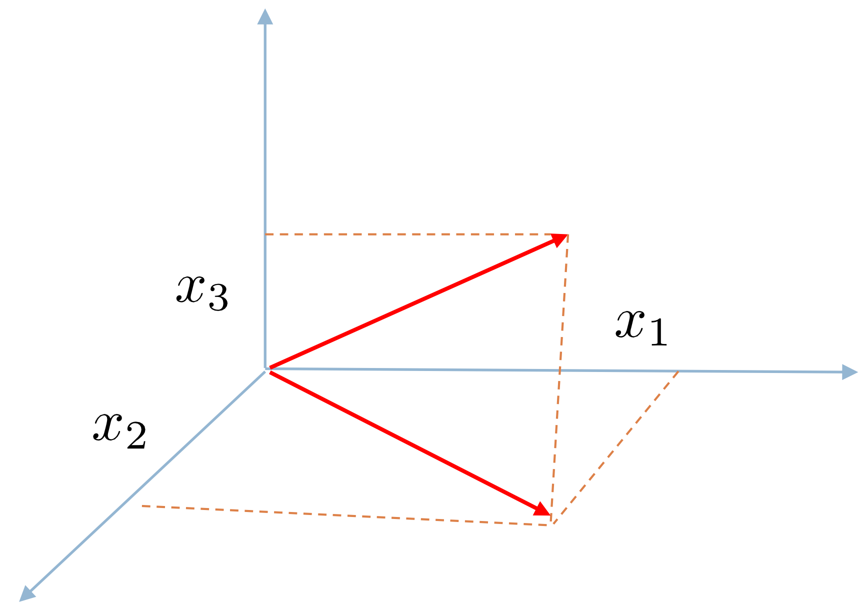
- p = 0 : how many elements in x are non-zeros (sparsity)
- p = 1 : sum of absolute values of elements
- p = 2 : length of the vector

Vector

Length of a vector

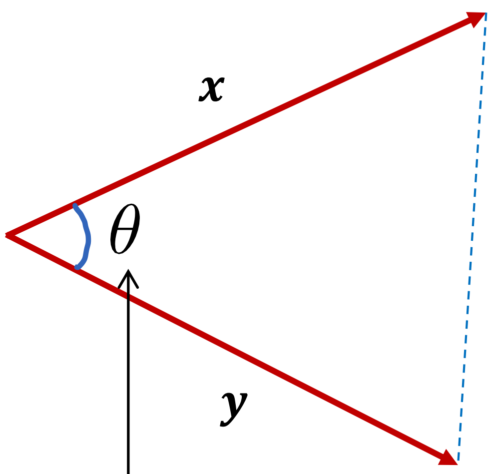
$$\text{length}(\mathbf{x}) = \sqrt{x_1^2 + \ldots + x_D^2}$$

$$||\mathbf{x}||_2 = \sqrt{x_1^2 + \ldots + x_D^2}$$



Vector

Distance between two vectors



Euclidean distance

$$||\boldsymbol{x} - \boldsymbol{y}|| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

Cosine similarity

$$\cos(\theta) = \frac{\boldsymbol{x}^T \boldsymbol{y}}{||\boldsymbol{x}||\,||\boldsymbol{y}||} = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\sqrt{\boldsymbol{x}^T \boldsymbol{x}}\sqrt{\boldsymbol{y}^T \boldsymbol{y}}}$$

Cosine distance is defined as

$$1 - \cos(\theta)$$

Cosine distance can be computed via Euclidean distance if vectors are made unit vectors! (why?)

Matrix

number of columns

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

number of rows

number of columns

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

number of rows

column vector

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\begin{array}{cc} A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} & \text{row vector} \end{array}$$

Vector Space Model

	vocabulary	Doc1	Doc2
1	goal	1	0
2	data	1	2
3	information	2	2
4	insight	1	0
5	you	0	2

Document 1

"The goal is to turn data into information, and information into insight"
Carly Fiorina

Document 2

"You can have data without information, but you cannot have information without data."
Daniel Keys Moran

terms

documents

$$A = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 \\ 0 & 2 & 2 & 0 & 2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \longleftarrow x_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}^{\mathrm{T}}$$

Vector Space Model

	vocabulary	Doc1	Doc2
1	goal	1	0
2	data	1	2
3	information	2	2
4	insight	1	0
5	you	0	2

Document 1

"The goal is to turn data into information, and information into insight"

Carly Fiorina

Document 2

"You can have data without information, but you cannot have information without data."

