Assignment

Foundations of Mathematics for Deep Learning

Jiksa K
M.Tech CSE(DS&AI)[PT]
CUSAT

Contents

Set Theory

Definition of Set

A set is a well-defined collection of distinct objects. These objects are called elements or members of the set. Sets are denoted by capital letters (e.g., A, B, C), and their elements are written inside curly braces, separated by commas.

Examples:

- $\{1, 2, 3\}$: A set of integers.
- $\{\text{cat, dog, rabbit}\}$: A set of animals.
- $\{a, e, i, o, u\}$: A set of vowels in the English alphabet.

A set is considered *well-defined* if there is a clear rule to determine whether an object belongs to the set. For example, the set of prime numbers less than 10 is $\{2, 3, 5, 7\}$.

—

Representation of Sets

There are two standard methods to represent a set:

1. **Roster Form:** In this method, all the elements of the set are listed explicitly, separated by commas, and enclosed within curly braces.

$$A = \{1, 2, 3, 4\}.$$

   Example: The set of even numbers less than 10 can be written as:

$$E = \{2, 4, 6, 8\}.$$

2. **Set-builder Form:** In this method, a set is described by a property that its elements satisfy. i.e, in Set-builder form, elements are shown or represented in statements expressing relations among elements.

$$A = \{x \mid x > 0 \text{ and } x \in \mathbb{Z}\},$$

   where $x \in \mathbb{Z}$ means x belongs to the set of integers.

   Example: The set of all natural numbers greater than 5 is:

$$N = \{x \mid x > 5, x \in \mathbb{N}\}.$$

—

Types of Sets

Sets can be classified into various types based on their properties:

- **Singleton Set:** A set with exactly one element.

$$A = \{5\}.$$

- **Empty Set:** A set with no elements, denoted as $\emptyset$ or $\{\}$.

$$C = \emptyset.$$

- **Finite Set:** A set with a limited number of elements.

$$A = \{1, 2, 3, 4\}.$$

- **Infinite Set:** A set with an unlimited number of elements.

$$B = \{x \mid x \in \mathbb{N}\} = \{1, 2, 3, \dots\}.$$

- **Equal Sets:** Two sets are equal if they contain exactly the same elements.

  Example:

$$A = \{1, 2, 3\}, \quad B = \{3, 2, 1\}, \quad A = B.$$

- **Equivalent Sets:** Two sets are equivalent if they have the same number of elements.

  Example:

$$A = \{1, 2, 3\}, \quad B = \{x, y, z\}, \quad A \sim B.$$

- **Subset:** A set A is a subset of B if every element of A is also an element of B, denoted as $A \subseteq B$.

  Example: If $B = \{1, 2, 3, 4\}$, then $A = \{1, 2\}$ is a subset of B.

- **Power Set:** The set of all subsets of a set A, denoted as $P(A)$.

  Example: If $A = \{1, 2\}$, then:

$$P(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}.$$

- **Universal Set:** A set that contains all possible elements in a particular context.

$$U = \{\text{all natural numbers}\}.$$

- **Disjoint Sets:** Two sets are disjoint if they have no elements in common.

  Example:

$$A = \{1, 2, 3\}, \quad B = \{4, 5, 6\}, \quad A \cap B = \emptyset.$$

—

Set Operations

Operations on sets help combine or compare sets to derive new sets.

- **Union:** The union of sets A and B, denoted as $A \cup B$, is the set of all elements that belong to either A or B.

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Example:

$$A = \{1, 2, 3\}, \quad B = \{3, 4, 5\}, \quad A \cup B = \{1, 2, 3, 4, 5\}.$$

- **Intersection:** The intersection of sets A and B, denoted as $A \cap B$, is the set of all elements that belong to both A and B.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

Example:

$$A = \{1, 2, 3\}, \quad B = \{3, 4, 5\}, \quad A \cap B = \{3\}.$$

- **Difference:** The difference of sets A and B, denoted as $A - B$, is the set of all elements that belong to A but not to B.

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}.$$

Example:

$$A = \{1, 2, 3\}, \quad B = \{3, 4, 5\}, \quad A - B = \{1, 2\}.$$

- **Complement:** The complement of a set A, denoted as A^c, is the set of all elements in the universal set that are not in A.

—

Axioms and Theorems of Set Theory

The axioms and theorems of set theory provide the foundation for mathematical reasoning involving sets. Below are the key axioms and some fundamental theorems.

