COMP 551 - Applied Machine Learning
A Brief Tutorial on Linear Algebra

Safa Alver
(slides adapted from Stanford's CS229 Linear Algebra Review)

McGill University

January 16, 2021

About this tutorial

- This is not an comprehensive review of Linear Algebra.

About this tutorial

- This is not an comprehensive review of Linear Algebra.
- The focus is on the subset related to COMP 551.

About this tutorial

- This is not an comprehensive review of Linear Algebra.
- The focus is on the subset related to COMP 551.
- More references can be found at the end of the slides.

About this tutorial

- This is not an comprehensive review of Linear Algebra.
- The focus is on the subset related to COMP 551.
- More references can be found at the end of the slides.
- Also please shoot me an email if you find any typos or mistakes!

Outline

Outline

Scalars

- A scalar is just a single number (integers, rational numbers, . . .).

Scalars

- A scalar is just a single number (integers, rational numbers, ...).
- Examples: $1, 2, \pi, e, -112, \frac{1}{4}, \ldots$

Vectors

- A vector is a 1D array of numbers.

Vectors

- A vector is a 1D array of numbers.
- Examples:

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad y = \begin{bmatrix} -1 \\ 2 \\ 5 \\ 6 \end{bmatrix}, \qquad \ldots$$

Vectors

- A vector is a 1D array of numbers.
- Examples:

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad y = \begin{bmatrix} -1 \\ 2 \\ 5 \\ 6 \end{bmatrix}, \qquad \dots$$

- In this course we will mostly use real vectors living in an n dimensional space:

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

Vectors

- A vector is a 1D array of numbers.
- Examples:

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad y = \begin{bmatrix} -1 \\ 2 \\ 5 \\ 6 \end{bmatrix}, \qquad \ldots$$

- In this course we will mostly use real vectors living in an n dimensional space:

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

- We use $x \in \mathbb{R}^n$ to denote this.

Matrices

- A matrix is a 2D array of numbers.

Matrices

- A matrix is a 2D array of numbers.
- Examples:

$$X = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}, \qquad Y = \begin{bmatrix} -1 & 9 \\ 2 & -1 \\ 5 & 5 \end{bmatrix}, \qquad \ldots$$

Matrices

- A matrix is a 2D array of numbers.
- Examples:

$$X = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}, \qquad Y = \begin{bmatrix} -1 & 9 \\ 2 & -1 \\ 5 & 5 \end{bmatrix}, \qquad \dots$$

- In this course we will mostly use real matrices living in an $m \times n$ dimensional space:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

Here, m is the number of rows and n is the number of columns.

Matrices

- A matrix is a 2D array of numbers.
- Examples:

$$X = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}, \qquad Y = \begin{bmatrix} -1 & 9 \\ 2 & -1 \\ 5 & 5 \end{bmatrix}, \qquad \dots$$

- In this course we will mostly use real matrices living in an $m \times n$ dimensional space:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

  Here, m is the number of rows and n is the number of columns.
- We use $A \in \mathbb{R}^{m \times n}$ to denote this.

Tensors

- A tensor is the generalization of the idea of having an array of numbers.

Tensors

- A tensor is the generalization of the idea of having an array of numbers.
- It may be:
  - 0D and be a scalar,

Tensors

- A tensor is the generalization of the idea of having an array of numbers.
- It may be:
  - 0D and be a scalar,
  - 1D and be a vector,

Tensors

- A tensor is the generalization of the idea of having an array of numbers.
- It may be:
  - 0D and be a scalar,
  - 1D and be a vector,
  - 2D and be a matrix,

Tensors

- A tensor is the generalization of the idea of having an array of numbers.
- It may be:
  - 0D and be a scalar,
  - 1D and be a vector,
  - 2D and be a matrix,
  - nD and just be an nD tensor.

Some important matrices - The Indentity Matrix

- A square matrix is a matrix with the same number of rows and columns, i.e. $m = n$.

