AIFS Lecture 1: Bare Minimum Linear Algebra

Suraj Narayanan Sasikumar

Hessian AI Labs

February 16, 2020

Overview

Table of Contents

Sets

- A *set* is a well-defined collection of distinct objects.
- Two ways to write a set: Explicit or Implicit
- *Explicit*: Write the set by explicitly mentioning its members within curly brackets.
  Eg: $\{Honda, Ferrari, Renault\}$, $\{3, 5, 17\}$, $\emptyset$, $\{1, 3, \{1, 5\}\}$
- *Implicit*: Write in set-builder notation.
  Eg: $A = \{x : x \in H \text{ and } x < 10\}$ where $H = \{94, 324, 1, 9, 50\}$
  $A = \{1, 9\}$
- The set builder notation is read out loud as:
  x such that x in H and x less than 10

Properties of Sets

- All elements of a set are unique.
- The order of the elements in the set does not matter.
- Sets need not be homogeneous.
- *Cardinality*: The number of elements in the set. Eg: Let A={3,4}, then $|A| = 2$
- *Membership*: Given any object x and a set A, we can ask if x is a member of A. If x is a member then we write $x \in A$, if not we write $x \notin A$. Eg: Let $A = \{4, 7\}$, then $4 \in A$ and $9 \notin A$
- *Subset*: Given some sets B and A, B is a subset of A if all the members of B is in A. It's written as $B \subset A$. Eg: Let $A = \{3, 5, 7\}$ and $B = \{3, 7\}$, then $B \subset A$
- *Power Set*: Given a set S, the power-set of S, denoted by $\mathcal{P}(S)$, is the set of all subsets of S. Eg: Let $S = \{3, 5, 7\}$, then $\mathcal{P}(S) = \{\emptyset, \{3\}, \{5\}, \{7\}, \{3, 5\}, \{3, 7\}, \{5, 7\}, \{3, 5, 7\}\}$

Tuples

- A tuple is a finite ordered list of elements. Eg: $(1, 3), (1, 3, 1), (5, 5)$
- The ordering is important in tuples, and hence the uniqueness property does not hold.
- Based on the length of the tuple it's often referred to as an *n*-tuple. Eg: $(1, 3)$ is a 2-tuple. A 2-tuple is also known as an *ordered pair*.

Operations on Sets

Let $A = \{3, 5, 7, 11, 13\}$, and $B = \{1, 3, 5, 7, 9\}$

- Union: $\cup$ (think joining)
  $A \cup B = \{1, 3, 5, 7, 9, 11, 13\}$
- Intersection: $\cap$ (think commonality)
  $A \cap B = \{3, 5, 7\}$
- Difference: $\setminus$ (think subtraction)
  $A \setminus B = \{11, 13\}$
  $B \setminus A = \{1, 9\}$
- Cartesian Product: $\times$ (think combination)
  $A \times B = \{(x, y) : x \in A \text{ and } x \in B\}$
  $A \times B = \{(3, 1), (3, 3), (3, 5), \ldots, (5, 1), (5, 3), (5, 5), \ldots\}$

Number Systems

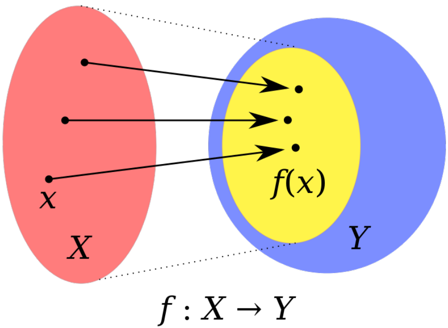
- Natural Numbers: $\mathbb{N} = \{1, 2, 3, 4 \ldots\}$ or $\{0, 1, 2, 3, 4 \ldots\}$
- Integers: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- Rational Numbers: The set of all numbers that can be written in the form $\frac{a}{b}$ where $b \neq 0$ and $a, b \in \mathbb{Z}$. Denoted by $\mathbb{Q}$
- Real Numbers: The set of all rational numbers and irrational numbers (such as $\sqrt{2}, \sqrt{5}, \pi$, etc.). Denoted by $\mathbb{R}$