—

Axioms of Set Theory

The axioms define the basic rules for constructing and manipulating sets:

- **Axiom of Extensionality:** Two sets are equal if they contain exactly the same elements.

$$A = \{1, 2, 3\}, \quad B = \{3, 2, 1\}, \quad \text{so } A = B.$$

- **Axiom of Empty Set:** There exists a set with no elements, called the empty set, denoted as:

$$\emptyset = \{\}.$$

- **Axiom of Pairing:** For any two sets A and B, there exists a set containing exactly A and B.

$$\text{For } A = \{1\}, B = \{2\}, \quad C = \{\{1\}, \{2\}\}.$$

- **Axiom of Union:** For any collection of sets, there exists a set that contains all elements of these sets (the union).

$$A = \{1, 2\}, \quad B = \{3, 4\}, \quad A \cup B = \{1, 2, 3, 4\}.$$

- **Axiom of Power Set:** For any set A, there exists a set (called the power set) that contains all subsets of A.

$$A = \{1, 2\}, \quad P(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}.$$

- **Axiom of Subsets (Separation):** For any set A, one can form a subset B consisting of all elements of A that satisfy a given condition.

$$B = \{x \in A \mid x > 2\}, \quad \text{if } A = \{1, 2, 3, 4\}, \quad B = \{3, 4\}.$$

- **Axiom of Infinity:** There exists a set containing the natural numbers $\mathbb{N}$, which is an infinite set.

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

- **Axiom of Replacement:** The image of a set under a function is also a set.

$$A = \{1, 2, 3\}, \quad f(x) = x^2, \quad B = \{1, 4, 9\}.$$

- **Axiom of Choice:** For any collection of non-empty, disjoint sets, there exists a set containing exactly one element from each set.

$$A = \{\{a, b\}, \{1, 2\}, \{x, y\}\}, \quad C = \{b, 2, y\}.$$

—

Theorems of Set Theory

The theorems of set theory derive from the axioms and are used to establish further properties and relationships between sets:

- **Theorem of Union:** The union of two sets A and B contains all elements that are in A, B, or both.

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

  Example:

$$A = \{1, 2\}, \quad B = \{2, 3\}, \quad A \cup B = \{1, 2, 3\}.$$

- **Theorem of Intersection:** The intersection of two sets A and B contains all elements that are in both A and B.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

  Example:

$$A = \{1, 2, 3\}, \quad B = \{2, 3, 4\}, \quad A \cap B = \{2, 3\}.$$

- **Theorem of Difference:** The difference of two sets A and B contains all elements that are in A but not in B.

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}.$$

Example:
$$A = \{1, 2, 3\}, \quad B = \{2, 3, 4\}, \quad A - B = \{1\}.$$

- **Theorem of Complement:** The complement of a set A contains all elements in the universal set U that are not in A.

$$A^c = \{x \in U \mid x \notin A\}.$$

Example:
$$U = \{1, 2, 3, 4, 5\}, \quad A = \{2, 4\}, \quad A^c = \{1, 3, 5\}.$$

- **Theorem of De Morgan's Laws:** The following hold for any two sets A and B:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$

Example:
$$A = \{1, 2\}, \quad B = \{2, 3\}, \quad (A \cup B)^c = A^c \cap B^c.$$

- **Theorem of Power Set Size:** If a set A has n elements, then its power set $P(A)$ has 2^n elements. **Example:**

$$A = \{1, 2\}, \quad P(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}, \quad \text{so } |P(A)| = 2^2 = 4.$$

—

Cartesian Product of Sets

The Cartesian product of two sets A and B is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$. It is denoted as:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

Key Points:
1. **Ordered Pairs:** - An ordered pair (a, b) consists of two elements where the order matters. - Two ordered pairs are equal if and only if their first elements and second elements are equal:

$$(a, b) = (c, d) \iff a = c \text{ and } b = d.$$

2. **Size of Cartesian Products:** - If set A has m elements and set B has n elements, then the Cartesian product $A \times B$ contains $m \times n$ elements.

$$|A \times B| = |A| \cdot |B|.$$

3. **Higher-Dimensional Cartesian Products:** - Cartesian products can extend to more than two sets to form ordered triples, quadruples, and so on. - For example:

$$A \times B \times C = \{(a, b, c) \mid a \in A, b \in B, c \in C\}.$$

Example: If $A = \{1, 2\}$ and $B = \{x, y\}$, then:

$$A \times B = \{(1, x), (1, y), (2, x), (2, y)\}.$$

If $C = \{p, q\}$, then:

$$A \times B \times C = \{(1, x, p), (1, x, q), (1, y, p), (1, y, q), (2, x, p), (2, x, q), (2, y, p), (2, y, q)\}.$$

Application: - Cartesian products are used to define coordinates in geometry. For example, the Cartesian plane $\mathbb{R} \times \mathbb{R}$ represents all points (x, y), where x and y are real numbers. - In higher dimensions, Cartesian products extend to describe points in three-dimensional space ($\mathbb{R}^3$) or higher-dimensional spaces.

Properties of Cartesian Products: 1. $A \times B \neq B \times A$, unless $A = B$. The order of elements matters in ordered pairs. 2. $A \times \emptyset = \emptyset$, because there are no pairs when one of the sets is empty. 3. Cartesian products distribute over unions:

$$A \times (B \cup C) = (A \times B) \cup (A \times C).$$

4. Cartesian products distribute over intersections:

$$A \times (B \cap C) = (A \times B) \cap (A \times C).$$

—

Relations

A relation R from set A to set B is a subset of $A \times B$. It defines how elements of A are related to elements of B.