Some important matrices - The Indentity Matrix

- A square matrix is a matrix with the same number of rows and columns, i.e. $m = n$.
- The identity matrix, denoted $I \in R^{n \times n}$, is a special square matrix with ones on the diagonal and zeros elsewhere:

$$I_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

Some important matrices - The Indentity Matrix

- A square matrix is a matrix with the same number of rows and columns, i.e. $m = n$.
- The identity matrix, denoted $I \in R^{n \times n}$, is a special square matrix with ones on the diagonal and zeros elsewhere:

$$I_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

- Examples:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \dots$$

Some important matrices - The Indentity Matrix

- A square matrix is a matrix with the same number of rows and columns, i.e. $m = n$.
- The identity matrix, denoted $I \in R^{n\times n}$, is a special square matrix with ones on the diagonal and zeros elsewhere:

$$I_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

- Examples:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \ldots$$

- Its most important property is that for any $A \in \mathbb{R}^{n\times n}$:

$$AI = A = IA.$$

Some important matrices - Diagonal Matrices

- A diagonal matrix, denoted $D \in \mathbb{R}^{n \times n}$ is a special square matrix where all non-diagonal elements are zero:

$$D_{ij} = \begin{cases} d_i, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

Some important matrices - Diagonal Matrices

- A diagonal matrix, denoted $D \in \mathbb{R}^{n \times n}$ is a special square matrix where all non-diagonal elements are zero:

$$D_{ij} = \begin{cases} d_i, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

- Examples:

$$I = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, \qquad I = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 9 \end{bmatrix}, \qquad \dots$$

Some important matrices - Diagonal Matrices

- A diagonal matrix, denoted $D \in \mathbb{R}^{n \times n}$ is a special square matrix where all non-diagonal elements are zero:

$$D_{ij} = \begin{cases} d_i, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

- Examples:

$$I = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, \qquad I = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 9 \end{bmatrix}, \qquad \dots$$

- These matrices can also be denoted as

$$D = \operatorname{diag}(d_1, d_2, \dots, d_n).$$

Some important matrices - Diagonal Matrices

- A diagonal matrix, denoted $D \in \mathbb{R}^{n \times n}$ is a special square matrix where all non-diagonal elements are zero:

$$D_{ij} = \begin{cases} d_i, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

- Examples:

$$I = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, \qquad I = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 9 \end{bmatrix}, \qquad \ldots$$

- These matrices can also be denoted as

$$D = \text{diag}(d_1, d_2, \ldots, d_n).$$

- It should also be noted that $I = \text{diag}(1, 1, \ldots, 1)$.

Outline

Vector-Vector Products

- The **inner product** (or **dot product**) between two vectors:

$$x^\top z = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \sum_{i=1}^{n} x_i z_i.$$

Vector-Vector Products

- The **inner product** (or **dot product**) between two vectors:

$$x^\top z = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \sum_{i=1}^{n} x_i z_i .$$

- Example:

$$\begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \end{bmatrix} = 1 \times 2 + 3 \times 6 = 20.$$

Vector-Vector Products...

- The **outer product** between two vectors:

$$xz^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} = \begin{bmatrix} x_1 z_1 & x_1 z_2 & \dots & x_1 z_n \\ x_2 z_1 & x_2 z_2 & \dots & x_2 z_n \\ \vdots & \vdots & \vdots & \vdots \\ x_n z_1 & x_n z_2 & \dots & x_n z_n \end{bmatrix}.$$

Vector-Vector Products...

- The **outer product** between two vectors:

$$xz^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} = \begin{bmatrix} x_1 z_1 & x_1 z_2 & \dots & x_1 z_n \\ x_2 z_1 & x_2 z_2 & \dots & x_2 z_n \\ \vdots & \vdots & \vdots & \vdots \\ x_n z_1 & x_n z_2 & \dots & x_n z_n \end{bmatrix}.$$

- Example:

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 6 \end{bmatrix} = \begin{bmatrix} 1 \times 2 & 1 \times 6 \\ 3 \times 2 & 3 \times 6 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}.$$

Matrix-Vector Products

- If we write A in terms of its row vectors, the product Ax can be expressed as:

$$Ax = \begin{bmatrix} \text{—}a_1^\top\text{—} \\ \text{—}a_2^\top\text{—} \\ \vdots \\ \text{—}a_m^\top\text{—} \end{bmatrix} x = \begin{bmatrix} a_1^\top x \\ a_2^\top x \\ \vdots \\ a_m^\top x \end{bmatrix}.$$

Matrix-Vector Products

- If we write A in terms of its row vectors, the product Ax can be expressed as:

$$Ax = \begin{bmatrix} \text{—} a_1^\top \text{—} \\ \text{—} a_2^\top \text{—} \\ \vdots \\ \text{—} a_m^\top \text{—} \end{bmatrix} x = \begin{bmatrix} a_1^\top x \\ a_2^\top x \\ \vdots \\ a_m^\top x \end{bmatrix}.$$

- It should be noted that Ax is just a vector whose elements are the dot products between A's row vectors and the vector x.

