Introduction to Machine Learning and Data Science

Hayk Aprikyan, Hayk Tarkhanyan

What is Machine Learning?

- Machine Learning is a field of artificial intelligence that focuses on the development of algorithms allowing computers to learn and make predictions or decisions based on data.

What is Machine Learning?

- Machine Learning is a field of artificial intelligence that focuses on the development of algorithms allowing computers to learn and make predictions or decisions based on data.
- It's about creating models that can generalize patterns from data.

Machine Learning as a School

Metaphor: A School for Machines

- Imagine building a school for machines.
- Our goal: Teach machines to learn and make decisions autonomously.
- Similar to how children come to school, machines learn from data.
- After training, machines leave with knowledge and competences.

Key Components of the "School"

- **Students (Machines):** Our machines are the students of the school.
- **Teachers (Data):** The data serves as the teachers, providing information for the machines to learn.
- **Curriculum (Algorithms):** Algorithms act as the curriculum, guiding the learning process.
- **Graduation (Model Deployment):** Once trained, machines graduate and are ready to apply their knowledge in real-world scenarios.

Why Machine Learning?

- Automation of complex tasks.

Why Machine Learning?

- Automation of complex tasks.
- Handling large and diverse datasets.

Why Machine Learning?

- Automation of complex tasks.
- Handling large and diverse datasets.
- Making predictions and decisions in real-time.

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)
- Text to speech

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)
- Text to speech
- Autonomous vehicles

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)
- Text to speech
- Autonomous vehicles
- Fraud detection in finance

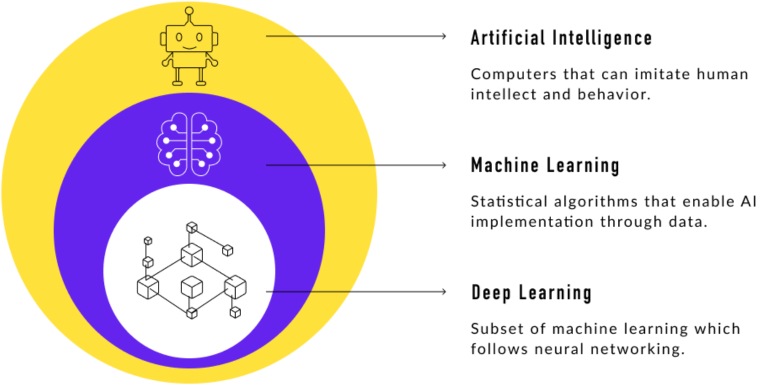
Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)
- Text to speech
- Autonomous vehicles
- Fraud detection in finance
- Medical diagnosis and treatment planning

Examples of Machine Learning Applications

- Email filtering, spam detection (e.g. Gmail)
- Language translation services (e.g. Google Translate)
- Image and speech recognition (e.g. Google Photos, Facebook, Gboard)
- Recommender systems (e.g. Netflix, Amazon)
- Virtual personal assistants (e.g. Siri, Google Assistant)
- Text to speech
- Autonomous vehicles
- Fraud detection in finance
- Medical diagnosis and treatment planning
- & many more...

Difference between AI, ML, and DL



Vectors

Hayk Aprikyan, Hayk Tarkhanyan

- Communication:
    - For questions, memes, announcements, homeworks: Slack
    - For lecture slides and books: Google Drive, GitHub
- Main books (in English):
    - Poole, "Linear Algebra: A Modern Introduction"
    - Johnston, "Introduction to Linear and Matrix Algebra"
    - Stewart, "Calculus"
    - Blitzstein, Hwang, "Introduction to Probability"
    - Grimmett, Welsh, "Probability: An Introduction"
- Supplementary books (in Armenian):
    - Ohanyan, "Probability Theory", Lectures
    - Gevorgyan, Sahakyan, "Algebra and Elements of Mathematical Analysis 11, 12"
    - Musoyan, "Mathematical Analysis", parts 1-2
    - Movsisyan, "Higher Algebra and Number Theory"

Vectors

- Suppose one has 2 x 10 dram, 1 x 20 dram, 2 x 50 dram, and 1 x 200 dram coins in his pocket.

Vectors

- Suppose one has 2 x 10 dram, 1 x 20 dram, 2 x 50 dram, and 1 x 200 dram coins in his pocket.
- How can we denote that?

Vectors

- Suppose one has 2 x 10 dram, 1 x 20 dram, 2 x 50 dram, and 1 x 200 dram coins in his pocket.
- How can we denote that?
- Using a table:

Coins	**Quantity**
10	2
20	1
50	2
100	0
200	1

Vectors

- Suppose one has 2 x 10 dram, 1 x 20 dram, 2 x 50 dram, and 1 x 200 dram coins in his pocket.
- How can we denote that?
- Using a table:

Coins	Quantity
10	2
20	1
50	2
100	0
200	1

- Or by taking the two columns of the table: $\begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$

Vectors

Definition

An ordered set of n real numbers is called a **vector** (or **column vector**) in $\mathbb{R}^n$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_1, v_2, \ldots, v_n$ are the **components** of the vector.
A vector written horizontally is called a **row vector**:

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

We will denote $\mathbf{v} \in \mathbb{R}^n$ to indicate that $\mathbf{v}$ is a vector in $\mathbb{R}^n$.

Vectors

Definition

An ordered set of n real numbers is called a **vector** (or **column vector**) in $\mathbb{R}^n$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_1, v_2, \ldots, v_n$ are the **components** of the vector.
A vector written horizontally is called a **row vector**:

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

We will denote $\mathbf{v} \in \mathbb{R}^n$ to indicate that $\mathbf{v}$ is a vector in $\mathbb{R}^n$.
Vectors in $\mathbb{R}^1$

Vectors

Definition

An ordered set of n real numbers is called a **vector** (or **column vector**) in $\mathbb{R}^n$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_1, v_2, \ldots, v_n$ are the **components** of the vector.
A vector written horizontally is called a **row vector**:

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

We will denote $\mathbf{v} \in \mathbb{R}^n$ to indicate that $\mathbf{v}$ is a vector in $\mathbb{R}^n$.
Vectors in $\mathbb{R}^1$ are real numbers: $[v] \in \mathbb{R}$.

Examples of Vectors in $\mathbb{R}^n$

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad \text{(3-dimensional column vector)}$$

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{(4-dimensional column vector)}$$

$$\mathbf{v}_3 = \begin{bmatrix} -3 \\ 2 \end{bmatrix} \quad \text{(2-dimensional column vector)}$$

$$\mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{(Zero vector in 3-dimensional space)}$$

$$\mathbf{v}_5 = \begin{bmatrix} 1 & -1 & 2 \end{bmatrix} \quad \text{(3-dimensional row vector)}$$

Addition of vectors

Let's denote $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$.

Addition of vectors

Let's denote $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$. $\mathbf{a}, \mathbf{b} \in \mathbb{R}^5$.

Addition of vectors

Let's denote $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$. $\mathbf{a}, \mathbf{b} \in \mathbb{R}^5$. What if someone gave us

3×100 drams and 1×200 drams?

Addition of vectors

Let's denote $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$. $\mathbf{a}, \mathbf{b} \in \mathbb{R}^5$. What if someone gave us 3×100 drams and 1×200 drams? Denote $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 1 \end{bmatrix}$.

Addition of vectors

Let's denote $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$. $\mathbf{a}, \mathbf{b} \in \mathbb{R}^5$. What if someone gave us 3×100 drams and 1×200 drams? Denote $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 1 \end{bmatrix}$.

We would have the following coins:

$$\mathbf{b} + \mathbf{c} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2+0 \\ 1+0 \\ 2+0 \\ 0+3 \\ 1+1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 3 \\ 2 \end{bmatrix}$$

Addition of vectors

Definition

To add two vectors $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ in $\mathbb{R}^n$, add their corresponding components:

$$\mathbf{v} + \mathbf{u} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{bmatrix}$$

Addition of vectors

Definition

To add two vectors $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ in $\mathbb{R}^n$, add their corresponding components:

$$\mathbf{v} + \mathbf{u} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{bmatrix}$$

Note that we can only add two vectors if they are of the same length!

Multiplication of vector by scalar

What if the money in our pockets doubled?

Multiplication of vector by scalar

What if the money in our pockets doubled? We would have:

$$2 \cdot \mathbf{b} = 2 \cdot \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 4 \\ 0 \\ 2 \end{bmatrix}$$

from each coin.

Multiplication of vector by scalar

Definition

To multiply a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ by a scalar c in $\mathbb{R}^n$, multiply each component of the vector by the scalar:

$$c \cdot \mathbf{v} = \begin{bmatrix} c \cdot v_1 \\ c \cdot v_2 \\ \vdots \\ c \cdot v_n \end{bmatrix}$$

Properties of Vectors

Associativity and Commutativity of Vector Addition

For any vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the vector addition is commutative and associative:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

Properties of Vectors

Associativity and Commutativity of Vector Addition

For any vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the vector addition is commutative and associative:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

Associativity and Commutativity of Scalar Multiplication

For any scalar c and vectors $\mathbf{v}$ and $\mathbf{u}$ in $\mathbb{R}^n$, scalar multiplication is associative and commutative:

$$c \cdot (\mathbf{v} + \mathbf{u}) = c \cdot \mathbf{v} + c \cdot \mathbf{u}$$

$$(c \cdot d) \cdot \mathbf{v} = c \cdot (d \cdot \mathbf{v})$$

Vector Subtraction

What if we buy something and spend 2 x 50 drams?

Vector Subtraction

What if we buy something and spend 2 x 50 drams?

Definition

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, the **negative** of $\mathbf{v}$, denoted as $-\mathbf{v}$, is obtained by negating each component:

$$-\mathbf{v} = \begin{bmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{bmatrix}$$

Vector Subtraction

The subtraction of vectors **u** and **v** in $\mathbb{R}^n$ is defined as the sum of **u** and the negative of **v**:

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}) = \begin{bmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{bmatrix}$$

Vector Subtraction

Vector Subtraction

The subtraction of vectors **u** and **v** in $\mathbb{R}^n$ is defined as the sum of **u** and the negative of **v**:

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}) = \begin{bmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{bmatrix}$$

Example

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 2-1 \\ -1-4 \\ 3-0 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ 3 \end{bmatrix}$$

Vector Transposition

Definition

For a column vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, the **transpose**, denoted as $\mathbf{v}^T$, is a row vector:

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

Vector Transposition

Definition

For a column vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, the **transpose**, denoted as $\mathbf{v}^T$, is a row vector:

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

For a row vector $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$ in $\mathbb{R}^n$, the **transpose**, denoted as $\mathbf{u}^T$, is a column vector:

$$\mathbf{u}^T = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Vector Transposition

Examples:

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}$$

Vector Transposition

Examples:

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}$$

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

Vector Transposition

Examples:

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}$$

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

Vector Transposition

Examples:

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}$$

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

Transpose Properties

- For any vector $\mathbf{v}$ in $\mathbb{R}^n$, $(\mathbf{v}^T)^T = \mathbf{v}$
- For any scalar c, $(c \cdot \mathbf{v})^T = c \cdot \mathbf{v}^T$

Dot Product of Vectors

In our example we had $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ coins of $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$ nominations (values) respectively.

Dot Product of Vectors

In our example we had $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ coins of $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$ nominations (values) respectively.

How much money do we have in total?

Dot Product of Vectors

In our example we had $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ coins of $\mathbf{a} = \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix}$ nominations (values) respectively.

How much money do we have in total?

Definition

The **dot product** of two vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$ is:

$$\mathbf{u} \cdot \mathbf{v} = u_1 \cdot v_1 + u_2 \cdot v_2 + \cdots + u_n \cdot v_n$$

Dot Product of Vectors

Example

If $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}$, then:

$$\mathbf{u} \cdot \mathbf{v} = (2 \cdot 1) + (-1 \cdot 4) + (3 \cdot 0) = 2 - 4 + 0 = -2$$

Dot Product of Vectors

Example

If $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}$, then:

$$\mathbf{u} \cdot \mathbf{v} = (2 \cdot 1) + (-1 \cdot 4) + (3 \cdot 0) = 2 - 4 + 0 = -2$$

Going back to our example, we can calculate our money with the dot product of **a** and **b**:

$$\mathbf{a} \cdot \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 20 \\ 50 \\ 100 \\ 200 \end{bmatrix} = 2 \cdot 10 + 1 \cdot 20 + 2 \cdot 50 + 0 \cdot 100 + 1 \cdot 200 = 340$$

Dot Product of Vectors

Remark 1

The dot product of two vectors is defined if and only if the vectors have the same number of components (i.e. are of the same length).

Remark 2

The dot product of two vectors is a *number* (scalar), not a vector.

Dot Product of Vectors

Remark 1

The dot product of two vectors is defined if and only if the vectors have the same number of components (i.e. are of the same length).

Remark 2

The dot product of two vectors is a *number* (scalar), not a vector.

This is why the dot product is often called **scalar product**.

Properties of Dot Product

Properties

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. The dot product has the following properties:

1. Commutativity:
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$$
2. Distributivity over Vector Addition:
$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$$
3. Scalar Multiplication:
$$(c \cdot \mathbf{u}) \cdot \mathbf{v} = c \cdot (\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c \cdot \mathbf{v})$$
4. Non-negativity:
$$\mathbf{u} \cdot \mathbf{u} \geq 0 \text{ and } \mathbf{u} \cdot \mathbf{u} = 0 \text{ if and only if } \mathbf{u} = \mathbf{0}$$

Examples

Consider vectors $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$.

Let's calculate $(5\mathbf{u} - \mathbf{v}) \cdot \mathbf{w}$:

Consider vectors $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$.

Let's calculate $(5\mathbf{u} - \mathbf{v}) \cdot \mathbf{w}$:

$$
\begin{aligned}
(5\mathbf{u} - \mathbf{v}) \cdot \mathbf{w} &= \left(5 \cdot \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} \right) \cdot \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} \\
&= \left(\begin{bmatrix} 5 \\ -10 \\ 15 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} \right) \cdot \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} 5 \\ -14 \\ 16 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} = 5 \cdot (-2) + (-14) \cdot 1 + 16 \cdot 2 = 8
\end{aligned}
$$

Geometric interpretation of vectors

- So far, we were treating vectors in $\mathbb{R}^n$ as lists of numbers only.

Geometric interpretation of vectors

- So far, we were treating vectors in $\mathbb{R}^n$ as lists of numbers only.
- For example: $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, a 2-dimensional column vector.

Geometric interpretation of vectors

- So far, we were treating vectors in $\mathbb{R}^n$ as lists of numbers only.
- For example: $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, a 2-dimensional column vector.

Geometric interpretation of vectors

- So far, we were treating vectors in $\mathbb{R}^n$ as lists of numbers only.
- For example: $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, a 2-dimensional column vector.

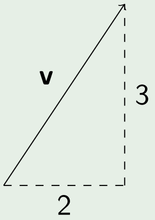
Geometric Interpretation

- In addition to their algebraic representation, vectors have a geometric interpretation.
- We can think of a vector $\mathbf{v}$ as a point in the 2d space,
- Or we can imagine it as an arrow in space, starting from the origin ($O(0,0)$) and pointing to the mentioned point.
- The components of $\mathbf{v}$ are the **coordinates** of the point in the plane.

Geometric interpretation of vectors

Example

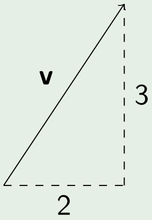
- Consider the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.
- This vector points to the point (2, 3) in the plane.



In general, the vector with coordinates $\begin{bmatrix} x \\ y \end{bmatrix}$ is represented by the point with coordinates (x, y).

Example

- Consider the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.
- This vector points to the point (2, 3) in the plane.

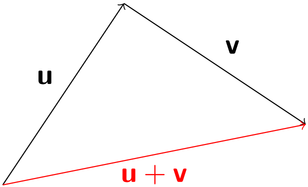


In general, the vector with coordinates $\begin{bmatrix} x \\ y \end{bmatrix}$ is represented by the point with coordinates (x, y).
What do you think happens in the 3d space?

Addition of vectors

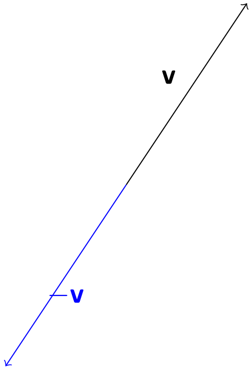
Let's interpret some of our vector operations geometrically.

- **Addition:** To add vectors $\mathbf{u}$ and $\mathbf{v}$, place the tail of $\mathbf{v}$ at the head of $\mathbf{u}$. The sum $\mathbf{u} + \mathbf{v}$ is the vector pointing from the tail of $\mathbf{u}$ to the head of $\mathbf{v}$.



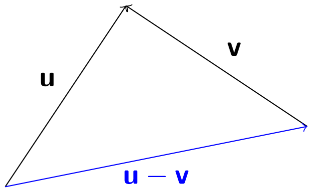
Negative of vectors

- **Negation:** The negative of a vector $\mathbf{v}$, denoted $-\mathbf{v}$, is a vector with the same magnitude but opposite direction.



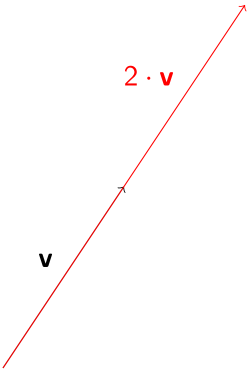
Subtraction of vectors

- **Subtraction:** To subtract $\mathbf{v}$ from $\mathbf{u}$, place the tail of $\mathbf{v}$ at the head of $\mathbf{u}$. The result $\mathbf{u} - \mathbf{v}$ is the vector pointing from the head of $\mathbf{v}$ to the head of $\mathbf{u}$.



Multiplication by scalar

- **Scalar Multiplication:** Scaling a vector **v** by a scalar c stretches or compresses the vector. The result $c \cdot \mathbf{v}$ has the same direction as **v** but a different magnitude.



Example

Example

Let $\mathbf{a} = [3, 2]$ and $\mathbf{b} = [2, 0]$. We want to find $3\mathbf{a} + \mathbf{b}$.

Example

Example

Let $\mathbf{a} = [3, 2]$ and $\mathbf{b} = [2, 0]$. We want to find $3\mathbf{a} + \mathbf{b}$.

Vector Operations

- $3\mathbf{a} + \mathbf{b} = 3 \cdot [3, 2] + [2, 0]$
- $3\mathbf{a} + \mathbf{b} = [9, 6] + [2, 0]$
- $3\mathbf{a} + \mathbf{b} = [11, 6]$

Example

Example

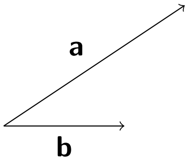
Let $\mathbf{a} = [3, 2]$ and $\mathbf{b} = [2, 0]$. We want to find $3\mathbf{a} + \mathbf{b}$.

Vector Operations

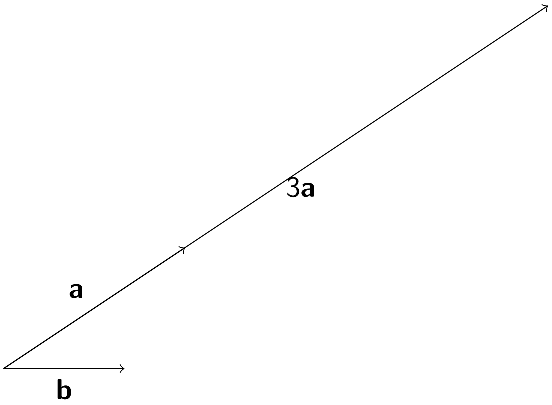
- $3\mathbf{a} + \mathbf{b} = 3 \cdot [3, 2] + [2, 0]$
- $3\mathbf{a} + \mathbf{b} = [9, 6] + [2, 0]$
- $3\mathbf{a} + \mathbf{b} = [11, 6]$

How can we interpret it geometrically?

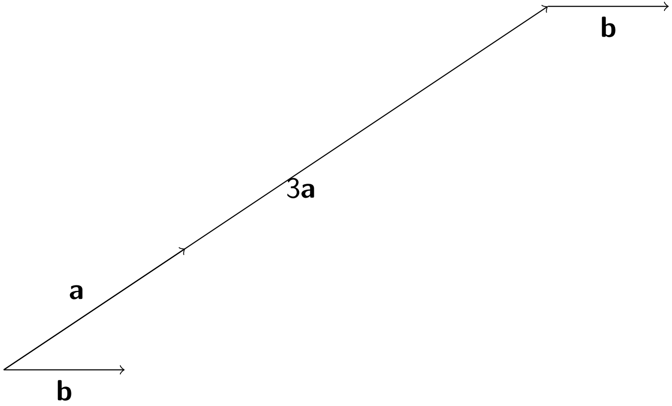
Example



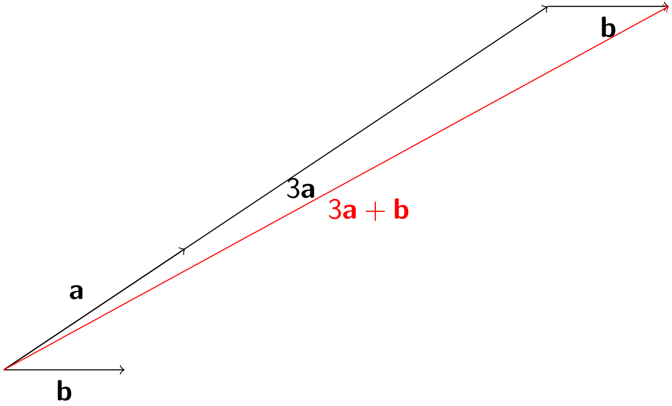
Example



Example



Example

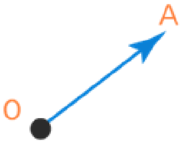


Geometry of Vectors, Matrices

Hayk Aprikyan, Hayk Tarkhanyan

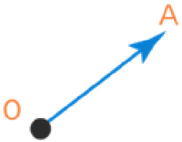
Norm

What if we want to measure the length of some vector?



Norm

What if we want to measure the length of some vector?



What we can say, is that

the length of the vector

$$=$$

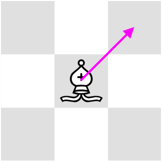
the distance between O and A.

Norm

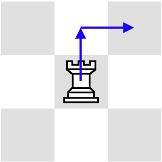
But how to measure distance?

Norm

But how to measure distance?



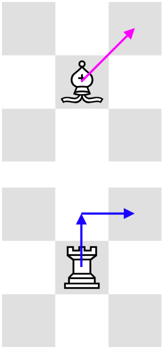
For a bishop, the distance to its upper-right neighbor is 1.



While for a rook, it is 2.

Norm

But how to measure distance?



For a bishop, the distance to its upper-right neighbor is 1.

While for a rook, it is 2.

So there are different ways to measure distance and length.

Norm

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, its **Euclidean norm** or **L2 norm** is:

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

or, equivalently,

$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Norm

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, its **Euclidean norm** or **L2 norm** is:

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

or, equivalently,

$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Example

Let $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. The Euclidean norm of $\mathbf{v}$ is:

$$\|\mathbf{v}\|_2 = \sqrt{3^2 + 4^2} = 5$$

Norm

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, its **Euclidean norm** or **L2 norm** is:

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

or, equivalently,

$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Example

Let $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. The Euclidean norm of $\mathbf{v}$ is:

$$\|\mathbf{v}\|_2 = \sqrt{3^2 + 4^2} = 5$$

Euclidean norm is the standard length we use in classic geometry.
Sometimes we omit the little "2" and just write $\|\mathbf{v}\|$ instead of $\|\mathbf{v}\|_2$.

Norm

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, its **Manhattan norm** or **L1 norm** is:

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \cdots + |v_n|$$

Norm

For a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, its **Manhattan norm** or **L1 norm** is:

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \cdots + |v_n|$$

Example

Let $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. The Manhattan norm of $\mathbf{v}$ is:

$$\|\mathbf{v}\|_1 = |3| + |4| = 7$$

Norm

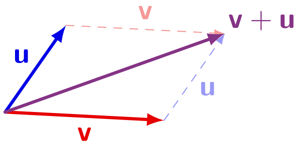
As we have seen, there are different types of norms (=many different ways to calculate the length of a vector), and one of them is chosen depending on the problem.

Norm

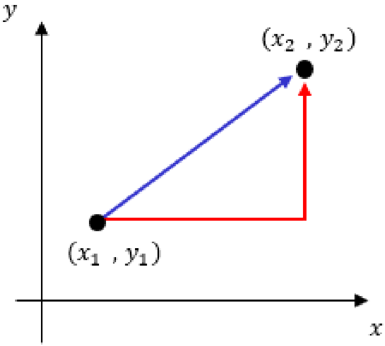
As we have seen, there are different types of norms (=many different ways to calculate the length of a vector), and one of them is chosen depending on the problem.

Notice, however, that independently of which one we take, all norms always satisfy the following three properties:

1. $\|\mathbf{v}\| \geq 0$, and equals 0 if only if $\mathbf{v} = \mathbf{0}$,
2. $\|c\mathbf{v}\| = |c| \cdot \|\mathbf{v}\|$,
3. $\|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{v}\| + \|\mathbf{u}\|$.



Norm



Manhattan Distance L^1

Euclidean Distance L^2

$$L^1 = |x_2 - x_1| + |y_2 - y_1|$$

$$L^2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Angle between vectors

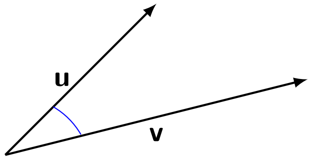
Geometrically, a vector is an arrow in space, that is, it has both a length and direction.

Angle between vectors

Geometrically, a vector is an arrow in space, that is, it has both a length and direction. How do we describe a direction in mathematics?

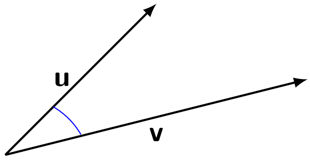
Angle between vectors

Geometrically, a vector is an arrow in space, that is, it has both a length and direction. How do we describe a direction in mathematics? By angles!



Angle between vectors

Geometrically, a vector is an arrow in space, that is, it has both a length and direction. How do we describe a direction in mathematics? By angles!



Remember the formula from high school geometry:
$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \alpha$, where α is the angle between **a** and **b**.

Angle between vectors

Definition

The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the angle $0 \leq \theta \leq \pi$ for which:

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

Angle between vectors

Definition

The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the angle $0 \leq \theta \leq \pi$ for which:

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

Example

Let $\mathbf{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$. Find the angle θ between $\mathbf{u}$ and $\mathbf{v}$.

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|} = \frac{(3 \cdot 7) + (4 \cdot 1)}{\sqrt{3^2 + 4^2} \cdot \sqrt{7^2 + 1^2}} = \frac{25}{\sqrt{25} \cdot \sqrt{50}} = \frac{1}{\sqrt{2}}$$

$$\frac{1}{\sqrt{2}} = \cos\frac{\pi}{4} \quad \Rightarrow \quad \theta = \arccos\frac{1}{\sqrt{2}} = \frac{\pi}{4} = 45^\circ$$

Angle between vectors

Corollary 1

For any vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \cdot \|\mathbf{u}\| \cdot \cos\theta,$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{u}$.

Angle between vectors

Corollary 1

For any vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \cdot \|\mathbf{u}\| \cdot \cos\theta,$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{u}$.

Corollary 2

The dot product of two vectors equals 0 if and only if they are perpendicular to each other (form a 90° angle).

Angle between vectors

Corollary 1

For any vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \cdot \|\mathbf{u}\| \cdot \cos\theta,$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{u}$.

Corollary 2

The dot product of two vectors equals 0 if and only if they are perpendicular to each other (form a 90° angle).

Corollary 3

Any vector $\mathbf{v} \in \mathbb{R}^n$ forms an angle of 0° with itself and 180° with its negative.

Vector Space

Finally, we are left to notice two things. Take, for example,

- the set $D = \{0, 1, 2, \ldots, 9\}$ of digits, and
- the set $P = \{x \in \mathbb{R} : x > 0\}$ of positive numbers.

Vector Space

Finally, we are left to notice two things. Take, for example,

- the set $D = \{0, 1, 2, \ldots, 9\}$ of digits, and
- the set $P = \{x \in \mathbb{R} : x > 0\}$ of positive numbers.

Notice that,

- while the sum of two digits may not be a digit (e.g. $5 + 7 = 12$), the sum of two vectors is *always* a vector,

Vector Space

Finally, we are left to notice two things. Take, for example,

- the set $D = \{0, 1, 2, \ldots, 9\}$ of digits, and
- the set $P = \{x \in \mathbb{R} : x > 0\}$ of positive numbers.

Notice that,

- while the sum of two digits may not be a digit (e.g. $5 + 7 = 12$), the sum of two vectors is *always* a vector,

and

- while the product of a positive number with an arbitrary scalar c may not be positive (e.g. $4 \cdot (-1) = -4$), the product of a vector with a scalar is *always* a vector.

Vector Space

Finally, we are left to notice two things. Take, for example,

- the set $D = \{0, 1, 2, \ldots, 9\}$ of digits, and
- the set $P = \{x \in \mathbb{R} : x > 0\}$ of positive numbers.

Notice that,

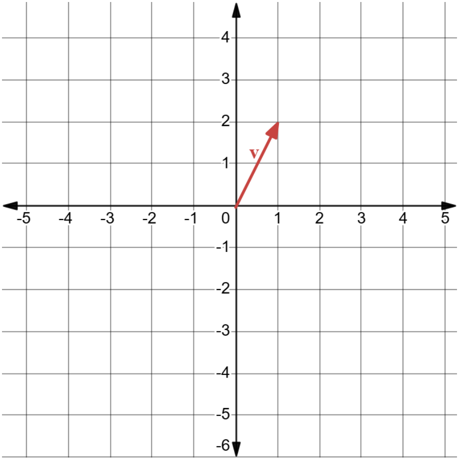
- while the sum of two digits may not be a digit (e.g. $5 + 7 = 12$), the sum of two vectors is *always* a vector,

and

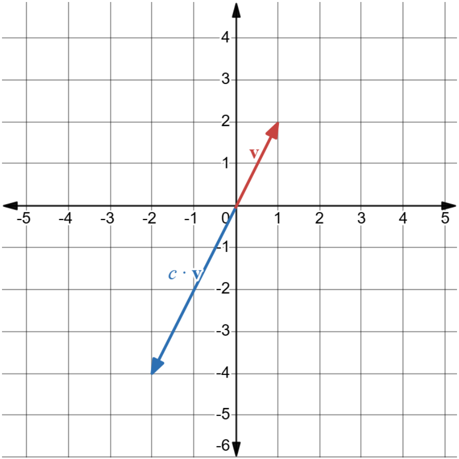
- while the product of a positive number with an arbitrary scalar c may not be positive (e.g. $4 \cdot (-1) = -4$), the product of a vector with a scalar is *always* a vector.

In this case we say that the set of vectors is **closed under addition and scalar multiplication**, while D or P are not (P is closed under addition only).

Furthermore, take the line $y = 2x$ and choose any vector on it:



After multiplying it with any number c, it will still stay on the line $y = 2x$:



Vector Space

Similarly, if we add two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ which both lie on the line $y = 2x$, their sum would again be on the same line.

Vector Space

Similarly, if we add two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ which both lie on the line $y = 2x$, their sum would again be on the same line.

In other words, the line $y = 2x$ is **closed under addition and scalar multiplication**, just like the whole set of vectors $\mathbb{R}^2$. This motivates us to give a special name to the good sets like the line $y = 2x$ and $\mathbb{R}^2$.

Vector Space

Similarly, if we add two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ which both lie on the line $y = 2x$, their sum would again be on the same line.

In other words, the line $y = 2x$ is **closed under addition and scalar multiplication**, just like the whole set of vectors $\mathbb{R}^2$. This motivates us to give a special name to the good sets like the line $y = 2x$ and $\mathbb{R}^2$.

We say that $\mathbb{R}^2$ is a **vector space**, and the set of vectors lying on the line $y = 2x$ are a **vector subspace** of $\mathbb{R}^2$.

Vector Space

Definition

A set V is called a **vector space** if

1. it is closed under addition and scalar multiplication,
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
4. There exists a vector $\mathbf{0}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$
5. For every $\mathbf{v} \in V$, there exists a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
6. $(cd) \cdot \mathbf{v} = c \cdot (d \cdot \mathbf{v})$
7. $1 \cdot \mathbf{v} = \mathbf{v}$
8. $c \cdot (\mathbf{u} + \mathbf{v}) = c \cdot \mathbf{u} + c \cdot \mathbf{v}$
9. $(c + d) \cdot \mathbf{v} = c \cdot \mathbf{v} + d \cdot \mathbf{v}$

No need to memorize the properties–just the natural laws of addition and scalar multiplication.

Vector Space

Definition

A subset U of a vector space V is called a **subspace** of V if U itself is a vector space.

Vector Space

Definition

A subset U of a vector space V is called a **subspace** of V if U itself is a vector space.

How to check if a subset U of V is a sub*space*? Just make sure that it is closed under addition and scalar multiplication!

Definition

A subset U of a vector space V is called a **subspace** of V if U itself is a vector space.

How to check if a subset U of V is a sub*space*? Just make sure that it is closed under addition and scalar multiplication!

Theorem

Assume V is a vector space, and U is a subset of V. Then U is a subspace of V if and only if

1. $\mathbf{x} + \mathbf{y} \in U$, for all $\mathbf{x}, \mathbf{y} \in U$,
2. $c\mathbf{x} \in U$, for all $\mathbf{x} \in U$ and $c \in \mathbb{R}$.

Vector Space

Definition

A subset U of a vector space V is called a **subspace** of V if U itself is a vector space.

How to check if a subset U of V is a sub*space*? Just make sure that it is closed under addition and scalar multiplication!

Theorem

Assume V is a vector space, and U is a subset of V. Then U is a subspace of V if and only if

1. $\mathbf{x} + \mathbf{y} \in U$, for all $\mathbf{x}, \mathbf{y} \in U$,
2. $c\mathbf{x} \in U$, for all $\mathbf{x} \in U$ and $c \in \mathbb{R}$.

- So $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$, ... are all vector spaces.
- The set of all vectors that lie on the same line (e.g. $y = kx$) form a subspace (on the condition that the line also contains the $\mathbf{0}$ vector).

Matrices

Definition

An $m \times n$ tuple A of elements a_{ij} ($i = 1, \ldots, m$ and $j = 1, \ldots, n$), is called a real-valued (m, n) **matrix**:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \qquad a_{ij} \in \mathbb{R}.$$

The set of all real-valued (m, n) matrices is denoted by $\mathbb{R}^{m \times n}$.

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3} \qquad B = \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

Matrices

Definition

An $m \times n$ tuple A of elements a_{ij} ($i = 1, \ldots, m$ and $j = 1, \ldots, n$), is called a real-valued (m, n) **matrix**:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \qquad a_{ij} \in \mathbb{R}.$$

The set of all real-valued (m, n) matrices is denoted by $\mathbb{R}^{m \times n}$.

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3} \qquad B = \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

Note that the first number in (m, n) always shows rows, second: columns.

Matrix Addition

The vectors are practically 1-column matrices: $\mathbb{R}^n = \mathbb{R}^{n\times 1}$. Similar to vectors, we define the following operations with the matrices:

Matrix Addition

The vectors are practically 1-column matrices: $\mathbb{R}^n = \mathbb{R}^{n\times 1}$. Similar to vectors, we define the following operations with the matrices:

Definition

The sum of two matrices A and B, denoted as $A + B$, is obtained by adding corresponding elements. If A is of size $m \times n$ and B is of the same size, then $A + B$ is also of size $m \times n$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

Matrix Addition

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix}$$

$$A + B = \begin{bmatrix} -1 & 2 \\ 4 & 7 \end{bmatrix}$$

Matrix Addition

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix}$$

$$A + B = \begin{bmatrix} -1 & 2 \\ 4 & 7 \end{bmatrix}$$

Remark

Matrix addition is only defined for matrices of the same size.

Scalar Multiplication of a Matrix

Definition

The product of a scalar c and a matrix A, denoted as cA, is obtained by multiplying each element of the matrix by the scalar.

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

$$= \begin{bmatrix} c \cdot a_{11} & c \cdot a_{12} & \dots & c \cdot a_{1n} \\ c \cdot a_{21} & c \cdot a_{22} & \dots & c \cdot a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c \cdot a_{m1} & c \cdot a_{m2} & \dots & c \cdot a_{mn} \end{bmatrix}$$

Scalar Multiplication of a Matrix

Definition

The product of a scalar c and a matrix A, denoted as cA, is obtained by multiplying each element of the matrix by the scalar.

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

$$= \begin{bmatrix} c \cdot a_{11} & c \cdot a_{12} & \dots & c \cdot a_{1n} \\ c \cdot a_{21} & c \cdot a_{22} & \dots & c \cdot a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c \cdot a_{m1} & c \cdot a_{m2} & \dots & c \cdot a_{mn} \end{bmatrix}$$

Scalar multiplication can be performed for any scalar c and any matrix A.

Negative of a Matrix

Definition

The negative of a matrix A, denoted as $-A$, is obtained by changing the sign of each element in the matrix.

$$-A = -\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} -a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & -a_{22} & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ -a_{m1} & -a_{m2} & \dots & -a_{mn} \end{bmatrix}$$

Negative of a Matrix

Definition

The negative of a matrix A, denoted as $-A$, is obtained by changing the sign of each element in the matrix.

$$-A = -\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} -a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & -a_{22} & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ -a_{m1} & -a_{m2} & \dots & -a_{mn} \end{bmatrix}$$

Remark

The negative of a matrix equals (-1) times the matrix.

Matrix Subtraction

Definition

The difference of two matrices A and B, denoted as $A - B$, is obtained by subtracting corresponding elements, or by adding A and $-B$. If A and B are both of size $m \times n$, then $A - B$ is also of size $m \times n$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$A - B = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \dots & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & \dots & a_{2n} - b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} - b_{m1} & a_{m2} - b_{m2} & \dots & a_{mn} - b_{mn} \end{bmatrix}$$

Matrix Subtraction

Definition

The difference of two matrices A and B, denoted as $A - B$, is obtained by subtracting corresponding elements, or by adding A and $-B$. If A and B are both of size $m \times n$, then $A - B$ is also of size $m \times n$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$A - B = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \dots & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & \dots & a_{2n} - b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} - b_{m1} & a_{m2} - b_{m2} & \dots & a_{mn} - b_{mn} \end{bmatrix}$$

Matrix subtraction is only defined for matrices of the same size.