Table of Contents

Functions

- A function is a mapping from one set to another where each element of the first set is mapped exactly to one element in the second
- *Notation*: A function, f, which maps elements in set X to set Y is written as:
$$f : X \to Y$$
- *Notation*: The mapping is explicitly specified as $y = f(x)$ read out loud as *"f of x"*. Here y is the *value* or *output* of the function and x is the *argument* or *input* to the function.
- *Example*: Let $X = \{-1, -2, 2, 0, 3\}$ and $Y = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, then forall $x \in X$, $f(x) = x^2$ is a function from X to Y

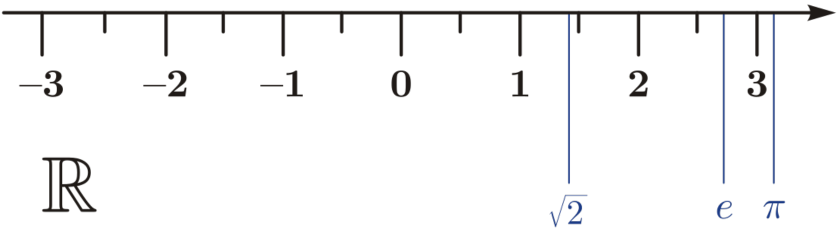
Functions



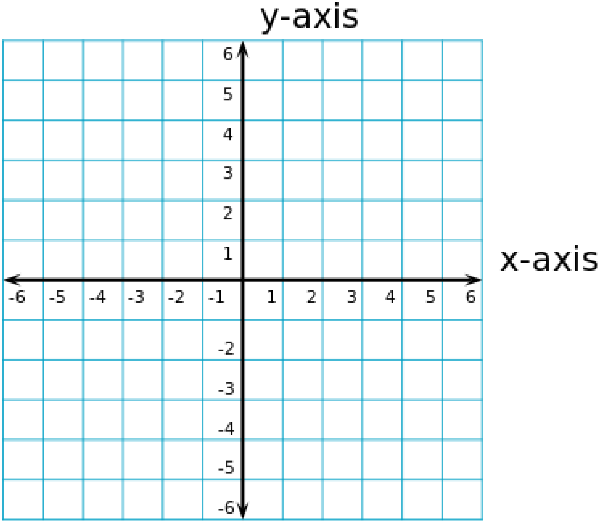
- Red set is called the *Domain* of the function.
- Blue set is called the *Codomain* of the function.
- Yellow set is a subset of the Codomain called the *range* of the function.

Table of Contents

1-D Space

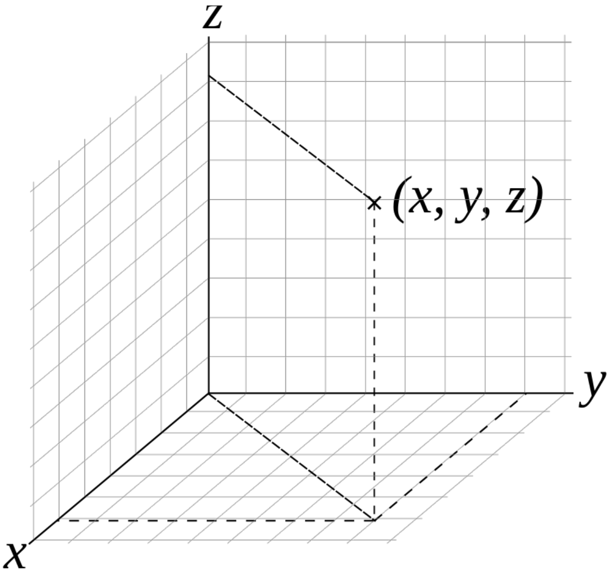


$\mathbb{R}^2$



3-D Space

$\mathbb{R}^3$



N-D Space

- A point in $\mathbb{R}^n$ space would have n coordinates.
- A line in $\mathbb{R}^2$ would be a plane in $\mathbb{R}^3$, and a hyper-plane in $\mathbb{R}^n$
- To visualize an N-D space, think of a 3-D space and call it N-D :)