Types of Relations

Relations on a set A can be categorized into various types based on their properties. Below are the key types of relations with examples:

- **Empty Relation:** A relation R on A is empty if no elements of A are related to any other elements in A.

  $$R = \emptyset \quad \text{or} \quad R \subseteq A \times A.$$

  Example: If $A = \{\text{Alice, Bob}\}$ and the relation is "is a sibling," and no one in A has siblings, $R = \emptyset$.

- **Universal Relation:** A relation R on A is universal if it contains all possible pairs from $A \times A$.

  $$R = A \times A.$$

  Example: If $A = \{1, 2\}$, the universal relation is:

  $$R = \{(1, 1), (1, 2), (2, 1), (2, 2)\}.$$

- **Identity Relation:** A relation R on A is identity if every element of A is related only to itself.
$$R = \{(a, a) \mid a \in A\}.$$
Example: If $A = \{1, 2\}$, the identity relation is:
$$R = \{(1, 1), (2, 2)\}.$$

- **Reflexive Relation:** A relation R on A is reflexive if every element is related to itself.
$$(a, a) \in R \quad \forall a \in A.$$
Example: For $A = \{1, 2\}$, $R = \{(1, 1), (2, 2), (1, 2)\}$ is reflexive.

- **Irreflexive Relation:** A relation R on A is irreflexive if no element is related to itself.
$$(a, a) \notin R \quad \forall a \in A.$$
Example: For $A = \{1, 2\}$, $R = \{(1, 2), (2, 1)\}$ is irreflexive.

- **Symmetric Relation:** A relation R on A is symmetric if:
$$(a, b) \in R \implies (b, a) \in R.$$
Example: If $R = \{(1, 2), (2, 1)\}$ on $A = \{1, 2\}$, R is symmetric.

- **Anti-Symmetric Relation:** A relation R on A is anti-symmetric if:
$$(a, b) \in R \text{ and } (b, a) \in R \implies a = b.$$
Example: For $A = \{1, 2\}$, $R = \{(1, 1), (2, 2), (1, 2)\}$ is anti-symmetric.

- **Transitive Relation:** A relation R on A is transitive if:
$$(a, b) \in R \text{ and } (b, c) \in R \implies (a, c) \in R.$$
Example: If $R = \{(1, 2), (2, 3), (1, 3)\}$ on $A = \{1, 2, 3\}$, R is transitive.

- **Equivalence Relation:** A relation is an equivalence relation if it is reflexive, symmetric, and transitive. **Example:** The relation "is equal to" $(=)$ on numbers is an equivalence relation.

- **Inverse Relation:** The inverse of a relation R contains all pairs with the order of elements reversed.
$$R^{-1} = \{(b, a) \mid (a, b) \in R\}.$$
Example: If $R = \{(1, 2), (3, 4)\}$, then $R^{-1} = \{(2, 1), (4, 3)\}$.

Composition of Relations

The composition of two relations R and S connects elements of one set to another through an intermediate set. It is defined as:

$$S \circ R = \{(a, c) \mid \exists b \in B, (a, b) \in R \text{ and } (b, c) \in S\}.$$

Explanation: 1. Start with a pair (a, b) from R, where $a \in A$ and $b \in B$. 2. Find a pair (b, c) in S, where $b \in B$ and $c \in C$. 3. Combine these pairs to form (a, c), which is in $S \circ R$.

Properties of Composition of Relations:

1. **Associativity:**

$$T \circ (S \circ R) = (T \circ S) \circ R$$

The composition of relations is associative, meaning the order of grouping does not affect the result.

2. **Identity Relation:** The identity relation acts as a neutral element in composition. For any relation R:

$$R \circ I_A = R \quad \text{and} \quad I_B \circ R = R,$$

where $I_A = \{(a, a) \mid a \in A\}$.

3. **Distributivity:** Composition distributes over union:

$$R \circ (S \cup T) = (R \circ S) \cup (R \circ T),$$

and

$$(R \cup S) \circ T = (R \circ T) \cup (S \circ T).$$

4. **Inverse Relation:** If R is a relation, its inverse R^{-1} is defined as:

$$R^{-1} = \{(b, a) \mid (a, b) \in R\}.$$

The composition with its inverse has properties similar to functions:

$$R^{-1} \circ R = I_A \quad \text{and} \quad R \circ R^{-1} = I_B.$$

—

Example: Let $A = \{1, 2\}$, $B = \{x, y\}$, and $C = \{A, B\}$. Define the relations:

$$R = \{(1, x), (2, y)\}, \quad S = \{(x, A), (y, B)\}.$$

To find the composition $S \circ R$: 1. From R, take the pair $(1, x)$. Since $(x, A) \in S$, form the pair $(1, A)$. 2. From R, take the pair $(2, y)$. Since $(y, B) \in S$, form the pair $(2, B)$.

Thus:

$$S \circ R = \{(1, A), (2, B)\}.$$

—

Conclusion: The composition of relations provides a way to connect elements across sets via an intermediate set. It is a fundamental concept used in various fields like graph theory, database queries, and function composition.