Zero Matrix

Definition

The **zero matrix**, denoted as O or $O_{m \times n}$, is a matrix where all elements are zero.

Zero Matrix

Definition

The **zero matrix**, denoted as O or $O_{m\times n}$, is a matrix where all elements are zero.

Example

$$O_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad O_{3\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Zero Matrix

Definition

The **zero matrix**, denoted as O or $O_{m \times n}$, is a matrix where all elements are zero.

Example

$$O_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad O_{3\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Remark

$A + O = O + A = A$ for any matrix A.

Transpose of a Matrix

Definition

The **transpose** of a matrix A, denoted as A^T, is obtained by swapping its rows and columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix}$$

Transpose of a Matrix

Definition

The **transpose** of a matrix A, denoted as A^T, is obtained by swapping its rows and columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix}$$

Example

$$A = \begin{bmatrix} 7 & 4 & 2 \\ 0 & 1 & -3 \end{bmatrix} \qquad A^T = \begin{bmatrix} 7 & 0 \\ 4 & 1 \\ 2 & -3 \end{bmatrix}$$

Transpose of a Matrix

Definition

The **transpose** of a matrix A, denoted as A^T, is obtained by swapping its rows and columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix}$$

Example

$$A = \begin{bmatrix} 7 & 4 & 2 \\ 0 & 1 & -3 \end{bmatrix} \qquad A^T = \begin{bmatrix} 7 & 0 \\ 4 & 1 \\ 2 & -3 \end{bmatrix}$$

Remark

The transpose of an (m, n) matrix is an (n, m) matrix.

Matrices

Matrices can be added together and multiplied by numbers, and these operations share the same "good" properties (e.g. $A + B = B + A$) with vectors.

Matrices

Matrices can be added together and multiplied by numbers, and these operations share the same "good" properties (e.g. $A + B = B + A$) with vectors.

In that sense, it is not difficult to prove that:

Theorem

For each $m, n \in \mathbb{N}$ the set of real-valued matrices $\mathbb{R}^{m \times n}$ forms a vector space.

Matrix-Vector Multiplication

Definition

Let A be an $m \times n$ matrix and $\mathbf{v}$ be a column vector of size $n \times 1$. The product $A\mathbf{v}$ is a column vector of size $m \times 1$ obtained by multiplying each row of A by the corresponding element of $\mathbf{v}$ and summing the results.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \dots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \dots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \dots + a_{mn}v_n \end{bmatrix}$$

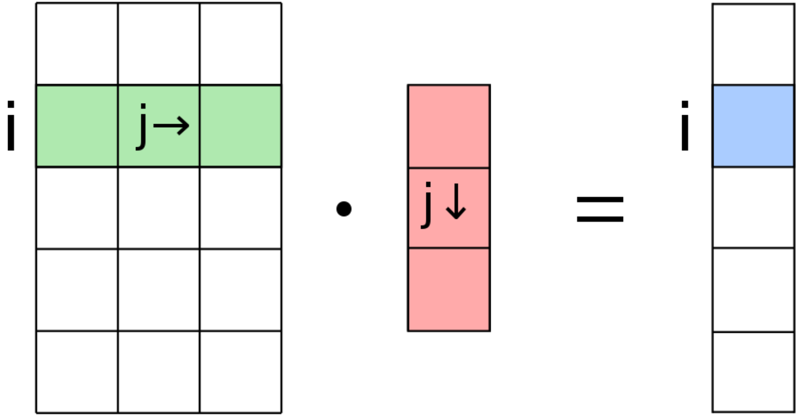
Matrix-Vector Multiplication

Or, in other words, if we denote the rows of A by $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_m$, the product $A\mathbf{v}$ will be a column vector of size $m \times 1$ obtained by taking the dot product of each row of A with the vector $\mathbf{v}$:

$$A = \begin{bmatrix} \ldots & \mathbf{A}_1 & \ldots \\ \ldots & \mathbf{A}_2 & \ldots \\ & \vdots & \\ \ldots & \mathbf{A}_m & \ldots \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \ldots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \ldots + a_{mn}v_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \cdot \mathbf{v} \\ \mathbf{A}_2 \cdot \mathbf{v} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{v} \end{bmatrix}$$

Matrix-Vector Multiplication



Matrix-Vector Multiplication

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) + 3 \cdot 3 \\ 4 \cdot 2 + 5 \cdot (-1) + 6 \cdot 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 21 \end{bmatrix}$$

Matrix-Vector Multiplication

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) + 3 \cdot 3 \\ 4 \cdot 2 + 5 \cdot (-1) + 6 \cdot 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 21 \end{bmatrix}$$

Example

$$A = \begin{bmatrix} -2 & 1 \\ 0 & 3 \\ 1 & -1 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} (-2) \cdot 4 + 1 \cdot 2 \\ 0 \cdot 4 + 3 \cdot 2 \\ 1 \cdot 4 + (-1) \cdot 2 \end{bmatrix} = \begin{bmatrix} -6 \\ 6 \\ 2 \end{bmatrix}$$

Matrix-Vector Multiplication

Matrix-vector multiplication shares properties with scalar multiplication and addition of vectors.

- **Distributive Property:**
  For a matrix A and vectors $\mathbf{v}$ and $\mathbf{w}$ of appropriate sizes:

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w}$$

- **Scalar Multiplication:**
  For a matrix A and a scalar c:

$$A(c\mathbf{v}) = c(A\mathbf{v})$$

Matrix-Vector Multiplication

Matrix-vector multiplication shares properties with scalar multiplication and addition of vectors.

- **Distributive Property:**
  For a matrix A and vectors $\mathbf{v}$ and $\mathbf{w}$ of appropriate sizes:

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w}$$

- **Scalar Multiplication:**
  For a matrix A and a scalar c:

$$A(c\mathbf{v}) = c(A\mathbf{v})$$

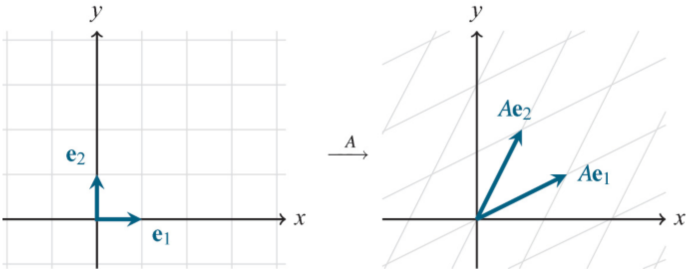
Note that we can only multiply a matrix by a vector if the number of columns of the matrix equals the length of the vector.

Geometric Interpretation

Why do we define the matrix-vector multiplication this way? Turns out, it has a beautiful geometrical interpretation.

Why do we define the matrix-vector multiplication this way? Turns out, it has a beautiful geometrical interpretation.

Think this way: when you multiply, say, a 2×2 matrix A by a vector $\mathbf{v} \in \mathbb{R}^2$, what you get is another vector $\mathbf{u} = A\mathbf{v} \in \mathbb{R}^2$. We call this $\mathbf{u}$ the **transformed version** of $\mathbf{v}$ (and we say that A is a linear transformation).



Geometric Interpretation

As we will see later, the resulting "transformed version" $\mathbf{u}$ is just the same old $\mathbf{v}$ except it is **rotated** and **scaled** to become longer or shorter (and possibly, flipped).

Geometric Interpretation

As we will see later, the resulting "transformed version" $\mathbf{u}$ is just the same old $\mathbf{v}$ except it is **rotated** and **scaled** to become longer or shorter (and possibly, flipped).

In this sense, all matrices are either just rotating vectors by some degree, or flipping them horizontally/vertically, or scale them, or do all three.

The key thing is: whatever a matrix "does" to one vector, it does the same to all other vectors too (when being multiplied with them).

Check different matrices yourself:
- `visualize-it.github.io/linear_transformations/simulation.html`
- `www.shad.io/MatVis`

We will learn more about this later–now back to matrices∼

Matrix Multiplication

Definition

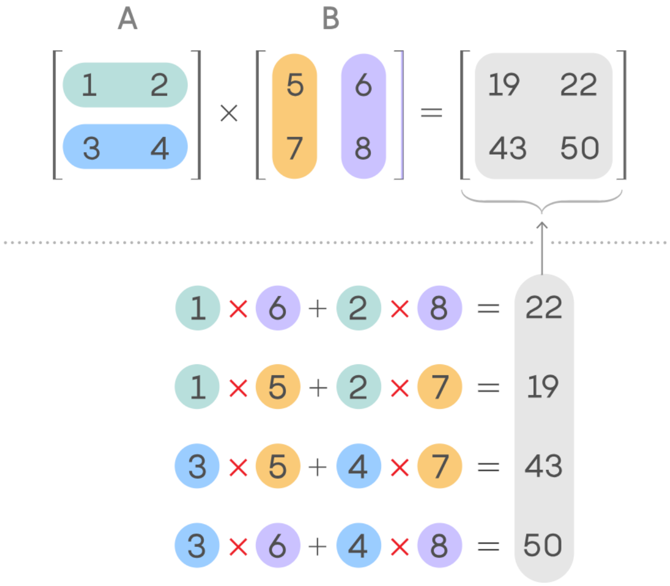
Let A be an $m \times n$ matrix, and let B be an $n \times k$ matrix. The product $C = AB$ is an $m \times k$ matrix, where each element c_{ij} is obtained by taking the dot product of the i-th row of A and the j-th column of B:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{bmatrix}$$

$$C = AB = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mk} \end{bmatrix}$$

$$\text{where } c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj} = \sum_{p=1}^{n} a_{ip}b_{pj}$$

Matrix Multiplication



Matrix Multiplication

Matrix multiplication shares properties with scalar multiplication and addition of vectors, as well as matrix-vector multiplication.

- **Distributive Property:**
  For matrices A, B, and C of appropriate sizes:
  $$A(B + C) = AB + AC \quad \text{and} \quad (A + B)C = AC + BC$$
- **Associativity Property:**
  For matrices A, B, and C of appropriate sizes:
  $$A(BC) = (AB)C$$
- **Scalar Multiplication:**
  For matrices A, B of appropriate sizes and a scalar c:
  $$A(cB) = c(AB) = (cA)B$$

Matrix Multiplication

Matrix multiplication shares properties with scalar multiplication and addition of vectors, as well as matrix-vector multiplication.

- **Distributive Property:**
  For matrices A, B, and C of appropriate sizes:
  $$A(B + C) = AB + AC \quad \text{and} \quad (A + B)C = AC + BC$$
- **Associativity Property:**
  For matrices A, B, and C of appropriate sizes:
  $$A(BC) = (AB)C$$
- **Scalar Multiplication:**
  For matrices A, B of appropriate sizes and a scalar c:
  $$A(cB) = c(AB) = (cA)B$$

Note that we can only multiply two matrices if the number of columns of the first matrix equals the number of rows of the second matrix: $(m \times n)$ with $(n \times k)$.

Matrix Multiplication

Example

Let

$$C = \begin{bmatrix} -1 & 0 \\ 2 & -3 \\ 4 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2} \qquad D = \begin{bmatrix} 5 & -2 & 1 \\ 3 & 0 & 7 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

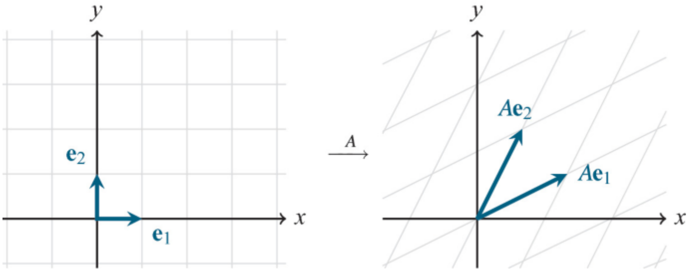
$$CD = \begin{bmatrix} -1\cdot 5 + 0\cdot 3 & -1\cdot(-2) + 0\cdot 0 & -1\cdot 1 + 0\cdot 7 \\ 2\cdot 5 + (-3)\cdot 3 & 2\cdot(-2) + (-3)\cdot 0 & 2\cdot 1 + (-3)\cdot 7 \\ 4\cdot 5 + 1\cdot 3 & 4\cdot(-2) + 1\cdot 0 & 4\cdot 1 + 1\cdot 7 \end{bmatrix}$$

$$= \begin{bmatrix} -5 & 2 & -1 \\ 1 & -4 & -19 \\ 23 & -8 & 11 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

Inverse, Determinant

Hayk Aprikyan, Hayk Tarkhanyan

Recap:
When you multiply, say, a 2×2 matrix A by a vector $\mathbf{v} \in \mathbb{R}^2$, what you get is another vector $\mathbf{u} = A\mathbf{v} \in \mathbb{R}^2$. We call this $\mathbf{u}$ the **transformed version** of $\mathbf{v}$ (and we say that A is a linear transformation).



Geometric Interpretation

As we will see later, the resulting "transformed version" **u** is just the same old **v** except it is **rotated** and **scaled** to become longer or shorter (and possibly, flipped).

Geometric Interpretation

As we will see later, the resulting "transformed version" **u** is just the same old **v** except it is **rotated** and **scaled** to become longer or shorter (and possibly, flipped).

In this sense, all matrices are either just rotating vectors by some degree, or flipping them horizontally/vertically, or scale them, or do all three.

The key thing is: whatever a matrix "does" to one vector, it does the same to all other vectors too (when being multiplied with them).

Check different matrices yourself:

- `visualize-it.github.io/linear_transformations/simulation.html`
- `www.shad.io/MatVis`

Geometric Interpretation

As we will see later, the resulting "transformed version" **u** is just the same old **v** except it is **rotated** and **scaled** to become longer or shorter (and possibly, flipped).

In this sense, all matrices are either just rotating vectors by some degree, or flipping them horizontally/vertically, or scale them, or do all three.

The key thing is: whatever a matrix "does" to one vector, it does the same to all other vectors too (when being multiplied with them).

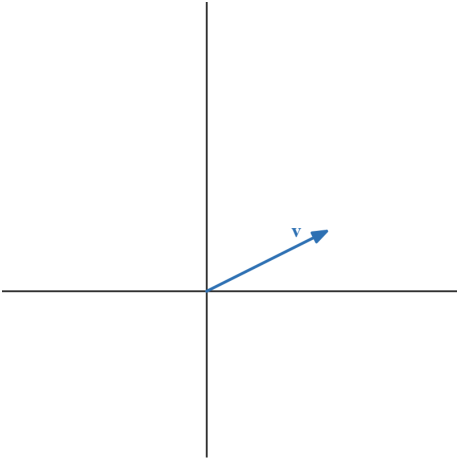
Check different matrices yourself:

- `visualize-it.github.io/linear_transformations/simulation.html`
- `www.shad.io/MatVis`

Now that we know what matrix $\times$ vector multiplication means, what about matrix $\times$ matrix multiplication? Why is it defined the way it is?

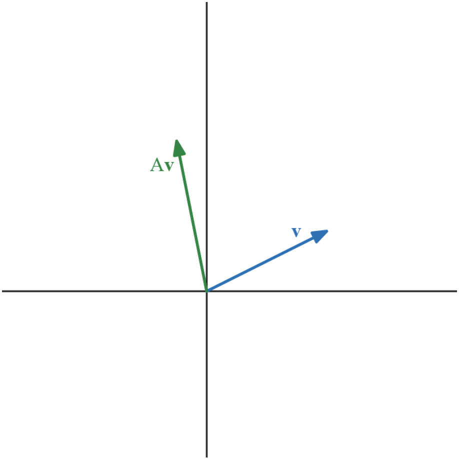
Geometric Interpretation

Suppose $\mathbf{v} \in \mathbb{R}^2$, $A \in \mathbb{R}^{2\times 2}$, $B \in \mathbb{R}^{2\times 2}$:



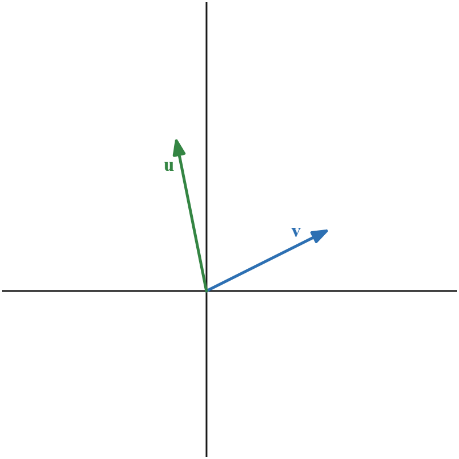
Geometric Interpretation

If we apply A on $\mathbf{v}$, we get a transformed version of $\mathbf{v}$,



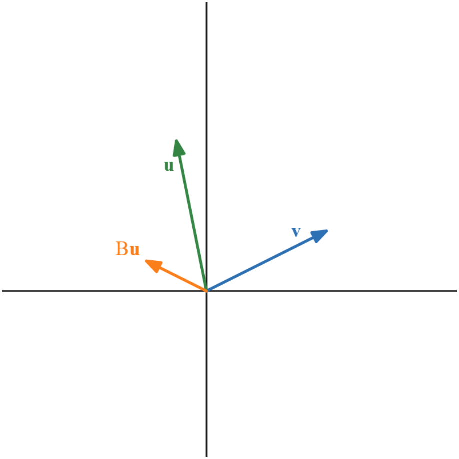
Geometric Interpretation

If we apply A on $\mathbf{v}$, we get a transformed version of $\mathbf{v}$, say $\mathbf{u}$:



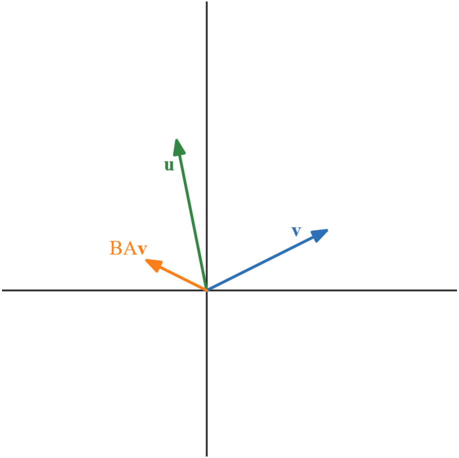
Geometric Interpretation

Now applying B on $\mathbf{u}$, we get a transformed version of $\mathbf{u}$, i.e. $B\mathbf{u}$



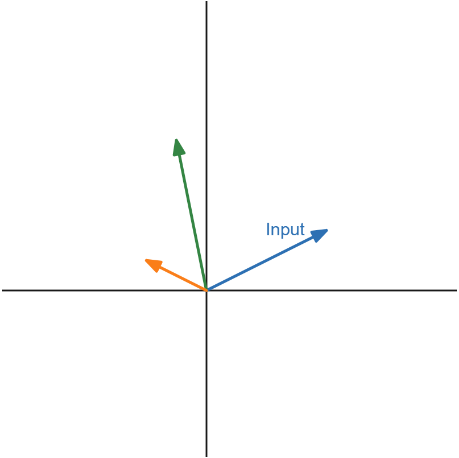
Geometric Interpretation

Now applying B on $\mathbf{u}$, we get a transformed version of $\mathbf{u}$, i.e. $B\mathbf{u} = BA\mathbf{v}$



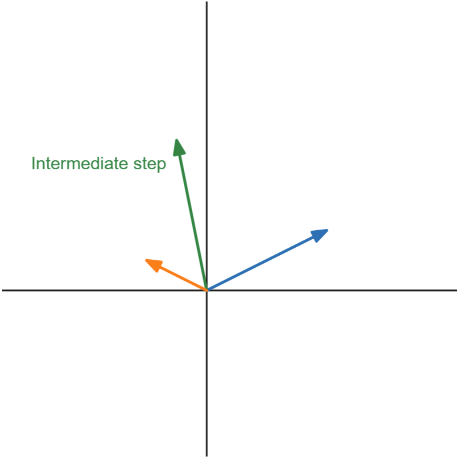
Geometric Interpretation

So what is the product BA? To get $(BA)(\mathbf{v})$, we do:



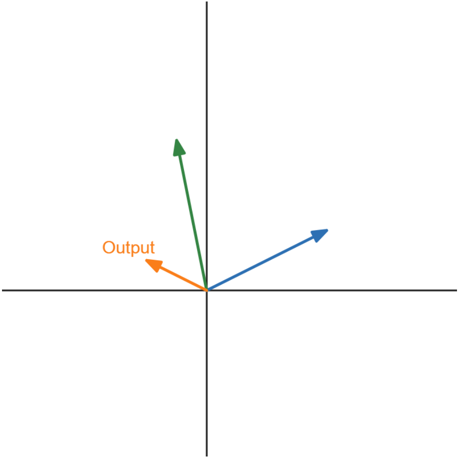
Geometric Interpretation

So what is the product BA? To get $(BA)(\mathbf{v})$, we do:



Geometric Interpretation

So what is the product BA? To get $(BA)(\mathbf{v})$, we do:



Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!

Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!
(Note that this means that often AB is *not the same* as BA)

Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!
(Note that this means that often AB is *not the same* as BA)

Question

Suppose A is the matrix that rotates the vectors by 30°, B the one that rotates by 50°, and C by 260°.

Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!
(Note that this means that often AB is *not the same* as BA)

Question

Suppose A is the matrix that rotates the vectors by 30°, B the one that rotates by 50°, and C by 260°.
What would the product matrix BA be?

Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!
(Note that this means that often AB is *not the same* as BA)

Question

Suppose A is the matrix that rotates the vectors by 30°, B the one that rotates by 50°, and C by 260°.
What would the product matrix BA be? What about CBA?

Geometric Interpretation

Hence BA is the linear transformation that we get by **first applying** A, and **then applying** B!
(Note that this means that often AB is *not the same* as BA)

Question

Suppose A is the matrix that rotates the vectors by 30°, B the one that rotates by 50°, and C by 260°.
What would the product matrix BA be? What about CBA?

Which matrix leaves everything in its place (does not touch anything)?

Identity Matrix

Definition

A matrix is said to be **square** if it has the same number of rows and columns. In other words, an $n \times n$ matrix is a square matrix.

Identity Matrix

Definition

A matrix is said to be **square** if it has the same number of rows and columns. In other words, an $n \times n$ matrix is a square matrix.

Definition

A square matrix A is **symmetric** if $A^T = A$, i.e., the transpose of A is equal to A.

Identity Matrix

Definition

A matrix is said to be **square** if it has the same number of rows and columns. In other words, an $n \times n$ matrix is a square matrix.

Definition

A square matrix A is **symmetric** if $A^T = A$, i.e., the transpose of A is equal to A.

Example

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -3 & 4 \\ 1 & 4 & 6 \end{bmatrix}$$

This matrix is both symmetric and (of course) square.

Identity Matrix

Definition

The **main diagonal** (or just the **diagonal**) of a matrix A are the terms a_{ii} for which the row and column indices are the same (a_{11}, a_{22}, $\ldots$), so from the upper left element to the lower right.

Identity Matrix

Definition

The **main diagonal** (or just the **diagonal**) of a matrix A are the terms a_{ii} for which the row and column indices are the same $(a_{11}, a_{22}, \dots)$, so from the upper left element to the lower right.

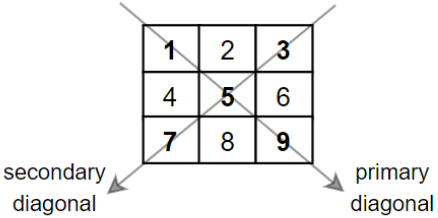
Similarly, the other diagonal from the upper right element to the lower left is called the **secondary diagonal**.

Identity Matrix

Definition

The **main diagonal** (or just the **diagonal**) of a matrix A are the terms a_{ii} for which the row and column indices are the same $(a_{11}, a_{22}, \ldots)$, so from the upper left element to the lower right.

Similarly, the other diagonal from the upper right element to the lower left is called the **secondary diagonal**.



Identity Matrix

For example, here the main diagonal is marked with red:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Identity Matrix

For example, here the main diagonal is marked with red:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The secondary diagonal is marked with red:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Identity Matrix

Definition

The **identity matrix** of $\mathbb{R}^{n \times n}$, denoted as I_n, is the square matrix with ones on the main diagonal and zeros elsewhere.

Applying the identity matrix on vectors does not change them.

Identity Matrix

Definition

The **identity matrix** of $\mathbb{R}^{n\times n}$, denoted as I_n, is the square matrix with ones on the main diagonal and zeros elsewhere.

Applying the identity matrix on vectors does not change them.

Example

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ is the } 3 \times 3 \text{ identity matrix.}$$

Identity Matrix

Definition

The **identity matrix** of $\mathbb{R}^{n\times n}$, denoted as I_n, is the square matrix with ones on the main diagonal and zeros elsewhere.

Applying the identity matrix on vectors does not change them.

Example

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ is the } 3 \times 3 \text{ identity matrix.}$$

Therefore, we can say:

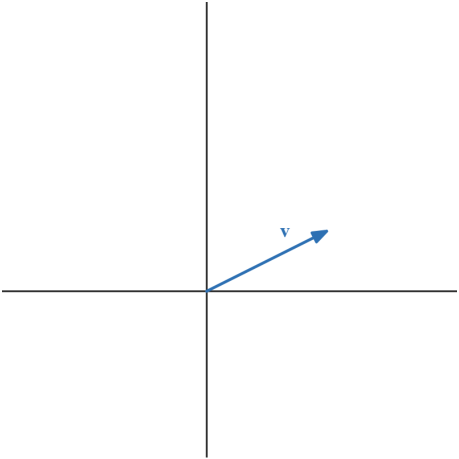
Property

For any matrix $A \in \mathbb{R}^{m\times n}$,

$$I_m A = A I_n = A$$

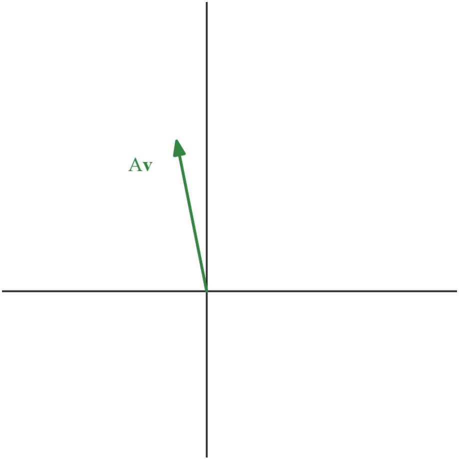
Inverse Matrix

Finally, what if we have a vector in $\mathbb{R}^n$,



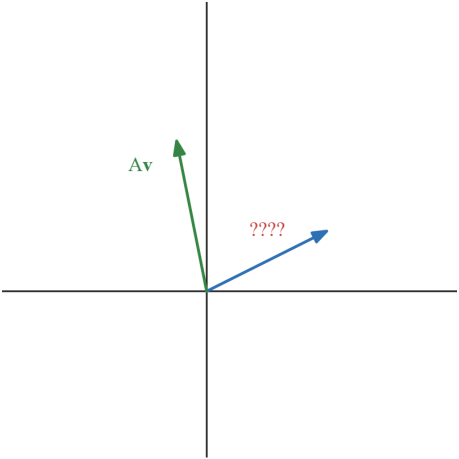
Inverse Matrix

Finally, what if we have a vector in $\mathbb{R}^n$, and we accidentally transform it?



Inverse Matrix

How to get back to the original vector?



Inverse Matrix

In other words, in terms of what we learned about matrix multiplication,

$$\textbf{what} \times A = I \quad ?$$

Inverse Matrix

In other words, in terms of what we learned about matrix multiplication,

$$\textbf{what} \times A = I \quad ?$$

We call that matrix the **inverse** of A, and we denote it by A^{-1}.

Inverse Matrix

In other words, in terms of what we learned about matrix multiplication,

$$A^{-1} \times A = I \quad !$$

We call that matrix the **inverse** of A, and we denote it by A^{-1}.

Inverse Matrix

Question

Assume the matrix $A \in \mathbb{R}^{n \times n}$ does the following when applied on a vector:

1. scales the vector up 2 times in the horizontal direction,
2. then rotates it by 30° clockwise,
3. then squishes it down 3 times in the vertical direction,
4. and then flips it horizontally (around the x-axis) $\sim$

Inverse Matrix

Question

Assume the matrix $A \in \mathbb{R}^{n\times n}$ does the following when applied on a vector:

1. scales the vector up 2 times in the horizontal direction,
2. then rotates it by 30° clockwise,
3. then squishes it down 3 times in the vertical direction,
4. and then flips it horizontally (around the x-axis) $\sim$

Given $\mathbf{v} = A\mathbf{u}$, could we recover the original $\mathbf{u}$?

Inverse Matrix

Question

Assume the matrix $A \in \mathbb{R}^{n \times n}$ does the following when applied on a vector:

1. scales the vector up 2 times in the horizontal direction,
2. then rotates it by 30° clockwise,
3. then squishes it down 3 times in the vertical direction,
4. and then flips it horizontally (around the x-axis) $\sim$

Given $\mathbf{v} = A\mathbf{u}$, could we recover the original $\mathbf{u}$?

The answer is yes, i.e. the matrix A has an inverse. As we will see soon, only some square matrices actually have an inverse.

Trace

Definition

The **trace** of a square matrix A, denoted as $\text{tr}(A)$, is the sum of the elements on its main diagonal.

$$\text{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}$$

Trace

Definition

The **trace** of a square matrix A, denoted as $\mathrm{tr}(A)$, is the sum of the elements on its main diagonal.

$$\mathrm{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}$$

Example

If

$$A = \begin{bmatrix} 2 & 5 & 1 \\ 0 & -3 & 4 \\ 7 & 2 & 6 \end{bmatrix}$$

then

$$\mathrm{tr}(A) = 2 + (-3) + 6 = 5$$

Trace

Definition

The **trace** of a square matrix A, denoted as $\text{tr}(A)$, is the sum of the elements on its main diagonal.

$$\text{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}$$

Example

If

$$A = \begin{bmatrix} 2 & 5 & 1 \\ 0 & -3 & 4 \\ 7 & 2 & 6 \end{bmatrix}$$

then

$$\text{tr}(A) = 2 + (-3) + 6 = 5$$

Note that only square matrices have a trace.

Trace

Trace Properties

For any matrices A and B, and any scalar c, the trace of a matrix satisfies the following properties:

- $\mathrm{tr}(cA) = c \cdot \mathrm{tr}(A)$
- $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
- $\mathrm{tr}(A^T) = \mathrm{tr}(A)$

Determinant of a 2×2 Matrix

Determinant Formula

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the determinant is given by

$$\det(A) = ad - bc$$

Determinant of a 2×2 Matrix

Determinant Formula

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the determinant is given by

$$\det(A) = ad - bc$$

Example

For the matrix

$$A = \begin{bmatrix} 2 & 5 \\ -3 & 4 \end{bmatrix}$$

the determinant is $\det(A) = (2)(4) - (5)(-3) = 8 + 15 = 23$.

Determinant of a 3×3 Matrix

Determinant Formula

For a 3×3 matrix

$$C = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

the determinant is given by

$$\det(C) = aei + bfg + cdh - ceg - bdi - afh$$

Determinant of a 3×3 Matrix

Determinant Formula

For a 3×3 matrix

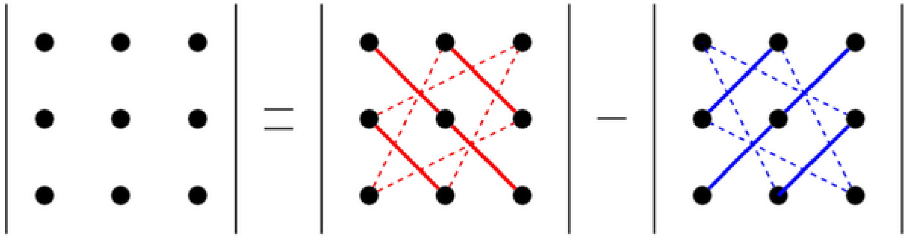
$$C = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

the determinant is given by

$$\det(C) = aei + bfg + cdh - ceg - bdi - afh$$

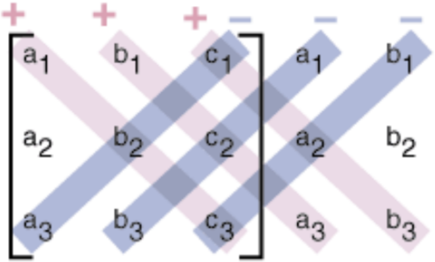
Forget that formula–remember the algorithm!

Determinant of a 3 × 3 Matrix



Determinant of a 3×3 Matrix

Alternatively,



$$\det A = (a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3) - (a_3 b_2 c_1 + b_3 c_2 a_1 + c_3 a_2 b_1)$$

Determinant of a 3×3 Matrix

Example

For the matrix

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

$$\det(C) = 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 3 \cdot 5 \cdot 7 - 2 \cdot 4 \cdot 9 - 1 \cdot 6 \cdot 8 = 0$$

Determinant of a 3×3 Matrix

Example

For the matrix

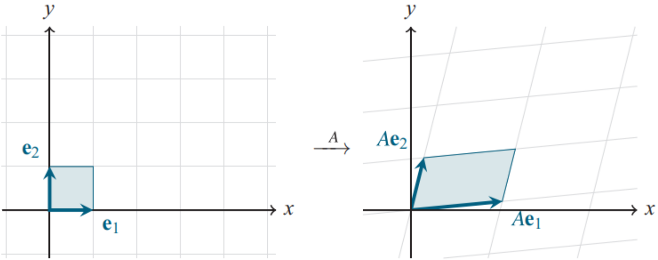
$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\det(C) = 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 3 \cdot 5 \cdot 7 - 2 \cdot 4 \cdot 9 - 1 \cdot 6 \cdot 8 = 0$$

But what does the determinant show, and how do we need it?

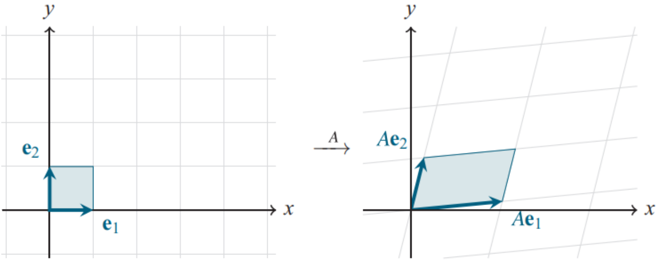
Determinant

If we take, for example, the so-called "unit square" formed by the vectors $\mathbf{e}_1 = [1 \;\; 0]$ and $\mathbf{e}_2 = [0 \;\; 1]$, we can see that their transformed versions, $A\mathbf{e}_1$ and $A\mathbf{e}_2$, form a parallelogram:



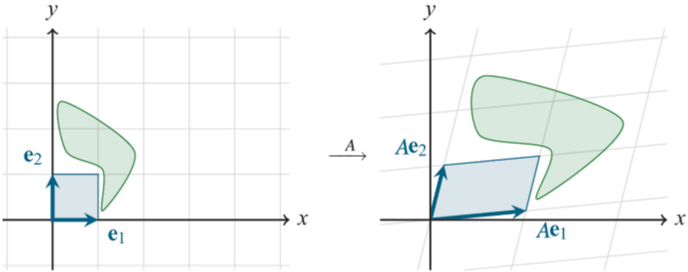
Determinant

If we take, for example, the so-called "unit square" formed by the vectors $\mathbf{e}_1 = [1 \;\; 0]$ and $\mathbf{e}_2 = [0 \;\; 1]$, we can see that their transformed versions, $A\mathbf{e}_1$ and $A\mathbf{e}_2$, form a parallelogram:



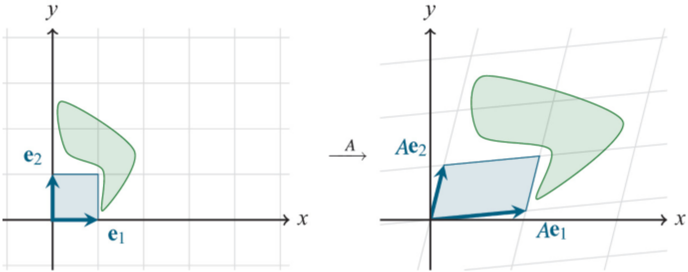
Then $\det(A)$ is the area of that parallelogram.

More generally, after we apply the transformation A (play that animation in your head), the area of *any shape* gets scaled by the factor of $\det(A)$:



Determinant

More generally, after we apply the transformation A (play that animation in your head), the area of *any shape* gets scaled by the factor of $\det(A)$:



So the determinant shows how much the matrix scales up everything in average. Note that it is defined **only** for square matrices.

Determinant

Determinant Properties

Let $A, B \in \mathbb{R}^{n \times n}$ be square matrices of the same size, and let $c \in \mathbb{R}$ be any scalar. Then:

- $\det(cA) = c^n \cdot \det(A)$ *(where n is the size of the matrix)*
- $\det(AB) = \det(A) \cdot \det(B)$ *(multiplicativity)*
- $\det(I) = 1$
- If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A^T) = \det(A)$ *(invariance under transpose)*
- If all numbers on some row or some column of A are zero, then $\det(A) = 0$
- If $\det(A) < 0$, then A flips the space around.

Determinant

Determinant Properties

Let $A, B \in \mathbb{R}^{n \times n}$ be square matrices of the same size, and let $c \in \mathbb{R}$ be any scalar. Then:

- $\det(cA) = c^n \cdot \det(A)$ *(where n is the size of the matrix)*
- $\det(AB) = \det(A) \cdot \det(B)$ *(multiplicativity)*
- $\det(I) = 1$
- If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A^T) = \det(A)$ *(invariance under transpose)*
- If all numbers on some row or some column of A are zero, then $\det(A) = 0$
- If $\det(A) < 0$, then A flips the space around.

It would be an exercise of huge importance to attempt proving these properties (except the last three) by playing the matrices in your head.