Matrix-Vector Products

- If we write A in terms of its row vectors, the product Ax can be expressed as:

$$Ax = \begin{bmatrix} —a_1^\top— \\ —a_2^\top— \\ \vdots \\ —a_m^\top— \end{bmatrix} x = \begin{bmatrix} a_1^\top x \\ a_2^\top x \\ \vdots \\ a_m^\top x \end{bmatrix}.$$

- It should be noted that Ax is just a vector whose elements are the dot products between A's row vectors and the vector x.
- Example:

$$\begin{bmatrix} 1 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \times 2 + 4 \times 6 \\ 4 \times 2 + 3 \times 6 \end{bmatrix} = \begin{bmatrix} 26 \\ 16 \end{bmatrix}.$$

Matrix-Vector Products (Another view)

- We can also write A in terms of its columns, then we would have:

$$Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ a^1 \\ | \end{bmatrix} x_1 + \cdots + \begin{bmatrix} | \\ a^n \\ | \end{bmatrix} x_n.$$

Matrix-Vector Products (Another view)

- We can also write A in terms of its columns, then we would have:

$$Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ a^1 \\ | \end{bmatrix} x_1 + \cdots + \begin{bmatrix} | \\ a^n \\ | \end{bmatrix} x_n.$$

- In this view, the product is a **linear combination** of the columns of A.

Matrix-Matrix Multiplication

- Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$.
- If we write A in terms of its row vectors and B in terms of it column vectors, their multiplication AB can be expressed as:

$$
\begin{aligned}
AB &= \begin{bmatrix} —a_1^\top— \\ —a_2^\top— \\ \vdots \\ —a_m^\top— \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^p \\ | & | & & | \end{bmatrix} \\
&= \begin{bmatrix} a_1^\top b^1 & a_1^\top b^2 & \cdots & a_1^\top b^p \\ a_2^\top b^1 & a_2^\top b^2 & \cdots & a_2^\top b^p \\ \vdots & \vdots & \vdots & \vdots \\ a_m^\top b^1 & a_m^\top b^2 & \cdots & a_m^\top b^p \end{bmatrix}.
\end{aligned}
$$

Matrix-Matrix Multiplication

- Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$.
- If we write A in terms of its row vectors and B in terms of it column vectors, their multiplication AB can be expressed as:

$$AB = \begin{bmatrix} — a_1^\top — \\ — a_2^\top — \\ \vdots \\ — a_m^\top — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^p \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} a_1^\top b^1 & a_1^\top b^2 & \cdots & a_1^\top b^p \\ a_2^\top b^1 & a_2^\top b^2 & \cdots & a_2^\top b^p \\ \vdots & \vdots & \vdots & \vdots \\ a_m^\top b^1 & a_m^\top b^2 & \cdots & a_m^\top b^p \end{bmatrix}.$$

- Here, AB is just a matrix whose entries are the dots products between A's row vectors and B's column vectors.

Properties of Matrix-Matrix Multiplication

- Associative: $A(BC) = (AB)C$.

Properties of Matrix-Matrix Multiplication

- Associative: $A(BC) = (AB)C$.
- Distributive: $A(B + C) = AB + AC$.

Properties of Matrix-Matrix Multiplication

- Associative: $A(BC) = (AB)C$.
- Distributive: $A(B + C) = AB + AC$.
- Not Commutative (in general): $AB \neq BA$.

Outline

Matrix Transpose

- Taking the transpose of a matrix corresponds to "flipping" its rows and columns.