Probability

Random Experiments

A random experiment is a process or activity that leads to well-defined outcomes but cannot be predicted with certainty. These experiments have the following characteristics:

- The set of all possible outcomes is well-defined.
- Each experiment can be repeated under the same conditions.

Example:

- Tossing a coin: Possible outcomes are {Head, Tail}.
- Rolling a die: Possible outcomes are {1, 2, 3, 4, 5, 6}.

Sample Space and Outcomes

In probability theory, the sample space (S) is the set of all possible outcomes of a random experiment. Each element in the sample space is called an outcome or sample point. The sample space serves as the universal set from which events are defined.

Characteristics of Sample Space:

- **Mutually Exclusive Outcomes:** Each outcome in the sample space must be distinct. No two outcomes can occur simultaneously.
- **Collectively Exhaustive:** The sample space must include all possible outcomes of the experiment. Every potential result should be accounted for.
- **Granularity:** The sample space must be precise and include only the relevant outcomes for the given experiment.

Examples of Sample Spaces:

- **Single Coin Toss:**
$$S = \{\text{Head, Tail}\}$$
Outcomes: H (Head), T (Tail).
- **Two Coin Tosses:**
$$S = \{(H,H),(H,T),(T,H),(T,T)\}$$
Total outcomes: 4 (HH, HT, TH, TT).

- **Rolling a Single Die:**
$$S = \{1, 2, 3, 4, 5, 6\}$$
Outcomes represent the face value of the die.

- **Rolling Two Dice:**
$$S = \{(1, 1), (1, 2), \dots, (6, 6)\}$$
Total outcomes: $6 \times 6 = 36$.

- **Tossing Three Coins:**
$$S = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}$$
Total outcomes: $2^3 = 8$.

- **Rolling Two Dice and Summing the Results:**
$$S = \{2, 3, 4, \dots, 12\}$$
Each outcome corresponds to the sum of the numbers rolled on both dice.

Outcomes: An outcome is a specific result of an experiment within the sample space. Each element in the sample space represents a potential outcome. For example:

- Tossing a single coin: The outcomes are H (Head) and T (Tail).
- Rolling a single die: Each face value (1 through 6) is an outcome.

Conclusion: The concept of sample space is foundational in probability theory as it provides a complete representation of all possible outcomes of a random experiment. It ensures that probabilities can be assigned correctly to events and analyzed effectively.

Events and Types of Events

An event is a subset of the sample space, representing specific outcomes or sets of outcomes. Events can be classified into the following types:

- **Simple Event:** An event consisting of a single outcome from the sample space.
  - **Example:** Rolling a die and getting a 4, represented as $\{4\}$.
- **Compound Event:** An event that includes two or more simple events.
  - **Example:** Rolling a die and getting an even number, represented as $\{2, 4, 6\}$.
- **Sure Event:** An event that is guaranteed to occur and coincides with the entire sample space.
  - **Example:** When rolling a die, the event "a number between 1 and 6 appears" is a sure event.
- **Impossible Event:** An event that cannot occur and corresponds to the empty set.

  - **Example:** Rolling a die and getting a 7 is an impossible event, as the sample space includes only numbers 1 to 6.

- **Mutually Exclusive Events:** Events that cannot occur at the same time; their intersection is empty.
  - **Example:** When rolling a die, the events "rolling an odd number" and "rolling an even number" are mutually exclusive.

- **Independent Events:** Events where the occurrence of one does not affect the occurrence of the other.
  - **Example:** Flipping a coin and rolling a die are independent events.

- **Dependent Events:** Events where the occurrence of one affects the probability of the other occurring.
  - **Example:** Drawing cards from a deck without replacement; the outcome of the first draw affects the second.

- **Complementary Events:** The complement of an event A includes all outcomes in the sample space that are not in A. Denoted as A' or $\overline{A}$.
  - **Example:** If A is "rolling an even number," then A' is "rolling an odd number."

- **Joint Events:** Events involving two or more events occurring together.
  - **Example:** Tossing two coins and getting "heads on both coins" represents a joint event.

- **Exhaustive Events:** A set of events is exhaustive if at least one of them must occur; they cover all possible outcomes in the sample space.
  - **Example:** When rolling a die, the events $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ are exhaustive.

Probability of an Event

The probability of an event A is a measure of the likelihood of the event occurring. It is defined as the ratio of the number of favorable outcomes to the total number of equally likely outcomes in the sample space S. Mathematically, it is expressed as:

$$P(A) = \frac{n(A)}{n(S)},$$

where:

- $n(A)$**:** Number of favorable outcomes for event A.
- $n(S)$**:** Total number of outcomes in the sample space.

—

Key Properties of Probability:

- $0 \leq P(A) \leq 1$: The probability of any event lies between 0 and 1, inclusive.
  - $P(A) = 0$**:** The event is impossible.
  - $P(A) = 1$**:** The event is certain to occur.
  - $P(A) = 0.5$**:** The event is equally likely to occur or not occur.
- $P(S) = 1$**:** The probability of the sample space (a sure event) is always 1.
- $P(\emptyset) = 0$**:** The probability of the empty set (an impossible event) is always 0.