Determinant

Finally,

Question

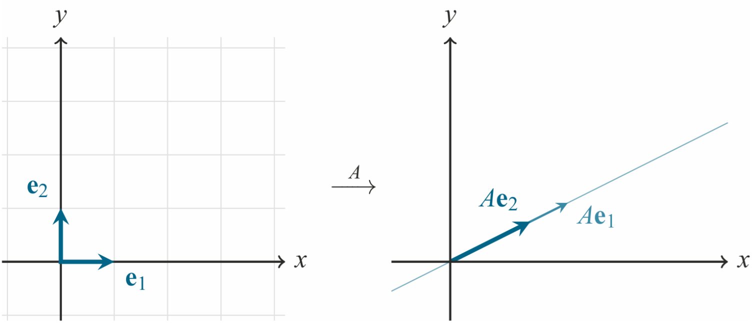
What does it mean if $\det A = 0$?

Determinant

Finally,

Question

What does it mean if $\det A = 0$?

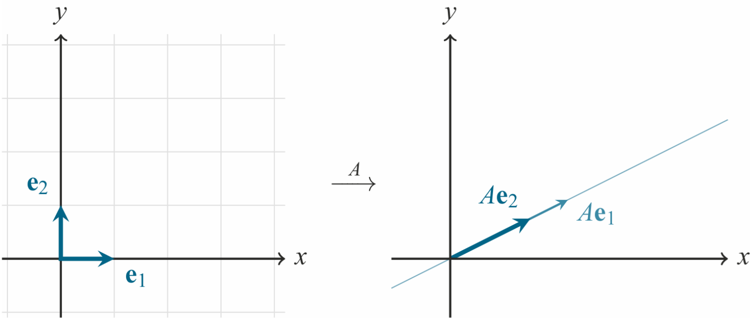


Determinant

Finally,

Question

What does it mean if $\det A = 0$?



Theorem

A square matrix A has an inverse if and only if its determinant is not zero.

Inverse Matrix

Formula for 2x2

For a 2×2 invertible matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse A^{-1} can be calculated using the formula:

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

Inverse Matrix

Formula for 2x2

For a 2×2 invertible matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse A^{-1} can be calculated using the formula:

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

Example

Given $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ with $\det A = (2 \times 4) - (3 \times 1) = 5$, we can calculate the inverse as follows:

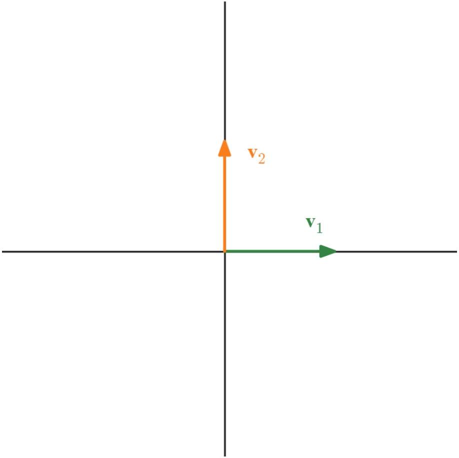
$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

Basis, Eigenvalues and Eigenvectors

Hayk Aprikyan, Hayk Tarkhanyan

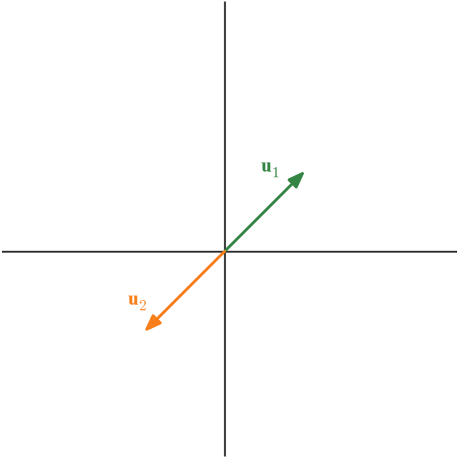
Motivation

When talking about vectors/matrices, why do we focus on these vectors?



Motivation

And not on these:

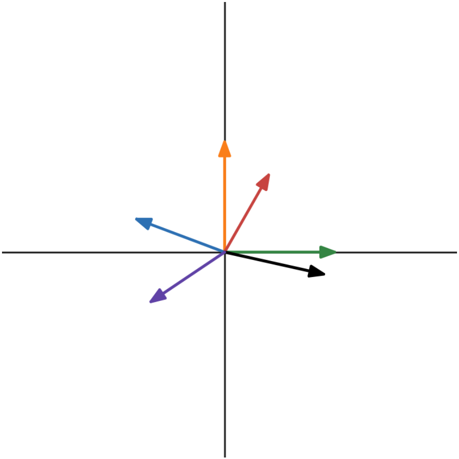


Motivation

Because you can get any vector with $\mathbf{v}_1$ and $\mathbf{v}_2$!

Motivation

Because you can get any vector with $\mathbf{v}_1$ and $\mathbf{v}_2$! Then why not these?



Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \ \ 0]$ and $\mathbf{v}_2 = [0 \ \ 1]$, we can express any vector using only those two.

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \quad 0]$ and $\mathbf{v}_2 = [0 \quad 1]$, we can express any vector using only those two.

For example, the vector $[4 \quad 7]$ can be written as:

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} =$$

Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \quad 0]$ and $\mathbf{v}_2 = [0 \quad 1]$, we can express any vector using only those two.

For example, the vector $[4 \quad 7]$ can be written as:

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} +$$

Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \;\; 0]$ and $\mathbf{v}_2 = [0 \;\; 1]$, we can express any vector using only those two.

For example, the vector $[4 \;\; 7]$ can be written as:

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 7 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} =$$

Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \quad 0]$ and $\mathbf{v}_2 = [0 \quad 1]$, we can express any vector using only those two.

For example, the vector $[4 \quad 7]$ can be written as:

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 7 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4\mathbf{v}_1 + 7\mathbf{v}_2$$

Motivation

Because now they are too much: 2 vectors are enough for $\mathbb{R}^2$.

Now, let's go step-by-step. For the first two vectors $\mathbf{v}_1 = [1 \;\; 0]$ and $\mathbf{v}_2 = [0 \;\; 1]$, we can express any vector using only those two.

For example, the vector $[4 \;\; 7]$ can be written as:

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 7 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4\mathbf{v}_1 + 7\mathbf{v}_2$$

We call expressions like these:

$$\text{something} \cdot \mathbf{v}_1 + \text{something} \cdot \mathbf{v}_2$$

the **linear combinations** of $\mathbf{v}_1$ and $\mathbf{v}_2$.

In our case, the vector $[4 \;\; 7]$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Span

More generally,

Definition

For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and for any scalars $c_1, c_2, \ldots, c_k$, the expression

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k$$

is called a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Span

More generally,

Definition

For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and for any scalars $c_1, c_2, \ldots, c_k$, the expression

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k$$

is called a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

So in this sense, all vectors of $\mathbb{R}^2$ can be written as linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$! In this case, we say that $\mathbb{R}^2$ is the **span** of $\mathbf{v}_1$ and $\mathbf{v}_2$:

Definition

The set of all possible linear combinations of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is called their **span**, i.e.

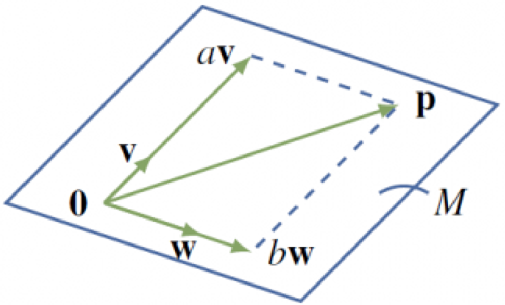
$$\text{span}\{\mathbf{v}_1,\mathbf{v}_2, \ldots, \mathbf{v}_n\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \mid c_1, c_2, \ldots, c_n \in \mathbb{R}\}.$$

Span

Geometrically, the span represents all the vectors that we can get by adding multiples of the given vectors.

Span

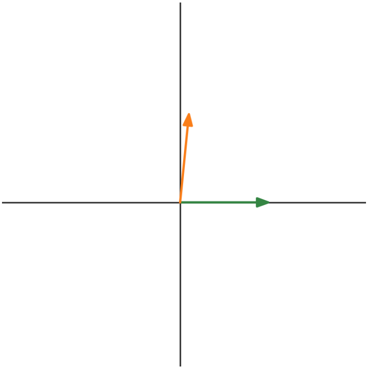
Geometrically, the span represents all the vectors that we can get by adding multiples of the given vectors.



Span

Question

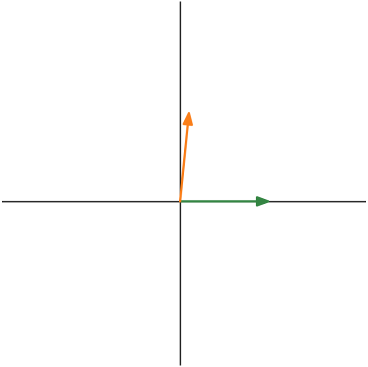
What is the span of vectors $[1 \quad 0]$ and $[0.1 \quad 1]$?



Span

Question

What is the span of vectors $[1 \quad 0]$ and $[0.1 \quad 1]$?

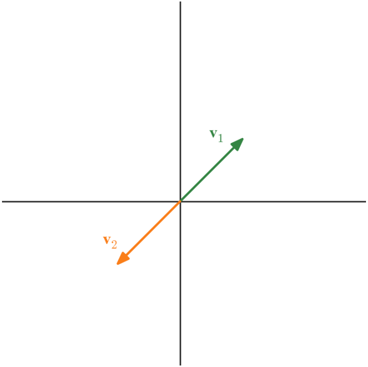


Again, it is the whole $\mathbb{R}^2$: We can express any vector using these two.

Span

Question

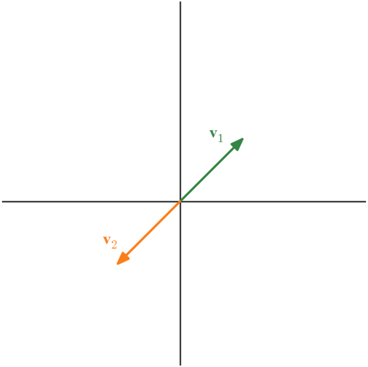
What is the span of vectors $[1 \quad 1]$ and $[-1 \quad -1]$?



Span

Question

What is the span of vectors $[1 \quad 1]$ and $[-1 \quad -1]$?

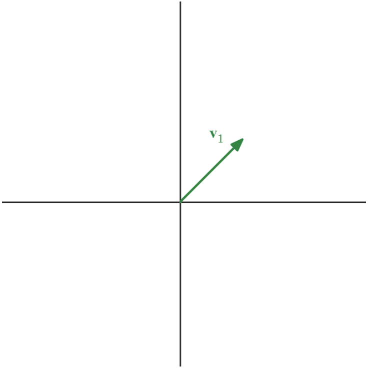


Since they both lie on the line $y = x$, their span is the line $y = x$ itself.

Span

Question

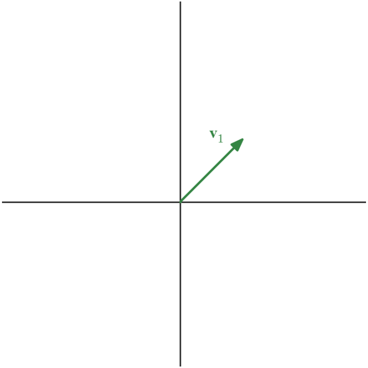
What about the span of the single vector $[1 \quad 1]$?



Span

Question

What about the span of the single vector $[1 \quad 1]$?



Again, the span of $[1 \quad 1]$ is the line $y = x$.

Span

Notice that in all cases so far, the span was either $\mathbb{R}^2$ or some subspace of $\mathbb{R}^2$. Indeed,

Span

Notice that in all cases so far, the span was either $\mathbb{R}^2$ or some subspace of $\mathbb{R}^2$. Indeed,

Theorem

The span of vectors **is a vector space** itself.

Span

Notice that in all cases so far, the span was either $\mathbb{R}^2$ or some subspace of $\mathbb{R}^2$. Indeed,

Theorem

The span of vectors **is a vector space** itself.

But what is the reason that

- in one case (e.g. $\mathbf{v}_1 = [1 \ \ 0]$ and $\mathbf{v}_2 = [0 \ \ 1]$) the span is the whole $\mathbb{R}^2$,
- but in another case (e.g. $\mathbf{u}_1 = [1 \ \ 1]$ and $\mathbf{u}_2 = [-1 \ \ -1]$) the span is only a line?

Span

Notice that in all cases so far, the span was either $\mathbb{R}^2$ or some subspace of $\mathbb{R}^2$. Indeed,

Theorem

The span of vectors **is a vector space** itself.

But what is the reason that

- in one case (e.g. $\mathbf{v}_1 = [1 \;\; 0]$ and $\mathbf{v}_2 = [0 \;\; 1]$) the span is the whole $\mathbb{R}^2$,
- but in another case (e.g. $\mathbf{u}_1 = [1 \;\; 1]$ and $\mathbf{u}_2 = [-1 \;\; -1]$) the span is only a line?

Because in the second case, one of the vectors can be expressed by another!

Linear Independence

Indeed, you can express $\mathbf{u}_2$ with $\mathbf{u}_1$:

$$\mathbf{u}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-1) \cdot \mathbf{u}_1,$$

but you **cannot express** $\mathbf{v}_2$ with $\mathbf{v}_1$ (or vice versa).

Linear Independence

Indeed, you can express $\mathbf{u}_2$ with $\mathbf{u}_1$:

$$\mathbf{u}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-1) \cdot \mathbf{u}_1,$$

but you **cannot express** $\mathbf{v}_2$ with $\mathbf{v}_1$ (or vice versa).

In this case, we say that

- the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are **linearly dependent,**
- while the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are **linearly independent**.

Linear Independence

Indeed, you can express $\mathbf{u}_2$ with $\mathbf{u}_1$:

$$\mathbf{u}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-1) \cdot \mathbf{u}_1,$$

but you **cannot express** $\mathbf{v}_2$ with $\mathbf{v}_1$ (or vice versa).

In this case, we say that

- the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are **linearly dependent,**
- while the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are **linearly independent**.

More generally,

Definition

The vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are called **linearly independent** if none of them can be written as a linear combination of the others.

Indeed, you can express $\mathbf{u}_2$ with $\mathbf{u}_1$:

$$\mathbf{u}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-1) \cdot \mathbf{u}_1,$$

but you **cannot express** $\mathbf{v}_2$ with $\mathbf{v}_1$ (or vice versa).

In this case, we say that

- the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are **linearly dependent,**
- while the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are **linearly independent**.

More generally,

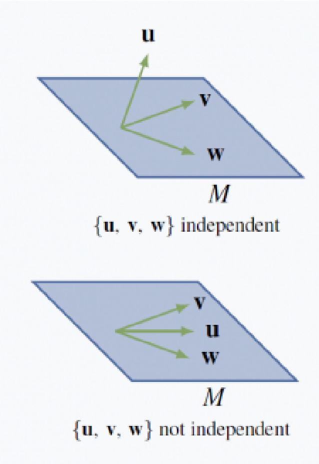
Definition

The vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are called **linearly independent** if none of them can be written as a linear combination of the others.

And we say that they are linearly dependent if one of them, say $\mathbf{v}_n$, can be written as

$$\mathbf{v}_n = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{n-1}\mathbf{v}_{n-1}$$

Linear Independence



Check these animations:

- www.desmos.com/calculator/9rnbn0ycdd
- www.desmos.com/calculator/aje8cboe0j

Linear Independence

Geometrically,

- Two vectors are linearly dependent if they lie on the same line,
- Three vectors are linearly dependent if they lie on the same plane,

and so on.

Linear Independence

Geometrically,

- Two vectors are linearly dependent if they lie on the same line,
- Three vectors are linearly dependent if they lie on the same plane,

and so on.

There is also another characterization of linear independence (try to prove it by yourself):

Theorem

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if and only if the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \mathbf{0}$$

is *only* true if $c_1 = c_2 = \ldots = c_n = 0$.

(i.e. if you plug in any numbers other than 0, the sum will not be $\mathbf{0}$).

Basis

So now we can say that the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are linearly independent and their span is $\mathbb{R}^2$.

We call the pairs of vectors like $\mathbf{v}_1$ and $\mathbf{v}_2$ the **basis** of the space $\mathbb{R}^2$.

Definition

The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are called a **basis** of the vector space V if:

1. $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent,
2. V is equal to the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Basis

So now we can say that the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are linearly independent and their span is $\mathbb{R}^2$.

We call the pairs of vectors like $\mathbf{v}_1$ and $\mathbf{v}_2$ the **basis** of the space $\mathbb{R}^2$.

Definition

The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are called a **basis** of the vector space V if:

1. $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent,
2. V is equal to the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

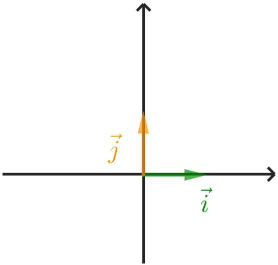
In other words, there are no "irrelevant", "redundant" vectors, and any vector of V can be expressed with $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
(In fact, such representation is always unique, i.e. there is only one way to express $[3 \;\; 4]$ with $\mathbf{v}_1 = [1 \;\; 0]$ and $\mathbf{v}_2 = [0 \;\; 1]$: $\quad 3 \cdot \mathbf{v}_1 + 4 \cdot \mathbf{v}_2$)

Example

The vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{R}^2$ and are called the **standard basis**. They are often denoted $\hat{i}, \hat{j}$.



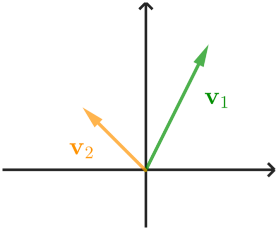
Basis

Example

The linearly independent vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{R}^2$ as these vectors are linearly independent and their span is $\mathbb{R}^2$.

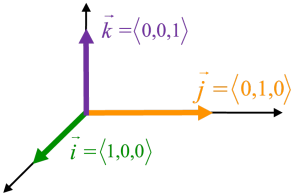


Example

The vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{R}^3$ and are called the **standard basis**. They are often denoted $\hat{i}, \hat{j}, \hat{k}$.



Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are linearly dependent.

Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are linearly dependent.

Example (not enough vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are linearly dependent.

Example (not enough vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^3$ because their span is not the whole $\mathbb{R}^3$.

Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are linearly dependent.

Example (not enough vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^3$ because their span is not the whole $\mathbb{R}^3$.

Example (horrible vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

Basis

Example (too many vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are linearly dependent.

Example (not enough vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^3$ because their span is not the whole $\mathbb{R}^3$.

Example (horrible vectors)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

are not a basis for $\mathbb{R}^2$ because they are dependent, their span is not $\mathbb{R}^2$.

Basis

So far we have noticed that

- There are many different bases for the same space,

Basis

So far we have noticed that

- There are many different bases for the same space,
- All bases have the same number of vectors,

Basis

So far we have noticed that

- There are many different bases for the same space,
- All bases have the same number of vectors,
- Any n independent vectors are a basis for $\mathbb{R}^n$.

Basis

So far we have noticed that

- There are many different bases for the same space,
- All bases have the same number of vectors,
- Any n independent vectors are a basis for $\mathbb{R}^n$.

The same is true for any vector space.

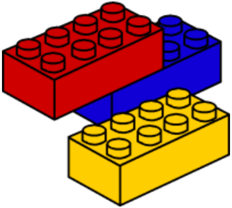
Basis

So far we have noticed that

- There are many different bases for the same space,
- All bases have the same number of vectors,
- Any n independent vectors are a basis for $\mathbb{R}^n$.

The same is true for any vector space.

The basis is our Lego set of "building blocks" out of which we build our castle (i.e. the vector space). Since they all have the same number of vectors, we call that number the **dimension** of the space.



Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- $\ldots$

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- ...
- The dimension of $\mathbb{R}^n$ is n,

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- ...
- The dimension of $\mathbb{R}^n$ is n,
- The dimension of a line is 1,

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- $\ldots$
- The dimension of $\mathbb{R}^n$ is n,
- The dimension of a line is 1,
- The dimension of a plane is 2,

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- ...
- The dimension of $\mathbb{R}^n$ is n,
- The dimension of a line is 1,
- The dimension of a plane is 2,
- etc.

Basis

Definition

The number of vectors in any basis of a vector space V is called the **dimension** of V and is denoted by $\dim V$.

For example,

- The dimension of $\mathbb{R}^1$ is 1,
- The dimension of $\mathbb{R}^2$ is 2,
- $\ldots$
- The dimension of $\mathbb{R}^n$ is n,
- The dimension of a line is 1,
- The dimension of a plane is 2,
- etc.

The dimension describes how **big** our vector space is. It shows how many linearly independent vectors are there in that vector space at most.

Geometric Interpretation of Matrices (last time)

Getting back to our matrices, our main question remains:

Question

How can we tell which transformation a matrix represents by just looking at it?

Geometric Interpretation of Matrices (last time)

Getting back to our matrices, our main question remains:

Question

How can we tell which transformation a matrix represents by just looking at it?

Turns out, the answer is hidden in the concept of basis.

Geometric Interpretation of Matrices (last time)

Consider the matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Geometric Interpretation of Matrices (last time)

Consider the matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Where do you think it takes the basis vectors $\mathbf{e}_1 = [1 \quad 0]$ and $\mathbf{e}_2 = [0 \quad 1]$?

Geometric Interpretation of Matrices (last time)

Consider the matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Where do you think it takes the basis vectors $\mathbf{e}_1 = [1 \quad 0]$ and $\mathbf{e}_2 = [0 \quad 1]$?

$$A\mathbf{e}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Geometric Interpretation of Matrices (last time)

Consider the matrix

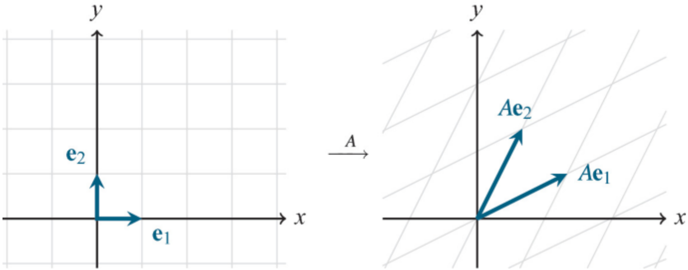
$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Where do you think it takes the basis vectors $\mathbf{e}_1 = [1 \;\; 0]$ and $\mathbf{e}_2 = [0 \;\; 1]$?

$$A\mathbf{e}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$A\mathbf{e}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Geometric Interpretation of Matrices (last time)



As we see, applying a matrix **transforms the basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$** into the **columns of the matrix**:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \text{1st column of } A$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \text{2nd column of } A$$

Geometric Interpretation of Matrices (last time)

Therefore any vector with coordinates $\begin{bmatrix} a \\ b \end{bmatrix}$ is transformed into

$$a \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

i.e. the linear transformation sends our basis vectors to its columns, which become a new basis for our space.

Geometric Interpretation of Matrices (last time)

Therefore any vector with coordinates $\begin{bmatrix} a \\ b \end{bmatrix}$ is transformed into

$$a \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

i.e. the linear transformation sends our basis vectors to its columns, which become a new basis for our space.

More precisely, if the columns of A are linearly independent, then they form a basis.

Definition

The number of linearly independent columns of the matrix A is called the **rank** of A.

Geometric Interpretation of Matrices (last time)

Therefore any vector with coordinates $\begin{bmatrix} a \\ b \end{bmatrix}$ is transformed into

$$a \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

i.e. the linear transformation sends our basis vectors to its columns, which become a new basis for our space.

More precisely, if the columns of A are linearly independent, then they form a basis.

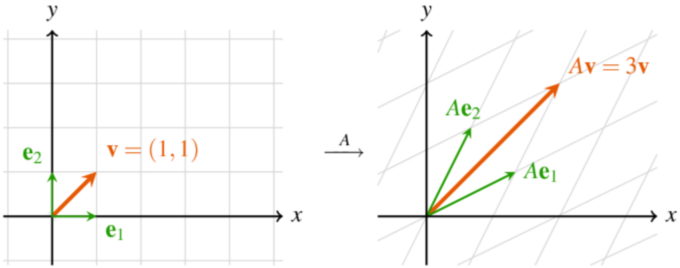
Definition

The number of linearly independent columns of the matrix A is called the **rank** of A.

Finally, peace.

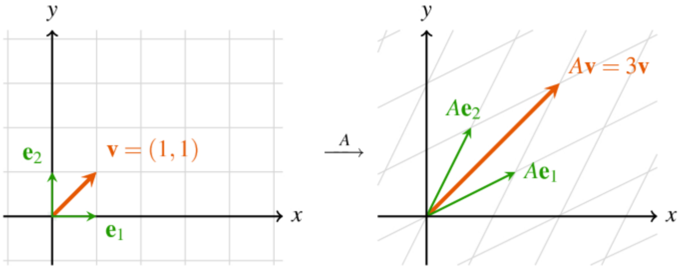
Geometric Interpretation of Matrices (last time)

Sometimes the matrix also has some vectors which **do not change their direction** when being multiplied (transformed) by that matrix, rather they get scaled by some number:



Geometric Interpretation of Matrices (last time)

Sometimes the matrix also has some vectors which **do not change their direction** when being multiplied (transformed) by that matrix, rather they get scaled by some number:



Vectors like these are of special interest to us, and we call them *eigenvectors*.

Definition

If for some number λ and some non-zero vector $\mathbf{v}$

$$A\mathbf{v} = \lambda\mathbf{v}$$

then we say

- λ is an **eigenvalue** of A,
- $\mathbf{v}$ is an **eigenvector** of A corresponding to the eigenvalue λ.

Eigenvalues and Eigenvectors

Definition

If for some number λ and some non-zero vector $\mathbf{v}$

$$A\mathbf{v} = \lambda\mathbf{v}$$

then we say

- λ is an **eigenvalue** of A,
- $\mathbf{v}$ is an **eigenvector** of A corresponding to the eigenvalue λ.

Example

For the matrix $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$ and vector $\mathbf{v} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 20 \\ 4 \end{bmatrix} = 4\mathbf{v}$$

so $\lambda = 4$ is an eigenvalue of A with eigenvector $\mathbf{v}$.

Eigenvalues and Eigenvectors

Definition

If for some number λ and some non-zero vector $\mathbf{v}$

$$A\mathbf{v} = \lambda\mathbf{v}$$

then we say

- λ is an **eigenvalue** of A,
- $\mathbf{v}$ is an **eigenvector** of A corresponding to the eigenvalue λ.

Example

For the matrix $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$ and vector $\mathbf{v} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 20 \\ 4 \end{bmatrix} = 4\mathbf{v}$$

so $\lambda = 4$ is an eigenvalue of A with eigenvector $\mathbf{v}$. What about $-3\mathbf{v}$?

Eigenvalues and Eigenvectors

Remark

If $\mathbf{v}$ is an eigenvector of A, then for any scalar $c \neq 0$, $c\mathbf{v}$ is also an eigenvector for A.

Eigenvalues and Eigenvectors

Remark

If $\mathbf{v}$ is an eigenvector of A, then for any scalar $c \neq 0$, $c\mathbf{v}$ is also an eigenvector for A.

Definition

For $A \in \mathbb{R}^{n \times n}$, the set of all eigenvectors of A associated with an eigenvalue λ, together with the zero vector, is called the **eigenspace** of A with respect to λ and is denoted by E_λ.

Remark

If $\mathbf{v}$ is an eigenvector of A, then for any scalar $c \neq 0$, $c\mathbf{v}$ is also an eigenvector for A.

Definition

For $A \in \mathbb{R}^{n \times n}$, the set of all eigenvectors of A associated with an eigenvalue λ, together with the zero vector, is called the **eigenspace** of A with respect to λ and is denoted by E_λ.

Definition

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

Eigenvalues and Eigenvectors

Remark

If $\mathbf{v}$ is an eigenvector of A, then for any scalar $c \neq 0$, $c\mathbf{v}$ is also an eigenvector for A.

Definition

For $A \in \mathbb{R}^{n \times n}$, the set of all eigenvectors of A associated with an eigenvalue λ, together with the zero vector, is called the **eigenspace** of A with respect to λ and is denoted by E_λ.

Definition

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

How can we find the eigenvalues and eigenvectors of a given matrix?

Eigenvalues and Eigenvectors

Suppose $A \in \mathbb{R}^{n \times n}$ is a matrix, and we want to find $x \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ such that:

$$
\begin{aligned}
A\mathbf{v} = x\mathbf{v} = I(x\mathbf{v}) &= xI\mathbf{v} \\
A\mathbf{v} - xI\mathbf{v} &= \mathbf{0} \\
(A - xI)\mathbf{v} &= \mathbf{0}
\end{aligned}
$$

Eigenvalues and Eigenvectors

Suppose $A \in \mathbb{R}^{n \times n}$ is a matrix, and we want to find $x \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ such that:

$$A\mathbf{v} = x\mathbf{v} = I(x\mathbf{v}) = xI\mathbf{v}$$
$$A\mathbf{v} - xI\mathbf{v} = \mathbf{0}$$
$$(A - xI)\mathbf{v} = \mathbf{0}$$

Since $\mathbf{v} \neq \mathbf{0}$,

$$\det(A - xI) = 0$$

Eigenvalues and Eigenvectors

Suppose $A \in \mathbb{R}^{n \times n}$ is a matrix, and we want to find $x \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ such that:

$$A\mathbf{v} = x\mathbf{v} = I(x\mathbf{v}) = xI\mathbf{v}$$
$$A\mathbf{v} - xI\mathbf{v} = \mathbf{0}$$
$$(A - xI)\mathbf{v} = \mathbf{0}$$

Since $\mathbf{v} \neq \mathbf{0}$,

$$\det(A - xI) = 0$$

The polynomial above is called the **characteristic polynomial** of A. Its roots are the eigenvalues of A.

Eigenvalues and Eigenvectors

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$:

Eigenvalues and Eigenvectors

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$:

$$A - xI =$$

Eigenvalues and Eigenvectors

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$:

$$A - xI = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} - \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} =$$

Eigenvalues and Eigenvectors

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$:

$$A - xI = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} - \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} = \begin{bmatrix} 3-x & 5 \\ 1 & -1-x \end{bmatrix}$$

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$:

$$A - xI = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} - \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} = \begin{bmatrix} 3-x & 5 \\ 1 & -1-x \end{bmatrix}$$

$$p_A(x) = \det(A - xI) = (3-x)(-1-x) - 5 = x^2 - 2x - 8$$

Eigenvalues and Eigenvectors

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$.

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix},$$

Eigenvalues and Eigenvectors

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix}, \quad 4\mathbf{v} = \begin{bmatrix} 4v_1 \\ 4v_2 \end{bmatrix}$$

$$A\mathbf{v} - 4\mathbf{v} = \begin{bmatrix} -v_1 + 5v_2 \\ v_1 - 5v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Eigenvalues and Eigenvectors

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix}, \quad 4\mathbf{v} = \begin{bmatrix} 4v_1 \\ 4v_2 \end{bmatrix}$$

$$A\mathbf{v} - 4\mathbf{v} = \begin{bmatrix} -v_1 + 5v_2 \\ v_1 - 5v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad v_1 = 5v_2$$

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix}, \quad 4\mathbf{v} = \begin{bmatrix} 4v_1 \\ 4v_2 \end{bmatrix}$$

$$A\mathbf{v} - 4\mathbf{v} = \begin{bmatrix} -v_1 + 5v_2 \\ v_1 - 5v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad v_1 = 5v_2$$

Setting $v_2 = a$ and $v_1 = 5a$ for any scalar $a \in \mathbb{R}$, we will get the solution. There are infinite solutions which are all multiplies of each other:

$$\mathbf{v} = a \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

Eigenvalues and Eigenvectors

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix}, \quad 4\mathbf{v} = \begin{bmatrix} 4v_1 \\ 4v_2 \end{bmatrix}$$

$$A\mathbf{v} - 4\mathbf{v} = \begin{bmatrix} -v_1 + 5v_2 \\ v_1 - 5v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad v_1 = 5v_2$$

Setting $v_2 = a$ and $v_1 = 5a$ for any scalar $a \in \mathbb{R}$, we will get the solution. There are infinite solutions which are all multiplies of each other:

$$\mathbf{v} = a \begin{bmatrix} 5 \\ 1 \end{bmatrix} \qquad E_4 = \left\{ a \begin{bmatrix} 5 \\ 1 \end{bmatrix} \mid \text{for any } a \in \mathbb{R} \right\}$$

Eigenvalues and Eigenvectors

Example

Hence, the roots of $p_A(x)$ are $\lambda = 4$ and $\lambda = -2$, so these are the eigenvalues of A. The spectrum of A is $\{4, -2\}$. For $\lambda = 4$,

$$A\mathbf{v} = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 + 5v_2 \\ v_1 - v_2 \end{bmatrix}, \quad 4\mathbf{v} = \begin{bmatrix} 4v_1 \\ 4v_2 \end{bmatrix}$$

$$A\mathbf{v} - 4\mathbf{v} = \begin{bmatrix} -v_1 + 5v_2 \\ v_1 - 5v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad v_1 = 5v_2$$

Setting $v_2 = a$ and $v_1 = 5a$ for any scalar $a \in \mathbb{R}$, we will get the solution. There are infinite solutions which are all multiplies of each other:

$$\mathbf{v} = a \begin{bmatrix} 5 \\ 1 \end{bmatrix} \qquad E_4 = \left\{ a \begin{bmatrix} 5 \\ 1 \end{bmatrix} \mid \text{for any } a \in \mathbb{R} \right\}$$

Similarly we get $E_{-2} = \{a \cdot [-1 \quad 1]^T \mid \text{for any } a \in \mathbb{R}\}$.

Eigenvalues and Eigenvectors

Again, what we are concerned with, is the *concept* and not the computation.

Eigenvalues and Eigenvectors

Again, what we are concerned with, is the *concept* and not the computation.

As a bonus, we have a surprising theorem:

Theorem

The determinant of a matrix is equal to the product of its eigenvalues:

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n$$

and the trace of a matrix is equal to the sum of its eigenvalues:

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$$

Systems of Linear Equations

Let us consider one last application of the matrices.

Question

Imagine scrolling Facebook, when you suddenly see the following problem: You have 2 types of fruits, apples and oranges. You buy 2 apples and 3 oranges for a total cost of 11 dollars. Additionally, you buy 1 apple and 4 oranges for a total cost of 7 dollars.
Only people with 140 IQ can find the prices of apples and oranges.

Systems of Linear Equations

Let us consider one last application of the matrices.

Question

Imagine scrolling Facebook, when you suddenly see the following problem: You have 2 types of fruits, apples and oranges. You buy 2 apples and 3 oranges for a total cost of 11 dollars. Additionally, you buy 1 apple and 4 oranges for a total cost of 7 dollars.
Only people with 140 IQ can find the prices of apples and oranges.

Let x be the cost of one apple and y be the cost of one orange. The problem can be represented as a 2×2 system of linear equations:

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

Solving this system will give us the prices of apples (x) and oranges (y).

Systems of Linear Equations

Definition

A **system of linear equations** is a collection of two or more linear equations (all with the same variables).

Systems of Linear Equations

Definition

A **system of linear equations** is a collection of two or more linear equations (all with the same variables).

A system of m linear equations with n variables can be written as:

$$
\begin{array}{ccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\
\vdots & & \vdots & & \dots & & \vdots & & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m
\end{array}
$$

Systems of Linear Equations

Definition

A **system of linear equations** is a collection of two or more linear equations (all with the same variables).

A system of m linear equations with n variables can be written as:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \dots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}$$

Definition

A **particular solution** to the system is a set of values for the variables $(x_1, x_2, \dots, x_n)$ that satisfies all equations simultaneously. The collection of all particular solutions is called the **general solution**.

Systems of Linear Equations

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

Systems of Linear Equations

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

Systems of Linear Equations

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

implicating

Systems of Linear Equations

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

implicating

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

Systems of Linear Equations

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

implicating

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} =$$

Going back to our example,

$$\begin{cases} 2x + 3y &= 11 \\ x + 4y &= 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

implicating

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ 7 \end{bmatrix}$$

Going back to our example,

$$\begin{cases} 2x + 3y & = 11 \\ x + 4y & = 7 \end{cases}$$

we may notice that

- the first row is the dot product of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,
- the second row is the dot product of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$,

implicating

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ 7 \end{bmatrix}$$

So Facebook is just asking: On which vector should you apply this matrix to get $[11 \quad 7]$?

Systems of Linear Equations

Let's consider three systems of linear equations:

a)

$$\begin{cases} 2x + 3y & = 7 \\ 4x - y & = 5 \end{cases}$$

b)

$$\begin{cases} 2x + 3y & = 7 \\ 4x + 6y & = 14 \end{cases}$$

c)

$$\begin{cases} 2x + 3y & = 7 \\ 4x + 6y & = 15 \end{cases}$$

a)

$$\begin{cases} 2x + 3y & = 7 \\ 4x - y & = 5 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

Systems of Linear Equations

a)

$$\begin{cases} 2x + 3y &= 7 \\ 4x - y &= 5 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) - y = 5 \Rightarrow 7 - 3y - y = 5 \Rightarrow -4y = -2 \Rightarrow y = \frac{1}{2}$$

a)

$$\begin{cases} 2x + 3y &= 7 \\ 4x - y &= 5 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) - y = 5 \Rightarrow 7 - 3y - y = 5 \Rightarrow -4y = -2 \Rightarrow y = \frac{1}{2}$$

$$x = \frac{7 - 3(\frac{1}{2})}{2} = 2$$

a)

$$\begin{cases} 2x + 3y & = 7 \\ 4x - y & = 5 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) - y = 5 \Rightarrow 7 - 3y - y = 5 \Rightarrow -4y = -2 \Rightarrow y = \frac{1}{2}$$

$$x = \frac{7 - 3(\frac{1}{2})}{2} = 2$$

Solution: $x = 2, \quad y = \frac{1}{2}$

Systems of Linear Equations

b)

$$\begin{cases} 2x + 3y & = 7 \\ 4x + 6y & = 14 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

b)

$$\begin{cases} 2x + 3y &= 7 \\ 4x + 6y &= 14 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) + 6y = 14 \Rightarrow 14 - 6y + 6y = 14 \Rightarrow 14 = 14$$

b)

$$\begin{cases} 2x + 3y &= 7 \\ 4x + 6y &= 14 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) + 6y = 14 \Rightarrow 14 - 6y + 6y = 14 \Rightarrow 14 = 14$$

y can be any number

b)

$$\begin{cases} 2x + 3y & = 7 \\ 4x + 6y & = 14 \end{cases}$$

$$2x + 3y = 7 \Rightarrow 2x = 7 - 3y \Rightarrow x = \frac{7 - 3y}{2}$$

$$4\left(\frac{7 - 3y}{2}\right) + 6y = 14 \Rightarrow 14 - 6y + 6y = 14 \Rightarrow 14 = 14$$

y can be any number

Infinite solutions: $x = \frac{7-3y}{2}, \quad \text{for any } y \in \mathbb{R}$

Systems of Linear Equations

c)

$$\begin{cases} 2x + 3y & = 7 \\ 4x + 6y & = 15 \end{cases}$$

Multiplying the first equation by 2 gives:

$$4x + 6y = 14$$

which contradicts the second equation.

c)

$$\begin{cases} 2x + 3y &= 7 \\ 4x + 6y &= 15 \end{cases}$$

Multiplying the first equation by 2 gives:

$$4x + 6y = 14$$

which contradicts the second equation.
No solution.