Matrix Transpose

- Taking the transpose of a matrix corresponds to "flipping" its rows and columns.
- Given an $m \times n$ matrix A, its transpose, denoted as $A^\top$, is an $n \times m$ matrix whose entries are given by:

$$(A^\top)_{ij} = A_{ji}.$$

Matrix Transpose

- Taking the transpose of a matrix corresponds to "flipping" its rows and columns.
- Given an $m \times n$ matrix A, its transpose, denoted as $A^\top$, is an $n \times m$ matrix whose entries are given by:

$$(A^\top)_{ij} = A_{ji}.$$

- Example:

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix}, \qquad A^\top = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}.$$

Matrix Transpose

- Taking the transpose of a matrix corresponds to "flipping" its rows and columns.
- Given an $m \times n$ matrix A, its transpose, denoted as $A^\top$, is an $n \times m$ matrix whose entries are given by:

$$(A^\top)_{ij} = A_{ji}.$$

- Example:

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix}, \qquad A^\top = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}.$$

- It has the following properties:
  - $(A^\top)^\top = A$
  - $(AB)^\top = B^\top A^\top$
  - $(A+B)^\top = A^\top + B^\top$

Matrix Trace

- The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

Matrix Trace

- The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

- Example:

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix}, \qquad \mathrm{Tr}(A) = 1 + 5 = 6.$$

Matrix Trace

- The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

- Example:

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix}, \qquad \mathrm{Tr}(A) = 1 + 5 = 6.$$

- It has the following properties (assuming $B \in \mathbb{R}^{n \times n}$):
  - $\mathrm{Tr}(A) = \mathrm{Tr}(A^\top)$.
  - $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$.
  - $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

Vector Norms

- Informally, the **norm** of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is a measure of how "large" it is.

Vector Norms

- Informally, the **norm** of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is a measure of how "large" it is.
- Formally, it is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies certain properties[1].

[1] See page 10 of CS229's Linear Algebra Review for these properties.

Vector Norms

- Informally, the **norm** of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is a measure of how "large" it is.
- Formally, it is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies certain properties[1].
- Commonly used norms are as follows:
  - Eucledean or ℓ_2 norm (most popular):

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

[1] See page 10 of CS229's Linear Algebra Review for these properties.

Vector Norms

- Informally, the **norm** of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is a measure of how "large" it is.
- Formally, it is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies certain properties[1].
- Commonly used norms are as follows:
  - Eucledean or ℓ_2 norm (most popular):

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

  - ℓ_1 norm:

$$||x||_1 = \sum_{i=1}^{n} |x_i|.$$

[1] See page 10 of CS229's Linear Algebra Review for these properties.

Vector Norms

- Informally, the **norm** of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is a measure of how "large" it is.
- Formally, it is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies certain properties[1].
- Commonly used norms are as follows:
  - Eucledean or ℓ_2 norm (most popular):

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

  - ℓ_1 norm:

$$||x||_1 = \sum_{i=1}^{n} |x_i|.$$

  - ℓ_∞ norm:

$$||x||_\infty = \max_i |x_i|.$$

[1] See page 10 of CS229's Linear Algebra Review for these properties.

Vector Norms

- All of these norms belong to the family of ℓ_p norms, parameterized by a real number $p \geq 1$, which is defined as:

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

Vector Norms

- All of these norms belong to the family of ℓ_p norms, parameterized by a real number $p \geq 1$, which is defined as:

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

- It should be noted that norms can also be defined for matrices (Frobenius norm), but they are out of the scope of this tutorial.

Linear Independence

- A set of vectors $\mathcal{X} = \{x_1, x_2, \ldots x_m\} \subset \mathbb{R}^n$ are called **linearly dependent** if any one of them can be represented as a linear combination of the others:

$$x_k = \sum_{x_i \in \mathcal{X} - \{x_k\}} c_i x_i,$$

where $1 \leq k \leq m$ and $c_i \in \mathbb{R}$.

Linear Independence

- A set of vectors $\mathcal{X} = \{x_1, x_2, \dots x_m\} \subset \mathbb{R}^n$ are called **linearly dependent** if any one of them can be represented as a linear combination of the others:

$$x_k = \sum_{x_i \in \mathcal{X} - \{x_k\}} c_i x_i,$$

  where $1 \leq k \leq m$ and $c_i \in \mathbb{R}$.
- Otherwise, they are called **linearly independent**.

