MATH5836: Data and Machine Learning

Week 0: Basics of Linear Algebra

Sarat Moka *UNSW, Sydney*

Key Topics

Reference:

- MIT Linear Algebra course by Prof. Gilbert Strang `https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/`

0.2.1 Vectors and Matrices

Notation

- $\mathbb{R}^n$: Space of n-dimensional real column vectors.
- $\mathbb{R}^{m\times n}$: Space of real matrices with m rows and n columns.
- I_n: The $n \times n$ identity matrix.
- $0_{m\times n}$: The $m \times n$ zero matrix.

Basic Operations

- **Matrix addition and scalar multiplication**:
$$A + B = [a_{ij} + b_{ij}], \qquad cA = [ca_{ij}]$$
For example, let
$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$
Then their sum is
$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$
And for a scalar $c \in \mathbb{R}$,
$$cA = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}$$
- **Matrix-vector product**: If $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, then $Ax \in \mathbb{R}^m$. For example, Let
$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \in \mathbb{R}^{3\times 2}, \qquad x = \begin{bmatrix} 7 \\ 8 \end{bmatrix} \in \mathbb{R}^2$$
Then
$$Ax = \begin{bmatrix} 1\cdot 7 + 2\cdot 8 \\ 3\cdot 7 + 4\cdot 8 \\ 5\cdot 7 + 6\cdot 8 \end{bmatrix} = \begin{bmatrix} 7 + 16 \\ 21 + 32 \\ 35 + 48 \end{bmatrix} = \begin{bmatrix} 23 \\ 53 \\ 83 \end{bmatrix} \in \mathbb{R}^3$$
- **Matrix-matrix product**: If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, then $AB \in \mathbb{R}^{m\times p}$. Let
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$
Then
$$AB = \begin{bmatrix} 1\cdot 7 + 2\cdot 9 + 3\cdot 11 & 1\cdot 8 + 2\cdot 10 + 3\cdot 12 \\ 4\cdot 7 + 5\cdot 9 + 6\cdot 11 & 4\cdot 8 + 5\cdot 10 + 6\cdot 12 \end{bmatrix} = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

Transpose

- The transpose of $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ is $A^\top \in \mathbb{R}^{n \times m}$ defined by $(A^\top)_{ij} = a_{ji}$. For example, when
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$
then, the transpose is
$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$
- Properties:
  - $(A^\top)^\top = A$
  - $(AB)^\top = B^\top A^\top$
  - $(A + B)^\top = A^\top + B^\top$

Symmetric Matrices

A square matrix A is symmetric if $A = A^\top$.

0.2.2 Linear Combinations, Span, and Independence

Linear Combination

A vector v is a **linear combination** of vectors $v_1, \dots, v_k$ if
$$v = \alpha_1 v_1 + \cdots + \alpha_k v_k, \quad \text{for some } \alpha_i \in \mathbb{R}.$$
For example, let
$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \quad \text{and } \alpha_1 = 2,\, \alpha_2 = -1.$$
Then a linear combination is:
$$\alpha_1 v_1 + \alpha_1 v_2 = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}.$$

Span

The **span** of $\{v_1, \dots, v_k\}$ is the set
$$\operatorname{span}(v_1, \dots, v_k) = \left\{ \sum_{i=1}^{k} \alpha_i v_i : \alpha_i \in \mathbb{R} \right\}.$$

For example, let

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Then the span of $\{v_1, v_2\}$ is the set of all vectors of the form:

$$\alpha_1 v_1 + \alpha_2 v_2 = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \text{where } \alpha_1, \alpha_2 \in \mathbb{R}.$$

Thus,

$$\operatorname{span}(v_1, v_2) = \mathbb{R}^2.$$

Linear Independence

Vectors $v_1, \ldots, v_k$ are **linearly independent** if the only solution to

$$\sum_{i=1}^{k} \alpha_i v_i = 0$$

is $\alpha_1 = \cdots = \alpha_k = 0$. Otherwise, they are **linearly dependent**.

For example, let

$$v_1 = \begin{bmatrix} 10 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ -3.3 \end{bmatrix}.$$

Consider the equation

$$\alpha_1 v_1 + \alpha_2 v_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This becomes

$$\alpha_1 \begin{bmatrix} 10 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ -3.3 \end{bmatrix} = \begin{bmatrix} 10\alpha_1 \\ -3.3\alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This implies $\alpha_1 = 0$ and $\alpha_2 = 0$. Therefore, v_1 and v_2 are linearly independent.

Exercise 0.2.2.1

Show that if $v_1, \ldots, v_k$ are linearly dependent, then one of them can be written as a linear combination of the others.

0.2.3 Matrix Rank and Systems of Equations

Matrix Rank

The **rank** r of a matrix $A \in \mathbb{R}^{m \times n}$ is the number of linearly independent columns (equivalently, the dimension of the column space of A). Note that $r \leq \min(m, n)$, and when $r = \min(m, n)$, we say A is **full rank**. A full rank square matrix is referred to as **non-singular** (or **invertible**).

Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

The third column is a linear combination of the first two:

$$\text{Column}_3 = -1 \cdot \text{Column}_1 + 2 \cdot \text{Column}_2.$$

Hence, the columns are linearly dependent, and the rank of A is 2.

Determinant

Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$. The determinant of A, denoted $\det(A)$, can be defined as

$$\det(A) \;=\; \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)},$$

where S_n is the set of all permutations of $\{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is the sign of σ.