Systems of Linear Equations

So as we saw, in general a system of linear equations can have a *unique solution*, *no solution*, or *infinitely many solutions*.

Definition

A system of linear equations is **consistent** if it has at least one solution. A system is *inconsistent* if it has no solutions.

Systems of Linear Equations

Consider the system of three linear equations:

$$
\begin{cases}
2x + y - z &= 5 \\
-3x - 2y + 2z &= -8 \\
x + 4y - 3z &= 1
\end{cases}
$$

Systems of Linear Equations

Consider the system of three linear equations:

$$\begin{cases} 2x + y - z & = 5 \\ -3x - 2y + 2z & = -8 \\ x + 4y - 3z & = 1 \end{cases}$$

We can write it in the form:

$$A\mathbf{x} = \mathbf{b},$$

where

$$A = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -2 & 2 \\ 1 & 4 & -3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ -8 \\ 1 \end{bmatrix}$$

Systems of Linear Equations

Consider the system of three linear equations:

$$\begin{cases} 2x + y - z & = 5 \\ -3x - 2y + 2z & = -8 \\ x + 4y - 3z & = 1 \end{cases}$$

We can write it in the form:

$$A\mathbf{x} = \mathbf{b},$$

where

$$A = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -2 & 2 \\ 1 & 4 & -3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ -8 \\ 1 \end{bmatrix}$$

Theorem (very fundamental)

The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for any vector $\mathbf{b} \in \mathbb{R}^n$, if and only if $\det A \neq 0$ (i.e. A is invertible).

Limit, Derivative, Extrema of a Function

Hayk Aprikyan, Hayk Tarkhanyan

Motivation

Suppose you run a supermarket and your profits from the sales of apples vary like this:

$$f(x) = 20\sqrt{x} - 3x^3$$

where x is the price of apples in dollars.

Motivation

Suppose you run a supermarket and your profits from the sales of apples vary like this:

$$f(x) = 20\sqrt{x} - 3x^3$$

where x is the price of apples in dollars.

As a decent businessperson, you want to make your profit as high as possible, so you are interested in the following question:
How much should x be so the profit $f(x)$ is maximized?

Motivation

Suppose you run a supermarket and your profits from the sales of apples vary like this:

$$f(x) = 20\sqrt{x} - 3x^3$$

where x is the price of apples in dollars.

As a decent businessperson, you want to make your profit as high as possible, so you are interested in the following question:
How much should x be so the profit $f(x)$ is maximized?

In real life where you sell more goods rather than apples, the situation looks more complicated.

Motivation

Suppose you run a supermarket and your profits from the sales of apples vary like this:

$$f(x) = 20\sqrt{x} - 3x^3$$

where x is the price of apples in dollars.

As a decent businessperson, you want to make your profit as high as possible, so you are interested in the following question:
How much should x be so the profit $f(x)$ is maximized?

In real life where you sell more goods rather than apples, the situation looks more complicated. For example, your profits look like

$$f(x,y,z,t) = 3xy^2 - y\log t - (1-y)\log(1-t) + \frac{z^3}{t}$$

with real-time values $x = 4$, $y = 0.4$, $z = 0.8$, $t = 55$, and you should decide whether to increase or decrease each of x, y, z, t (and how much).

Motivation

Suppose you run a supermarket and your profits from the sales of apples vary like this:

$$f(x) = 20\sqrt{x} - 3x^3$$

where x is the price of apples in dollars.

As a decent businessperson, you want to make your profit as high as possible, so you are interested in the following question:
How much should x be so the profit $f(x)$ is maximized?

In real life where you sell more goods rather than apples, the situation looks more complicated. For example, your profits look like

$$f(x, y, z, t) = 3xy^2 - y \log t - (1 - y) \log(1 - t) + \frac{z^3}{t}$$

with real-time values $x = 4$, $y = 0.4$, $z = 0.8$, $t = 55$, and you should decide whether to increase or decrease each of x, y, z, t (and how much).
In machine learning you often have $1.000.000+$ such parameters.

Limit of a Sequence

To begin our journey, let's start with the definition of a sequence of numbers.

Limit of a Sequence

To begin our journey, let's start with the definition of a sequence of numbers. A **sequence of numbers** is an unending (infinite)... well, sequence of numbers.

Example

- $1, 2, 3, 4, 5, \ldots$
- $1, -1, 1, -1, 1, \ldots$
- $0, 0.2, 0.4, 0.6, 0.8, \ldots$
- $6, 6, 6, 6, 6, \ldots$

Limit of a Sequence

To begin our journey, let's start with the definition of a sequence of numbers. A **sequence of numbers** is an unending (infinite)... well, sequence of numbers.

Example

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \ldots$
- $6,\ 6,\ 6,\ 6,\ 6,\ \ldots$

We usually fix a letter, say a, and denote the first term by a_1, the second term by a_2, and so on. In general, for the n^{th} term we write a_n, and to denote the whole sequence we use $\{a_n\}$ or $\{a_n\}_{n=1}^{\infty}$.

Limit of a Sequence

To begin our journey, let's start with the definition of a sequence of numbers. A **sequence of numbers** is an unending (infinite)... well, sequence of numbers.

Example

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \ldots$
- $6,\ 6,\ 6,\ 6,\ 6,\ \ldots$

We usually fix a letter, say a, and denote the first term by a_1, the second term by a_2, and so on. In general, for the n^{th} term we write a_n, and to denote the whole sequence we use $\{a_n\}$ or $\{a_n\}_{n=1}^{\infty}$.
Sometimes it also comes in handy to give the formula of the general n^{th} term, e.g. $a_n = n^2$ or $\{a_n\} = \{n^2\}$, which means:

$$a_1 = 1, \qquad a_2 = 4, \qquad a_3 = 9, \qquad \ldots$$

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \frac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \ldots$$

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \frac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \ldots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \dfrac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \ldots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Question

Is there a number to which $\{a_n\}$ approaches (gets closer and closer to)?

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \dfrac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \ldots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Question

Is there a number to which $\{a_n\}$ approaches (gets closer and closer to)?

Indeed, we can say that as $n \to \infty$, the sequence moves closer to 0.

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \dfrac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \dots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Question

Is there a number to which $\{a_n\}$ approaches (gets closer and closer to)?

Indeed, we can say that as $n \to \infty$, the sequence moves closer to 0. We say that 0 is **the limit** of the sequence.

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \frac{1}{n}$:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{5}, \ \ldots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Question

Is there a number to which $\{a_n\}$ approaches (gets closer and closer to)?

Indeed, we can say that as $n \to \infty$, the sequence moves closer to 0. We say that 0 is **the limit** of the sequence.

Question

Does the sequence become equal to 0 at some point?

Limit of a Sequence

There are many interesting examples of sequences. Take, for example, the sequence $a_n = \dfrac{1}{n}$:

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \frac{1}{5},\ \ldots$$

Note that as n becomes larger, the sequence gets smaller and smaller.

Question

Is there a number to which $\{a_n\}$ approaches (gets closer and closer to)?

Indeed, we can say that as $n \to \infty$, the sequence moves closer to 0. We say that 0 is **the limit** of the sequence.

Question

Does the sequence become equal to 0 at some point?

Interestingly, it does not: The numbers come arbitrarily close to 0 but they never actually become 0. This shows that the sequence may or may not eventually equal to its limit.

Limit of a Sequence

Definition

We say that $\{a_n\}$ **converges** to the number L (or that the number L is its **limit**), denoted as

$$\lim_{n\to\infty} a_n = L \qquad (\text{or } a_n \to L)$$

if as n becomes large enough, all terms a_n become arbitrarily close to L.

Limit of a Sequence

Definition

We say that $\{a_n\}$ **converges** to the number L (or that the number L is its **limit**), denoted as

$$\lim_{n\to\infty} a_n = L \qquad (\text{or } a_n \to L)$$

if as n becomes large enough, all terms a_n become arbitrarily close to L.

In other words, if you say:

"Will a_n be close enough to L after some point?"

we will say:

"Sure, how much close do you want it to be?"

Limit of a Sequence

Definition

We say that $\{a_n\}$ **converges** to the number L (or that the number L is its **limit**), denoted as

$$\lim_{n\to\infty} a_n = L \qquad (\text{or } a_n \to L)$$

if as n becomes large enough, all terms a_n become arbitrarily close to L.

In other words, if you say:

"Will a_n be close enough to L after some point?"

we will say:

"Sure, how much close do you want it to be?"

and then whatever number you say (e.g. "not further than 0.002"), we can point out some number N (say, $N = 1000$) such that after the N^{th} term, all others are close to L by 0.002, i.e.

$$|a_N - L| < 0.002, \quad |a_{N+1} - L| < 0.002, \quad |a_{N+2} - L| < 0.002, \quad \ldots$$

Limit of a Sequence

So more technically, $\lim_{n \to \infty} a_n = L$ means that

- for any positive number $\varepsilon > 0$
- there exists N large enough
- such that $|a_n - L| < \varepsilon$ for all $n = N, \; N+1, \; N+2, \; \ldots$

Limit of a Sequence

So more technically, $\lim_{n\to\infty} a_n = L$ means that

- for any positive number $\varepsilon > 0$
- there exists N large enough
- such that $|a_n - L| < \varepsilon$ for all $n = N,\ N+1,\ N+2,\ \ldots$

Definition

Similarly, we say $\lim_{n\to\infty} a_n = +\infty$ if a_n becomes greater than any number after some number of steps.

Limit of a Sequence

So more technically, $\lim_{n\to\infty} a_n = L$ means that

- for any positive number $\varepsilon > 0$
- there exists N large enough
- such that $|a_n - L| < \varepsilon$ for all $n = N, \; N+1, \; N+2, \; \ldots$

Definition

Similarly, we say $\lim_{n\to\infty} a_n = +\infty$ if a_n becomes greater than any number after some number of steps.

Definition

If $\{a_n\}$ has a *finite* limit, we say that it is **convergent**, otherwise it is **divergent**.

Limit of a Sequence

Example

In the previous examples,

- $1, \ 2, \ 3, \ 4, \ 5, \ \ldots$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$ does not have a limit

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \dots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \dots$ does not have a limit
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \dots$

Limit of a Sequence

Example

In the previous examples,

- $1, 2, 3, 4, 5, \ldots \to +\infty$
- $1, -1, 1, -1, 1, \ldots$ does not have a limit
- $0, 0.2, 0.4, 0.6, 0.8, \ldots \to +\infty$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$ does not have a limit
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \ldots \to +\infty$
- $6,\ 6,\ 6,\ 6,\ 6,\ \ldots$

Limit of a Sequence

Example

In the previous examples,

- $1, 2, 3, 4, 5, \ldots \to +\infty$
- $1, -1, 1, -1, 1, \ldots$ does not have a limit
- $0, 0.2, 0.4, 0.6, 0.8, \ldots \to +\infty$
- $6, 6, 6, 6, 6, \ldots \to 6$

Limit of a Sequence

Example

In the previous examples,

- $1, 2, 3, 4, 5, \ldots \to +\infty$
- $1, -1, 1, -1, 1, \ldots$ does not have a limit
- $0, 0.2, 0.4, 0.6, 0.8, \ldots \to +\infty$
- $6, 6, 6, 6, 6, \ldots \to 6$

In general,

- $n^k \to \infty$ for any $k > 0$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$ does not have a limit
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \ldots \to +\infty$
- $6,\ 6,\ 6,\ 6,\ 6,\ \ldots \to 6$

In general,

- $n^k \to \infty$ for any $k > 0$
- $\dfrac{1}{n^k} \to 0$ for any $k > 0$

Limit of a Sequence

Example

In the previous examples,

- $1, 2, 3, 4, 5, \ldots \to +\infty$
- $1, -1, 1, -1, 1, \ldots$ does not have a limit
- $0, 0.2, 0.4, 0.6, 0.8, \ldots \to +\infty$
- $6, 6, 6, 6, 6, \ldots \to 6$

In general,

- $n^k \to \infty$ for any $k > 0$
- $\frac{1}{n^k} \to 0$ for any $k > 0$
- $c^n \to +\infty$ if $c > 1$, but $c^n \to 0$ if $|c| < 1$

Limit of a Sequence

Example

In the previous examples,

- $1,\ 2,\ 3,\ 4,\ 5,\ \ldots \to +\infty$
- $1,\ -1,\ 1,\ -1,\ 1,\ \ldots$ does not have a limit
- $0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ \ldots \to +\infty$
- $6,\ 6,\ 6,\ 6,\ 6,\ \ldots \to 6$

In general,

- $n^k \to \infty$ for any $k > 0$
- $\dfrac{1}{n^k} \to 0$ for any $k > 0$
- $c^n \to +\infty$ if $c > 1$, but $c^n \to 0$ if $|c| < 1$
- If a sequence consists of the same number (or if it becomes constant starting from some point), the limit is that number.

Limit of a Sequence

More examples (we will not go further into details):

Example

Consider the sequence $\{\frac{1}{n}\}$. We claim that $\lim\limits_{n \to \infty} \frac{1}{n} = 0$.

Limit of a Sequence

More examples (we will not go further into details):

Example

Consider the sequence $\{\frac{1}{n}\}$. We claim that $\lim\limits_{n\to\infty} \frac{1}{n} = 0$.

Proof: For any $\varepsilon > 0$, choose N such that $\frac{1}{N} < \varepsilon$. Then, for all $n \geq N$, we have

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \varepsilon.$$

Therefore, the sequence converges to 0, as n approaches infinity.

Limit of a Sequence

More examples (we will not go further into details):

Example

Consider the sequence $\{\frac{1}{n}\}$. We claim that $\lim_{n\to\infty} \frac{1}{n} = 0$.

Proof: For any $\varepsilon > 0$, choose N such that $\frac{1}{N} < \varepsilon$. Then, for all $n \geq N$, we have

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \varepsilon.$$

Therefore, the sequence converges to 0, as n approaches infinity.

Example

Consider the sequence $\left\{(\frac{2}{3})^n\right\}$. $\lim_{n\to\infty} (\frac{2}{3})^n = 0$.

Limit of a Sequence

More examples (we will not go further into details):

Example

Consider the sequence $\{\frac{1}{n}\}$. We claim that $\lim\limits_{n\to\infty} \frac{1}{n} = 0$.

Proof: For any $\varepsilon > 0$, choose N such that $\frac{1}{N} < \varepsilon$. Then, for all $n \geq N$, we have

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \varepsilon.$$

Therefore, the sequence converges to 0, as n approaches infinity.

Example

Consider the sequence $\left\{\left(\frac{2}{3}\right)^n\right\}$. $\lim\limits_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$.

Example

Consider the sequence $\{0.3n\}$. $\lim\limits_{n\to\infty} 0.3n = \infty$ (it is divergent).

Limit of a Sequence

Properties

1. If $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = M$, then

$$\lim_{n\to\infty} (a_n + b_n) = L + M$$

$$\lim_{n\to\infty} (a_n - b_n) = L - M$$

$$\lim_{n\to\infty} (a_n \cdot b_n) = L \cdot M$$

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{L}{M} \qquad (\text{if } M \neq 0)$$

Limit of a Sequence

Properties

1. If $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = M$, then

$$\lim_{n\to\infty} (a_n + b_n) = L + M$$

$$\lim_{n\to\infty} (a_n - b_n) = L - M$$

$$\lim_{n\to\infty} (a_n \cdot b_n) = L \cdot M$$

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{L}{M} \qquad (\text{if } M \neq 0)$$

2. If $\lim_{n\to\infty} a_n = L$, then for any constant c,

$$\lim_{n\to\infty} (c \cdot a_n) = c \cdot L.$$

Limit of a Sequence

Properties

1. If $\lim\limits_{n\to\infty} a_n = L$ and $\lim\limits_{n\to\infty} b_n = M$, then

$$\lim_{n\to\infty} (a_n + b_n) = L + M$$

$$\lim_{n\to\infty} (a_n - b_n) = L - M$$

$$\lim_{n\to\infty} (a_n \cdot b_n) = L \cdot M$$

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{L}{M} \qquad \text{(if } M \neq 0\text{)}$$

2. If $\lim\limits_{n\to\infty} a_n = L$, then for any constant c,

$$\lim_{n\to\infty} (c \cdot a_n) = c \cdot L.$$

3. If $a_n \leq b_n \leq c_n$ for all n and $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n = L$, then $\lim\limits_{n\to\infty} b_n = L$.

Limit of a Function

Now that we have the notion of

$$\lim_{n\to\infty}\left(2+\frac{1}{n}\right)^3$$

what do you think the expression

$$\lim_{x\to 0}(2+x)^3$$

would mean?

Limit of a Function

Now that we have the notion of

$$\lim_{n\to\infty}\left(2+\frac{1}{n}\right)^3$$

what do you think the expression

$$\lim_{x\to 0}(2+x)^3$$

would mean?

Similarly to sequences, we can define the limit of the above expression, i.e. of the function

$$f(x)=(2+x)^3,$$

as x approaches 0.

How do we do that?

Limit of a Function

Now that we have the notion of

$$\lim_{n\to\infty}\left(2+\frac{1}{n}\right)^3$$

what do you think the expression

$$\lim_{x\to 0}(2+x)^3$$

would mean?

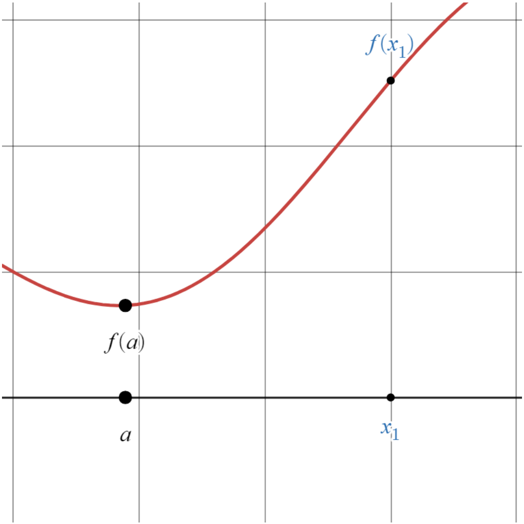
Similarly to sequences, we can define the limit of the above expression, i.e. of the function

$$f(x)=(2+x)^3,$$

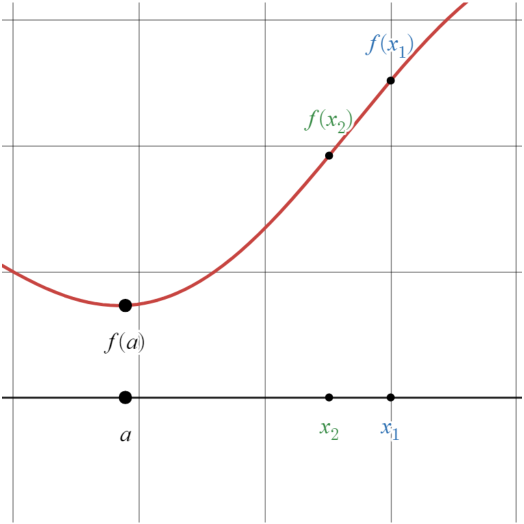
as x approaches 0.

How do we do that? We can say: take any sequence x_n that converges to a, calculate the values of $f(x)$ at $x_1, x_2, \ldots,$ and see what happens.

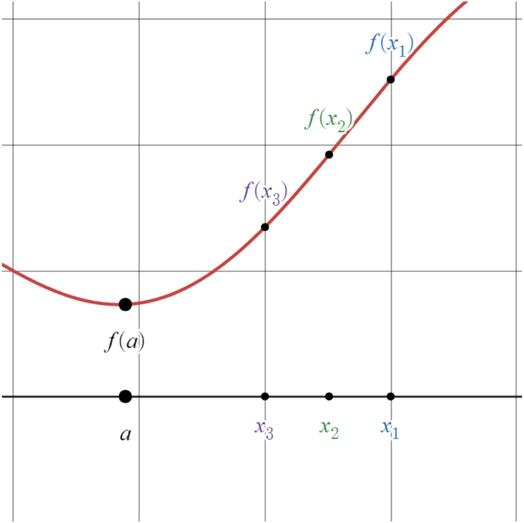
Limit of a Function



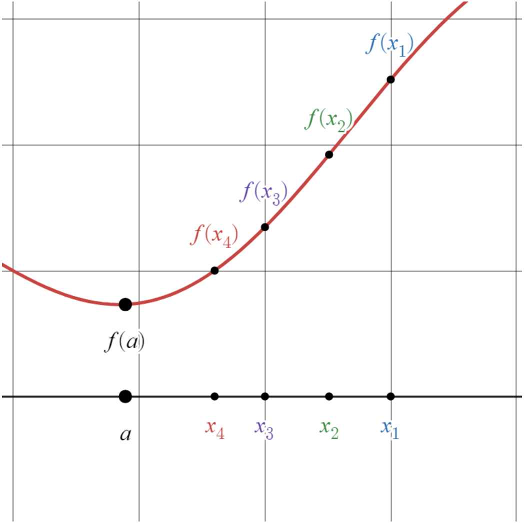
Limit of a Function



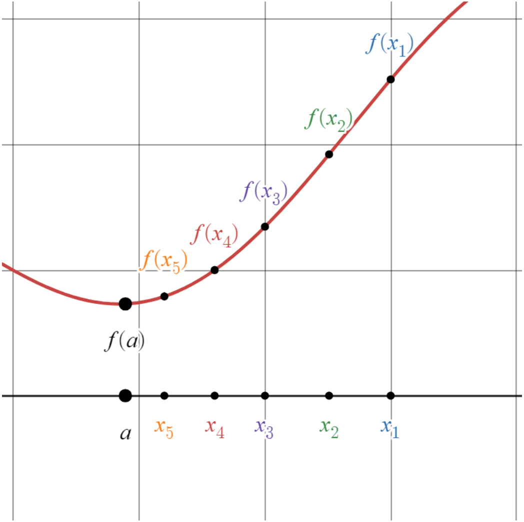
Limit of a Function

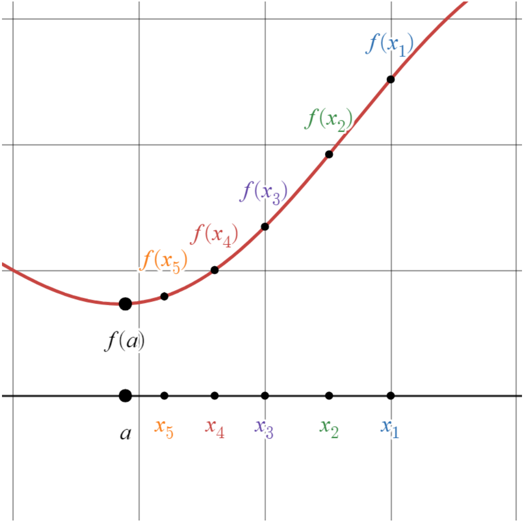


Limit of a Function



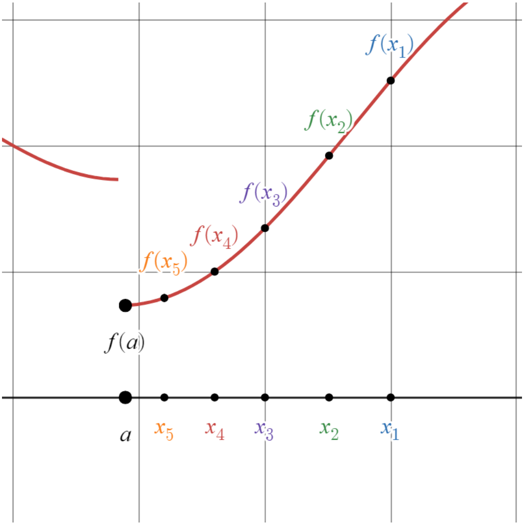
Limit of a Function





If numbers $f(x_1)$, $f(x_2)$, $\ldots$ approach some limit L, then $\lim_{x \to a} f(x) = L$

Limit of a Function



It may happen that if $x_n \to a$ from the other side, we get another "limit".

Limit of a Function

In the second case, we say that the limit does not exist and the function is **discontinuous** at that point.

In the second case, we say that the limit does not exist and the function is **discontinuous** at that point.

Definition

If for **all** sequences $x_n \to a$ (no matter from left or right), the sequence

$$f(x_1), \quad f(x_2), \quad f(x_3), \quad \ldots, \quad f(x_n), \quad \ldots$$

converges to a certain number L, then we say

$$\lim_{x \to a} f(x) = L$$

Limit of a Function

In the second case, we say that the limit does not exist and the function is **discontinuous** at that point.

Definition

If for **all** sequences $x_n \to a$ (no matter from left or right), the sequence

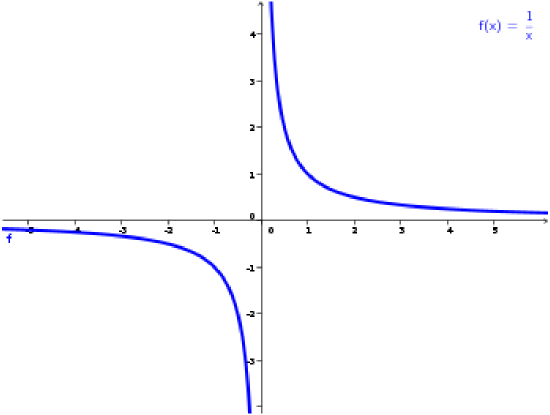
$$f(x_1),\ \ f(x_2),\ \ f(x_3),\ \ \ldots,\ \ f(x_n),\ \ \ldots$$

converges to a certain number L, then we say

$$\lim_{x \to a} f(x) = L$$

In other words, if the value of $f(x)$ always approaches L, as its input approaches a.

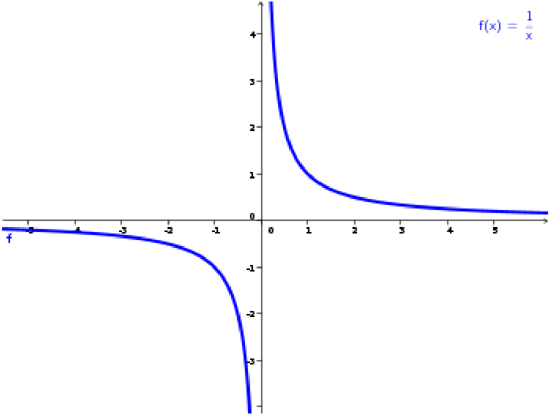
Limit of a Function



Example

$$\lim_{x \to 3} \frac{1}{x} = \frac{1}{3}$$

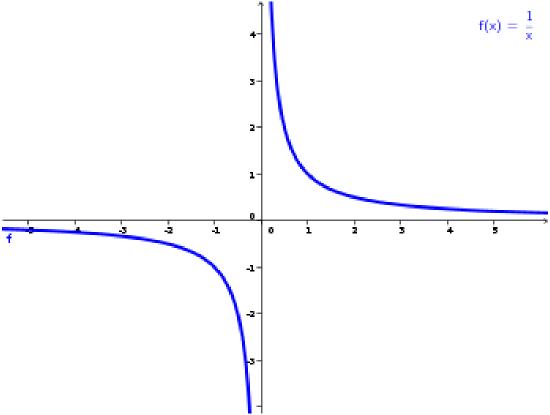
Limit of a Function



Example

$$\lim_{x\to 3}\frac{1}{x}=\frac{1}{3} \qquad \text{but} \qquad \lim_{x\to 0}\frac{1}{x}$$

Limit of a Function



Example

$$\lim_{x \to 3} \frac{1}{x} = \frac{1}{3} \qquad \text{but} \qquad \lim_{x \to 0} \frac{1}{x} \text{ does not exist.}$$

Continuity

Notice how the graph of $f(x) = \dfrac{1}{x}$ looks like it consists of two separate graphs put together at $x = 0$.

Continuity

Notice how the graph of $f(x) = \dfrac{1}{x}$ looks like it consists of two separate graphs put together at $x = 0$.

You can draw its graph without lifting pencil on $(-\infty, 0)$ and $(0, +\infty)$ but not at 0. That's why we say that it is **continuous** all points except for 0. We say that it is **discontinuous** at 0.

Continuity

Notice how the graph of $f(x) = \dfrac{1}{x}$ looks like it consists of two separate graphs put together at $x = 0$.

You can draw its graph without lifting pencil on $(-\infty, 0)$ and $(0, +\infty)$ but not at 0. That's why we say that it is **continuous** all points except for 0. We say that it is **discontinuous** at 0.

Definition

A function $f(x)$ is said to be **continuous** at the point c if it is defined at c and

1. $\lim\limits_{x \to c} f(x)$ exists,
2. $\lim\limits_{x \to c} f(x) = f(c)$ (i.e. limit $=$ value).

Continuity

Notice how the graph of $f(x) = \dfrac{1}{x}$ looks like it consists of two separate graphs put together at $x = 0$.

You can draw its graph without lifting pencil on $(-\infty, 0)$ and $(0, +\infty)$ but not at 0. That's why we say that it is **continuous** all points except for 0. We say that it is **discontinuous** at 0.

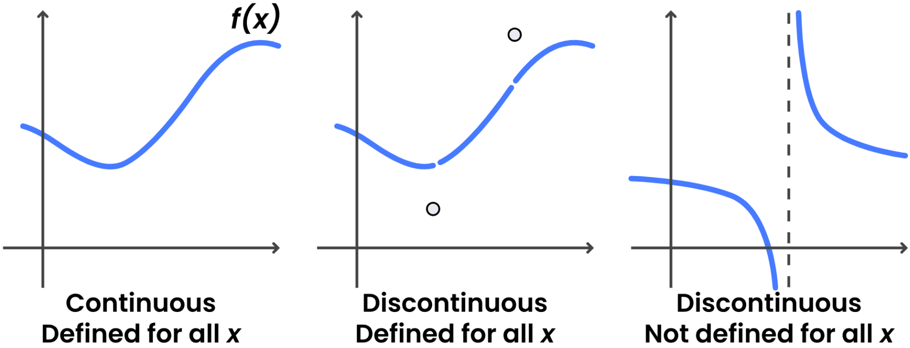
Definition

A function $f(x)$ is said to be **continuous** at the point c if it is defined at c and

1. $\lim\limits_{x \to c} f(x)$ exists,
2. $\lim\limits_{x \to c} f(x) = f(c)$ (i.e. limit = value).

If a function is continuous at all points, it is called a **continuous function**.

Continuity


f(x)
Continuous
Defined for all x
Discontinuous
Defined for all x
Discontinuous
Not defined for all x

Properties

If f and g are continuous at some point a, then

- $f + g$

Continuity

Properties

If f and g are continuous at some point a, then

- $f + g$
- $f - g$

Continuity

Properties

If f and g are continuous at some point a, then

- $f + g$
- $f - g$
- fg

Continuity

Properties

If f and g are continuous at some point a, then

- $f + g$
- $f - g$
- fg
- $\dfrac{f}{g}$ (if $g(a) \neq 0$)

Continuity

Properties

If f and g are continuous at some point a, then

- $f + g$
- $f - g$
- fg
- $\frac{f}{g}$ (if $g(a) \neq 0$)
- cf for any scalar c

Continuity

Properties

If f and g are continuous at some point a, then

- $f + g$
- $f - g$
- fg
- $\dfrac{f}{g}$ (if $g(a) \neq 0$)
- cf for any scalar c

are also continuous at a.

Continuity

Properties

If f and g are continuous at some point a, then

- $f+g$
- $f-g$
- fg
- $\dfrac{f}{g}$ (if $g(a)\neq 0$)
- cf for any scalar c

are also continuous at a.

In fact, most "good" functions are continuous (in their domains!):

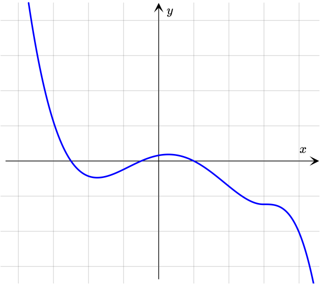
- Polynomials (e.g. x^2+7x-1, $xy-y^4+z$)
- Root functions (e.g. $\sqrt{x}$, $\sqrt[5]{x}$)
- Exponential and logarithmic functions (e.g. 2^x, e^{3x}, $\ln x$)
- Trigonometric functions and their inverses (e.g. $\cos(3x)$, $\arcsin x$)

Derivative

Continuous functions are better and more preferable to work with (continuous $= \heartsuit$), and many processes in the world are actually described by continuous functions.

Derivative

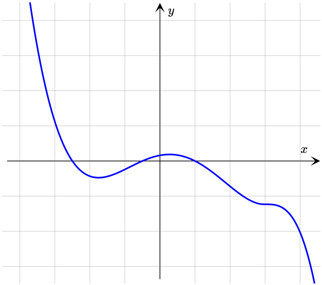
Continuous functions are better and more preferable to work with (continuous = ♡), and many processes in the world are actually described by continuous functions.



Now assume we want to maximize or minimize this function.

Derivative

Continuous functions are better and more preferable to work with (continuous $= \heartsuit$), and many processes in the world are actually described by continuous functions.



Now assume we want to maximize or minimize this function.
Notice how at some points it changes "faster" than at the others.
How can we measure that?

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a.

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a. If we move now slightly to the left or right of a, say by some small amount h, the function will change by

$$f(a+h) - f(a)$$

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a. If we move now slightly to the left or right of a, say by some small amount h, the function will change by

$$f(a+h) - f(a)$$

If we now divide it by the change we made to a, we get

$$\frac{f(a+h) - f(a)}{h}$$

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a. If we move now slightly to the left or right of a, say by some small amount h, the function will change by

$$f(a+h) - f(a)$$

If we now divide it by the change we made to a, we get

$$\frac{f(a+h) - f(a)}{h}$$

Intuitively, this value should express how quickly or slowly the function changes at the point a. We just have to make sure that h is really small enough (so the values of f far from a do not affect it):

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a. If we move now slightly to the left or right of a, say by some small amount h, the function will change by

$$f(a+h) - f(a)$$

If we now divide it by the change we made to a, we get

$$\frac{f(a+h) - f(a)}{h}$$

Intuitively, this value should express how quickly or slowly the function changes at the point a. We just have to make sure that h is really small enough (so the values of f far from a do not affect it):

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

Derivative

First, we fix a point where we want to measure the "speed" of the function, say a. If we move now slightly to the left or right of a, say by some small amount h, the function will change by

$$f(a+h) - f(a)$$

If we now divide it by the change we made to a, we get

$$\frac{f(a+h) - f(a)}{h}$$

Intuitively, this value should express how quickly or slowly the function changes at the point a. We just have to make sure that h is really small enough (so the values of f far from a do not affect it):

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$

Derivative

Definition

We say $f(x)$ is **differentiable at a point** a, if the following limit

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

or, equivalently,

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists and is finite. We denote it by $f'(a)$ or $\frac{df}{dx}(a)$ and call the **derivative** of function $f(x)$ at a.

Derivative

Definition

We say $f(x)$ is **differentiable at a point** a, if the following limit

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

or, equivalently,

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists and is finite. We denote it by $f'(a)$ or $\frac{df}{dx}(a)$ and call the **derivative** of function $f(x)$ at a.

The derivative of a function at a given point shows the **"speed" of the function** at that point, i.e. the **sensitivity of change** of its output with respect to its input.

Derivative

Definition

We say $f(x)$ is **differentiable at a point** a, if the following limit

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

or, equivalently,

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a}$$

exists and is finite. We denote it by $f'(a)$ or $\frac{df}{dx}(a)$ and call the **derivative** of function $f(x)$ at a.

The derivative of a function at a given point shows the **"speed" of the function** at that point, i.e. the **sensitivity of change** of its output with respect to its input.

Note that $f'(x)$ is a **function** itself and not a fixed number!

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) =$$

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} =$$

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h}$$

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x$$

Thus, $f(x)$ is differentiable at any point, and $f'(x) = 2x$.

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x$$

Thus, $f(x)$ is differentiable at any point, and $f'(x) = 2x$.

Remark

If a function is differentiable at some point, then it is also continuous, but the reverse is not always true.

Derivative

Example

Consider the function $f(x) = x^2$. The derivative of $f(x)$ is:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x$$

Thus, $f(x)$ is differentiable at any point, and $f'(x) = 2x$.

Remark

If a function is differentiable at some point, then it is also continuous, but the reverse is not always true.

Example

The function $f(x) = |x|$ is continuous but it is not differentiable at point $x = 0$.

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$
2. $(f \pm g)' = f' \pm g'$

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$
2. $(f \pm g)' = f' \pm g'$
3. $(fg)' = f'g + fg'$

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$
2. $(f \pm g)' = f' \pm g'$
3. $(fg)' = f'g + fg'$
4. $\left(\dfrac{1}{f}\right)' = -\dfrac{f'}{f^2}$, if $f \neq 0$

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$
2. $(f \pm g)' = f' \pm g'$
3. $(fg)' = f'g + fg'$
4. $\left(\dfrac{1}{f}\right)' = -\dfrac{f'}{f^2}$, if $f \neq 0$
5. $\left(\dfrac{f}{g}\right)' = \dfrac{f'g - fg'}{g^2}$, if $g \neq 0$

Derivative

Similarly, we can compute the derivative of $f'(x)$ itself (it will show the speed of the speed of $f(x)$, i.e. its *acceleration*).