Daniel Keys Moran

terms

documents

$$A = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 \\ \boxed{0 \quad 2 \quad 2 \quad 0 \quad 2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$x_2 = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}^{\mathrm{T}}$$

Vector Space Model

Document 1

"The goal is to turn data into information, and information into insight"

Carly Fiorina

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}^T$$

Document 2

"You can have data without information, but you cannot have information without data."

Daniel Keys Moran

$$x_2 = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}^T$$

- Each document is now represented as a vector.
- We can now compute distance between vectors to compute their similarity.
- We can then make quantitative statement about which document is how similar to other!
- This is the core of information retrieval/clustering!

Euclidean distance: $\sqrt{(1-0)^2+(1-2)^2+(2-2)^2+(1-0)^2+(0-2)^2}$

$= \sqrt{0+1+0+1+4} = \sqrt{6} \approx 2.45$

Feature Matrix

Term-by-document matrix for text analysis

Document 1

"The goal is to turn data into information, and information into insight"

Carly Fiorina

Document 2

"You can have data without information, but you cannot have information without data."

Daniel Keys Moran

Document 3

....

Document 4

....

Document 5

....

....

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}^{\mathrm{T}} \qquad x_2 = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}^{\mathrm{T}} \qquad x_3 \qquad x_4 \qquad x_5 \qquad \cdots\cdots$$

Feature Matrix

- In general, we create a vocabulary of features for all the instances in the dataset.
- Represent each instance as a vector on features listed in the vocabulary.
- Let us say our dataset has N instances, so we create N vectors x_1, x_2,…, x_N.
- Each of these vectors are called feature vector.
- We stack these vectors as a matrix A and call it feature matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Matrix Algebra

also applies to vector

- Matrix addition
- Matrix subtraction
- Matrix multiplication
  - Matrix and scalar multiplication
  - Matrix and matrix multiplication (usual and dot product)
- Other operations (Inverse, Determinant, Trace)
- Type of matrices (Symmetric, Orthogonal)

Matrix Algebra
Addition and subtraction

- You can add or subtract matrices if they have the same size.
- The elements in the corresponding positions are added or subtracted.
- Addition: $$\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 10 \\ 5 & 7 \\ 2 & 9 \end{bmatrix}$$
- Subtraction is similar

Matrix algebra
Multiplication and division

- To multiply a matrix A with scalar α, multiply each element of A with α as below:

- Scalar multiplication: $$3 \times \begin{bmatrix} 4 & 6 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 18 \\ 0 & 3 \\ 3 & 6 \end{bmatrix}$$

- Scalar division is similarly done except that division by 0 is not allowed for obvious reason.

Matrix algebra
Elementwise multiplication

- You can multiply any two matrices elementwise if they have the same size.

- Consider $A \odot B = C$, now $C(i,j)$ is computed as a product of $A(i,j)$ and $B(i,j)$.

- Elementwise multiplication: $\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} \odot \begin{bmatrix} 1 & 5 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 20 \\ 10 & 12 \\ 3 & 7 \end{bmatrix}$

Matrix algebra

Multiply two matrices

- You can multiply any two matrices if the #columns in the first matrix is equal to #rows in the second matrix.
- Consider $AB = C$, now $C(i,j)$ is computed as a dot product of $A(i,:)$ and $B(:,j)$.
- Multiplication: $\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 11 & 28 \\ 8 & 23 \end{bmatrix}$
- Matrix multiplication is NOT commutative. Multiplication order matters, in other words, in general:

$$AB \neq BA$$

- They may even not be size compatible if multiplied in other order!

Matrix algebra

Square and rectangular matrices

- If a matrix A has size $m \times n$ such that $m = n$, then it is called a square matrix otherwise it is a rectangular matrix.

- $A = \begin{bmatrix} 1 & 6 \\ 2 & 3 \end{bmatrix}$ is a square matrix of size 2×2.