Linear Independence

- A set of vectors $\mathcal{X} = \{x_1, x_2, \ldots x_m\} \subset \mathbb{R}^n$ are called **linearly dependent** if any one of them can be represented as a linear combination of the others:

$$x_k = \sum_{x_i \in \mathcal{X} - \{x_k\}} c_i x_i,$$

  where $1 \leq k \leq m$ and $c_i \in \mathbb{R}$.
- Otherwise, they are called **linearly independent**.
- Example:

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

  are linearly independent vectors as any linear combination of the two of them can't give the other.

Matrix Rank

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of column vectors of A that constitute a linearly independent set.

Matrix Rank

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of column vectors of A that constitute a linearly independent set.
- The **row rank** is the same thing, except for the row vectors of A.

Matrix Rank

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of column vectors of A that constitute a linearly independent set.
- The **row rank** is the same thing, except for the row vectors of A.
- For any matrix $A \in \mathbb{R}^{m \times n}$, these two are equal. So, both of them are collectively referred to as the **rank** of a A, and are denoted as $\text{rank}(A)$.

Matrix Rank

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of column vectors of A that constitute a linearly independent set.
- The **row rank** is the same thing, except for the row vectors of A.
- For any matrix $A \in \mathbb{R}^{m \times n}$, these two are equal. So, both of them are collectively referred to as the **rank** of a A, and are denoted as $\text{rank}(A)$.
- Properties:
  - $\text{rank}(A) \leq \min(m, n)$.
  - If $\text{rank}(A) = \min(m, n)$, A is called **full rank**.
  - $\text{rank}(A) = \text{rank}(A^\top)$.
  - For more see page 11 of CS229's Linear Algebra Review.

The Inverse of a Square Matrix

- The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted as A^{-1} and is unique such that:

$$A^{-1}A = I = AA^{-1}.$$

The Inverse of a Square Matrix

- The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted as A^{-1} and is unique such that:

$$A^{-1}A = I = AA^{-1}.$$

- A is called **invertible** (or **non-singular**) if A^{-1} exists, and **non-invertible** (or **singular**) otherwise.

The Inverse of a Square Matrix

- The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted as A^{-1} and is unique such that:

$$A^{-1}A = I = AA^{-1}.$$

- A is called **invertible** (or **non-singular**) if A^{-1} exists, and **non-invertible** (or **singular**) otherwise.
- In order for A^{-1} to exist, A must be full rank.

The Inverse of a Square Matrix

- The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted as A^{-1} and is unique such that:

$$A^{-1}A = I = AA^{-1}.$$

- A is called **invertible** (or **non-singular**) if A^{-1} exists, and **non-invertible** (or **singular**) otherwise.
- In order for A^{-1} to exist, A must be full rank.
- Properties:
  - $(A^{-1})^{-1} = A$.
  - $(AB)^{-1} = B^{-1}A^{-1}$.
  - $(A^{-1})^\top = (A^\top)^{-1}$.

Orthogonal Matrices

- Two vectors $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$ are **orthogonal** if $x^\top z = 0$.

Orthogonal Matrices

- Two vectors $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$ are **orthogonal** if $x^\top z = 0$.
- A vector $x \in \mathbb{R}^n$ is **normalized** if $||x||_2 = 1$.

Orthogonal Matrices

- Two vectors $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$ are **orthogonal** if $x^\top z = 0$.
- A vector $x \in \mathbb{R}^n$ is **normalized** if $||x||_2 = 1$.
- A square matrix $U \in \mathbb{R}^{n \times n}$ is **orthogonal** if all of its columns are orthogonal to each other and are normalized. Its columns are then referred to as being **orthonormal**.

Orthogonal Matrices

- Two vectors $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$ are **orthogonal** if $x^\top z = 0$.
- A vector $x \in \mathbb{R}^n$ is **normalized** if $||x||_2 = 1$.
- A square matrix $U \in \mathbb{R}^{n \times n}$ is **orthogonal** if all of its columns are orthogonal to each other and are normalized. Its columns are then referred to as being **orthonormal**.
- Properties:
  - The inverse of an orthogonal matrix is its transpose:

$$U^\top U = I = UU^\top.$$

  - Multiplying an n dimensional vector with an $n \times n$ orthogonal matrix will not change its Euclidean norm:

$$||Ux||_2 = ||x||_2.$$

Span and Projection

- The **span** of a set of vectors $\{x_1, x_2, \ldots, x_m\}$ is the set of all vectors that can be expressed as a linear combination of all of them:

$$\text{span}(\{x_1, x_2, \ldots, x_m\}) = \left\{ v : v = \sum_{i=1}^{m} c_i x_i, \quad c_i \in \mathbb{R} \right\}.$$