We denote the derivative of $f'(x)$ by $f''(x)$, that of $f''(x)$ by $f'''(x)$, and so on. The derivative taken of $f(x)$ n times is also denoted by $f^{(n)}(x)$.

For any differentiable functions f, g and a real number c, the following rules hold:

Properties

1. $(cf)' = c \cdot f'$
2. $(f \pm g)' = f' \pm g'$
3. $(fg)' = f'g + fg'$
4. $\left(\dfrac{1}{f}\right)' = -\dfrac{f'}{f^2}$, if $f \neq 0$
5. $\left(\dfrac{f}{g}\right)' = \dfrac{f'g - fg'}{g^2}$, if $g \neq 0$
6. $(f(g(x)))' = f'(g(x)) \cdot g'(x)$

Derivative

- The derivative of any constant $f(x) = c$ is:

$$(c)' = 0$$

Derivative

- The derivative of any constant $f(x) = c$ is:

$$(c)' = 0$$

- The derivative of $f(x) = x$ is:

$$(x)' = 1$$

Derivative

- The derivative of any constant $f(x) = c$ is:

$$(c)' = 0$$

- The derivative of $f(x) = x$ is:

$$(x)' = 1$$

- The derivative of $f(x) = x^2$ is:

$$(x^2)' = 2x$$

Derivative

- The derivative of any constant $f(x) = c$ is:

$$(c)' = 0$$

- The derivative of $f(x) = x$ is:

$$(x)' = 1$$

- The derivative of $f(x) = x^2$ is:

$$(x^2)' = 2x$$

- For any constant n, the derivative of $f(x) = x^n$ is:

$$(x^n)' = nx^{n-1}$$

Derivative

- The derivative of $f(x) = e^x$ is:

$$(e^x)' = e^x$$

Derivative

- The derivative of $f(x) = e^x$ is:

$$(e^x)' = e^x$$

- The derivative of $f(x) = a^x$ is:

$$(a^x)' = a^x \ln a$$

Derivative

- The derivative of $f(x) = e^x$ is:

$$(e^x)' = e^x$$

- The derivative of $f(x) = a^x$ is:

$$(a^x)' = a^x \ln a$$

- The derivative of $f(x) = \ln x$ is:

$$(\ln x)' = \frac{1}{x}$$

Derivative

- The derivative of $f(x) = e^x$ is:

$$(e^x)' = e^x$$

- The derivative of $f(x) = a^x$ is:

$$(a^x)' = a^x \ln a$$

- The derivative of $f(x) = \ln x$ is:

$$(\ln x)' = \frac{1}{x}$$

- The derivative of $f(x) = \sin x$ is:

$$(\sin x)' = \cos x$$

Derivative

- The derivative of $f(x) = e^x$ is:

$$(e^x)' = e^x$$

- The derivative of $f(x) = a^x$ is:

$$(a^x)' = a^x \ln a$$

- The derivative of $f(x) = \ln x$ is:

$$(\ln x)' = \frac{1}{x}$$

- The derivative of $f(x) = \sin x$ is:

$$(\sin x)' = \cos x$$

- The derivative of $f(x) = \cos x$ is:

$$(\cos x)' = -\sin x$$

Derivative is All You Need

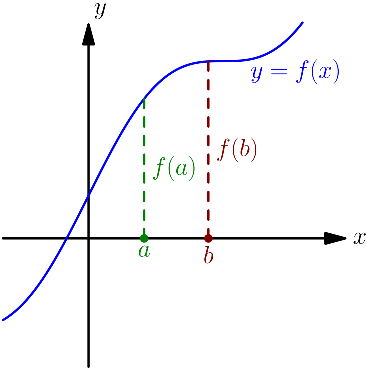
These formulas might seem to much, but you do not have to memorize them–after using them for a while, one begins to "feel" how fast or slow a given function is.

Derivative is All You Need

These formulas might seem to much, but you do not have to memorize them–after using them for a while, one begins to "feel" how fast or slow a given function is.

Derivatives tell about amazingly many interesting properties of the function.

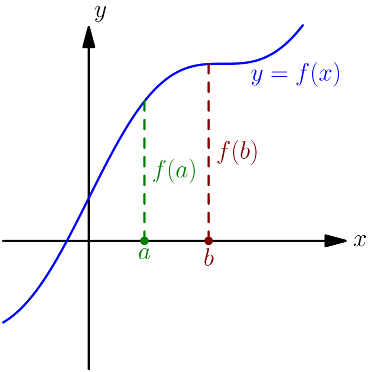
Derivative is All You Need



We say that a function $f(x)$ is **increasing** at the point a if

$$f(c) < f(a) < f(b) \qquad \text{when} \qquad c < a < b$$

Derivative is All You Need



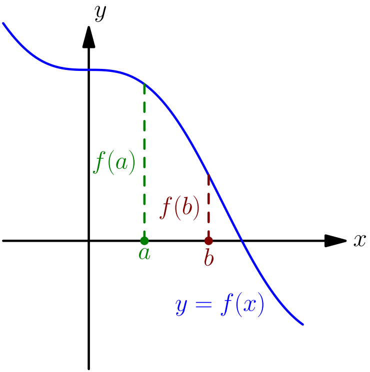
We say that a function $f(x)$ is **increasing** at the point a if

$$f(c) < f(a) < f(b) \qquad \text{when} \qquad c < a < b$$

and we say that it is just **non-decreasing** if

$$f(c) \leq f(a) \leq f(b) \qquad \text{when} \qquad c < a < b$$

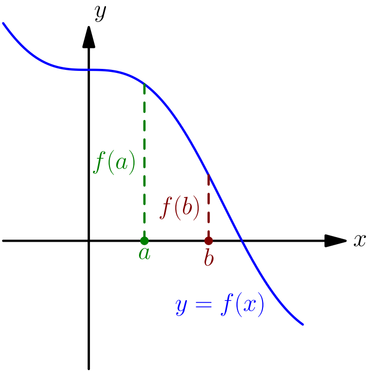
Derivative is All You Need



Similarly, we say that $f(x)$ is **decreasing** at the point a if

$$f(c) > f(a) > f(b) \qquad \text{when} \qquad c < a < b$$

Derivative is All You Need



Similarly, we say that $f(x)$ is **decreasing** at the point a if

$$f(c) > f(a) > f(b) \qquad \text{when} \qquad c < a < b$$

and we say that it is just **non-increasing** if

$$f(c) \geq f(a) \geq f(b) \qquad \text{when} \qquad c < a < b$$

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.
- If $f'(a) \geq 0$ then f is non-decreasing at a.

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.
- If $f'(a) \geq 0$ then f is non-decreasing at a.
- If $f'(a) < 0$ then f is decreasing at a.

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.
- If $f'(a) \geq 0$ then f is non-decreasing at a.
- If $f'(a) < 0$ then f is decreasing at a.
- If $f'(a) \leq 0$ then f is non-increasing at a.

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.
- If $f'(a) \geq 0$ then f is non-decreasing at a.
- If $f'(a) < 0$ then f is decreasing at a.
- If $f'(a) \leq 0$ then f is non-increasing at a.

That simple.

Derivative is All You Need

If your profit function $f(x)$ is increasing at some point, you might want to consider increasing x. How do you know whether f is increasing?

Theorem

- If $f'(a) > 0$ then f is increasing at a.
- If $f'(a) \geq 0$ then f is non-decreasing at a.
- If $f'(a) < 0$ then f is decreasing at a.
- If $f'(a) \leq 0$ then f is non-increasing at a.

That simple.

Question

What if $f'(a) = 0$?

Extrema of a Function

Finally, what about the largest/smallest values of the function?

Extrema of a Function

Finally, what about the largest/smallest values of the function?

Definition

We say x_0 is a **local maximum (minimum) point** of f, if the largest (smallest) value of f on a small enough interval $(x_0 - \delta, x_0 + \delta)$ is the value $f(x_0)$ itself.

Extrema of a Function

Finally, what about the largest/smallest values of the function?

Definition

We say x_0 is a **local maximum (minimum) point** of f, if the largest (smallest) value of f on a small enough interval $(x_0 - \delta, x_0 + \delta)$ is the value $f(x_0)$ itself.

More technically, if there exists small enough $\delta > 0$ such that ,

$$f(x) \leq f(x_0)$$

$$(f(x) \geq f(x_0))$$

for any x you take from $(x_0 - \delta, x_0 + \delta)$.

Extrema of a Function

Finally, what about the largest/smallest values of the function?

Definition

We say x_0 is a **local maximum (minimum) point** of f, if the largest (smallest) value of f on a small enough interval $(x_0 - \delta, x_0 + \delta)$ is the value $f(x_0)$ itself.

More technically, if there exists small enough $\delta > 0$ such that ,

$$f(x) \leq f(x_0)$$

$$(f(x) \geq f(x_0))$$

for any x you take from $(x_0 - \delta, x_0 + \delta)$.

In that case, the value $f(x_0)$ is called a **local maximum (minimum)** of f.

Extrema of a Function

Finally, what about the largest/smallest values of the function?

Definition

We say x_0 is a **local maximum (minimum) point** of f, if the largest (smallest) value of f on a small enough interval $(x_0 - \delta, x_0 + \delta)$ is the value $f(x_0)$ itself.

More technically, if there exists small enough $\delta > 0$ such that ,

$$f(x) \leq f(x_0)$$

$$(f(x) \geq f(x_0))$$

for any x you take from $(x_0 - \delta, x_0 + \delta)$.

In that case, the value $f(x_0)$ is called a **local maximum (minimum)** of f. Together, local minima and maxima are called local extrema of f.

Extrema of a Function

Definition

We say x_0 is a **global maximum (minimum) point** of f, if the largest (smallest) value of f (on whole $\mathbb{R}$) is the value $f(x_0)$ itself.

Technically, x_0 is a global maximum (minimum) point if for any x,

$$f(x) \leq f(x_0)$$

$$(f(x) \geq f(x_0))$$

Extrema of a Function

Definition

We say x_0 is a **global maximum (minimum) point** of f, if the largest (smallest) value of f (on whole $\mathbb{R}$) is the value $f(x_0)$ itself.

Technically, x_0 is a global maximum (minimum) point if for any x,

$$f(x) \leq f(x_0)$$

$$(f(x) \geq f(x_0))$$

Together, global minima and maxima are called global extrema of f.
Apparently, every global extrema is also a local extrema (and the converse is not true).

Extrema of a Function

Definition

We say x_0 is a **global maximum (minimum) point** of f, if the largest (smallest) value of f (on whole $\mathbb{R}$) is the value $f(x_0)$ itself.

Technically, x_0 is a global maximum (minimum) point if for any x,

$$f(x) \leq f(x_0)$$

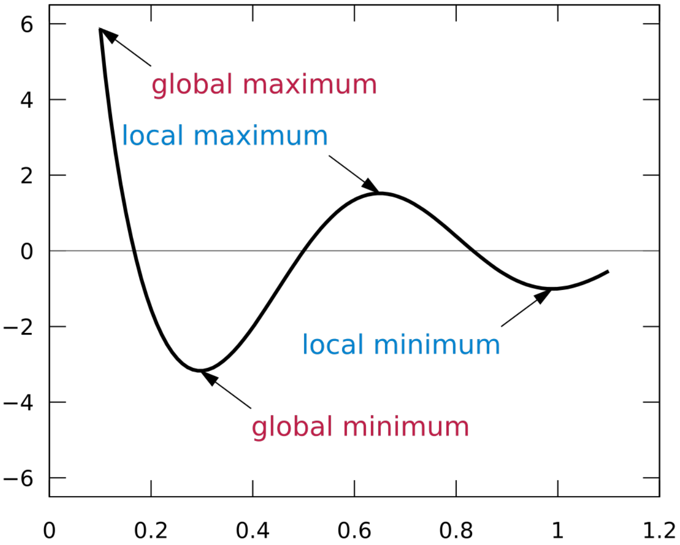
$$(f(x) \geq f(x_0))$$

Together, global minima and maxima are called global extrema of f. Apparently, every global extrema is also a local extrema (and the converse is not true).

Theorem

Every continuous function f has both a global maximum and a global minimum on any **closed** interval $[a, b]$.

Extrema of a Function



Extrema of a Function

Theorem

If x_0 is an extremum point of f and there exists $f'(x_0)$, then $f'(x_0) = 0$.

Extrema of a Function

Theorem

If x_0 is an extremum point of f and there exists $f'(x_0)$, then $f'(x_0) = 0$.

Example

Consider the function $f_1(x) = x^2$ which has a minimum at $x = 0$.

$$f_1'(x) = 2x, \qquad f_1'(0) = 0$$

Extrema of a Function

Theorem

If x_0 is an extremum point of f and there exists $f'(x_0)$, then $f'(x_0) = 0$.

Example

Consider the function $f_1(x) = x^2$ which has a minimum at $x = 0$.

$$f_1'(x) = 2x, \qquad f_1'(0) = 0$$

Moreover, the function $f_2(x) = -x^2$ has a maximum at $x = 0$ and again:

$$f_2'(x) = -2x, \qquad f_2'(0) = 0$$

Extrema of a Function

Theorem

If x_0 is an extremum point of f and there exists $f'(x_0)$, then $f'(x_0) = 0$.

Example

Consider the function $f_1(x) = x^2$ which has a minimum at $x = 0$.

$$f_1'(x) = 2x, \qquad f_1'(0) = 0$$

Moreover, the function $f_2(x) = -x^2$ has a maximum at $x = 0$ and again:

$$f_2'(x) = -2x, \qquad f_2'(0) = 0$$

The function $f_3(x) = x^3$ has no local extremum point, yet

$$f_3'(x) = 3x^2, \qquad f_3'(0) = 0$$

Extrema of a Function

Theorem

If x_0 is an extremum point of f and there exists $f'(x_0)$, then $f'(x_0) = 0$.

Example

Consider the function $f_1(x) = x^2$ which has a minimum at $x = 0$.

$$f_1'(x) = 2x, \qquad f_1'(0) = 0$$

Moreover, the function $f_2(x) = -x^2$ has a maximum at $x = 0$ and again:

$$f_2'(x) = -2x, \qquad f_2'(0) = 0$$

The function $f_3(x) = x^3$ has no local extremum point, yet

$$f_3'(x) = 3x^2, \qquad f_3'(0) = 0$$

Hence, the condition $f'(x) = 0$ is necessary but *not sufficient*.

Extrema of a Function

How to find the actual local extrema?

Extrema of a Function

How to find the actual local extrema?

Definition

If $f'(x_0)$ doesn't exist or $f'(x_0) = 0$, then we call x_0 a **critical point**.

Extrema of a Function

How to find the actual local extrema?

Definition

If $f'(x_0)$ doesn't exist or $f'(x_0) = 0$, then we call x_0 a **critical point**.

How can we tell if a critical point is a local minimum/maximum point?

Extrema of a Function

Theorem 1 (f'' at one point)

If $f'(x_0) = 0$ and there exists finite $f''(x_0)$, then

1. If $f''(x_0) > 0$, then x_0 is a local minimum point,
2. If $f''(x_0) < 0$, then x_0 is a local maximum point.

Extrema of a Function

Theorem 1 (f'' at one point)

If $f'(x_0) = 0$ and there exists finite $f''(x_0)$, then

1. If $f''(x_0) > 0$, then x_0 is a local minimum point,
2. If $f''(x_0) < 0$, then x_0 is a local maximum point.

Theorem 2 (f' at multiple points)

If for some $\delta > 0$, f is differentiable in the intervals $(x_0 - \delta, x_0)$ and $(x_0, x_0 + \delta)$ and continuous at x_0, then

1. If $f'(x) > 0$ for $x \in (x_0 - \delta, x_0)$ and $f'(x) < 0$ for $x \in (x_0, x_0 + \delta)$, then x_0 is a local maximum point.
2. If $f'(x) < 0$ for $x \in (x_0 - \delta, x_0)$ and $f'(x) > 0$ for $x \in (x_0, x_0 + \delta)$, then x_0 is a local minimum point.
3. If $f'(x)$ doesn't change its sign, then x_0 is not an extremum point.

Extrema of a Function

Wrapping up, how can we use our knowledge to find the local extrema of a given function $f(x)$?

Extrema of a Function

Wrapping up, how can we use our knowledge to find the local extrema of a given function $f(x)$?

Step 0: Make sure that $f(x)$ is continuous in the given set. Otherwise, divide it into intervals where $f(x)$ is continuous. *Example: For* $f(x) = \frac{1}{x}$ *we should consider it separately on* $(-\infty, 0)$ *and* $(0, +\infty)$

Extrema of a Function

Wrapping up, how can we use our knowledge to find the local extrema of a given function $f(x)$?

Step 0: Make sure that $f(x)$ is continuous in the given set. Otherwise, divide it into intervals where $f(x)$ is continuous. *Example: For* $f(x) = \frac{1}{x}$ *we should consider it separately on* $(-\infty, 0)$ *and* $(0, +\infty)$

Step 1: Find the critical points x_0 of $f(x)$.

Extrema of a Function

Wrapping up, how can we use our knowledge to find the local extrema of a given function $f(x)$?

Step 0: Make sure that $f(x)$ is continuous in the given set. Otherwise, divide it into intervals where $f(x)$ is continuous.
Example: For $f(x) = \frac{1}{x}$ *we should consider it separately on* $(-\infty, 0)$ *and* $(0, +\infty)$

Step 1: Find the critical points x_0 of $f(x)$.

Step 2: If $f(x)$ is given on a closed interval $[a, b]$, also check its values on endpoints a and b.
Example: When defined on a closed interval $[a, b]$, $f(x) = 2x$ *has its minimum at* $x = a$ *and maximum at* $x = b$

Extrema of a Function

Wrapping up, how can we use our knowledge to find the local extrema of a given function $f(x)$?

Step 0: Make sure that $f(x)$ is continuous in the given set. Otherwise, divide it into intervals where $f(x)$ is continuous.
Example: For $f(x) = \frac{1}{x}$ we should consider it separately on $(-\infty, 0)$ and $(0, +\infty)$

Step 1: Find the critical points x_0 of $f(x)$.

Step 2: If $f(x)$ is given on a closed interval $[a, b]$, also check its values on endpoints a and b.
Example: When defined on a closed interval $[a, b]$, $f(x) = 2x$ has its minimum at $x = a$ and maximum at $x = b$

Step 3: a) If there exists finite $f''(x_0) \neq 0$, use Theorem 1.
b) If you find the sign of $f'(x)$ on left and right "sides" of x_0, use Theorem 2.

Taylor Series, Integral

Hayk Aprikyan, Hayk Tarkhanyan

Convex and Concave Functions

We know that the sign of the derivative tells us whether a function is increasing or decreasing:

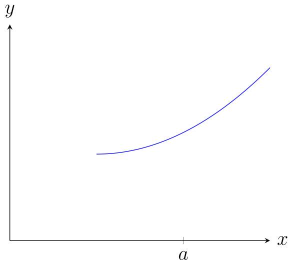
$$f' > 0 \quad \Rightarrow \quad \text{increasing}, \qquad\qquad f' < 0 \quad \Rightarrow \quad \text{decreasing}$$

Convex and Concave Functions

We know that the sign of the derivative tells us whether a function is increasing or decreasing:

$$f' > 0 \quad \Rightarrow \quad \text{increasing}, \qquad\qquad f' < 0 \quad \Rightarrow \quad \text{decreasing}$$

However, a function can increase like this

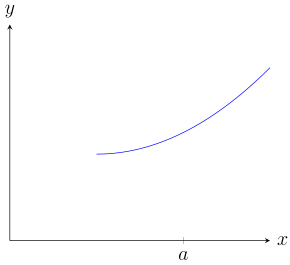


Convex and Concave Functions

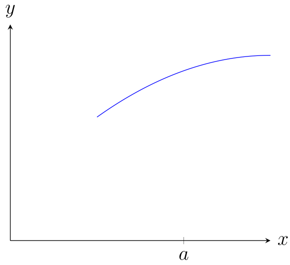
We know that the sign of the derivative tells us whether a function is increasing or decreasing:

$$f' > 0 \quad \Rightarrow \quad \text{increasing}, \qquad\qquad f' < 0 \quad \Rightarrow \quad \text{decreasing}$$

However, a function can increase like this



or like this

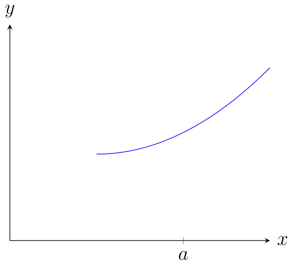


Convex and Concave Functions

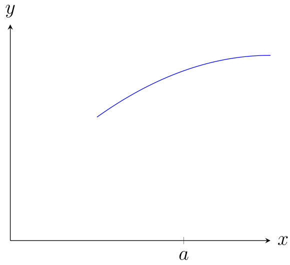
We know that the sign of the derivative tells us whether a function is increasing or decreasing:

$$f' > 0 \quad \Rightarrow \quad \text{increasing}, \qquad\qquad f' < 0 \quad \Rightarrow \quad \text{decreasing}$$

However, a function can increase like this

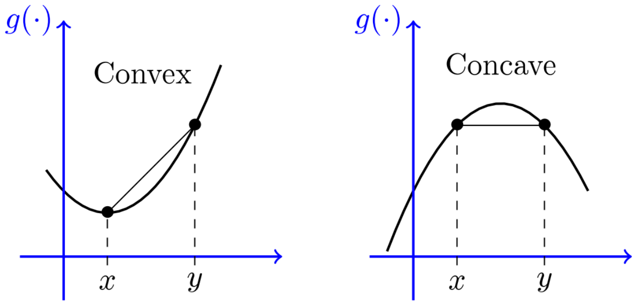


or like this



How can you determine which way it is?

Convex and Concave Functions



Definition

We say that $f(x)$ is **convex** on some interval if for any two points on its graph, the line connecting them always lies **above** the graph of $f(x)$.

Definition

We say that $f(x)$ is **concave** on some interval if for any two points on its graph, the line connecting them always lies **below** the graph of $f(x)$.

Convex and Concave Functions

More technically, a function is

- convex if

$$f(\alpha p + (1 - \alpha)q) \leq \alpha f(p) + (1 - \alpha)f(q)$$

- concave if

$$f(\alpha p + (1 - \alpha)q) \geq \alpha f(p) + (1 - \alpha)f(q)$$

for any two points p and q, and for any number $0 \leq \alpha \leq 1$.

Convex and Concave Functions

More technically, a function is

- convex if

$$f(\alpha p + (1-\alpha)q) \leq \alpha f(p) + (1-\alpha) f(q)$$

- concave if

$$f(\alpha p + (1-\alpha)q) \geq \alpha f(p) + (1-\alpha) f(q)$$

for any two points p and q, and for any number $0 \leq \alpha \leq 1$.

We will not prove this

Convex and Concave Functions

More technically, a function is

- convex if
$$f(\alpha p + (1-\alpha)q) \leq \alpha f(p) + (1-\alpha) f(q)$$
- concave if
$$f(\alpha p + (1-\alpha)q) \geq \alpha f(p) + (1-\alpha) f(q)$$

for any two points p and q, and for any number $0 \leq \alpha \leq 1$.

We will not prove this but open the link and play with α to get a feeling of what this means:

- www.desmos.com/calculator/ujoh0mh59d

Convex and Concave Functions

So, how to know if a function is convex or concave?

Convex and Concave Functions

So, how to know if a function is convex or concave?
Again, derivative is all you need!

Theorem

If $f(x)$ is twice-differentiable (i.e. there exists $f''(x)$), then:

- f is convex if and only if $f''(x) \geq 0$
- f is concave if and only if $f''(x) \leq 0$

Convex and Concave Functions

So, how to know if a function is convex or concave?
Again, derivative is all you need!

Theorem

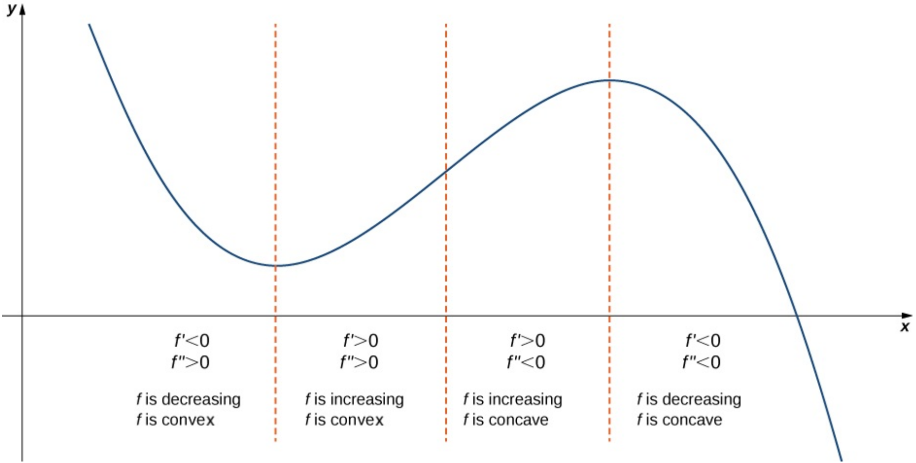
If $f(x)$ is twice-differentiable (i.e. there exists $f''(x)$), then:

- f is convex if and only if $f''(x) \geq 0$
- f is concave if and only if $f''(x) \leq 0$

Examples

1. $f(x) = x^2$ is convex on $(-\infty, \infty)$.
   $f''(x) = 2 > 0$
2. $f(x) = -x^2$ is concave on $(-\infty, \infty)$.
   $f''(x) = -2 < 0$
3. $f(x) = x$ is both convex and concave on $(-\infty, \infty)$.
   $f''(x) = 0$

Convex and Concave Functions



Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has:

Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has:

- the same value as f

Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has:

- the same value as f
- the same rate of change as f

Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has:

- the same value as f
- the same rate of change as f
- the same rate of the rate of change as f

Taylor Series

As we saw, the derivatives of a function tell so much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has:

- the same value as f
- the same rate of change as f
- the same rate of the rate of change as f
- ...

Taylor Series

As we saw, the derivatives of a function tell too much about them. Can they be used to approximate the function?

Suppose we have a function $f(x)$. Let us construct another function $g(x)$ which, at some point a has (or, equivalently):

- $g(a) = f(a)$
- $g'(a) = f'(a)$
- $g''(a) = f''(a)$
- $\ldots$

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Definition

If there exist derivatives f', f'', $\dots$, $f^{(k)}$, then the polynomial

$$P_k(x) = f(a)$$

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Definition

If there exist derivatives f', f'', $\ldots$, $f^{(k)}$, then the polynomial

$$P_k(x) = f(a) + f'(a)(x - a)$$

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Definition

If there exist derivatives f', f'', ..., $f^{(k)}$, then the polynomial

$$P_k(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2$$

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Definition

If there exist derivatives f', f'', ..., $f^{(k)}$, then the polynomial

$$P_k(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k$$

is called the **Taylor polynomial** of order k of the function $f(x)$ about the point a.

Taylor Series

For example, suppose $f(x) = e^x$. Then $f(0) = 1$ and $f'(0) = 1$.

Question

Which simple function has the value 1 at $x = 0$, and derivative 1 at $x = 0$?

The polynomial $g(x) = 1 + x$ has just that properties. Moreover, it is easy to work with, since it is a polynomial!

Definition

If there exist derivatives f', f'', ..., $f^{(k)}$, then the polynomial

$$P_k(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k$$

is called the **Taylor polynomial** of order k of the function $f(x)$ about the point a.

If we take $k = \infty$, we will get an infinite sum which is called the **Taylor series** of $f(x)$ about a.

Taylor Series

Taylor's Theorem

If a function $f(x)$ is k times differentiable at the point a, then

$$P_k(x) \to f(x), \qquad k \to \infty$$

around some (maybe small) interval around a.

Taylor Series

Taylor's Theorem

If a function $f(x)$ is k times differentiable at the point a, then

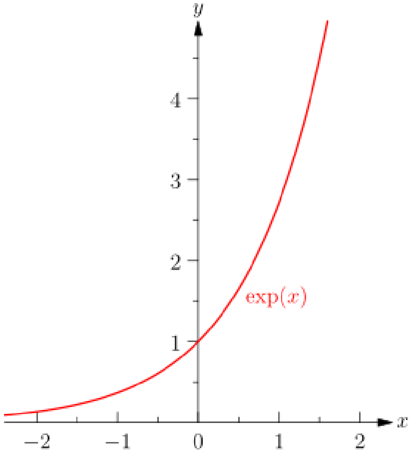
$$P_k(x) \to f(x), \qquad k \to \infty$$

around some (maybe small) interval around a.

Example

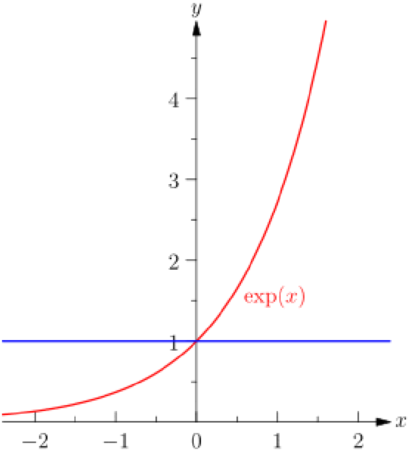
- The Taylor series for any polynomial is the polynomial itself.
- $\dfrac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \cdots$
- $\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \dfrac{x^{n+1}}{n+1} = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots$
- $e^x = \sum_{n=0}^{\infty} \dfrac{x^n}{n!} = 1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots$

Taylor Series



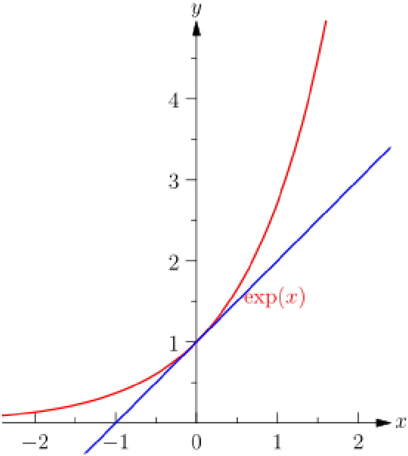
Taylor Series

1

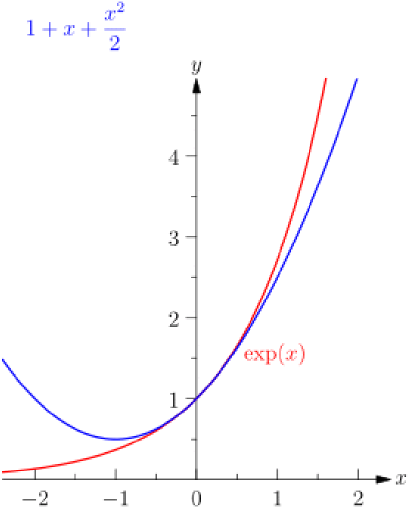


Taylor Series

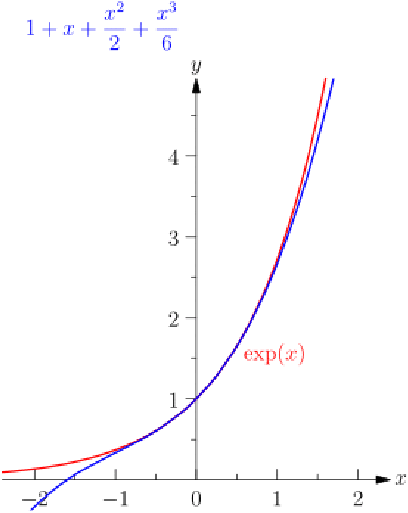
$1 + x$



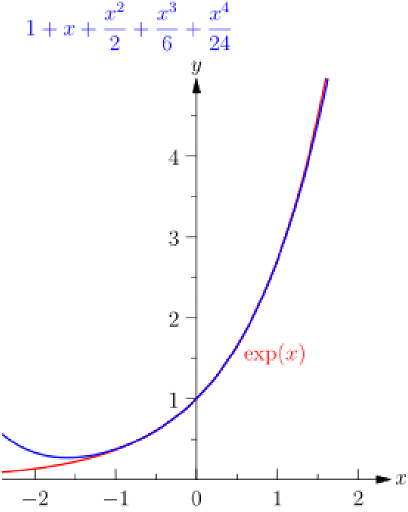
Taylor Series



Taylor Series

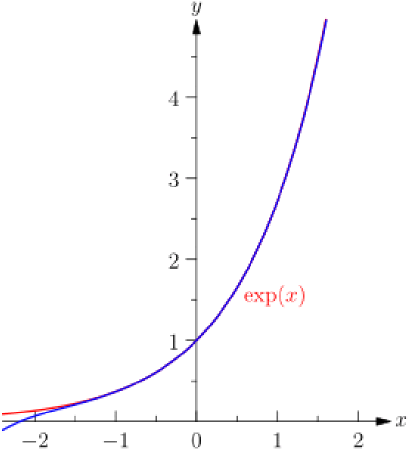


Taylor Series



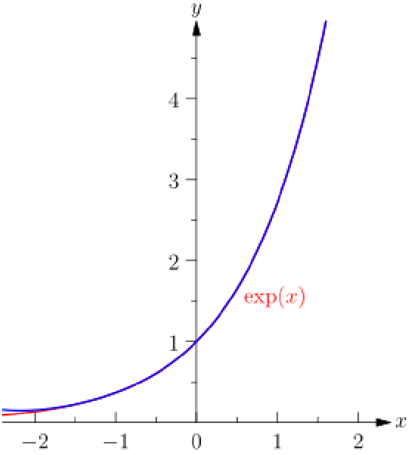
Taylor Series

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{5!}$$

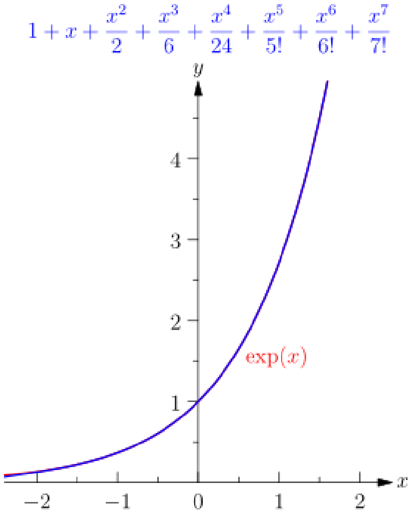


Taylor Series

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{5!} + \frac{x^6}{6!}$$



Taylor Series



- Try another function yourself!

Indefinite Integral

$2x$ is the derivative of x^2

Indefinite Integral

$2x$ is the derivative of x^2

So what is x^2?

Indefinite Integral

$$2x \qquad \text{is the derivative of} \qquad x^2$$

So what is x^2?

$$x^2 \qquad \text{is the } \textit{anti}\text{derivative of} \qquad 2x$$

Indefinite Integral

$$2x \qquad \text{is the derivative of} \qquad x^2$$

So what is x^2?

$$x^2 \qquad \text{is the } \textit{anti}\text{derivative of} \qquad 2x$$

Definition

F is called an **antiderivative** or **indefinite integral** of f if $F'(x) = f(x)$.

Indefinite Integral

$$2x \qquad \text{is the derivative of} \qquad x^2$$

So what is x^2?

$$x^2 \qquad \text{is the \textit{anti}derivative of} \qquad 2x$$

Definition

F is called an **antiderivative** or **indefinite integral** of f if $F'(x) = f(x)$.

The indefinite integral is denoted like this:

$$\int f(x)\,dx$$

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

- $(x^2 + 1)' = 2x$

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

- $(x^2 + 1)' = 2x$
- $(x^2 + 2)' = 2x$

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

- $(x^2 + 1)' = 2x$
- $(x^2 + 2)' = 2x$
- $(x^2 + c)' = 2x$

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

- $(x^2 + 1)' = 2x$
- $(x^2 + 2)' = 2x$
- $(x^2 + c)' = 2x$

If $F'(x) = f(x)$, then $(F(x) + c)' = f(x)$ for any number c, so actually:

$$\int 2x\,dx = x^2 + C$$

for any number C.

Indefinite Integral

So in this notation,

$$\int 2x\,dx = x^2$$

Can you name another function that also has derivative $2x$?

- $(x^2+1)' = 2x$
- $(x^2+2)' = 2x$
- $(x^2+c)' = 2x$

If $F'(x) = f(x)$, then $(F(x)+c)' = f(x)$ for any number c, so actually:

$$\int 2x\,dx = x^2 + C$$

for any number C.

$\int f(x)\,dx$ is not one function, but a *set* of functions.

Indefinite Integral

Properties

1. $$\int af(x)\,dx = a\int f(x)\,dx$$

Indefinite Integral

Properties

1. $$\int af(x)\,dx = a\int f(x)\,dx$$

2. $$\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$

Indefinite Integral

Properties

1. $$\int af(x)\,dx = a\int f(x)\,dx$$

2. $$\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$

3. $$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$$

Indefinite Integral

Properties

1. $$\int af(x)\,dx = a\int f(x)\,dx$$
2. $$\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$
3. $$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$$

Indefinite Integral

Properties

1. $$\int af(x)\,dx = a\int f(x)\,dx$$
2. $$\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$
3. $$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$$

   This can also be written as:

   $$\int f\,dg = fg - \int g\,df,$$

   where $\displaystyle\int f\,dg = \int fg'\,dx$.