—

Complementary Events: The probability of the complement of an event A (denoted as A' or $\overline{A}$) is given by:

$$P(A') = 1 - P(A).$$

—

Types of Events:

- **Independent Events:** The occurrence of one event does not affect the occurrence of another.
- **Dependent Events:** The occurrence of one event affects the probability of the other.
- **Mutually Exclusive Events:** Two events cannot occur simultaneously; their intersection is empty.

—

Experimental vs. Theoretical Probability:

- **Theoretical Probability:** Based on reasoning and known outcomes.
- **Experimental Probability:** Based on actual experiments and observed results.

—

Examples of Probability:

- **Rolling a Die:**
  - The probability of rolling a 4 is:

$$P(4) = \frac{1}{6}.$$

  - The probability of rolling an even number ($\{2, 4, 6\}$) is:

$$P(\text{even}) = \frac{3}{6} = \frac{1}{2}.$$

- **Drawing a Ball:** A bag contains 6 red balls and 4 blue balls. The probability of drawing a blue ball is:

$$P(\text{blue}) = \frac{4}{10} = 0.4.$$

- **Flipping Two Coins:** The probability of getting at least one head is:

$$P(\text{at least one Head}) = \frac{3}{4}.$$

—

Conclusion: Probability provides a numerical measure to assess the likelihood of events occurring. It is a powerful tool for analyzing risks, making decisions, and understanding patterns in various fields such as statistics, science, and everyday life.

Principles of Counting

The principles of counting, along with permutations and combinations, are fundamental concepts in combinatorics. They help determine the total number of possible outcomes in an experiment.

—

Fundamental Counting Principle

The Fundamental Counting Principle states that if one event can occur in m ways and a second independent event can occur in n ways, then the total number of ways both events can occur is:

$$\text{Total Outcomes} = m \times n.$$

This principle can be extended to any number of events. For example: - If there are three events that can occur in m, n, and p ways, then the total number of outcomes is:

$$\text{Total Outcomes} = m \times n \times p.$$

Example:

- If a person has 3 shirts and 2 pants, the total number of outfit combinations is:

$$3 \times 2 = 6.$$

—

Addition Principle (Rule of Sum)

The Addition Principle states that if a task can be performed in either one way or another (where the two methods cannot happen simultaneously), then the total number of ways to perform the task is the sum of the individual ways:

$$n(E) = n(A) + n(B).$$

Example:

- If a bakery offers 20 cupcakes and 10 doughnuts, the total number of choices is:

$$20 + 10 = 30.$$

—

Inclusion-Exclusion Principle

The Inclusion-Exclusion Principle is used to calculate the number of elements in the union of multiple sets by including the sizes of each set and excluding the sizes of their intersections. For two sets A and B:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Example:

- If there are 50 students who like math, 30 students who like science, and 10 students who like both subjects, the total number of students who like at least one subject is:
$$|A \cup B| = 50 + 30 - 10 = 70.$$

—

Permutations

Permutations refer to the arrangement of n distinct objects in a specific order. The total number of permutations of n objects is given by:

$$n! = n \times (n-1) \times \cdots \times 1.$$

If r objects are to be selected and arranged from n objects, the number of permutations is:

$$P(n,r) = \frac{n!}{(n-r)!}.$$

Example:

- Arranging 3 books on a shelf: The number of ways is:
$$3! = 6.$$

- Selecting 2 items from 4 and arranging them:
$$P(4,2) = \frac{4!}{(4-2)!} = \frac{24}{2} = 12.$$

- Arranging 3 books out of a shelf of 5 books:
$$P(5,3) = \frac{5!}{(5-3)!} = \frac{120}{2} = 60.$$

—

Combinations

Combinations refer to the selection of r objects from n objects, where the order of selection does not matter. The number of combinations is given by:

$$C(n, r) = \frac{n!}{r!(n-r)!}.$$

Example:

- Choosing 2 fruits from a basket of 5 fruits:

$$C(5, 2) = \frac{5!}{2!(5-2)!} = \frac{120}{2 \times 6} = 10.$$

- Choosing 3 books from a collection of 5 books:

$$C(5, 3) = \frac{5!}{3! \times (5-3)!} = \frac{120}{6 \times 2} = 10.$$

—

Conclusion: Understanding the principles of counting, permutations, and combinations allows for efficient analysis of outcomes in probability and combinatorics. These tools are crucial for solving mathematical and real-world problems involving arrangements and selections.

Linear Algebra

Matrix

A matrix is a rectangular array of numbers, symbols, or expressions, arranged in rows and columns. It is denoted as:

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

where a_{ij} represents the element in the i^{th} row and j^{th} column.

Example:

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

Types of Matrices

Matrices can be classified into various types based on their structure or properties:

- **Row Matrix:** A matrix with only one row.