Span and Projection

- The **span** of a set of vectors $\{x_1, x_2, \ldots, x_m\}$ is the set of all vectors that can be expressed as a linear combination of all of them:

$$\text{span}(\{x_1, x_2, \ldots, x_m\}) = \left\{ v : v = \sum_{i=1}^{m} c_i x_i, \quad c_i \in \mathbb{R} \right\}.$$

- The **projection** of a vector $z \in \mathbb{R}^p$ onto the span of $\{x_1, x_2, \ldots, x_m\}$ is the vector $v \in \text{span}(\{x_1, x_2, \ldots, x_m\})$, such that v is as close as possible to z, as measure by the Euclidean norm $||v - z||_2$:

  $\text{Proj}(z; \{x_1, x_2, \ldots, x_m\}) = \arg\min_{v \in \text{span}(\{x_1, x_2, \ldots, x_m\})} ||v - z||_2.$

Range and Nullspace of a Matrix

- The **range** of the columnspace of the matrix $A \in \mathbb{R}^{m \times n}$, denoted as $\mathcal{R}(A)$, is the span of the columns of A.

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}$$

Range and Nullspace of a Matrix

- The **range** of the columnspace of the matrix $A \in \mathbb{R}^{m\times n}$, denoted as $\mathcal{R}(A)$, is the span of the columns of A.

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}$$

- The **nullspace** of a matrix $A \in \mathbb{R}^{m\times n}$, denoted as $\mathcal{N}(A)$, is the set of all vectors that equal to 0 when multiplied by A:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

The Determinant of a Matrix

- The **determinant** of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and it is denoted as:

$$|A| \quad \text{or} \quad \det(A).$$

The Determinant of a Matrix

- The **determinant** of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and it is denoted as:

$$|A| \quad \text{or} \quad \det(A).$$

- Let $A_{-i-j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix that results from deleting the ith row and jth column of matrix A.

The Determinant of a Matrix

- The **determinant** of a square matrix $A \in \mathbb{R}^{n\times n}$ is a function $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$, and it is denoted as:

$$|A| \quad \text{or} \quad \det(A).$$

- Let $A_{-i-j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix that results from deleting the ith row and jth column of matrix A.
- Then the determinant can algebraically be computed with the following recursive formula:

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{-i-j}| \quad (\text{for any } j \in 1, \ldots, n)$$

The Determinant of a Matrix

- The **determinant** of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and it is denoted as:

$$|A| \quad \text{or} \quad \det(A).$$

- Let $A_{-i-j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix that results from deleting the ith row and jth column of matrix A.
- Then the determinant can algebraically be computed with the following recursive formula:

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{-i-j}| \quad (\text{for any } j \in 1, \ldots, n)$$

- However, this formula has too many terms for matrices bigger than 3×3. Thus for big matrices people hardly ever use it.

The Determinant of a Matrix

- The **determinant** of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and it is denoted as:

$$|A| \quad \text{or} \quad \det(A).$$

- Let $A_{-i-j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix that results from deleting the ith row and jth column of matrix A.
- Then the determinant can algebraically be computed with the following recursive formula:

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{-i-j}| \quad (\text{for any } j \in 1, \ldots, n)$$

- However, this formula has too many terms for matrices bigger than 3×3. Thus for big matrices people hardly ever use it.
- The determinant has a much more intuitive geometric interpretation however. See the the video on determinants in 3Blue1Brown's "Essence of LA".

The Determinant of a Matrix...

- Example:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}, \qquad |A| = 2 \times 5 - 1 \times 3 = 7$$

The Determinant of a Matrix...

- Example:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}, \qquad |A| = 2 \times 5 - 1 \times 3 = 7$$

- Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$.

The Determinant of a Matrix...

- Example:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}, \qquad |A| = 2 \times 5 - 1 \times 3 = 7$$

- Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$.
- Properties:
  - $A = |A^\top|$.
  - $|AB| = |A||B|$.
  - $|A|$ if and only if A is non-invertable.
  - For more see page 14 of CS229's Linear Algebra Review.