Indefinite Integral

- The antiderivative of any constant $f(x) = a$ is:

$$\int a\,dx = ax + C$$

Indefinite Integral

- The antiderivative of any constant $f(x) = a$ is:

$$\int a\,dx = ax + C$$

- The antiderivative of $f(x) = x$ is:

$$\int x\,dx = \frac{x^2}{2} + C$$

Indefinite Integral

- The antiderivative of any constant $f(x) = a$ is:

$$\int a\,dx = ax + C$$

- The antiderivative of $f(x) = x$ is:

$$\int x\,dx = \frac{x^2}{2} + C$$

- For any constant $n \neq 1$, the derivative of $f(x) = x^n$ is:

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$$

Indefinite Integral

- The antiderivative of any constant $f(x) = a$ is:

$$\int a\, dx = ax + C$$

- The antiderivative of $f(x) = x$ is:

$$\int x\, dx = \frac{x^2}{2} + C$$

- For any constant $n \neq 1$, the derivative of $f(x) = x^n$ is:

$$\int x^n\, dx = \frac{x^{n+1}}{n+1} + C$$

- The antiderivative of $f(x) = \frac{1}{x}$ is:

$$\int \frac{1}{x}\, dx = \ln|x| + C$$

Indefinite Integral

- The antiderivative of $f(x) = e^x$ is:

$$\int e^x \, dx = e^x + C$$

Indefinite Integral

- The antiderivative of $f(x) = e^x$ is:

$$\int e^x \, dx = e^x + C$$

- The antiderivative of $f(x) = \cos x$ is:

$$\int \cos x \, dx = \sin x + C$$

Indefinite Integral

- The antiderivative of $f(x) = e^x$ is:

$$\int e^x\, dx = e^x + C$$

- The antiderivative of $f(x) = \cos x$ is:

$$\int \cos x\, dx = \sin x + C$$

- The antiderivative of $f(x) = \sin x$ is:

$$\int \sin x\, dx = -\cos x + C$$

Indefinite Integral

- The antiderivative of $f(x) = e^x$ is:

$$\int e^x \, dx = e^x + C$$

- The antiderivative of $f(x) = \cos x$ is:

$$\int \cos x \, dx = \sin x + C$$

- The antiderivative of $f(x) = \sin x$ is:

$$\int \sin x \, dx = -\cos x + C$$

- The antiderivative of $f(x) = \frac{1}{1+x^2}$ is:

$$\int \frac{1}{1+x^2} \, dx = \operatorname{arctg} x + C$$

Example

Which function should you differentiate to get $x \cos x$?

Indefinite Integral

Example

Which function should you differentiate to get $x \cos x$?

Notice that $\cos x = (\sin x)'$, so by the Property 3,

$$\int x \cos x \, dx = \int x (\sin x)' \, dx = \int x \, d(\sin x) = x(\sin x) - \int (\sin x) \, dx$$

Indefinite Integral

Example

Which function should you differentiate to get $x \cos x$?

Notice that $\cos x = (\sin x)'$, so by the Property 3,

$$\int x \cos x \, dx = \int x(\sin x)' \, dx = \int x \, d(\sin x) = x(\sin x) - \int (\sin x) \, dx$$

and the last integral $= \cos x + C$ (as we know).

Example

Which function should you differentiate to get $x \cos x$?

Notice that $\cos x = (\sin x)'$, so by the Property 3,

$$\int x \cos x \, dx = \int x (\sin x)' \, dx = \int x \, d(\sin x) = x(\sin x) - \int (\sin x) \, dx$$

and the last integral $= \cos x + C$ (as we know).

All of this is, of course, computation and tricks (no need to memorize).

Indefinite Integral

Example

Which function should you differentiate to get $x \cos x$?

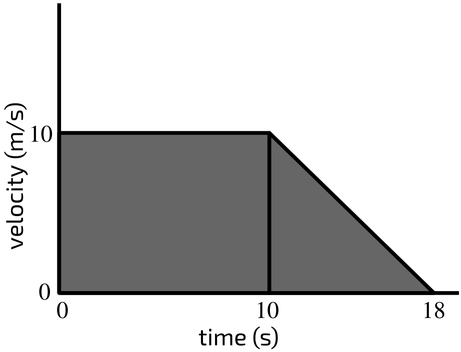
Notice that $\cos x = (\sin x)'$, so by the Property 3,

$$\int x \cos x \, dx = \int x (\sin x)' \, dx = \int x \, d(\sin x) = x(\sin x) - \int (\sin x) \, dx$$

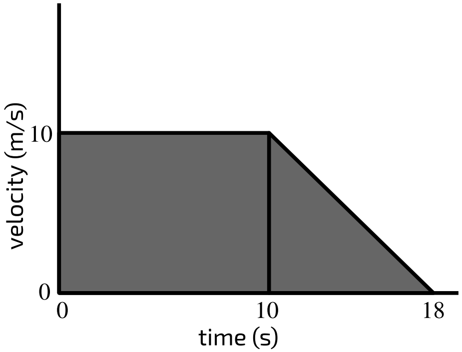
and the last integral $= \cos x + C$ (as we know).

All of this is, of course, computation and tricks (no need to memorize). How about we make an *actual* use of this?

Suppose we are given the velocity of a car at each timepoint. How can we calculate the distance travelled by the car?



Suppose we are given the velocity of a car at each timepoint. How can we calculate the distance travelled by the car?



According to physics, distance = area under the velocity curve.

Definite Integral

Question

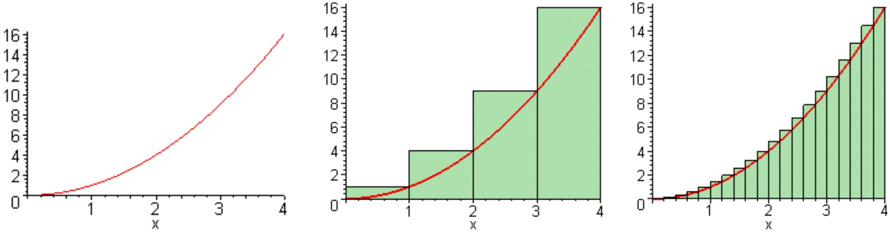
Suppose you have a continuous function f. How can we calculate the area under its graph?

Definite Integral

Question

Suppose you have a continuous function f. How can we calculate the area under its graph?

By dividing it into tiny rectangles and adding up their areas.



Definite Integral

Take the interval $[a, b]$ and divide (partition) it into n small parts with points $\{x_0, x_1, \ldots, x_n\}$. Let $\Delta x_i = x_i - x_{i-1}$ denote the length of $[x_{i-1}, x_i]$.

Definite Integral

Take the interval $[a, b]$ and divide (partition) it into n small parts with points $\{x_0, x_1, \ldots, x_n\}$. Let $\Delta x_i = x_i - x_{i-1}$ denote the length of $[x_{i-1}, x_i]$.

Definition

The **Riemann sum** of a function $f(x)$ is given by:

$$R_n = \sum_{i=1}^{n} f(c_i)\Delta x_i$$

where c_i is any point from $[x_{i-1}, x_i]$.

Definite Integral

Take the interval $[a, b]$ and divide (partition) it into n small parts with points $\{x_0, x_1, \ldots, x_n\}$. Let $\Delta x_i = x_i - x_{i-1}$ denote the length of $[x_{i-1}, x_i]$.

Definition

The **Riemann sum** of a function $f(x)$ is given by:

$$R_n = \sum_{i=1}^{n} f(c_i)\Delta x_i$$

where c_i is any point from $[x_{i-1}, x_i]$.

Making the Δs smaller, this sum will get closer to the area.

Definite Integral

Take the interval $[a, b]$ and divide (partition) it into n small parts with points $\{x_0, x_1, \ldots, x_n\}$. Let $\Delta x_i = x_i - x_{i-1}$ denote the length of $[x_{i-1}, x_i]$.

Definition

The **Riemann sum** of a function $f(x)$ is given by:

$$R_n = \sum_{i=1}^{n} f(c_i)\Delta x_i$$

where c_i is any point from $[x_{i-1}, x_i]$.

Making the Δs smaller, this sum will get closer to the area.

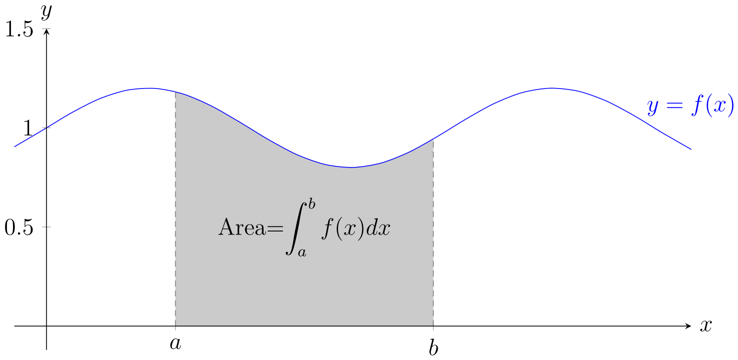
Definition

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(c_i)\, \Delta x_i$$

is called the **definite integral** of the function $f(x)$ on $[a, b]$.

Definite Integral

The definite integral $\int_a^b f(x)\,dx$ represents the **signed area** between the graph of $f(x)$ and the x-axis over the interval $[a, b]$.



- Play with Riemann sums!

Definite Integral

How can we calculate the definite integral without limits?

Definite Integral

How can we calculate the definite integral without limits?

Theorem (very important)

Suppose that $f(x)$ is continuous on the interval $[a, b]$. If $F(x)$ is an antiderivative of $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

Definite Integral

How can we calculate the definite integral without limits?

Theorem (very important)

Suppose that $f(x)$ is continuous on the interval $[a, b]$. If $F(x)$ is an antiderivative of $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

Example

- $$\int_0^2 x^2\,dx = \frac{1}{3}\cdot x^3\Big|_0^2 = \frac{1}{3}(2^3 - 0^3) = \frac{8}{3}$$
- $$\int_0^\pi \sin x\,dx = -\cos x\Big|_0^\pi = -\cos\pi - (-\cos 0) = 2$$

Definite Integral

1. $$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^a f(x)\,dx = 0$$

Definite Integral

1. $$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^a f(x)\,dx = 0$$

2. $$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$

Definite Integral

1. $$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^a f(x)\,dx = 0$$
2. $$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$
3. $$\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$$

Definite Integral

1. $$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^a f(x)\,dx = 0$$

2. $$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$

3. $$\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$$

4. $$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

Definite Integral

1. $$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^a f(x)\,dx = 0$$
2. $$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$
3. $$\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$$
4. $$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$
5. $$\int_a^b f(x)\,dx = \int_a^b f(y)\,dy$$

i.e. the name of the variable does not matter.

Functions of Several Variables

Hayk Aprikyan, Hayk Tarkhanyan

Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

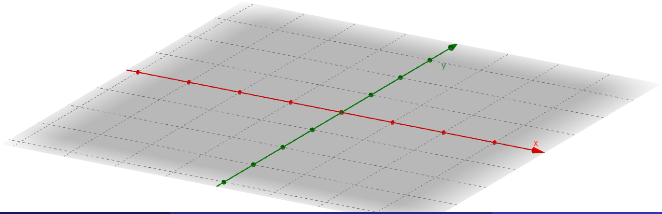
First, our input is now not a single number, but a pair of numbers (x, y).

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need the xy-plane



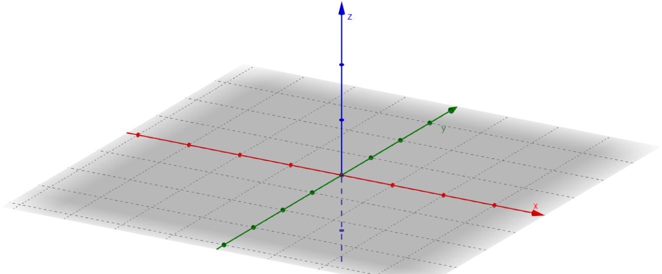
Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need xy-plane + a new axis



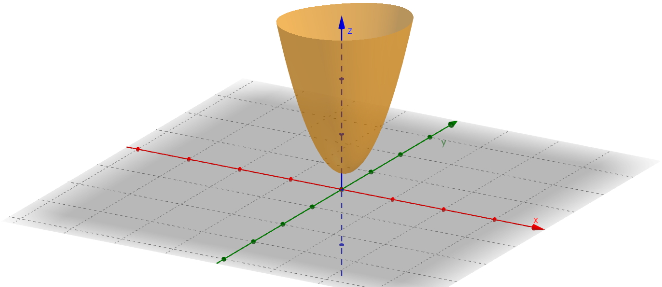
Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need xy-plane + a new axis (the graph is a surface):



A typical graph in this case looks much like a napkin hanging in the air.

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

A picture is worth a thousand words

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

A picture is worth a thousand words
(two pictures = two thousand words)

Partial Derivative

Again, suppose x and y are the costs of apples and peaches, and your profit is given by:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

Question

How can you measure the effect of increasing the cost of apple by a little (i.e. how quickly will f change if x changes)?

Partial Derivative

Again, suppose x and y are the costs of apples and peaches, and your profit is given by:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

Question

How can you measure the effect of increasing the cost of apple by a little (i.e. how quickly will f change if x changes)?

By fixing y and then doing the usual derivative stuff with x!

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}$$

then it is called the **partial derivative** of $f(x,y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

then it is called the **partial derivative** of $f(x, y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Similarly, we define the partial derivative of $f(x, y)$ with respect to y as:

$$f_y = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

then it is called the **partial derivative** of $f(x, y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Similarly, we define the partial derivative of $f(x, y)$ with respect to y as:

$$f_y = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

Example

If $f(x, y) = x^2 + y^2$, then:

$$f_x = 2x \qquad \text{and} \qquad f_y = 2y$$

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

So naturally, instead of 1 object to describe the speed, we have 2 objects. We can combine them together in one vector:

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

So naturally, instead of 1 object to describe the speed, we have 2 objects. We can combine them together in one vector:

Definition

The vector consisting of the partial derivatives of $f(x, y)$:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

is called the **gradient** of $f(x, y)$.

In the previous example, $\nabla f = [2x \quad 2y]$.

Partial Derivative

Similarly, for a function of n variables, $f(x_1, \dots, x_n) = f(\mathbf{x})$ we define partial derivatives as:

$$f_{x_1}(\mathbf{x}) = \frac{\partial f}{\partial x_1}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1 + h, x_2, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{h},$$

$$f_{x_2}(\mathbf{x}) = \frac{\partial f}{\partial x_2}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2 + h, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{h},$$

$$\vdots$$

$$f_{x_n}(\mathbf{x}) = \frac{\partial f}{\partial x_n}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2, \dots, x_n + h) - f(x_1, x_2, \dots, x_n)}{h}.$$

And the gradient of $f(\mathbf{x})$ as:

$$\nabla f = \frac{df}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1 \times n}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$

2. $$\frac{\partial}{\partial x_i}(c \cdot f(\mathbf{x})) = c \cdot \frac{\partial f(\mathbf{x})}{\partial x_i}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$
2. $$\frac{\partial}{\partial x_i}(c \cdot f(\mathbf{x})) = c \cdot \frac{\partial f(\mathbf{x})}{\partial x_i}$$
3. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \cdot g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \cdot g(\mathbf{x}) + f(\mathbf{x}) \cdot \frac{\partial g(\mathbf{x})}{\partial x_i}$$

Partial Derivative

Example

Let $f(x,y) = 2x^2$ and $g(x,y) = 4x + 6y$.

$$(f \cdot g)_x = 2x^2(4) + (4x + 6y)(4x) = 24x^2 + 24xy$$
$$(f \cdot g)_y = 2x^2(6) + (4x + 6y)(0) = 12x^2$$

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas.

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas. You notice that obviously, as temperature t gets higher, more people start consuming them. In particular,

$$x(t) = 400 + 3t, \qquad y(t) = 500 + 2t$$

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas. You notice that obviously, as temperature t gets higher, more people start consuming them. In particular,

$$x(t) = 400 + 3t, \qquad y(t) = 500 + 2t$$

How does a change of temperature affect your profit?

Chain Rule

In other words,

- if f depends on x and y
- and x (or y) depends on t
- how much does f change as t changes?

Chain Rule

In other words,

- if f depends on x and y
- and x (or y) depends on t
- how much does f change as t changes?

Turns out, there is a simple formula for that:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial t}$$

which is called the **chain rule**.

Chain Rule

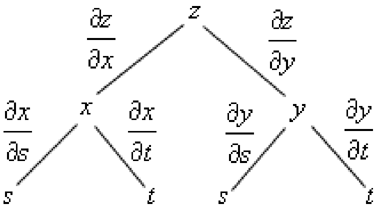
In other words,

- if f depends on x and y
- and x (or y) depends on t
- how much does f change as t changes?

Turns out, there is a simple formula for that:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial t}$$

which is called the **chain rule**.



Chain Rule

Example

Let $z = \sin(x^2 + y^2),\ x = t^2 + 3,\ y = t^3.$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2+y^2))\cdot(2t) + (2y\cos(x^2+y^2))\cdot(3t^2)$$
$$= 4xt\cos(x^2+y^2) + 6yt^2\cos(x^2+y^2)$$

Chain Rule

Example

Let $z = \sin(x^2 + y^2),\ x = t^2 + 3,\ y = t^3.$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2+y^2))\cdot(2t) + (2y\cos(x^2+y^2))\cdot(3t^2)$$
$$= 4xt\cos(x^2+y^2) + 6yt^2\cos(x^2+y^2)$$

The rule also works for one variable:

Example

Let $z(x) = x^2 + 4x,\ x(t) = 5t^3 + 2t.$

Chain Rule

Example

Let $z = \sin(x^2 + y^2)$, $x = t^2 + 3$, $y = t^3$.

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2 + y^2)) \cdot (2t) + (2y\cos(x^2 + y^2)) \cdot (3t^2)$$
$$= 4xt\cos(x^2 + y^2) + 6yt^2\cos(x^2 + y^2)$$

The rule also works for one variable:

Example

Let $z(x) = x^2 + 4x$, $x(t) = 5t^3 + 2t$. We can again use the chain rule:

$$\frac{dz}{dt} = \frac{dz}{dx} \cdot \frac{dx}{dt} = (2x + 4) \cdot (15t^2 + 2) = (2 \cdot (5t^3 + 2t) + 4) \cdot (15t^2 + 2)$$

$$= 150t^5 + 80t^3 + 60t^2 + 8t + 8$$

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$.

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$. If the following limit exists:

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

it is called the **directional derivative** of f **along the vector v**.

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$. If the following limit exists:

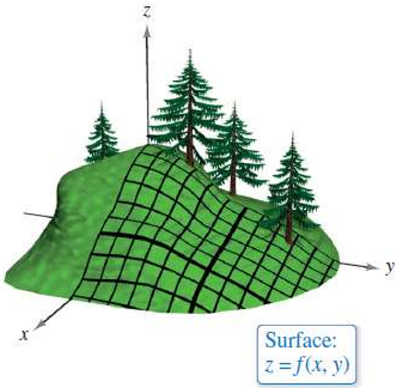
$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

it is called the **directional derivative** of f **along the vector v**.

Note that the directional derivative is a *number* (like partial derivatives), not a vector.

Directional Derivative

The directional derivative shows how much our function changes if we "walk" not only along the x or y-axis, but by an arbitrary direction of our choice.



Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Question

What happens if you change your variables along the direction $\mathbf{e}_1 = [1 \quad 0]$?
In other words,

$$\nabla_{\mathbf{e}_1} f =$$

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Question

What happens if you change your variables along the direction $\mathbf{e}_1 = [1 \quad 0]$?
In other words,

$$\nabla_{\mathbf{e}_1} f = f_x$$

Directional Derivative

How to actually compute the directional derivative?

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

- Play with directional derivative

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

- Play with directional derivative

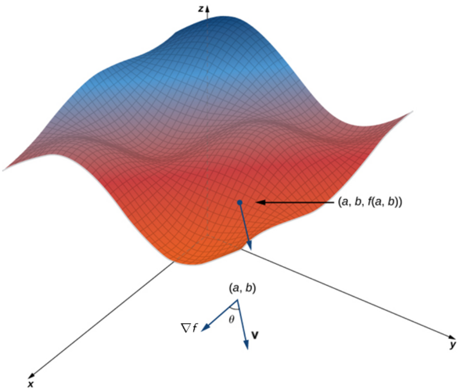
A particularly important question you might ask is:

Question

By which direction should I move, so the function increases the most?

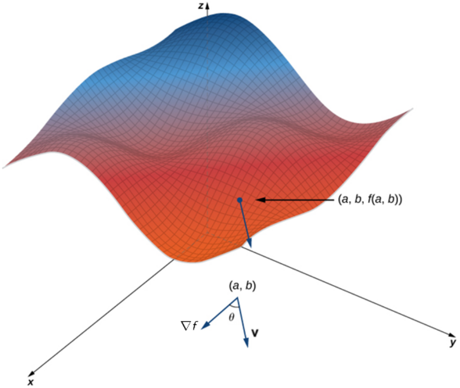
In other words, along which direction does $\nabla_{\mathbf{v}} f$ take its highest value?

Directional Derivative



Suppose $\mathbf{v}$ is any vector (with $\|\mathbf{v}\| = 1$).

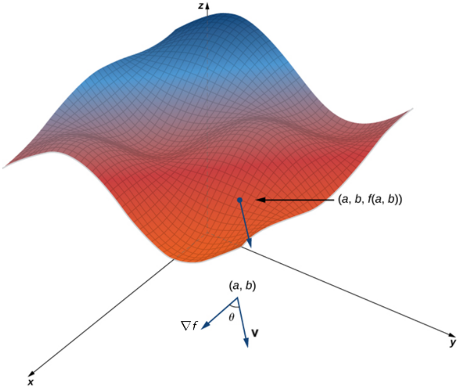
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} =$$

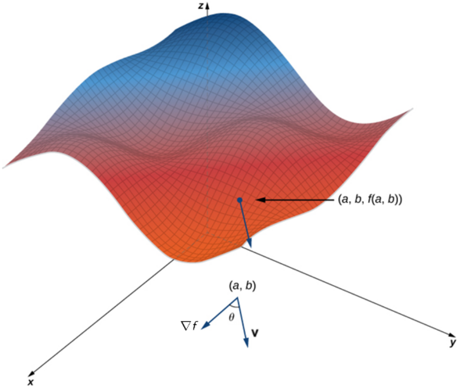
Directional Derivative



Suppose $\mathbf{v}$ is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta =$$

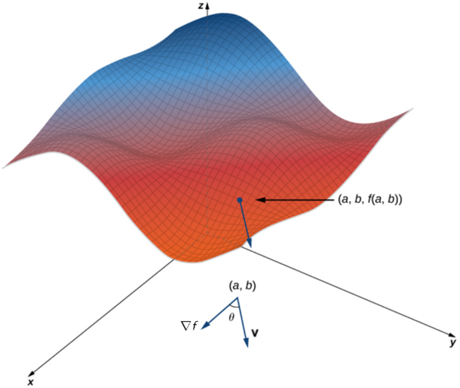
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta$$

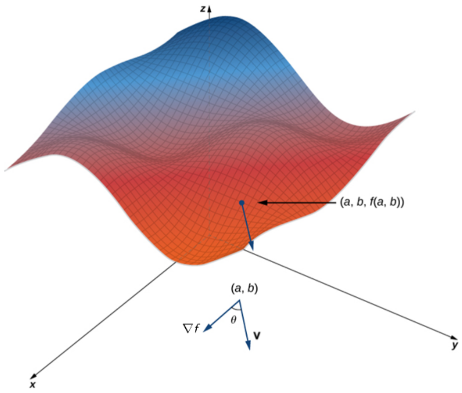
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq$$

Directional Derivative

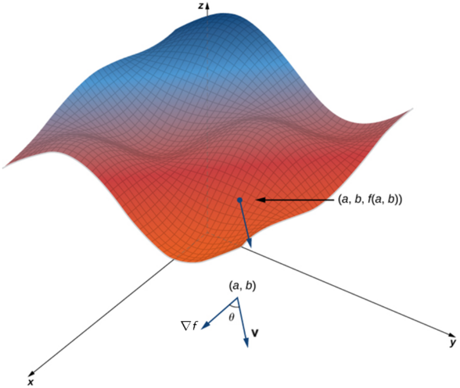


Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq \|\nabla f\|$$

where θ is the angle between **v** and ∇f.

Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq \|\nabla f\|$$

where θ is the angle between **v** and ∇f.

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Theorem

The gradient is the **fastest increasing direction** of the function.

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Theorem

The gradient is the **fastest increasing direction** of the function.

Similarly, $-\nabla f$ is the fastest decreasing direction of the function.

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Definition

$\mathbf{x}_0$ is called a **local maximum (minimum)** point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0) \quad (f(\mathbf{x}) \geq f(\mathbf{x}_0))$.

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Definition

$\mathbf{x}_0$ is called a **local maximum (minimum)** point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0) \;\; (f(\mathbf{x}) \geq f(\mathbf{x}_0))$.

Theorem

If $\mathbf{x}_0$ is a local extremum point of f and there exists $\nabla f(\mathbf{x}_0)$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$. (The converse is not true).

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Definition

$\mathbf{x}_0$ is called a **local maximum (minimum)** point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \le f(\mathbf{x}_0)$ $\;(f(\mathbf{x}) \ge f(\mathbf{x}_0))$.

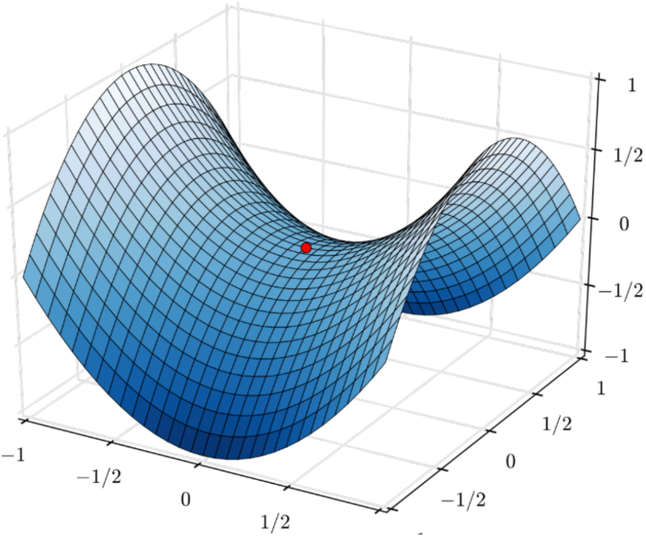
Theorem

If $\mathbf{x}_0$ is a local extremum point of f and there exists $\nabla f(\mathbf{x}_0)$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$. (The converse is not true).

Definition

$\mathbf{x}_0$ is called a **saddle point** of f if $\nabla f(\mathbf{x}_0) = \mathbf{0}$ but it's not an extremum point.

Extrema of a Function



Extrema of a Function

For the functions of one variable we looked at the sign of f''.
In case of two variables, we look at:

- f_{xx}
- and $D = f_{xx}f_{yy} - f_{xy}^2$

For the functions of one variable we looked at the sign of f''.
In case of two variables, we look at:

- f_{xx}
- and $D = f_{xx}f_{yy} - f_{xy}^2$

Theorem

If $\nabla f(a, b) = \mathbf{0}$ at some point (a, b), and

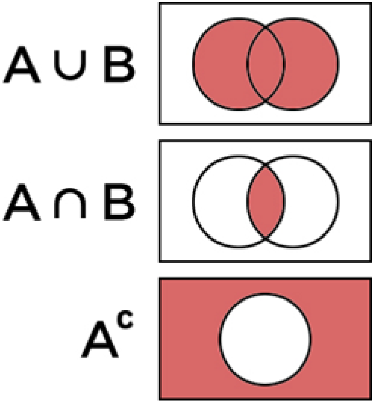
- $D > 0$ and $f_{xx} > 0$ $\Rightarrow$ local minimum
- $D > 0$ and $f_{xx} < 0$ $\Rightarrow$ local maximum
- $D < 0$ $\Rightarrow$ saddle point

Probability, Independence, Bayes Rule

Hayk Aprikyan, Hayk Tarkhanyan

Preliminaries

Recall the set operations:



Union $A \cup B$:
All elements that belong to A or B or both

Intersection $A \cap B$:
All elements that belong to *both* A and B

Complement A^c:
All elements that do not belong to A

What is probability?

- Oftentimes, we encounter a situation where we do not know the precise value or outcome of a particular event or circumstance.

What is probability?

- Oftentimes, we encounter a situation where we do not know the precise value or outcome of a particular event or circumstance.
- E.g. predicting the score of a football match, or the outcome of a tossed coin.

What is probability?

- Oftentimes, we encounter a situation where we do not know the precise value or outcome of a particular event or circumstance.
- E.g. predicting the score of a football match, or the outcome of a tossed coin.
- While we cannot tell the exact outcome of such an event, we can still speculate about the likely outcomes (e.g. which outcome is more likely than the others).

What is probability?

- Oftentimes, we encounter a situation where we do not know the precise value or outcome of a particular event or circumstance.
- E.g. predicting the score of a football match, or the outcome of a tossed coin.
- While we cannot tell the exact outcome of such an event, we can still speculate about the likely outcomes (e.g. which outcome is more likely than the others).
- The mathematical notion associated with the likeliness of a particular output to happen is called **probability**.

Outcomes

- A **random** (or **probabilistic**) **experiment** is a situation, where we are uncertain about the result.

Outcomes

- A **random** (or **probabilistic**) **experiment** is a situation, where we are uncertain about the result.
- The possible results of a random experiment are called the **outcomes**.

Outcomes

- A **random** (or **probabilistic**) **experiment** is a situation, where we are uncertain about the result.
- The possible results of a random experiment are called the **outcomes**.
- The set of all outcomes is called the **sample space** of the experiment.

Outcomes

- A **random** (or **probabilistic**) **experiment** is a situation, where we are uncertain about the result.
- The possible results of a random experiment are called the **outcomes**.
- The set of all outcomes is called the **sample space** of the experiment.

We denote the sample space with the letter Ω, so

$$\Omega = \{\text{all possible outcomes}\}$$

Outcomes

Example

A football match is a random experiment, where:

- Outcome is the score of the game.
- For example, one possible outcomes is "Pyunik 2 - 1 Alashkert". One way to denote it is $(2, 1)$.
- The sample space is:

$$\Omega = \{(0,0); (0,1); (1,0); (1,1); (2,0); \dots\}$$

Outcomes

Example

A football match is a random experiment, where:

- Outcome is the score of the game.
- For example, one possible outcomes is "Pyunik 2 - 1 Alashkert". One way to denote it is $(2, 1)$.
- The sample space is:

$$\Omega = \{(0,0); (0,1); (1,0); (1,1); (2,0); \dots\}$$

Example

Tossing a coin is a random experiment, where:

- Outcomes are **Heads** and **Tails**,
- The sample space is:

$$\Omega = \{\textbf{Heads}, \textbf{Tails}\} = \{H, T\}$$

Outcomes

Example

A football match is a random experiment, where:

- Outcome is the score of the game.
- For example, one possible outcomes is "Pyunik 2 - 1 Alashkert". One way to denote it is $(2, 1)$.
- The sample space is:

$$\Omega = \{(0,0); (0,1); (1,0); (1,1); (2,0); \dots\}$$

Example

Tossing a coin is a random experiment, where:

- Outcomes are **Heads** and **Tails**,
- The sample space is:

$$\Omega = \{\textbf{Heads}, \textbf{Tails}\} = \{H, T\}$$

Informally, we can say that both outcomes are equally likely: $50/50$.

Outcomes

Let's observe an example where the sample space contains infinite values.

Outcomes

Let's observe an example where the sample space contains infinite values.

Example

A student waits for the bus in the station. How much will she wait?

- Outcome is the time until the bus arrives.
- Practically, it can be *any* positive value, perhaps under 60 minutes. E.g. 2.3497 minutes is a possible outcome.
- The sample space is:

$$\Omega = [0, 60]$$

Outcomes

Let's observe an example where the sample space contains infinite values.

Example

A student waits for the bus in the station. How much will she wait?

- Outcome is the time until the bus arrives.
- Practically, it can be *any* positive value, perhaps under 60 minutes. E.g. 2.3497 minutes is a possible outcome.
- The sample space is:

$$\Omega = [0, 60]$$

What is the probability that the bus arrives at 20.230911 minutes?

Outcomes

Let's observe an example where the sample space contains infinite values.

Example

A student waits for the bus in the station. How much will she wait?

- Outcome is the time until the bus arrives.
- Practically, it can be *any* positive value, perhaps under 60 minutes. E.g. 2.3497 minutes is a possible outcome.
- The sample space is:

$$\Omega = [0, 60]$$

What is the probability that the bus arrives at 20.230911 minutes? How about exactly 20.0 minutes?

Outcomes

Let's observe an example where the sample space contains infinite values.

Example

A student waits for the bus in the station. How much will she wait?

- Outcome is the time until the bus arrives.
- Practically, it can be *any* positive value, perhaps under 60 minutes. E.g. 2.3497 minutes is a possible outcome.
- The sample space is:

$$\Omega = [0, 60]$$

What is the probability that the bus arrives at 20.230911 minutes? How about exactly 20.0 minutes? The probability is **zero**.

Outcomes

Let's observe an example where the sample space contains infinite values.

Example

A student waits for the bus in the station. How much will she wait?

- Outcome is the time until the bus arrives.
- Practically, it can be *any* positive value, perhaps under 60 minutes. E.g. 2.3497 minutes is a possible outcome.
- The sample space is:

$$\Omega = [0, 60]$$

What is the probability that the bus arrives at 20.230911 minutes? How about exactly 20.0 minutes? The probability is **zero**.

Whenever Ω is an interval (e.g. $(0, 1), [1, 10], [0, +\infty)$), the probability of each outcome is 0.
(of course, this does not mean that they are impossible – watch this!)

Events

Let's consider rolling a die.

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?
  - Is it any number?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?
  - Is it any number?
  - Was there no outcome after rolling the die?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?
  - Is it any number?
  - Was there no outcome after rolling the die?

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
    - Is the outcome equal to 4?
    - Is it smaller than 3?
    - Is it a prime number?
    - Is it a natural number?
    - Is it any number?
    - Was there no outcome after rolling the die?

  To formulate the questions which are of interest to us, we represent them by *sets*. For example, the third question can be represented by the set $\{2, 3, 5\}$.

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?
  - Is it any number?
  - Was there no outcome after rolling the die?

  To formulate the questions which are of interest to us, we represent them by *sets*. For example, the third question can be represented by the set $\{2, 3, 5\}$.
- The sets of outcomes of interest to us are called **events**. The set $\{2, 3, 5\}$ above is an event.

Events

Let's consider rolling a die.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Questions we might be interested in are:
  - Is the outcome equal to 4?
  - Is it smaller than 3?
  - Is it a prime number?
  - Is it a natural number?
  - Is it any number?
  - Was there no outcome after rolling the die?

  To formulate the questions which are of interest to us, we represent them by *sets*. For example, the third question can be represented by the set $\{2, 3, 5\}$.
- The sets of outcomes of interest to us are called **events**. The set $\{2, 3, 5\}$ above is an event.
- The set of all events is called the **event space** and denoted by $\mathcal{F}$.

Events

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

Events

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

The event of having an odd number is:

$$B = \{1, 3, 5\}$$

Events

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

The event of having an odd number is:

$$B = \{1, 3, 5\}$$

and the event of an even number is:

$$C = \{2, 4, 6\}$$

Events

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

The event of having an odd number is:

$$B = \{1, 3, 5\}$$

and the event of an even number is:

$$C = \{2, 4, 6\}$$

Can A and B happen at the same time?

Events

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

The event of having an odd number is:

$$B = \{1, 3, 5\}$$

and the event of an even number is:

$$C = \{2, 4, 6\}$$

Can A and B happen at the same time? How would you denote the event of both A and B happening?

Example

If you roll a die, the event that a number smaller than 3 shows up is the set:

$$A = \{1, 2\} \in \mathcal{F}$$

The event of having an odd number is:

$$B = \{1, 3, 5\}$$

and the event of an even number is:

$$C = \{2, 4, 6\}$$

Can A and B happen at the same time? How would you denote the event of both A and B happening? What about B and C?

Events

Definition

Two events A and B of the same experiment are called **disjoint** or **mutually exclusive** if $A \cap B = \varnothing$.

In other words, events are disjoint if they cannot occur at the same time.

Events

Definition

Two events A and B of the same experiment are called **disjoint** or **mutually exclusive** if $A \cap B = \varnothing$.

In other words, events are disjoint if they cannot occur at the same time.

Example

When rolling a die, the events $A = \{1, 4\}$ and $B = \{2, 5\}$ are disjoint, while A and $C = \{3, 4, 5\}$ are not.

Example

When waiting for a bus, the events $A = [0, 20]$ and $B = [30, 40]$ are disjoint, but none of them is disjoint with $C = [10, 40]$.

Probability

Question

How can we measure the possibilities of different events?

Probability

Question

How can we measure the possibilities of different events?

It depends on the problem/experiment itself.

Probability

Question

How can we measure the possibilities of different events?

It depends on the problem/experiment itself.

Suppose you roll a fair die. Since each of the outcomes has the same likelihood as the others, you would expect the probability of rolling, say, 3 to be equal to $\frac{1}{6}$:

$$\mathbb{P}(3) = \frac{1}{6}$$

Probability

Question

How can we measure the possibilities of different events?

It depends on the problem/experiment itself.

Suppose you roll a fair die. Since each of the outcomes has the same likelihood as the others, you would expect the probability of rolling, say, 3 to be equal to $\frac{1}{6}$:

$$\mathbb{P}(3) = \frac{1}{6}$$

Similarly, each of the outcomes $\{1, 2, 3, 4, 5, 6\}$ also has the probability

$$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}$$

of showing up.

Question

How can we measure the possibilities of different events?

It depends on the problem/experiment itself.

Suppose you roll a fair die. Since each of the outcomes has the same likelihood as the others, you would expect the probability of rolling, say, 3 to be equal to $\frac{1}{6}$:

$$\mathbb{P}(3) = \frac{1}{6}$$

Similarly, each of the outcomes $\{1, 2, 3, 4, 5, 6\}$ also has the probability

$$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}$$

of showing up.

Experiments like this (where all outcomes have the same probability of occuring) are called **equiprobable**.

Probability

The probability of an event, for example, $A = \{3, 5, 6\}$, will be

$$\mathbb{P}(A) =$$

Probability

The probability of an event, for example, $A = \{3, 5, 6\}$, will be

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$$

since A has 3 elements and the total number of outcomes is 6.

In equiprobable case, the probability of any event $A \in \mathcal{F}$ is:

$$\mathbb{P}(A) = \frac{\text{number of elements in } A}{\text{total number of outcomes}} = \frac{|A|}{|\Omega|}$$

Probability

The probability of an event, for example, $A = \{3, 5, 6\}$, will be

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$$

since A has 3 elements and the total number of outcomes is 6.