A matrix A is invertible if and only if $\det(A) \neq 0$.

Example

Compute the determinant of

$$B \;=\; \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}$$

by expansion along the first row:

$$\det(B) = 1 \cdot \det \begin{pmatrix} 1 & 4 \\ 6 & 0 \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 0 & 4 \\ 5 & 0 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 0 & 1 \\ 5 & 6 \end{pmatrix}.$$

Each 2×2 determinant is

$$\det \begin{pmatrix} 1 & 4 \\ 6 & 0 \end{pmatrix} = 1 \cdot 0 - 4 \cdot 6 = -24, \quad \det \begin{pmatrix} 0 & 4 \\ 5 & 0 \end{pmatrix} = 0 \cdot 0 - 4 \cdot 5 = -20, \quad \det \begin{pmatrix} 0 & 1 \\ 5 & 6 \end{pmatrix} = 0 \cdot 6 - 1 \cdot 5 = -5.$$

Hence

$$\det(B) = 1 \cdot (-24) - 2 \cdot (-20) + 3 \cdot (-5) = -24 + 40 - 15 = 1.$$

Systems of Linear Equations

A linear system $Ax = b$ satisfies:

- A solution exists if and only if b lies in the column space of A, i.e., $b \in \operatorname{col}(A) = \operatorname{span}(\text{columns of } A)$.

- If A is square and full rank, then the solution is unique and given by $x = A^{-1}b$.

For example, let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 11 \end{bmatrix}.$$

Since A is square and has full rank (rank 2), a unique solution exists:

$$x = A^{-1}b = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Inverse of a Matrix

Let $A \in \mathbb{R}^{n\times n}$. A matrix $B \in \mathbb{R}^{n\times n}$ is called a (two-sided) inverse of A if

$$B\,A \;=\; A\,B \;=\; I_n.$$

If such a B exists, then A is said to be *invertible* (or *nonsingular*); this occurs exactly when

$$\mathrm{rank}(A) = n \quad \Longleftrightarrow \quad \det(A) \neq 0.$$

In that case the inverse is unique and denoted A^{-1}.

Example 1 (Symbolic, 2×2): Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad ad - bc \neq 0.$$

One checks directly that

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

since $(A^{-1}A)_{ij} = \delta_{ij}$.

Example 2 (Numeric): Take

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

so $\det(A) = -2$. Then

$$A^{-1} = -\tfrac{1}{2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix},$$

and one verifies $A^{-1}A = I_2$.

Exercise 0.2.3.1

Show that if A is invertible, then $Ax = b$ has a unique solution.

Eigenvalues and Eigenvectors

Let $A \in \mathbb{R}^{n\times n}$. A nonzero vector $v \in \mathbb{R}^n$ is called an **eigenvector** of A if there exists a scalar $\lambda \in \mathbb{R}$ such that

$$A\,v = \lambda\,v.$$

The scalar λ is then called an **eigenvalue** of A corresponding to v. Equivalently, λ is an eigenvalue of A if and only if

$$\det\big(A - \lambda I\big) = 0.$$

The set of all eigenvalues of A is often called the **spectrum** of A.

For example, consider the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

1. Compute the characteristic polynomial:

$$\det(A - \lambda I) = \det\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3.$$

2. Solve $\lambda^2 - 4\lambda + 3 = 0$ to get

$$\lambda_1 = 1, \quad \lambda_2 = 3.$$

3. For each λ_i find a nonzero v_i with $(A - \lambda_i I)v_i = 0$:

- $\lambda_1 = 1$:

$$(A - I)v = 0 \implies \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} v = 0 \implies v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

- $\lambda_2 = 3$:

$$(A - 3I)v = 0 \implies \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} v = 0 \implies v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Hence the eigenpairs are

$$(\lambda_1, v_1) = \big(1, (1, -1)^\top\big), \quad (\lambda_2, v_2) = \big(3, (1, 1)^\top\big).$$

Exercise 0.2.3.2

Let $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$. Find its eigenvalues and eigenvectors.

0.2.4 Singular Value Decomposition (SVD)

Any real matrix $A \in \mathbb{R}^{m \times n}$ can be decomposed as

$$A = U \Sigma V^T$$

where:

- $U \in \mathbb{R}^{m \times m}$ is an orthogonal matrix (its columns are the left singular vectors of A),
- $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix with non-negative entries (the singular values of A),
- $V \in \mathbb{R}^{n \times n}$ is an orthogonal matrix (its columns are the right singular vectors of A).

Example

Let

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{2 \times 2}$$

A possible SVD of A (approximately) is:

$$U = \begin{bmatrix} 0.957 & -0.290 \\ 0.290 & 0.957 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3.257 & 0 \\ 0 & 1.842 \end{bmatrix}, \quad V = \begin{bmatrix} 0.882 & 0.472 \\ -0.472 & 0.882 \end{bmatrix},$$

so that

$$A = U \Sigma V^T$$

This decomposition helps in understanding the structure of A and is widely used in applications such as dimensionality reduction, data compression, and solving ill-posed problems.

Spectral Decomposition of Symmetric Matrices

If $A \in \mathbb{R}^{n \times n}$ is a *symmetric* matrix (i.e., $A = A^\top$), then it can be decomposed as

$$A = Q \Lambda Q^\top$$

where:

- $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix whose columns are the eigenvectors of A,
- $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal entries are the eigenvalues of A.

This decomposition is a special case of the SVD and is known as the **spectral decomposition** or **eigendecomposition** of A.