Quadratic Forms

- Given a square matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, the scalar $x^\top Ax$ is called a **quadratic form**, and can explicitly be written as:

$$x^\top Ax = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j .$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is:
  - **Positive Definite (PD)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top A x > 0.$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is:
  - **Positive Definite (PD)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top A x > 0.$$
  - **Positive Semidefinite (PSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top A x \geq 0.$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is:
  - **Positive Definite (PD)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
$$x^\top Ax > 0.$$
  - **Positive Semidefinite (PSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
$$x^\top Ax \geq 0.$$
  - **Negative Definite (ND)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
$$x^\top Ax < 0.$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is:
  - **Positive Definite (PD)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
    $$x^\top A x > 0.$$
  - **Positive Semidefinite (PSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
    $$x^\top A x \geq 0.$$
  - **Negative Definite (ND)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
    $$x^\top A x < 0.$$
  - **Negative Semidefinite (NSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
    $$x^\top A x \leq 0.$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is:
  - **Positive Definite (PD)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top Ax > 0.$$
  - **Positive Semidefinite (PSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top Ax \geq 0.$$
  - **Negative Definite (ND)** if for all *nonzero* vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top Ax < 0.$$
  - **Negative Semidefinite (NSD)** if for all vectors $x \in \mathbb{R}^n$, we have:
  $$x^\top Ax \leq 0.$$
- An important property of PD and ND matrices is that they are always full rank, and hence, invertible.

Eigenvalues and Eigenvectors

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A and $x \in \mathbb{R}^n$ is the corresponding **eigenvector** if:
$$Ax = \lambda x, \quad x \neq 0.$$

Eigenvalues and Eigenvectors

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A and $x \in \mathbb{R}^n$ is the corresponding **eigenvector** if:

$$Ax = \lambda x, \quad x \neq 0.$$

- Intuitively, this definition means that eigenvectors are special vectors that when multiplied by A, they just get scaled by a factor of λ (without its direction getting changed).

Finding the Eigenvalues and Eigenvectors

- The equation in the previous slide can be written as:

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

Finding the Eigenvalues and Eigenvectors

- The equation in the previous slide can be written as:

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

- For this equation to have a solution, $(\lambda I - A)$ should have a nonempty nullspace, which is only the case if $(\lambda I - A)$ is not invertible. That is it should have determinant of zero:

$$|(\lambda I - A)| = 0.$$

Finding the Eigenvalues and Eigenvectors

- The equation in the previous slide can be written as:

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

- For this equation to have a solution, $(\lambda I - A)$ should have a nonempty nullspace, which is only the case if $(\lambda I - A)$ is not invertible. That is it should have determinant of zero:

$$|(\lambda I - A)| = 0.$$

- This equation can be expanded into a polynomial in λ with a degree of n. This polynomial is called the characteristic equation of A and its solution gives the eigenvalues.

Finding the Eigenvalues and Eigenvectors

- The equation in the previous slide can be written as:

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

- For this equation to have a solution, $(\lambda I - A)$ should have a nonempty nullspace, which is only the case if $(\lambda I - A)$ is not invertible. That is it should have determinant of zero:

$$|(\lambda I - A)| = 0.$$

- This equation can be expanded into a polynomial in λ with a degree of n. This polynomial is called the characteristic equation of A and its solution gives the eigenvalues.
- After obtaining the eigenvalues, the eigenvectors can easily be obtained by plugging the λ values to the equation at the top of this slide.

Properties of Eigenvalues and Eigenvectors

- The trace of A is equal to the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

Properties of Eigenvalues and Eigenvectors

- The trace of A is equal to the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

- The determinant of A is equal to the product of its eigenvalues:

$$|A| = \prod_{i=1}^{n} \lambda_i.$$

Properties of Eigenvalues and Eigenvectors

- The trace of A is equal to the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i .$$

- The determinant of A is equal to the product of its eigenvalues:

$$|A| = \prod_{i=1}^{n} \lambda_i .$$

- The rank of A is equal to the number of nonzero eigenvalues of A.