In equiprobable case, the probability of any event $A \in \mathcal{F}$ is:

$$\mathbb{P}(A) = \frac{\text{number of elements in } A}{\text{total number of outcomes}} = \frac{|A|}{|\Omega|}$$

Example

Now we roll two fair dice. Our sample space will be
$\Omega = \{(x, y) \mid 1 \leq x, y \leq 6\}$.
Assuming that all of 36 outcomes are equally likely to show up, the probability of each (x, y) outcome is:

Probability

The probability of an event, for example, $A = \{3, 5, 6\}$, will be

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$$

since A has 3 elements and the total number of outcomes is 6.

In equiprobable case, the probability of any event $A \in \mathcal{F}$ is:

$$\mathbb{P}(A) = \frac{\text{number of elements in } A}{\text{total number of outcomes}} = \frac{|A|}{|\Omega|}$$

Example

Now we roll two fair dice. Our sample space will be
$\Omega = \{(x, y) \mid 1 \leq x, y \leq 6\}$.
Assuming that all of 36 outcomes are equally likely to show up, the probability of each (x, y) outcome is:

$$\mathbb{P}((x, y)) = \frac{1}{36} \qquad \text{for any } (x, y) \in \Omega$$

Probability

Question

What if instead of rolling a die, we go outside, choose a random person and check if they are left-handed? Would the probability still be $\frac{1}{2}$?

Probability

Question

What if instead of rolling a die, we go outside, choose a random person and check if they are left-handed? Would the probability still be $\frac{1}{2}$?

Since only about 10% of people are left-handed, it is 9 times less likely to pick a lefty than a righty, hence the probability is about $\frac{1}{10}$, not $\frac{1}{2}$.

Probability

Question

What if instead of rolling a die, we go outside, choose a random person and check if they are left-handed? Would the probability still be $\frac{1}{2}$?

Since only about 10% of people are left-handed, it is 9 times less likely to pick a lefty than a righty, hence the probability is about $\frac{1}{10}$, not $\frac{1}{2}$.

In this case, the experiment was not equiprobable. Instead, we had 2 outcomes, say,

$$\Omega = \{L, R\}$$

Probability

Question

What if instead of rolling a die, we go outside, choose a random person and check if they are left-handed? Would the probability still be $\frac{1}{2}$?

Since only about 10% of people are left-handed, it is 9 times less likely to pick a lefty than a righty, hence the probability is about $\frac{1}{10}$, not $\frac{1}{2}$.

In this case, the experiment was not equiprobable. Instead, we had 2 outcomes, say,

$$\Omega = \{L, R\}$$

with their probabilities being

$$\mathbb{P}(L) = \frac{1}{10}, \qquad \mathbb{P}(R) = \frac{9}{10}$$

Question

What if instead of rolling a die, we go outside, choose a random person and check if they are left-handed? Would the probability still be $\frac{1}{2}$?

Since only about 10% of people are left-handed, it is 9 times less likely to pick a lefty than a righty, hence the probability is about $\frac{1}{10}$, not $\frac{1}{2}$.

In this case, the experiment was not equiprobable. Instead, we had 2 outcomes, say,

$$\Omega = \{L, R\}$$

with their probabilities being

$$\mathbb{P}(L) = \frac{1}{10}, \qquad \mathbb{P}(R) = \frac{9}{10}$$

As it shows, probability can also be not equiprobable.

Probability

So in general, in different problems the *probability measure* $\mathbb{P}$ can be different, and it depends on the specifics of the problem how it is computed.

There are, however, three properties which the probability measure $\mathbb{P}$ always satisfies:

1. $\mathbb{P}(A) \geq 0$ for any event $A \in \mathcal{F}$,

Probability

So in general, in different problems the *probability measure* $\mathbb{P}$ can be different, and it depends on the specifics of the problem how it is computed.

There are, however, three properties which the probability measure $\mathbb{P}$ always satisfies:

1. $\mathbb{P}(A) \geq 0$ for any event $A \in \mathcal{F}$,
2. $\mathbb{P}(\Omega) = 1$,

Probability

So in general, in different problems the *probability measure* $\mathbb{P}$ can be different, and it depends on the specifics of the problem how it is computed.

There are, however, three properties which the probability measure $\mathbb{P}$ always satisfies:

1. $\mathbb{P}(A) \geq 0$ for any event $A \in \mathcal{F}$,
2. $\mathbb{P}(\Omega) = 1$,
3. For any disjoint events $A_1, A_2, \ldots$,

$$\mathbb{P}\left(A_1 \cup A_2 \cup \ldots\right) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots$$

i.e.

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$$

Probability

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Probability

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$

Probability

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

Probability

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. For any (maybe *not* disjoint) events $A_1, A_2, \dots$:

$$\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$$

Probability

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. For any (maybe *not* disjoint) events $A_1, A_2, \ldots$:

$$\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$$

5. **(Inclusion-Exclusion Rule)** For any events A and B:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. For any (maybe *not* disjoint) events $A_1, A_2, \ldots$:

$$\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$$

5. **(Inclusion-Exclusion Rule)** For any events A and B:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Probability

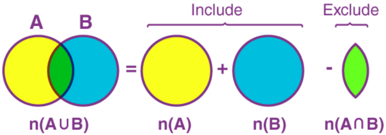
Other useful properties

1. For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. For any (maybe *not* disjoint) events $A_1, A_2, \ldots$:

$$\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$$

5. **(Inclusion-Exclusion Rule)** For any events A and B:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$



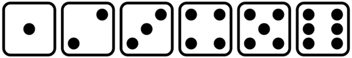
Conditional Probability

We role a fair die. Consider these two problems:

Conditional Probability

We role a fair die. Consider these two problems:

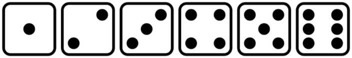
a. What is the probability that the outcome is even?



Conditional Probability

We role a fair die. Consider these two problems:

a. What is the probability that the outcome is even?



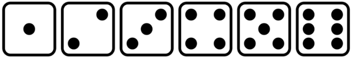
Since there are 6 possible outcomes, out of which only 3 are even,

$$A = \{2, 4, 6\}, \qquad \mathbb{P}(A) = \frac{1}{2}$$

Conditional Probability

We role a fair die. Consider these two problems:

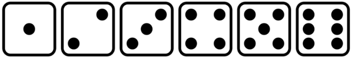
a. What is the probability that the outcome is even?



Since there are 6 possible outcomes, out of which only 3 are even,

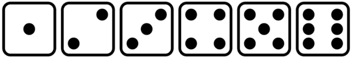
$$A = \{2, 4, 6\}, \qquad \mathbb{P}(A) = \frac{1}{2}$$

b. What is the probability that the outcome is even, given that we know that it is prime?



We role a fair die. Consider these two problems:

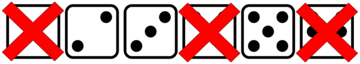
a. What is the probability that the outcome is even?



Since there are 6 possible outcomes, out of which only 3 are even,

$$A = \{2, 4, 6\}, \qquad \mathbb{P}(A) = \frac{1}{2}$$

b. What is the probability that the outcome is even, given that we know that it is prime?

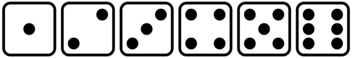


Here we know that the outcome **cannot be** 1, 4 or 6.

Conditional Probability

We role a fair die. Consider these two problems:

a. What is the probability that the outcome is even?



Since there are 6 possible outcomes, out of which only 3 are even,

$$A = \{2, 4, 6\}, \qquad \mathbb{P}(A) = \frac{1}{2}$$

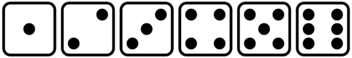
b. What is the probability that the outcome is even, given that we know that it is prime?

Here we know that the outcome **cannot be** 1, 4 or 6. Therefore we are left with only three possible outcomes $B = \{2, 3, 5\}$, out of which only 2 is even.

Conditional Probability

We role a fair die. Consider these two problems:

a. What is the probability that the outcome is even?



Since there are 6 possible outcomes, out of which only 3 are even,

$$A = \{2, 4, 6\}, \qquad \mathbb{P}(A) = \frac{1}{2}$$

b. What is the probability that the outcome is even, given that we know that it is prime?

Here we know that the outcome **cannot be** 1, 4 or 6. Therefore we are left with only three possible outcomes $B = \{2, 3, 5\}$, out of which only 2 is even. So in this case the probability of being even is $\frac{1}{3}$.

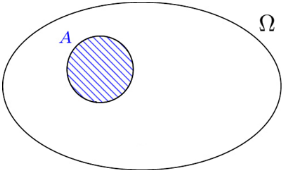
Conditional Probability

Definition

For any events A and B (if $\mathbb{P}(B) \neq 0$), the following number:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

is called the **conditional probability** of A **given** B
(or the probability of A under the condition of B).



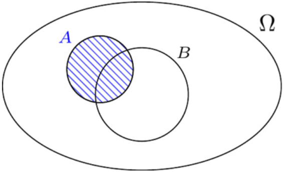
Conditional Probability

Definition

For any events A and B (if $\mathbb{P}(B) \neq 0$), the following number:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

is called the **conditional probability** of A **given** B
(or the probability of A under the condition of B).



more information
is given

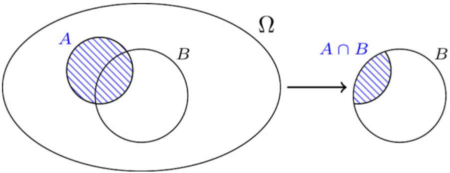
Conditional Probability

Definition

For any events A and B (if $\mathbb{P}(B) \neq 0$), the following number:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

is called the **conditional probability** of A **given** B
(or the probability of A under the condition of B).



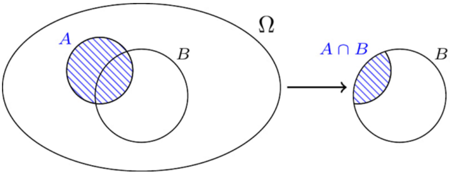
Conditional Probability

Definition

For any events A and B (if $\mathbb{P}(B) \neq 0$), the following number:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

is called the **conditional probability** of A **given** B
(or the probability of A under the condition of B).



In our problem, if A means being even and B means being prime, we had:

$$\mathbb{P}(\text{even} \mid \text{given that prime}) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Conditional Probability

Example

Suppose we roll two fair dice. What is the probability that the first one is 2, given that their sum is no greater than 5?

Conditional Probability

Example

Suppose we roll two fair dice. What is the probability that the first one is 2, given that their sum is no greater than 5?

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Usually, there would be 36 total outcomes, but since we *know* the sum is 5, they are only 10 possible outcomes left. Out of them only three outcomes ("2-1", "2-2" and "2-3") are desired. So the probability is $\frac{3}{10}$.

Law of Total Probability

Question

In some university, 3/5 of all students are women and the rest are men. It is known that 15% of men are over left-handed, while only 10% of women are. If you choose a random student, what is the probability of the student being left-handed?

Law of Total Probability

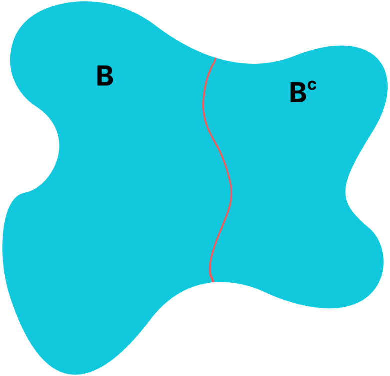
Question

In some university, 3/5 of all students are women and the rest are men. It is known that 15% of men are over left-handed, while only 10% of women are. If you choose a random student, what is the probability of the student being left-handed?

Let B denote the event that the randomly selected student is a woman, and B^c that he is a man.

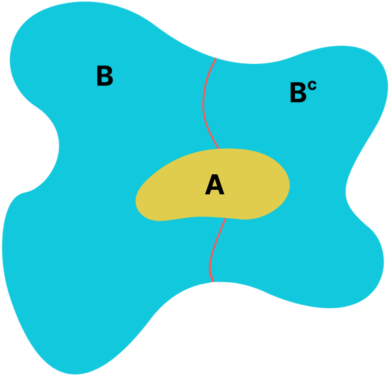
With another letter, say A, let us denote the event of being left-handed.

Law of Total Probability



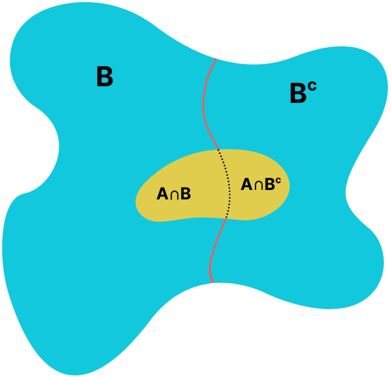
$\mathbb{P}(A) =$

Law of Total Probability



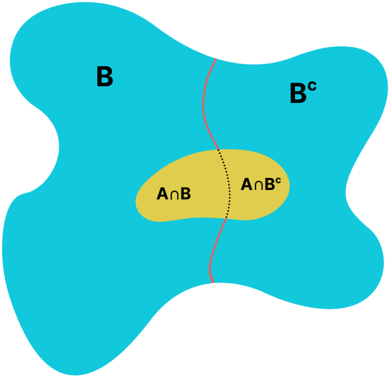
$\mathbb{P}(A) =$

Law of Total Probability



$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) =$$

Law of Total Probability



$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(B) \cdot \mathbb{P}(A|B) + \mathbb{P}(B^c) \cdot \mathbb{P}(A|B^c)$$

Law of Total Probability

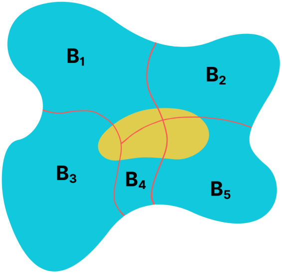
What we get is known as the Law of Total Probability:

Theorem

If A and B are some events such that $\mathbb{P}(B) \neq 0$, then

$$\mathbb{P}(A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B) + \mathbb{P}(B^c) \cdot \mathbb{P}(A|B^c)$$

Law of Total Probability



We can also generalize it to the case of three or more subgroups:

Theorem

If $B_1, B_2, \ldots, B_n$ are some disjoint events such that $A \subset \bigcup_{k=1}^{n} B_k$, then

$$\mathbb{P}(A) = \mathbb{P}(B_1) \cdot \mathbb{P}(A|B_1) + \cdots + \mathbb{P}(B_n) \cdot \mathbb{P}(A|B_n)$$

Law of Total Probability

Example

There are 52 cards in a deck. One of the cards is randomly removed from the deck, 51 are left. What is the probability that if we take a card, it will be a diamond?

Law of Total Probability

Example

There are 52 cards in a deck. One of the cards is randomly removed from the deck, 51 are left. What is the probability that if we take a card, it will be a diamond?

If we denote by A the probability of taking a diamond, and by B the probability that the removed card was a diamond, then either:

Law of Total Probability

Example

There are 52 cards in a deck. One of the cards is randomly removed from the deck, 51 are left. What is the probability that if we take a card, it will be a diamond?

If we denote by A the probability of taking a diamond, and by B the probability that the removed card was a diamond, then either:

- the taken card was a diamond, and there are 12 left,
- the taken card wasn't a diamond, and there are 13 left.

Law of Total Probability

Example

There are 52 cards in a deck. One of the cards is randomly removed from the deck, 51 are left. What is the probability that if we take a card, it will be a diamond?

If we denote by A the probability of taking a diamond, and by B the probability that the removed card was a diamond, then either:

- the taken card was a diamond, and there are 12 left,
- the taken card wasn't a diamond, and there are 13 left.

The total probability will be:

$$\mathbb{P}(A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B) + \mathbb{P}(B^c) \cdot \mathbb{P}(A|B^c)$$

$$\mathbb{P}(A) = \frac{13}{52} \cdot \frac{12}{51} + \frac{39}{52} \cdot \frac{13}{51} = \frac{1}{4}$$

Bayes Rule

One more simple yet powerful tool is the so called Bayes Rule:

Theorem

If A and B are some events (with non-zero probabilities), then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Bayes Rule

One more simple yet powerful tool is the so called Bayes Rule:

Theorem

If A and B are some events (with non-zero probabilities), then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Example

In nature, dangerous fires are rare (about 1% chance) but smoke is fairly common (10% chance) due to barbecues, and 90% of dangerous fires make smoke.

Bayes Rule

One more simple yet powerful tool is the so called Bayes Rule:

Theorem

If A and B are some events (with non-zero probabilities), then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Example

In nature, dangerous fires are rare (about 1% chance) but smoke is fairly common (10% chance) due to barbecues, and 90% of dangerous fires make smoke. You see a smoke. What is the probability that there is dangerous fire?

Bayes Rule

One more simple yet powerful tool is the so called Bayes Rule:

Theorem

If A and B are some events (with non-zero probabilities), then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Example

In nature, dangerous fires are rare (about 1% chance) but smoke is fairly common (10% chance) due to barbecues, and 90% of dangerous fires make smoke. You see a smoke. What is the probability that there is dangerous fire?

Let F denote dangerous fire, S smoke. Then:

$$\mathbb{P}(F|S) = \frac{\mathbb{P}(F) \cdot \mathbb{P}(S|F)}{\mathbb{P}(S)} = \frac{1}{100} \cdot \frac{90}{100} : \frac{10}{100} = 0.09$$

Independence

Now let's consider another simple experiment: flipping a fair coin and rolling a six-sided die. The probability of getting **Heads** on the coin flip is independent of the probability of rolling a specific number on the die, as these two events *do not affect* each other.

Independence

Now let's consider another simple experiment: flipping a fair coin and rolling a six-sided die. The probability of getting **Heads** on the coin flip is independent of the probability of rolling a specific number on the die, as these two events *do not affect* each other.

Definition

Events A and B of the same experiment are called **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Independence

Now let's consider another simple experiment: flipping a fair coin and rolling a six-sided die. The probability of getting **Heads** on the coin flip is independent of the probability of rolling a specific number on the die, as these two events *do not affect* each other.

Definition

Events A and B of the same experiment are called **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

or, equivalently,

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

Independence

Now let's consider another simple experiment: flipping a fair coin and rolling a six-sided die. The probability of getting **Heads** on the coin flip is independent of the probability of rolling a specific number on the die, as these two events *do not affect* each other.

Definition

Events A and B of the same experiment are called **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

or, equivalently,

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

The events are called **dependent** if they are not independent.

Independence

Now let's consider another simple experiment: flipping a fair coin and rolling a six-sided die. The probability of getting **Heads** on the coin flip is independent of the probability of rolling a specific number on the die, as these two events *do not affect* each other.

Definition

Events A and B of the same experiment are called **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

or, equivalently,

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

The events are called **dependent** if they are not independent.

This definition makes sense because it means that the probability of both A and B occurring together is simply the product of their individual probabilities: the outcome of one event has no effect on the other.

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

$$\mathbb{P}(A) =$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2},$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event A = the first die is even, B = the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) =$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event A = the first die is even, B = the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(A \cap B) =$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4}$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event A = the first die is even, B = the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$$

Independence

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Example

Suppose we are rolling two fair dice (for the last time).
Event $A =$ the first die is even, $B =$ the second die is odd.

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint.
   (Hence disjoint events are dependent.)

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint. (Hence disjoint events are dependent.)
2. If A and B are independent events, then

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint. (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint.
   (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c
   - A^c and B

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint. (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c
   - A^c and B
   - A^c and B^c

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint.
   (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c
   - A^c and B
   - A^c and B^c

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint. (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c
   - A^c and B
   - A^c and B^c

   are also independent.

Independence

Suppose A and B are some events with $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$.

Properties

1. If A and B are independent, they cannot be disjoint. (Hence disjoint events are dependent.)
2. If A and B are independent events, then
   - A and B^c
   - A^c and B
   - A^c and B^c

   are also independent.

When solving a problem, principles like total probability, Bayes rule, and independence help us a lot. Let's explore one more useful problem solving technique.

Geometric Probability

When thinking about probabilities of some events, it is often easier for us to think about them geometrically.

Geometric Probability

When thinking about probabilities of some events, it is often easier for us to think about them geometrically.

Example

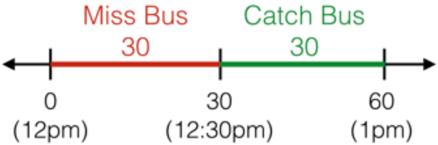
Your bus is coming at a random time between 12 pm and 1 pm. If you show up at 12:30 pm, how likely are you to catch the bus?

Geometric Probability

When thinking about probabilities of some events, it is often easier for us to think about them geometrically.

Example

Your bus is coming at a random time between 12 pm and 1 pm. If you show up at 12:30 pm, how likely are you to catch the bus?



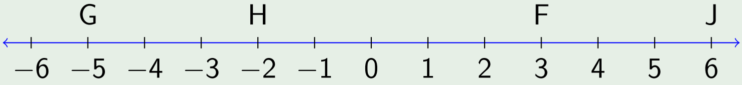
If we draw the hours on a number line and mark the time points for which we will miss or catch the bus, we see that the *length* of the segment for catching the bus is half the total length:

$$\mathbb{P}(\text{catching the bus}) = \frac{30}{30 + 30} = \frac{1}{2}$$

Geometric Probability

Example

A selection is to be made between points G and J as seen below

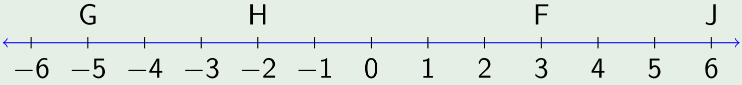


The probability that selection falls in HF is

Geometric Probability

Example

A selection is to be made between points G and J as seen below

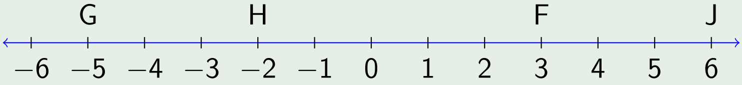


The probability that selection falls in HF is $\frac{HF}{GH} = \frac{5}{11}$.

Geometric Probability

Example

A selection is to be made between points G and J as seen below

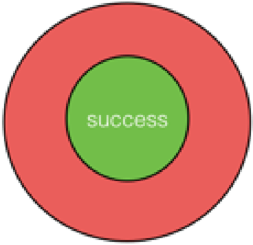


The probability that selection falls in HF is $\frac{HF}{GH} = \frac{5}{11}$.

So in general, we draw the sample space and the set of desired outcomes as lines or line segments, and divide the length of the "desired outcomes" by the length of the "sample space" to get the probability.

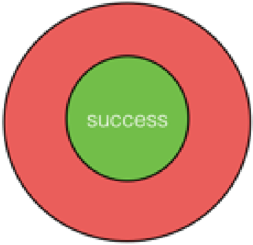
Geometric Probability

Imagine now that we are playing darts with two circles, and the smaller circle has two times smaller radius than the bigger one.



A dart is thrown at a random. What is the probability that it lands in the smaller circle?

Imagine now that we are playing darts with two circles, and the smaller circle has two times smaller radius than the bigger one.



A dart is thrown at a random. What is the probability that it lands in the smaller circle?

In this case, since the sample space and the set of desired outcomes is 2-dimensional, we should divide the *area* of the smaller circle by the *area* of the bigger circle.

The probability that the dart falls in the smaller circle, is $\frac{\pi r^2}{\pi(2r)^2} = \frac{1}{4}$.

Geometric Probability

One of the most powerful uses of geometric probability is applying it to problems that are not inherently geometric.

Geometric Probability

One of the most powerful uses of geometric probability is applying it to problems that are not inherently geometric.

Example

Both the bus and you get to the bus stop at random times between 12 pm and 1 pm. When the bus arrives, it waits for 5 minutes before leaving. When you arrive, you wait for 20 minutes before leaving if the bus doesn't come. What is the probability that you catch the bus?

Geometric Probability

One of the most powerful uses of geometric probability is applying it to problems that are not inherently geometric.

Example

Both the bus and you get to the bus stop at random times between 12 pm and 1 pm. When the bus arrives, it waits for 5 minutes before leaving. When you arrive, you wait for 20 minutes before leaving if the bus doesn't come. What is the probability that you catch the bus?

Let b denote the time that the bus arrives (after 12 pm), and y, the time you arrive. The set of all possible outcomes will be $[0, 60] \times [0, 60]$.

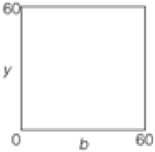
Geometric Probability

One of the most powerful uses of geometric probability is applying it to problems that are not inherently geometric.

Example

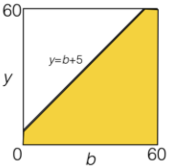
Both the bus and you get to the bus stop at random times between 12 pm and 1 pm. When the bus arrives, it waits for 5 minutes before leaving. When you arrive, you wait for 20 minutes before leaving if the bus doesn't come. What is the probability that you catch the bus?

Let b denote the time that the bus arrives (after 12 pm), and y, the time you arrive. The set of all possible outcomes will be $[0, 60] \times [0, 60]$.

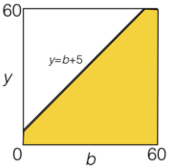


Geometric Probability

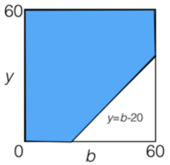
Then, we need to determine the region of "success"; that is, the points where we catch the bus. Since the bus will wait for 5 minutes, you need to arrive within 5 minutes of the bus' arrival, so $y \leq b + 5$.



Then, we need to determine the region of "success"; that is, the points where we catch the bus. Since the bus will wait for 5 minutes, you need to arrive within 5 minutes of the bus' arrival, so $y \leq b + 5$.

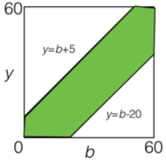


However, you only wait for 20 minutes, so you can't arrive more than 20 minutes before the bus, so $y \geq b - 20$.



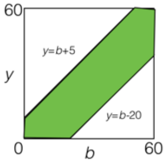
Geometric Probability

Combining our two conditions, we can draw the region of success:

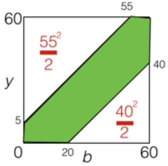


Geometric Probability

Combining our two conditions, we can draw the region of success:

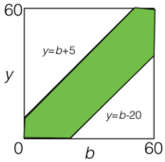


Now, we just need to find the area of this region. A simple method is to find the remaining area, and then subtract that from the total area:

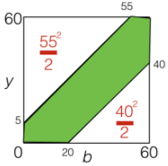


Geometric Probability

Combining our two conditions, we can draw the region of success:



Now, we just need to find the area of this region. A simple method is to find the remaining area, and then subtract that from the total area:



$$\mathbb{P}(\text{catching the bus}) = \frac{60^2 - \frac{55^2}{2} - \frac{40^2}{2}}{60^2} = \frac{103}{288}$$

Random Variables

Hayk Aprikyan, Hayk Tarkhanyan

April 15, 2025

Random Variables

Recap:

During the previous lecture we had a problem like this:

Example

Both the bus and you get to the bus stop at random times between 12 pm and 1 pm. When the bus arrives, it waits for 5 minutes before leaving. When you arrive, you wait for 20 minutes before leaving if the bus doesn't come. What is the probability that you catch the bus?

Here we did not know the **exact values** of the arrival times, so in order to compare them, we **denoted them** by y and b – and calculated their probabilities.

In this case, we say that y and b are **random variables**.

Random Variables

Definition

A quantity whose value is unknown and depends on a random experiment is called a **random variable**.

Random Variables

Definition

A quantity whose value is unknown and depends on a random experiment is called a **random variable**.

Example

You throw a fair dice and win \$1 if it is prime, lose \$1 if it is composite (4 or 6), and stay even otherwise. In other words, if X is your profit, then

$$X = \begin{cases} 1, & \text{if } \omega \in \{2,3,5\} \\ -1, & \text{if } \omega \in \{4,6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

Random Variables

Definition

A quantity whose value is unknown and depends on a random experiment is called a **random variable**.

Example

You throw a fair dice and win \$1 if it is prime, lose \$1 if it is composite (4 or 6), and stay even otherwise. In other words, if X is your profit, then

$$X = \begin{cases} 1, & \text{if } \omega \in \{2, 3, 5\} \\ -1, & \text{if } \omega \in \{4, 6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

Since the exact value you win/lose cannot be known, it is a random variable.

Definition

A quantity whose value is unknown and depends on a random experiment is called a **random variable**.

Example

You throw a fair dice and win \$1 if it is prime, lose \$1 if it is composite (4 or 6), and stay even otherwise. In other words, if X is your profit, then

$$X = \begin{cases} 1, & \text{if } \omega \in \{2, 3, 5\} \\ -1, & \text{if } \omega \in \{4, 6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

Since the exact value you win/lose cannot be known, it is a random variable.

What is the probability that $X = 1$?

Random Variables

Definition

A quantity whose value is unknown and depends on a random experiment is called a **random variable**.

Example

You throw a fair dice and win \$1 if it is prime, lose \$1 if it is composite (4 or 6), and stay even otherwise. In other words, if X is your profit, then

$$X = \begin{cases} 1, & \text{if } \omega \in \{2, 3, 5\} \\ -1, & \text{if } \omega \in \{4, 6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

Since the exact value you win/lose cannot be known, it is a random variable.

What is the probability that $X = 1$? (We'll get to this later.)

Random Variables

Example

Two fair dice are thrown. The sum of their numbers is a random variable.

Random Variables

Example

Two fair dice are thrown. The sum of their numbers is a random variable.

Example

Two teams play football.

- The number of goals they score is a random variable.
- The difference between their scores is also a random variable.

Random Variables

Example

Two fair dice are thrown. The sum of their numbers is a random variable.

Example

Two teams play football.

- The number of goals they score is a random variable.
- The difference between their scores is also a random variable.

Example

You call a random number and ask the number of children in their family.

Random Variables

Example

Two fair dice are thrown. The sum of their numbers is a random variable.

Example

Two teams play football.

- The number of goals they score is a random variable.
- The difference between their scores is also a random variable.

Example

You call a random number and ask the number of children in their family.

Example

You are waiting for the bus 62. The time until it arrives is a random variable.

Random Variables

Example

Two fair dice are thrown. The sum of their numbers is a random variable.

Example

Two teams play football.

- The number of goals they score is a random variable.
- The difference between their scores is also a random variable.

Example

You call a random number and ask the number of children in their family.

Example

You are waiting for the bus 62. The time until it arrives is a random variable.

What values can each of the random variables above take?

Random Variables (optional)

Not a random variable

Two fair dice are thrown,

- who threw the dice,
- the color of the sky,
- what will show up if you had thrown a third die

are not random variables, as their values *do not depend* on the outcome of the dice (even if you knew the outcome, you could not answer these).

Random Variables (optional)

Not a random variable

Two fair dice are thrown,

- who threw the dice,
- the color of the sky,
- what will show up if you had thrown a third die

are not random variables, as their values *do not depend* on the outcome of the dice (even if you knew the outcome, you could not answer these).

In practice, every unknown thing in the given problem is a random variable.

Technically speaking, although, the value of the random variable should somehow depend on the outcome ω of the experiment. In other words, if you ask "When is $X = 4$?", there should be a way to measure that probability, i.e. $\{X = 4\}$ should be an event.

In your textbook you will find it as $\{\omega \text{ for which } X = 4\} \in \mathcal{F}$, but we skip the technicalities.

Random Variables

Notice that in some cases the possible values look like $\{2, 3, 4, \ldots, 12\}$ or $\{0, 1, 2, \ldots\}$, while in others they are intervals like $(0, +\infty)$.

Random Variables

Notice that in some cases the possible values look like $\{2, 3, 4, \ldots, 12\}$ or $\{0, 1, 2, \ldots\}$, while in others they are intervals like $(0, +\infty)$.

Definition

If the values of X can be represented as a list (say, x_1, x_2, x_3, $\ldots$ with probabilities 0.2, 0.1, 0.4, $\ldots$), it is called a **discrete random variable.**

Random Variables

Notice that in some cases the possible values look like $\{2, 3, 4, \dots, 12\}$ or $\{0, 1, 2, \dots\}$, while in others they are intervals like $(0, +\infty)$.

Definition

If the values of X can be represented as a list (say, $x_1,\ x_2,\ x_3,\ \dots$ with probabilities $0.2,\ 0.1,\ 0.4,\ \dots$), it is called a **discrete random variable.**

Otherwise,

Definition

If the values of X cannot be represented as a list (e.g. they are an interval), it is called a **continuous random variable.**

PMF

Of course, what we are interested in the most, is the probability that X takes a certain value.

For example:

$$X = \begin{cases} 1, & \text{if } \omega \in \{2,3,5\} \\ -1, & \text{if } \omega \in \{4,6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

and the question is:

$$\mathbb{P}(X = 1) = ?$$

PMF

Of course, what we are interested in the most, is the probability that X takes a certain value.

For example:

$$X = \begin{cases} 1, & \text{if } \omega \in \{2,3,5\} \\ -1, & \text{if } \omega \in \{4,6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

and the question is:

$$\mathbb{P}(X = 1) = ?$$

Since $X = 1$ corresponds to the outcomes $\{2,3,5\}$, the probability $\mathbb{P}(X = 1)$ is the same as $\mathbb{P}(\{2,3,5\})$

PMF

Of course, what we are interested in the most, is the probability that X takes a certain value.

For example:

$$X = \begin{cases} 1, & \text{if } \omega \in \{2, 3, 5\} \\ -1, & \text{if } \omega \in \{4, 6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

and the question is:

$$\mathbb{P}(X = 1) = ?$$

Since $X = 1$ corresponds to the outcomes $\{2, 3, 5\}$, the probability $\mathbb{P}(X = 1)$ is the same as $\mathbb{P}(\{2, 3, 5\}) = \frac{3}{6} = 0.5$.

PMF

Of course, what we are interested in the most, is the probability that X takes a certain value.

For example:

$$X = \begin{cases} 1, & \text{if } \omega \in \{2,3,5\} \\ -1, & \text{if } \omega \in \{4,6\} \\ 0, & \text{if } \omega = 1 \end{cases}$$

and the question is:

$$\mathbb{P}(X = 1) = ?$$

Since $X = 1$ corresponds to the outcomes $\{2,3,5\}$, the probability $\mathbb{P}(X = 1)$ is the same as $\mathbb{P}(\{2,3,5\}) = \dfrac{3}{6} = 0.5.$

The thing above, $\mathbb{P}(X = \text{something})$, is called the **probability mass function** or just the **PMF** of X.

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam.

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let $X =$ number of questions the student gets correct.

$$\mathbb{P}(X = 0) =$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14}$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let $X =$ number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) =$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14}$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14} = \frac{44}{105}$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14} = \frac{44}{105}$$

$$\mathbb{P}(X = 2) =$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let $X =$ number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14} = \frac{44}{105}$$

$$\mathbb{P}(X = 2) = \frac{4}{15} \cdot \frac{3}{14}$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X=0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X=1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14} = \frac{44}{105}$$

$$\mathbb{P}(X=2) = \frac{4}{15} \cdot \frac{3}{14} = \frac{6}{105} \qquad \left(\text{or just } 1 - \frac{11}{21} - \frac{44}{105}\right)$$

PMF

Example

A pitiful student knows 4 out of 15 topics, and there will be 2 questions on the exam. Let X = number of questions the student gets correct.

$$\mathbb{P}(X = 0) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21} \qquad \text{(rip student)}$$

$$\mathbb{P}(X = 1) = \frac{4}{15} \cdot \frac{11}{14} + \frac{11}{15} \cdot \frac{4}{14} = \frac{44}{105}$$

$$\mathbb{P}(X = 2) = \frac{4}{15} \cdot \frac{3}{14} = \frac{6}{105} \qquad \left(\text{or just } 1 - \frac{11}{21} - \frac{44}{105}\right)$$

k (right answers)	$\mathbb{P}(X = k)$
0	$11/21$
1	$44/105$
2	$6/105$

CDF

While interested in gaining knowledge and stuff, one thing the student cares most is if they pass or not, i.e.

- Not only the probabilities that $X = 0$ or $X = 1$ or $X = 2$,

CDF

While interested in gaining knowledge and stuff, one thing the student cares most is if they pass or not, i.e.

- Not only the probabilities that $X = 0$ or $X = 1$ or $X = 2$,
- But importantly, the probability that $X > 0$,

CDF

While interested in gaining knowledge and stuff, one thing the student cares most is if they pass or not, i.e.

- Not only the probabilities that $X = 0$ or $X = 1$ or $X = 2$,
- But importantly, the probability that $X > 0$,
  (or the opposite, $X \leq 2$ or $X < 2$).

CDF

While interested in gaining knowledge and stuff, one thing the student cares most is if they pass or not, i.e.

- Not only the probabilities that $X = 0$ or $X = 1$ or $X = 2$,
- But importantly, the probability that $X > 0$,
(or the opposite, $X \leq 2$ or $X < 2$).

In our case, $\mathbb{P}(X < 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \dfrac{99}{105}$.

CDF

While interested in gaining knowledge and stuff, one thing the student cares most is if they pass or not, i.e.

- Not only the probabilities that $X = 0$ or $X = 1$ or $X = 2$,
- But importantly, the probability that $X > 0$,
  (or the opposite, $X \leq 2$ or $X < 2$).

In our case, $\mathbb{P}(X < 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \dfrac{99}{105}$.

If there were, say, 5 questions and more than 3 points needed to pass, then we would be interested in

$$\mathbb{P}(X \leq 3) \qquad \text{(probability of failing)}$$

or

$$\mathbb{P}(X > 3) \qquad \text{(probability of passing)}$$

CDF

The quantity

$$\mathbb{P}(X \leq k) \qquad \text{(where } k \text{ is any number)}$$

is called the **cumulative distribution function** or the **CDF** of X.

CDF

The quantity

$$\mathbb{P}(X \leq k) \qquad \text{(where } k \text{ is any number)}$$

is called the **cumulative distribution function** or the **CDF** of X.

It is conveniently denoted by $F_X(k)$ or $F(k)$:

$$F_X(k) = \mathbb{P}(X \leq k)$$

CDF

The quantity

$$\mathbb{P}(X \leq k) \qquad \text{(where } k \text{ is any number)}$$

is called the **cumulative distribution function** or the **CDF** of X.