- $A = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 3 & 8 \end{bmatrix}$ is a rectangular matrix of size 2×3.

Matrix algebra

Transpose of a matrix

- The transpose of a matrix A is obtained by putting all the elements on matrix rows on its columns.

- Transpose of A is denoted by A^T, then $\mathrm{A}^{\mathrm{T}}(i,j) = A(j,i)$.

- If $A = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 3 & 8 \end{bmatrix}$ then $A^T = \begin{bmatrix} 1 & 2 \\ 6 & 3 \\ 7 & 8 \end{bmatrix}$.

Matrix algebra

Symmetric matrices

- A matrix A is called symmetric if it is equal to its transpose, that is, $A = A^T$.
- Example: $A = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}$ is a symmetric matrix of size 2×2.
- Symmetric matrix is always a square matrix.

Matrix algebra
Diagonal matrices

- A matrix A is called a diagonal matrix if $A(i, j) = 0$ if $i \neq j$.
- Example: $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 8 \end{bmatrix}$
- Diagonal matrix is always a square matrix.

Matrix algebra

Identity matrix

- A matrix I is called an identity matrix if it is a diagonal matrix and $I(i,i) = 1$.

- Example: $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- Note that we often use $I_{n\times n}$ to denote an identity matrix of size $n \times n$.

Matrix algebra

Inverse of a matrix

- A matrix A is called as inverse of matrix B, if and only if $AB = BA = I$.

- Since $AB = BA$, both A and B need to be square matrices.

- If A is inverse of B , we denote it as $A = B^{-1}$

Matrix algebra

Orthogonal matrix

- A square matrix U is called an orthogonal matrix if its transpose is equal to its inverse, i.e.

$$U^T = U^{-1}$$

- Any identity matrix is orthogonal.
- The other examples of orthogonal matrices are rotation matrices.

$$A = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

Matrix algebra

Other related concepts

- Determinant of a matrix
- Trace of a matrix
- Linear Independence
- Rank of a matrix
- Eigen values/Eigen vectors of a matrix

Matrix algebra

Eigen analysis

- Given a square matrix A, a number λ and a vector u that satisfies
$$Au = \lambda u$$
are called an eigenvalue and the corresponding eigenvector of A.
- For a matrix A of size $d \times d$, there are d eigenvectors and eigenvalue pairs.
- It is possible to have only k (which is less than or equal to d) nonzero eigenvalues for A.
- The number of nonzero eigenvalues are equal to the rank of the matrix.

Matrix algebra

Eigen analysis

- If $U = [u_1\ u_2, \dots, u_d]$ are the d eigenvectors of matrix A and $\lambda_1, \dots, \lambda_d$ are the corresponding eigenvalues, then we have

$$Au_1 = \lambda_1 u_1, Au_2 = \lambda_2 u_2, \ \dots \ , Au_d = \lambda_d u_d.$$

- These can be collectively written as

$$AU = U \begin{bmatrix} \lambda_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{bmatrix} = UD$$

- The matrix U is always orthogonal meaning $u_i^T u_j = 0$ if $i \neq j$ and 1 otherwise. Clearly, $U^T = U^{-1}$. Therefore, we have

$$A = UDU^T$$

Matrix algebra
Eigen analysis

- Eigenvalues of a matrix $\boldsymbol{A}$ can be found by solving the characteristic polynomial in λ:

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$$

- The roots of the polynomial are the eigenvalues of the matrix $\boldsymbol{A}$.

- Once all the eigenvalues are obtained, a eigenvector corresponding to a particular eigenvalue can be obtained by solving

$$\boldsymbol{A}\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1$$

Matrix algebra

Singular value decomposition (SVD)

- Given any $n \times d$ matrix X, its SVD is given as

$$X = USV^T$$

  where $\boldsymbol{U}$ is an $n \times d$ orthogonal matrix, $\boldsymbol{S}$ is a $d \times d$ diagonal matrix with elements $S(i,i) = \sigma_i$ and V is an $d \times d$ orthogonal matrix.

- The diagonal elements of S, σ_i's are called singular values of the matrix X.

- The number of nonzero singular values is less than or equal to $\mathbf{min}(n, d)$ and is also equal to the rank of the matrix X.