Properties of Eigenvalues and Eigenvectors

- The trace of A is equal to the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i .$$

- The determinant of A is equal to the product of its eigenvalues:

$$|A| = \prod_{i=1}^{n} \lambda_i .$$

- The rank of A is equal to the number of nonzero eigenvalues of A.
- The eigenvalues of a diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ are just its diagonal entries.

Properties of Eigenvalues and Eigenvectors

- The trace of A is equal to the sum of its eigenvalues:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

- The determinant of A is equal to the product of its eigenvalues:

$$|A| = \prod_{i=1}^{n} \lambda_i.$$

- The rank of A is equal to the number of nonzero eigenvalues of A.
- The eigenvalues of a diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ are just its diagonal entries.
- For more see page 19 of CS229's Linear Algebra Review.

Eigenvalues and Eigenvectors of Symmetric Matrices

- In general, the structures of the eigenvalues and eigenvectors of a general square matrix can be subtle to characterize.

Eigenvalues and Eigenvectors of Symmetric Matrices

- In general, the structures of the eigenvalues and eigenvectors of a general square matrix can be subtle to characterize.
- Fortunately, in most of the cases in machine learning, it suffices to deal with symmetric real matrices, whose eigenvalues and eigenvectors have remarkable properties.

Eigenvalues and Eigenvectors of Symmetric Matrices

- In general, the structures of the eigenvalues and eigenvectors of a general square matrix can be subtle to characterize.
- Fortunately, in most of the cases in machine learning, it suffices to deal with symmetric real matrices, whose eigenvalues and eigenvectors have remarkable properties.
- However, for the sake of the brevity of the tutorial we will not go into the details of this special case. For more details on this see pages 19-22 of CS229's Linear Algebra Review.

Outline

The Gradient

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes as input a vector $x \in \mathbb{R}^n$ and and returns a scalar value. Then the **gradient** of f with respect to x is the vector of partial derivatives:

$$\nabla_x f(x) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}.$$

The Gradient

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes as input a vector $x \in \mathbb{R}^n$ and and returns a scalar value. Then the **gradient** of f with respect to x is the vector of partial derivatives:
$$\nabla_x f(x) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}.$$
- Note that the size of $\nabla_x f(x)$ is always the same size of x.

The Gradient

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes as input a vector $x \in \mathbb{R}^n$ and and returns a scalar value. Then the **gradient** of f with respect to x is the vector of partial derivatives:
$$\nabla_x f(x) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}.$$
- Note that the size of $\nabla_x f(x)$ is always the same size of x.
- Properties:
  - $\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$.
  - For $c \in \mathbb{R}$, $\nabla_x(cf(x)) = c\nabla_x f(x)$

The Hessian

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes as input a vector $x \in \mathbb{R}^n$ and and returns a scalar value. Then the **Hessian** of f with respect to x is a $n \times n$ matrix of partial derivatives:

$$\nabla_x^2 f(x) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \cdots & \partial^2 f/\partial x_2 \partial x_n \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \ldots & \partial^2 f/\partial x_n^2 \end{bmatrix}.$$

The Hessian

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes as input a vector $x \in \mathbb{R}^n$ and and returns a scalar value. Then the **Hessian** of f with respect to x is a $n \times n$ matrix of partial derivatives:

$$\nabla_x^2 f(x) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \cdots & \partial^2 f/\partial x_2 \partial x_n \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \ldots & \partial^2 f/\partial x_n^2 \end{bmatrix}.$$

- Note that the Hessian is always symmetric as:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

For more on Matrix Calculus

- If not familiar with these concepts, try going over the exercises of taking the gradients and Hessians of linear and quadratic functions in pages 26-27 of CS229's Linear Algebra Review.

For more on Matrix Calculus

- If not familiar with these concepts, try going over the exercises of taking the gradients and Hessians of linear and quadratic functions in pages 26-27 of CS229's Linear Algebra Review.
- Also for more details on
  - least squares
  - gradients of the determinant
  - eigenvalues and optimization

  see pages 27-29 of CS229's Linear Algebra Review.

References

- This tutorial is mainly adapted from Stanford's CS229 Linear Algebra Review. However, it doesn't give all the details. For the details please refer to this source.
- Here are a couple of good references that you might want to check out:
  - 3Blue1Brown's "Essence of Linear Algebra"
  - The legendary Gilbert Strang's Linear Algebra Course