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}.$$

- **Column Matrix:** A matrix with only one column.

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

- **Square Matrix:** A matrix with the same number of rows and columns ($m = n$).

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

- **Zero Matrix (Null Matrix):** A matrix where all elements are zero.

$$D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

- **Diagonal Matrix:** A square matrix where all off-diagonal elements are zero.

$$E = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}.$$

- **Scalar Matrix:** A diagonal matrix where all diagonal elements are the same.

$$F = \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix}.$$

- **Identity Matrix:** A diagonal matrix where all diagonal elements are 1.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- **Upper Triangular Matrix:** A square matrix where all elements below the main diagonal are zero.

$$G = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{bmatrix}.$$

- **Lower Triangular Matrix:** A square matrix where all elements above the main diagonal are zero.

$$H = \begin{bmatrix} h_{11} & 0 \\ h_{21} & h_{22} \end{bmatrix}.$$

- **Symmetric Matrix:** A square matrix that is equal to its transpose ($A = A^T$).

$$J = \begin{bmatrix} j_{11} & j_{12} \\ j_{12} & j_{22} \end{bmatrix}.$$

- **Skew-Symmetric Matrix:** A square matrix where the transpose equals the negative of the matrix ($A^T = -A$).

$$K = \begin{bmatrix} 0 & k_{12} \\ -k_{12} & 0 \end{bmatrix}.$$

- **Singular Matrix:** A square matrix with a determinant of zero. Such matrices do not have an inverse.

- **Non-Singular Matrix:** A square matrix with a non-zero determinant. These matrices are invertible.

Determinants

The determinant of a square matrix is a scalar value that provides insights into the matrix's properties, such as invertibility and linear transformations. It is denoted as $\det(A)$, $|A|$, or simply $\det A$.

—

Definition and Calculation

- For a 2×2 matrix:
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det(A) = ad - bc.$$
Example:
$$A = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}, \quad \det(A) = (3 \cdot 5) - (4 \cdot 2) = 15 - 8 = 7.$$
- For a 3×3 matrix:
$$B = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$
the determinant is calculated as:
$$\det(B) = a(ei - fh) - b(di - fg) + c(dh - eg).$$
Example:
$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \det(B) = 1(45 - 48) - 2(27 - 42) + 3(32 - 35) = 0.$$

For larger matrices, determinants can be computed using cofactor expansion or row reduction methods.

—

Properties of Determinants

- **Multiplicative Property:** For matrices A and B,
$$\det(AB) = \det(A) \cdot \det(B).$$
- **Invariance under Transpose:**
$$\det(A) = \det(A^T).$$
- **Row Operations:**
  - Multiplying a row by a scalar k multiplies the determinant by k.
  - Swapping two rows changes the sign of the determinant.
  - Adding a multiple of one row to another does not change the determinant.
- **Zero Rows or Columns:** If any row or column is entirely zero, the determinant is zero.
- **Proportional Rows or Columns:** If two rows or columns are proportional, the determinant is zero.
- **Determinant of Triangular Matrices:** For an upper or lower triangular matrix, the determinant is the product of the diagonal entries.

—

Significance of Determinants

- **Invertibility:** A square matrix is invertible if and only if its determinant is non-zero. If $\det(A) = 0$, the matrix is singular and does not have an inverse.
- **Linear Independence:** The columns (or rows) of a matrix are linearly independent if $\det(A) \neq 0$.
- **Volume Interpretation:** The absolute value of the determinant represents the scaling factor for volume under the linear transformation defined by the matrix.
- **Solving Linear Systems:** Determinants are used in Cramer's Rule to solve systems of linear equations.

—

Conclusion

Determinants encapsulate critical properties of matrices, enabling efficient analysis of linear transformations, invertibility, and more. Understanding their calculation and properties is essential in various applications of linear algebra.

Inverse of a Matrix

A matrix A is invertible (or non-singular) if there exists another matrix A^{-1} such that:

$$A \cdot A^{-1} = A^{-1} \cdot A = I,$$

where I is the identity matrix.

—

Inverse of a 2×2 Matrix

For a 2×2 matrix:

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

the inverse is given by:

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

provided that $\det(A) \neq 0$, where:

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

Example:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \quad \det(A) = (2 \cdot 4) - (3 \cdot 1) = 8 - 3 = 5.$$

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.6 \\ -0.2 & 0.4 \end{bmatrix}.$$

—

Inverse of a 3×3 Matrix

For a 3×3 matrix, the steps to find the inverse are as follows:

1. **Check Invertibility:** Compute $\det(A)$. If $\det(A) = 0$, the matrix is not invertible.
2. **Calculate Determinant:** For:

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

the determinant is:

$$\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg).$$

3. **Cofactor Matrix:** Calculate cofactors for each element of the matrix.
4. **Adjugate Matrix:** Transpose the cofactor matrix to get the adjugate ($\text{adj}(A)$).
5. **Find the Inverse:** Use the formula:

$$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A).$$

Example: Consider:

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

1. Calculate the determinant:

$$\det(A) = 1(1 \cdot 1 - 2 \cdot 2) - 2(2 \cdot 1 - (-1) \cdot 2) + (-1)(2 \cdot 2 - (-1) \cdot 1).$$

$$\det(A) = 1(-3) - 2(4) - 1(5) = -3 - 8 - 5 = -16.$$

2. Calculate the cofactor matrix and transpose it to find the adjugate ($\text{adj}(A)$).
3. Compute the inverse:

$$A^{-1} = \frac{1}{-16} \cdot \text{adj}(A).$$

—

Properties of Inverse Matrices

- **Uniqueness:** If a matrix A is invertible, its inverse A^{-1} is unique.
- **Inverse of the Transpose:**

$$(A^T)^{-1} = (A^{-1})^T.$$

- **Inverse of a Product:**

$$(AB)^{-1} = B^{-1}A^{-1}.$$

- **Inverse of the Inverse:**

$$(A^{-1})^{-1} = A.$$

—

Applications of Inverse Matrices

- **Solving Systems of Linear Equations:** Inverse matrices are used in solving systems of equations represented as $AX = B$, where:

$$X = A^{-1}B.$$

- **Linear Transformations:** The inverse matrix reverses the effects of a linear transformation.
- **Computer Graphics:** Used in geometric transformations like rotations, scaling, and projections.
- **Engineering and Statistics:** Widely used in numerical methods and data analysis.

—

Conclusion

The inverse of a matrix is an essential concept in linear algebra, with applications spanning various fields. Understanding its computation and properties enables efficient problem-solving in mathematics, physics, engineering, and computer science.

Rank of a Matrix

The rank of a matrix is the maximum number of linearly independent rows or columns in the matrix. It represents the dimensionality of the vector space spanned by the rows (row rank) or columns (column rank). For any matrix, the row rank is always equal to the column rank.

Key Points:

- A row (or column) is said to be linearly independent if it cannot be expressed as a linear combination of other rows (or columns).
- The rank is equal to the number of non-zero rows in the row-reduced echelon form (RREF) of the matrix.

Finding the Rank of a Matrix: The rank can be determined using **row reduction**: 1. Convert the matrix to its row echelon form using Gaussian elimination. 2. Count the number of non-zero rows in the row echelon form. This count gives the rank.

—

Example: Consider the matrix:

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

Step 1: Row Reduction To find the rank, perform Gaussian elimination to reduce the matrix to its row echelon form: 1. Subtract $4 \times$ Row 1 from Row 2:

$$\text{Row } 2 = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} - 4 \cdot \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & -3 & -6 \end{bmatrix}.$$

2. Subtract $7 \times$ Row 1 from Row 3:

$$\text{Row } 3 = \begin{bmatrix} 7 & 8 & 9 \end{bmatrix} - 7 \cdot \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & -6 & -12 \end{bmatrix}.$$

3. Divide Row 2 by -3 to simplify:

$$\text{Row } 2 = \begin{bmatrix} 0 & -3 & -6 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}.$$

4. Subtract $6 \times$ Row 2 from Row 3:

$$\text{Row } 3 = \begin{bmatrix} 0 & -6 & -12 \end{bmatrix} - 6 \cdot \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

The resulting row echelon form is:

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

—

Step 2: Count Non-Zero Rows In the row-reduced form, the number of non-zero rows is 2:

$$\text{Non-zero rows: } \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}.$$

Therefore, the rank of the matrix is:

$$\text{Rank}(A) = 2.$$

—

Step 3: Verify Linear Independence The third row of the original matrix, $\begin{bmatrix} 7 & 8 & 9 \end{bmatrix}$, is a linear combination of the first two rows:

$$\text{Row } 3 = (\text{Row } 1) + (\text{Row } 2).$$

This shows that only two rows are linearly independent, confirming that the rank is 2.

—

Conclusion: The rank of a matrix is determined by:

- Performing row reduction to bring the matrix into row echelon form.
- Counting the number of non-zero rows in the reduced matrix.

For the given matrix, the rank is 2, as only two rows are linearly independent. This means the matrix spans a 2-dimensional vector space.

Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors of a matrix A satisfy the equation:

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

where:

- λ is an eigenvalue, and

- $\mathbf{v}$ is the corresponding eigenvector.

How to Calculate Eigenvalues and Eigenvectors

Step 1: Find the Characteristic Equation The eigenvalues of a matrix are calculated by solving the characteristic equation:

$$\det(A - \lambda I) = 0,$$

where I is the identity matrix and λ is a scalar.

Step 2: Solve for λ The determinant of $A - \lambda I$ gives a polynomial equation in λ. Solve this equation to find the eigenvalues.

Step 3: Find Eigenvectors For each eigenvalue λ, substitute it back into the equation $(A - \lambda I)\mathbf{v} = 0$ to solve for the eigenvector $\mathbf{v}$.