It is conveniently denoted by $F_X(k)$ or $F(k)$:

$$F_X(k) = \mathbb{P}(X \leq k)$$

Properties

- $0 \leq F_X(x) \leq 1$, for any $x \in \mathbb{R}$

CDF

The quantity

$$\mathbb{P}(X \leq k) \qquad \text{(where } k \text{ is any number)}$$

is called the **cumulative distribution function** or the **CDF** of X.

It is conveniently denoted by $F_X(k)$ or $F(k)$:

$$F_X(k) = \mathbb{P}(X \leq k)$$

Properties

- $0 \leq F_X(x) \leq 1, \;$ for any $x \in \mathbb{R}$
- $F_X(x)$ is a non-decreasing function

CDF

The quantity

$$\mathbb{P}(X \le k) \qquad \text{(where } k \text{ is any number)}$$

is called the **cumulative distribution function** or the **CDF** of X.

It is conveniently denoted by $F_X(k)$ or $F(k)$:

$$F_X(k) = \mathbb{P}(X \le k)$$

Properties

- $0 \le F_X(x) \le 1, \;$ for any $x \in \mathbb{R}$
- $F_X(x)$ is a non-decreasing function
- $\lim\limits_{x \to -\infty} F_X(x) = 0 \qquad$ and $\qquad \lim\limits_{x \to +\infty} F_X(x) = 1$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) =$$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$
$$F_X(6) =$$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$
$$F_X(6) = 1$$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$
$$F_X(6) = 1$$
$$F_X(5.9) =$$

CDF

Example

Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$
$$F_X(6) = 1$$
$$F_X(5.9) = \frac{5}{6}$$

CDF

Example

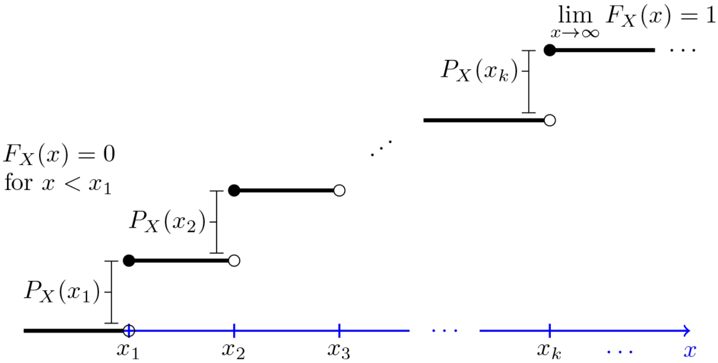
Let X indicate the number on a fair die. Then

$$F_X(6.1) = 1$$
$$F_X(6) = 1$$
$$F_X(5.9) = \frac{5}{6}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1, \\ \frac{1}{6}, & \text{if } 1 \le x < 2, \\ \frac{2}{6}, & \text{if } 2 \le x < 3, \\ \frac{3}{6}, & \text{if } 3 \le x < 4, \\ \frac{4}{6}, & \text{if } 4 \le x < 5, \\ \frac{5}{6}, & \text{if } 5 \le x < 6, \\ 1, & \text{if } x \ge 6 \end{cases}$$

CDF

Graphs of CDFs usually usually look like this:



CDF

Suppose we have a second pitiful student who has also studied for 4 topics out of 15. Let $Y =$ number of questions the second student gets correct.

CDF

Suppose we have a second pitiful student who has also studied for 4 topics out of 15. Let $Y =$ number of questions the second student gets correct.

- While the students may be asked different questions,
- $\Rightarrow$ it may happen that X and Y are not equal,
- still Y has the same PMF and CDF as X (check it!)

In this case, we say that X and Y are identically distributed:

Definition

Two random variables are said to be **identically distributed** if their PMFs/CDFs are equal.

CDF

Suppose we have a second pitiful student who has also studied for 4 topics out of 15. Let $Y =$ number of questions the second student gets correct.

- While the students may be asked different questions,
- $\Rightarrow$ it may happen that X and Y are not equal,
- still Y has the same PMF and CDF as X (check it!)

In this case, we say that X and Y are identically distributed:

Definition

Two random variables are said to be **identically distributed** if their PMFs/CDFs are equal.

Yet, identically distributed $\neq$ equal. E.g. if you toss a fair coin, and set

$$X = \begin{cases} 0, & \omega = \mathbf{H}, \\ 1, & \omega = \mathbf{T}, \end{cases} \qquad Y = \begin{cases} 1, & \omega = \mathbf{H}, \\ 0, & \omega = \mathbf{T}, \end{cases}$$

They both have the save PMFs:

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = \frac{1}{2}$$

CDF

Suppose we have a second pitiful student who has also studied for 4 topics out of 15. Let Y = number of questions the second student gets correct.

- While the students may be asked different questions,
- $\Rightarrow$ it may happen that X and Y are not equal,
- still Y has the same PMF and CDF as X (check it!)

In this case, we say that X and Y are identically distributed:

Definition

Two random variables are said to be **identically distributed** if their PMFs/CDFs are equal.

Yet, identically distributed $\neq$ equal. E.g. if you toss a fair coin, and set

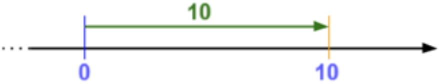
$$X = \begin{cases} 0, & \omega = \mathbf{H}, \\ 1, & \omega = \mathbf{T}, \end{cases} \qquad Y = \begin{cases} 1, & \omega = \mathbf{H}, \\ 0, & \omega = \mathbf{T}, \end{cases}$$

They both have the save PMFs:

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = \frac{1}{2} \qquad \text{but } X \neq Y$$

PDF

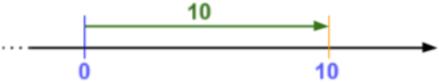
Finally, what if X is a continuous random variable?



Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$?

PDF

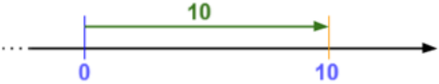
Finally, what if X is a continuous random variable?



Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$? What about $\mathbb{P}(X = 0.589)$?

PDF

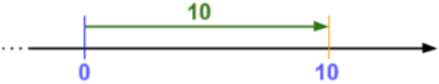
Finally, what if X is a continuous random variable?



Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$? What about $\mathbb{P}(X = 0.589)$? As previously, it is 0.

PDF

Finally, what if X is a continuous random variable?



Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$? What about $\mathbb{P}(X = 0.589)$? As previously, it is 0.

What can we say about the PMF of X?

PDF

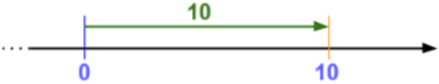
Finally, what if X is a continuous random variable?

Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$? What about $\mathbb{P}(X = 0.589)$? As previously, it is 0.

What can we say about the PMF of X? It is always zero!

PDF

Finally, what if X is a continuous random variable?



Say you randomly put your finger on some number X in $(0, 10)$. What is the probability that $\mathbb{P}(X = 0.5)$? What about $\mathbb{P}(X = 0.589)$? As previously, it is 0.

What can we say about the PMF of X? It is always zero!

So another question we might ask is:

Question

What do you think $\mathbb{P}(X \leq 5)$ is?

PDF

While for discrete random variables both PMF and CDF make sense, for continuous random variables PMF is useless, but still can rely on CDF:

$$\mathbb{P}(X \leq 5) = \frac{1}{2}$$

PDF

While for discrete random variables both PMF and CDF make sense, for continuous random variables PMF is useless, but still can rely on CDF:

$$\mathbb{P}(X \leq 5) = \frac{1}{2}$$

Fortunately, we have another tool in case of continuous random variables.

PDF

While for discrete random variables both PMF and CDF make sense, for continuous random variables PMF is useless, but still can rely on CDF:

$$\mathbb{P}(X \leq 5) = \frac{1}{2}$$

Fortunately, we have another tool in case of continuous random variables.

Definition

If X is a continuous random variable, then there exists a nonnegative function $f(x)$ such that for any $c \in \mathbb{R}$,

$$F_X(c) = \mathbb{P}(X \leq c) = \int_{-\infty}^{c} f(t)\,dt$$

The function f is called the **probability density function** or **PDF** of X.

PDF

While for discrete random variables both PMF and CDF make sense, for continuous random variables PMF is useless, but still can rely on CDF:

$$\mathbb{P}(X \leq 5) = \frac{1}{2}$$

Fortunately, we have another tool in case of continuous random variables.

Definition

If X is a continuous random variable, then there exists a nonnegative function $f(x)$ such that for any $c \in \mathbb{R}$,

$$F_X(c) = \mathbb{P}(X \leq c) = \int_{-\infty}^{c} f(t)\, dt$$

The function f is called the **probability density function** or **PDF** of X.

The PDF of a continuous random variable plays essentially the same role as the PMF of a discrete random variable.

PDF

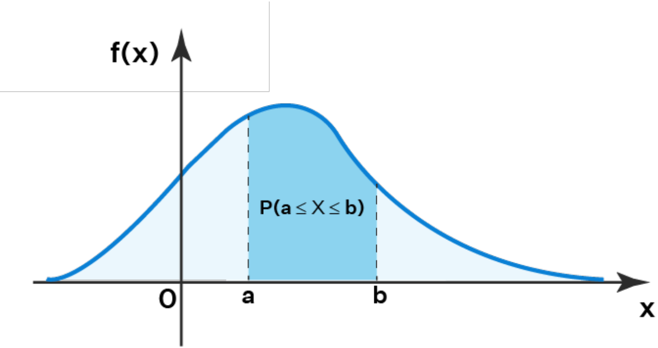
Just like the density of an object measures the concentration of mass (per unit volume), the probability density function captures the density of *probability* at point x:

$$f_X(x) = \lim_{h \to 0} \frac{\mathbb{P}(x < X \leq x + h)}{h}$$

PDF

Just like the density of an object measures the concentration of mass (per unit volume), the probability density function captures the density of *probability* at point x:

$$f_X(x) = \lim_{h \to 0} \frac{\mathbb{P}(x < X \leq x + h)}{h}$$



PDF

Properties

1. $f_X(x) \geq 0 \quad \text{for all } x \in \mathbb{R}$

PDF

Properties

1. $f_X(x) \geq 0 \quad \text{for all } x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

PDF

Properties

1. $f_X(x) \geq 0 \quad \text{for all } x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. If F_X is continuous at x, $F'_X(x) = f_X(x)$

PDF

Properties

1. $f_X(x) \geq 0 \quad \text{for all } x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. If F_X is continuous at x, $F_X'(x) = f_X(x)$
4. $\displaystyle\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$

PDF

Properties

1. $f_X(x) \geq 0 \quad \text{for all } x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. If F_X is continuous at x, $F_X'(x) = f_X(x)$
4. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \displaystyle\int_a^b f_X(x)\,dx$
5. $\mathbb{P}(X = a) = \mathbb{P}(X \leq a) - \mathbb{P}(X < a)$

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

Let's find the CDF and PDF of X.

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

Let's find the CDF and PDF of X.

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

Let's find the CDF and PDF of X.

For $x < a$, $F_X(x) = \mathbb{P}(X \leq x) = 0$,

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

Let's find the CDF and PDF of X.

For $x < a$, $F_X(x) = \mathbb{P}(X \leq x) = 0$,

For $x \geq b$, $F_X(x) = \mathbb{P}(X \leq x) = 1$,

PDF

Example

Ani chooses a random real number X uniformly from the interval $[a, b]$.

By "uniformly" we mean that for any two intervals of the same length (e.g. $(1.3, 1.5)$ and $(4.7, 4.9)$) X can belong to them with the same probability.

Let's find the CDF and PDF of X.

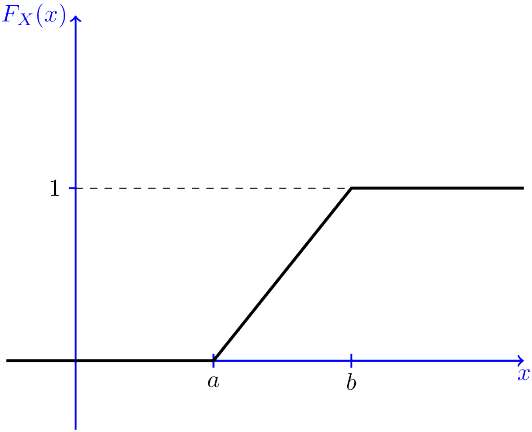
For $x < a$, $F_X(x) = \mathbb{P}(X \leq x) = 0$,

For $x \geq b$, $F_X(x) = \mathbb{P}(X \leq x) = 1$,

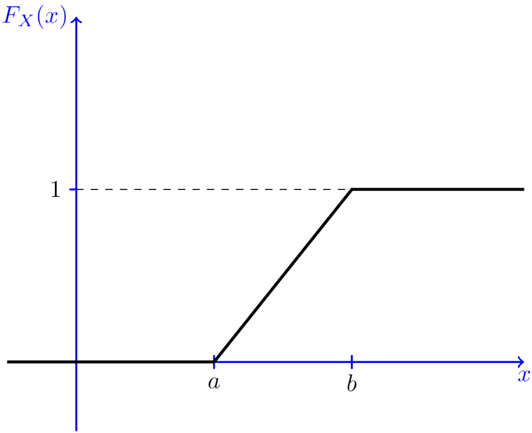
For $a \leq x \leq b$, we have:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in [a, x]) = \frac{x - a}{b - a}$$

PDF



PDF



Thus,

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

PDF

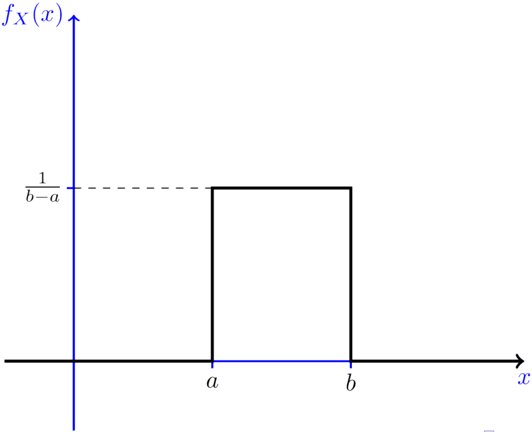
To find $f_X(x)$, we take the derivative of $F_X(x)$:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{if } x < a \text{ or } x > b \end{cases}$$

PDF

To find $f_X(x)$, we take the derivative of $F_X(x)$:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{if } x < a \text{ or } x > b \end{cases}$$



Independence

Recall that two events A and B are called independent if they do not affect each other, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Independence

Recall that two events A and B are called independent if they do not affect each other, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Similarly, we will say that two random variables X and Y are independent if their values do not affect each other (i.e. if you know the value of X, it gives you no information about the value of Y).

Independence

Recall that two events A and B are called independent if they do not affect each other, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Similarly, we will say that two random variables X and Y are independent if their values do not affect each other (i.e. if you know the value of X, it gives you no information about the value of Y).

Definition

X and Y are called **independent** if

$$\mathbb{P}(X \leq a \text{ and } Y \leq b) = \mathbb{P}(X \leq a) \cdot \mathbb{P}(Y \leq b)$$

for any $a, b \in \mathbb{R}$.

So the probability of both X and Y simultaneously being less than some numbers is just their *separate* probabilities multiplied together.

Expected Value, Variance, Distributions

Hayk Aprikyan, Hayk Tarkhanyan

April 18, 2025

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Would you play this game?

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Would you play this game? What if instead of \$36, you won \$150 if it fell on 8?

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet $1.

- If it falls on 8, you win $36,
- Otherwise, you lose $1.

Would you play this game? What if instead of $36, you won $150 if it fell on 8? How about $37.01?

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$$$

Expected Value

Since the chance of winning is only $\dfrac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Similarly, if you were again waiting for a bus and someone asked you "When will it come?", what would you say would be the average number of minutes the bus is expected to arrive.

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Similarly, if you were again waiting for a bus and someone asked you "When will it come?", what would you say would be the average number of minutes the bus is expected to arrive.

These examples motivate the notion of the **mean** or **expected value** of a random variable.

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Definition

If X is a continuous random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Definition

If X is a continuous random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

In words, the expected value is the weighted average of all its possible values, each of the values being weighted by its probability.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$. In general,

Definition

If X is a random variable and $g(x)$ is a continuous function, then $g(X)$ is a random variable too.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$. In general,

Definition

If X is a random variable and $g(x)$ is a continuous function, then $g(X)$ is a random variable too.

Theorem

If X is a continuous random variable, then for any continuous function g,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx$$

If X is a discrete random variable, then for any continuous function g,

$$\mathbb{E}(g(X)) = \sum_{x_i} g(x_i) \cdot \mathbb{P}(X = x_i)$$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$

Expected Value

Properties

1. If a is a constant,

$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$

2. If b is a constant,

$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Theorem

If X and Y are independent random variables, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

The converse is not necessarily true.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

In both games, your overall winnings is on average \$0. However in the first game the amounts differ from \$0 by a small amount, while in the second game, they differ by such a large amount you might not take a risk of playing at all.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

In both games, your overall winnings is on average \$0. However in the first game the amounts differ from \$0 by a small amount, while in the second game, they differ by such a large amount you might not take a risk of playing at all.

In this case, we say that the winnings of the second game have a **higher variance** than those of the first one.

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Definition

$\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Definition

$\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

The standard deviation shows how much, in average, do the values of the random variable deviate from their average ($\mathbb{E}(X)$).

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) =$$

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

To calculate $\text{Var}(X)$, we need $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

To calculate $\mathrm{Var}(X)$, we need $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \approx 2.92$$

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$, instead:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

Variance

Properties

1. $\text{Var}(X) \geq 0$,
2. If X is constant, $\text{Var}(X) = 0$,
3. $\text{Var}(aX) = a^2 \cdot \text{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$, instead:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Variance

Properties

1. $\text{Var}(X) \geq 0$,
2. If X is constant, $\text{Var}(X) = 0$,
3. $\text{Var}(aX) = a^2 \cdot \text{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$, instead:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Variance

Properties

1. $\text{Var}(X) \geq 0$,
2. If X is constant, $\text{Var}(X) = 0$,
3. $\text{Var}(aX) = a^2 \cdot \text{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$, instead:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

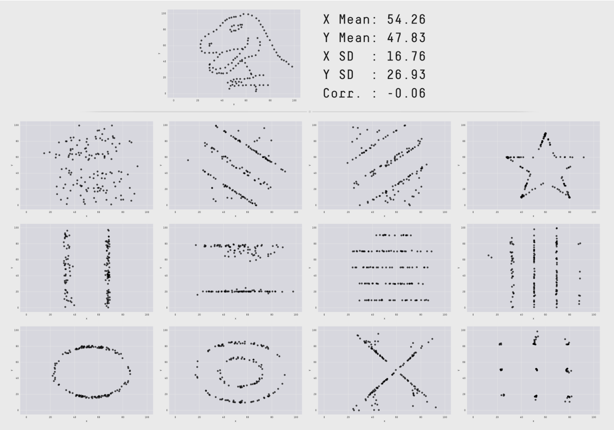
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Why do you think the 4th point makes sense?

Variance

Warning

Expected value and variance are very useful to describe random variables, **but they are not everything!** They do not replace CDF/PDF/PMF!



[source]

Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.

Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.

Due to some reasons, the stock of Tesla starts to decrease. Naturally, you are interested in how could that affect your Yeraz stocks, i.e.

Question

How would a change of X affect Y?

Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.

Due to some reasons, the stock of Tesla starts to decrease. Naturally, you are interested in how could that affect your Yeraz stocks, i.e.

Question

How would a change of X affect Y?

More specifically, if X goes up by 1 unit, how much would Y change?

Covariance

Definition

The **covariance** between random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Covariance

Definition

The **covariance** between random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Note that the formula for covariance can be simplified to

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Covariance

Definition

The **covariance** between random variables X and Y is defined by

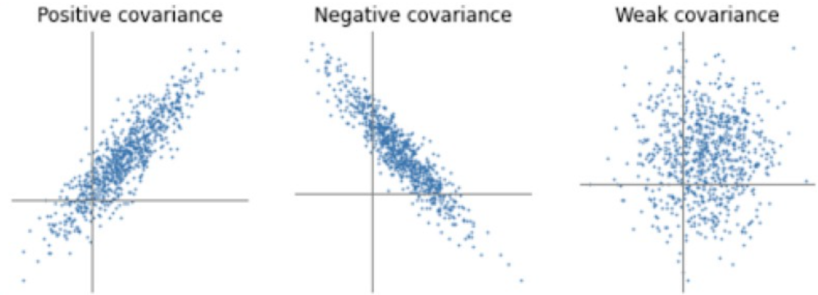
$$\text{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Note that the formula for covariance can be simplified to

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Covariance shows how much the linear growth of one RV is related to the linear growth of the other RV. It is very similar to the concept of dot product of the two vectors.

Covariance


Positive covariance
Negative covariance
Weak covariance

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
3. $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X, Y)$,

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
3. $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X, Y)$,
4. $\mathrm{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}(X_1 + X_2, Y_1 + Y_2) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\,(X_1 + X_2, Y_1 + Y_2) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}\,(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\text{Cov}(X, Y) = 0$ (converse isn't true).

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}\left(X_1 + X_2\right) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\text{Cov}(X, Y) = 0$ (converse isn't true).

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
3. $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X, Y)$,
4. $\mathrm{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)$,
5. $\mathrm{Var}\left(X_1 + X_2\right) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2 \cdot \mathrm{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$ (converse isn't true).

$\mathrm{Cov}(X, Y)$ can be *any* number (positive/negative, large/small, zero, etc). What if we want a normalized, universal method to measure the relatedness level of two random variables?

Correlation

Definition

For two non-constant ($\mathrm{Var}(X), \mathrm{Var}(Y) \neq 0$) random variables X and Y, the **correlation** (or **Pearson correlation coefficient**) between them is defined as:

$$\rho(X, Y) = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

Correlation

Definition

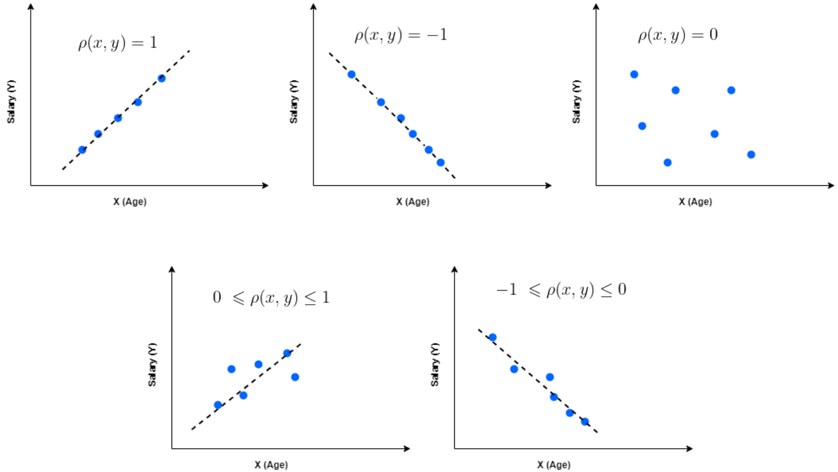
For two non-constant ($\mathrm{Var}(X), \mathrm{Var}(Y) \neq 0$) random variables X and Y, the **correlation** (or **Pearson correlation coefficient**) between them is defined as:

$$\rho(X, Y) = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

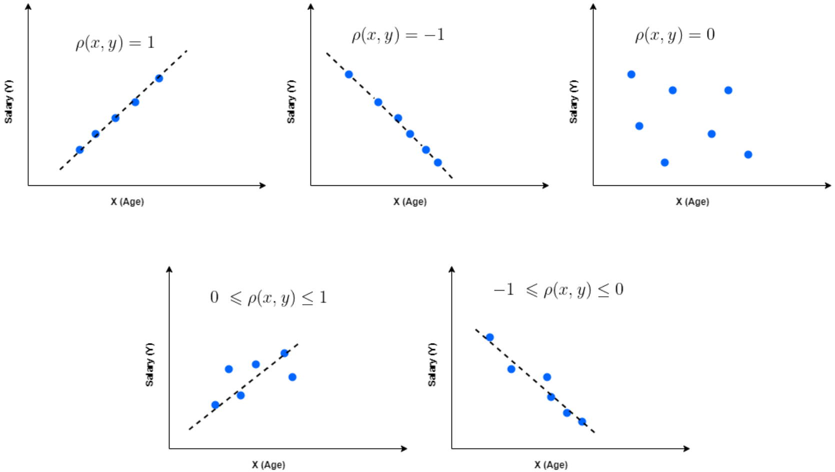
Properties

1. $\rho(X, Y) = \rho(Y, X)$,
2. $-1 \leq \rho(X, Y) \leq 1$,
3. $Y = aX + b$ for some constants a, b if and only if $\rho(X, Y) = \pm 1$.

Correlation

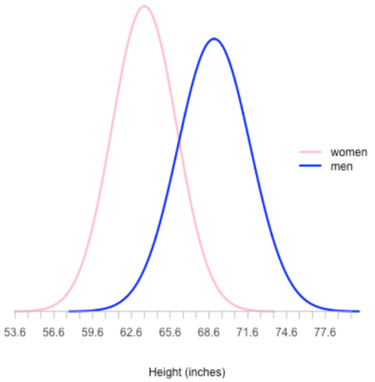


Correlation

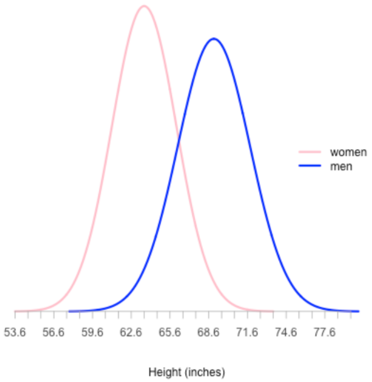


- Play with this correlation visualization!

Very often in practice, many random variables share similar properties. In particular, the probabilities of their values seem to follow a common pattern, i.e. their CDFs (or PMFs/PDFs) are similar to each other:



Very often in practice, many random variables share similar properties. In particular, the probabilities of their values seem to follow a common pattern, i.e. their CDFs (or PMFs/PDFs) are similar to each other:



The way the values of an RV are distributed is called a *distribution*.

Bernoulli Distribution

Consider these situations:

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Definition

A random variable X is said to be a **Bernoulli** random variable with parameter p, denoted by $X \sim Bernoulli(p)$, if it only takes two values and its PMF is given by:

$$\mathbb{P}(X = x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Definition

A random variable X is said to be a **Bernoulli** random variable with parameter p, denoted by $X \sim Bernoulli(p)$, if it only takes two values and its PMF is given by:

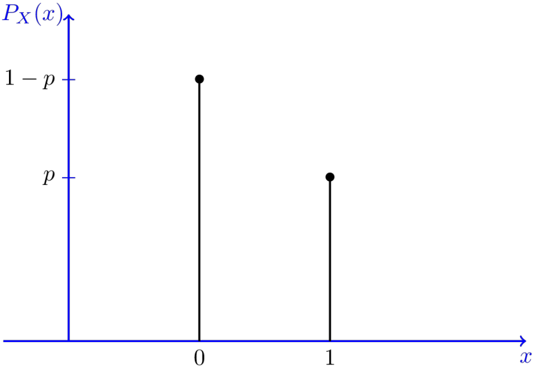
$$\mathbb{P}(X = x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

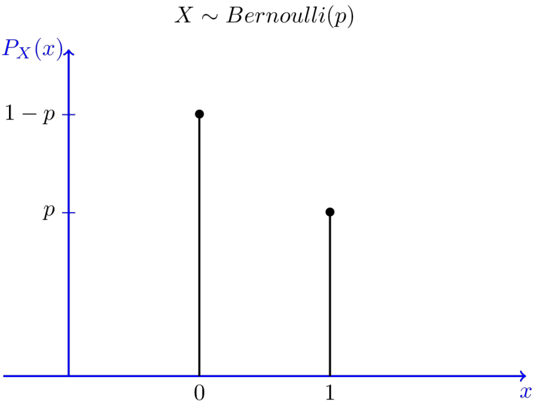
A series of n independent experiments all following Bernoulli distribution $Bernoulli(p)$, is called **Bernoulli trials**.

Bernoulli Distribution

$$X \sim Bernoulli(p)$$



Bernoulli Distribution



If $X \sim Bernoulli(p)$,

$$\mathbb{E}(X) = p, \qquad \mathrm{Var}(X) = p(1-p)$$

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.
For any natural number $k \geq 1$:

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p$$

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.

For any natural number $k \geq 1$:

$$\mathbb{P}(X = k) = (1-p)^{k-1}p$$

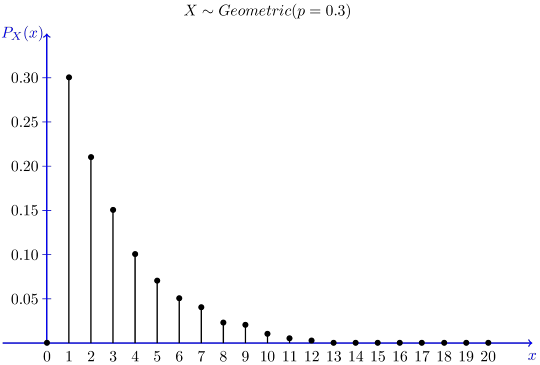
Definition

A random variable X is said to be a **geometric** random variable with parameter p, denoted by $X \sim Geo(p)$, if its PMF is given by:

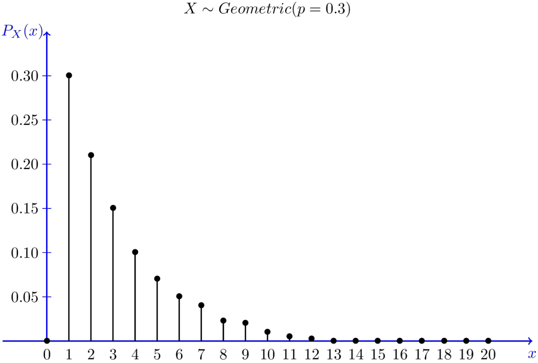
$$\mathbb{P}(X = k) = \begin{cases} (1-p)^{k-1}p & \text{for } k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Geometric Distribution



Geometric Distribution



If $X \sim Geo(p)$,

$$\mathbb{E}(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.
For any whole number $k \geq 0$:

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$$

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.
For any whole number $k \geq 0$:

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$$

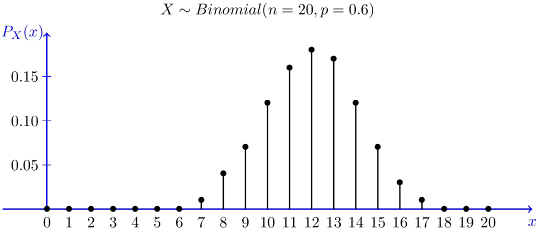
Definition

A random variable X is said to be a **binomial** random variable with parameters n and p, denoted by $X \sim B(n, p)$, if its PMF is given by:

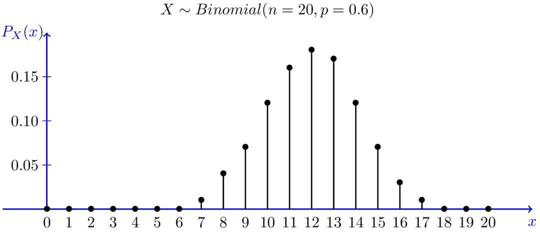
$$\mathbb{P}(X = k) = \begin{cases} C_n^k p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Binomial Distribution



Binomial Distribution



If $X \sim B(n, p)$,

$$\mathbb{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p)$$

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.
Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store.

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.

Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store. In practice, the number of customers tends to follow the following pattern:

Poisson Distribution

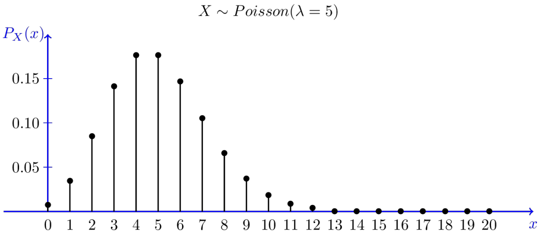
The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.
Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store. In practice, the number of customers tends to follow the following pattern:

Definition

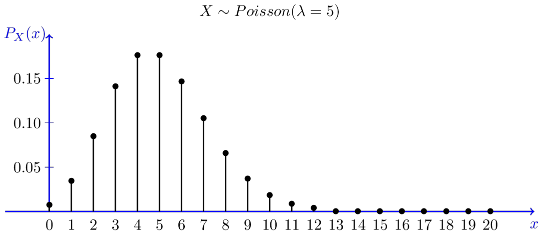
A random variable X is said to be a **Poisson** random variable with parameter λ, denoted by $X \sim Poisson(\lambda)$, if its PMF is given by:

$$\mathbb{P}(X = k) = \begin{cases} \dfrac{\lambda^k}{k!} e^{-\lambda} & \text{for } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution



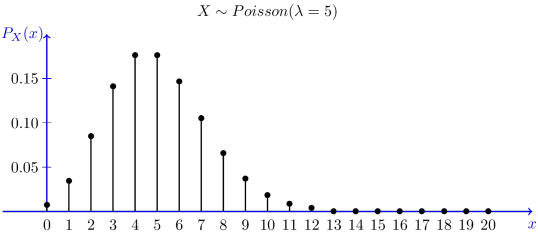
Poisson Distribution



If $X \sim Poisson(\lambda)$,

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda$$

Poisson Distribution



If $X \sim Poisson(\lambda)$,

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda$$

Remark

If $X \sim B(n, p)$, then

$$\mathbb{P}(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \text{where } \lambda = np$$

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

Definition

A random variable X is said to be a **uniform** random variable over the interval $[a, b]$, denoted by $X \sim U(a, b)$, if its PDF is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$$

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.
If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

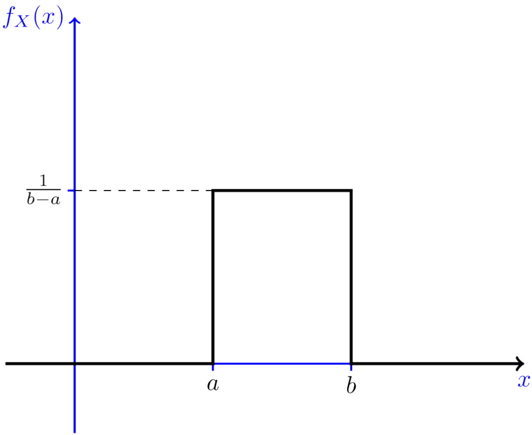
Definition

A random variable X is said to be a **uniform** random variable over the interval $[a, b]$, denoted by $X \sim U(a, b)$, if its PDF is given by:

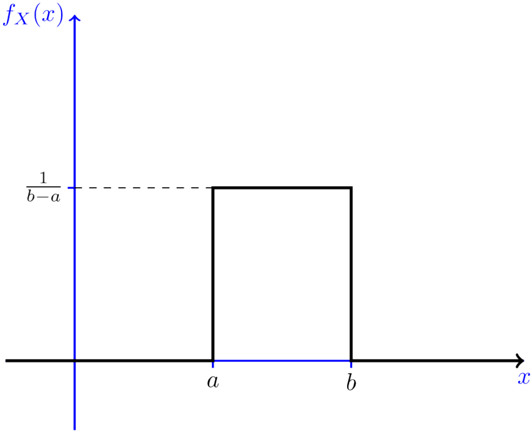
$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$$

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

Uniform Distribution



Uniform Distribution



If $X \sim U(a, b)$,

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Definition

A random variable X is said to be a **exponential** random variable with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its PDF is given by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Definition

A random variable X is said to be a **exponential** random variable with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its PDF is given by:

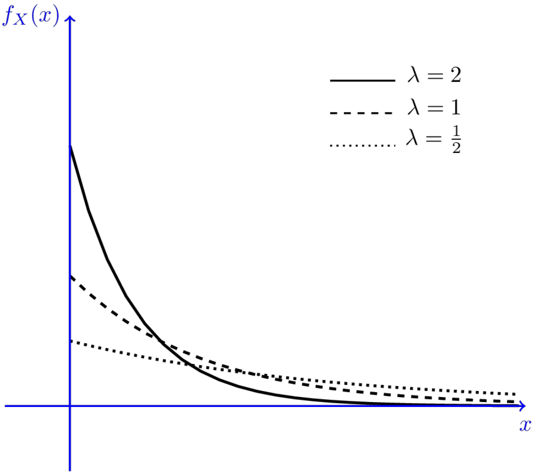
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Remark

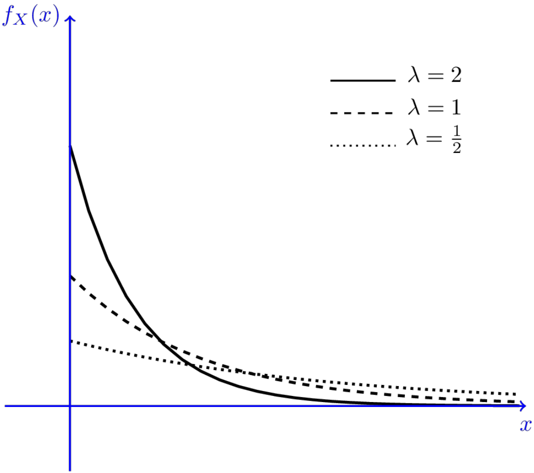
If $X \sim Exp(\lambda)$, then X is a **memoryless** random variable, that is

$$\mathbb{P}(X > x + a \mid X > a) = \mathbb{P}(X > x), \quad \text{for } a, x \geq 0.$$

Exponential Distribution



Exponential Distribution



If $X \sim Exp(\lambda)$,

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Normal Distribution

The *normal* distribution is by far the most important probability distribution.

Normal Distribution

The *normal* distribution is by far the most important probability distribution.

One of the main reasons for that is that if you add a large number of random variables, the distribution of the sum will be approximately normal under certain conditions. In real life, many RVs can be expressed as the sum of a large number of random variables, and their distribution is normal.

Normal Distribution

The *normal* distribution is by far the most important probability distribution.
One of the main reasons for that is that if you add a large number of random variables, the distribution of the sum will be approximately normal under certain conditions. In real life, many RVs can be expressed as the sum of a large number of random variables, and their distribution is normal.

Definition

A random variable X is said to be a **normal** random variable with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$, if its PDF is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}$$

Normal Distribution