—

Example: Let A be the matrix:

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

Step 1: Compute $A - \lambda I$ The identity matrix I for a 2×2 matrix is:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus,

$$A - \lambda I = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}.$$

Step 2: Find the Determinant The determinant of $A - \lambda I$ is:

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}.$$

Using the determinant formula for a 2×2 matrix:

$$\det(A - \lambda I) = (2-\lambda)(2-\lambda) - (1)(1) = (2-\lambda)^2 - 1.$$

Simplify:

$$\det(A - \lambda I) = 4 - 4\lambda + \lambda^2 - 1 = \lambda^2 - 4\lambda + 3.$$

Set the determinant equal to zero to find the eigenvalues:

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

Factorize:

$$(\lambda - 3)(\lambda - 1) = 0.$$

Thus, the eigenvalues are:

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

Step 3: Find Eigenvectors For each eigenvalue, solve $(A - \lambda I)\mathbf{v} = 0$.

—

For $\lambda_1 = 3$: Substitute $\lambda_1 = 3$ into $A - \lambda I$:

$$A - 3I = \begin{bmatrix} 2-3 & 1 \\ 1 & 2-3 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Solve $(A - 3I)\mathbf{v} = 0$, where $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$:

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives the equations:

$$-1x_1 + 1x_2 = 0 \quad \Rightarrow \quad x_1 = x_2.$$

The eigenvector corresponding to $\lambda_1 = 3$ is:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

—

For $\lambda_2 = 1$: Substitute $\lambda_2 = 1$ into $A - \lambda I$:

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

Solve $(A - I)\mathbf{v} = 0$, where $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives the equations:

$$x_1 + x_2 = 0 \quad \Rightarrow \quad x_1 = -x_2.$$

The eigenvector corresponding to $\lambda_2 = 1$ is:

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

—

Final Results:

- Eigenvalue $\lambda_1 = 3$, Eigenvector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- Eigenvalue $\lambda_2 = 1$, Eigenvector $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Vectors and Scalars

Vectors and scalars are fundamental concepts in linear algebra and play a crucial role in deep learning. They serve as the building blocks for representing data, weights, and operations in neural networks.

—

Scalars

A scalar is a single numerical value that represents magnitude. Scalars are used to measure quantities like learning rates, biases, or thresholds in deep learning.

Properties of Scalars:

- Scalars can be real numbers ($\mathbb{R}$), integers ($\mathbb{Z}$), or complex numbers ($\mathbb{C}$).
- Basic operations such as addition, subtraction, multiplication, and division are defined for scalars.

Examples in Deep Learning:

- A scalar 0.01 can represent the learning rate for gradient descent.
- A scalar 1.0 can denote the activation threshold for a ReLU function.

—

Vectors

A vector is an ordered list of numbers, typically represented as a column or row. Vectors are used in deep learning to represent data points, feature embeddings, and weight parameters.

Representation:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \text{or} \quad \mathbf{v} = [v_1, v_2, \ldots, v_n].$$

Examples in Deep Learning:

- Input to a neural network: A vector representing a data sample with features (e.g., pixel values for an image).
- Weight vector: A vector representing weights for a single layer in a neural network.

—

Types of Vectors

- **Zero Vector:** A vector with all elements equal to zero.

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

- **Unit Vector:** A vector with a magnitude of 1, often used to represent directions.

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|},$$

  where $\|\mathbf{v}\|$ is the magnitude of $\mathbf{v}$.

- **Row Vector:** A vector represented as a single row.

$$\mathbf{v} = [v_1, v_2, v_3].$$

- **Column Vector:** A vector represented as a single column.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

—

Vector Operations in Deep Learning

Addition and Subtraction:

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

Scalar Multiplication:

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

where c is a scalar. This is commonly used in gradient descent to scale gradients during optimization.

Dot Product (Inner Product):

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

Example: For $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$,

$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 4 + 10 + 18 = 32.$$

Cross Product: (For 3D vectors only)

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are unit vectors along the x, y, and z-axes.

—

Magnitude (Norm) of a Vector

The magnitude or norm of a vector $\mathbf{v}$ is given by:

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

Example: For $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$,

$$\|\mathbf{v}\| = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = 5.$$

Norms are essential in deep learning for regularization (e.g., L2 norm for weight decay) and measuring vector distances.

—

Applications in Deep Learning

Vectors and scalars are indispensable in deep learning:

- **Data Representation:** Input data is represented as vectors (e.g., pixel intensities, word embeddings).
- **Weights and Biases:** Parameters of neural networks are stored as vectors or matrices.
- **Activation Functions:** Operations like $\mathbf{W} \cdot \mathbf{x} + \mathbf{b}$ involve scalar and vector calculations.
- **Gradient Descent:** Gradients, represented as vectors, guide the optimization process.
- **Distance Measures:** Vectors are used to compute distances (e.g., Euclidean or cosine similarity) in embedding spaces.

—

Conclusion

Vectors and scalars form the foundation of mathematical operations in deep learning. They are crucial for representing data, optimizing parameters, and understanding complex transformations in neural networks. A strong grasp of these concepts enables better comprehension of deep learning models and their underlying mathematics.