Table of Contents

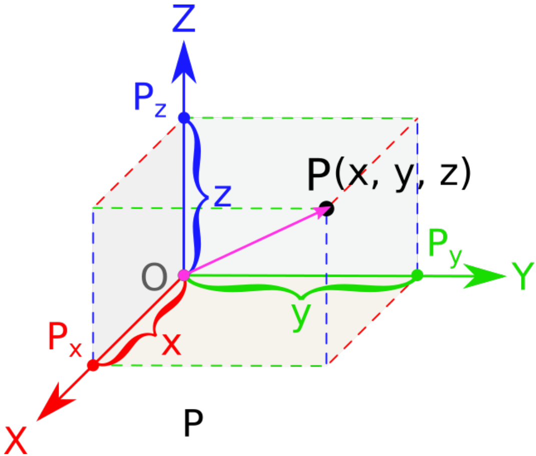
Scalars and Vectors

- *Scalar*: A single number. Usually written with lower case non-bold variable names. Eg: $3, 4.5, \frac{1}{3}, a$ for all $a \in \mathbb{N}$
- *Vector*: An array of ordered numbers. Can be written as a column of numbers, or as an n-tuple. Usually written with lower case bold variable names.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ OR } \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- A vector can be thought of as a 1-D array or numbers or a 2-D array with one column.

Cartesian Coordinate View



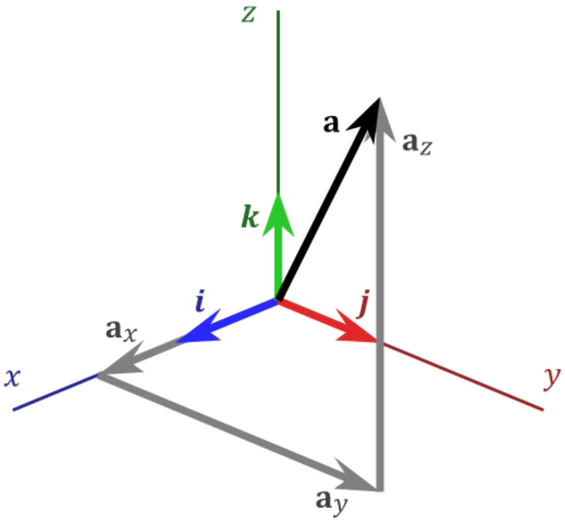
A vector can be visualized as an object identifying points in space, with each component of the vector giving the coordinate along a different axis.

Vector space

- A real valued vector space is a set V on which two operations $+$ and $\cdot$ are defined, called vector addition and scalar multiplication, and satisfy the following three properties for all scalars $c \in \mathbb{R}$.
- Closure under addition: If $\mathbf{u}$ and $\mathbf{v}$ are vectors in V then $\mathbf{u} + \mathbf{v}$ should also be in V.
- Closure under multiplication: If c is any scalar in $\mathbb{R}$ and $\mathbf{u}$ is a vector in V then $c \cdot \mathbf{u}$ is also in V.
- The zero vector $\mathbf{0}$ is in V.

Vector space

Vector space in $\mathbb{R}^3$



Matrices

- A *matrix* is a 2-D array of numbers arranged in rows and columns for which the addition ($+$) and multiplication ($\times$) operations are defined.

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

- If a matrix $\boldsymbol{A}$ has m rows and n columns then we call it an $m \times n$ matrix, read as "*m by n matrix*" where m and n are called it's *dimensions*. $m \times n$ is also referred to as the *shape* of the matrix.
- *Notations*: $\boldsymbol{A}_{m \times n}$ specifies the dimensions of the matrix as subscript. $A_{i,j}$ refers to the value in the i^{th} row and j^{th} column.
- Eg. of a 2×3 matrix

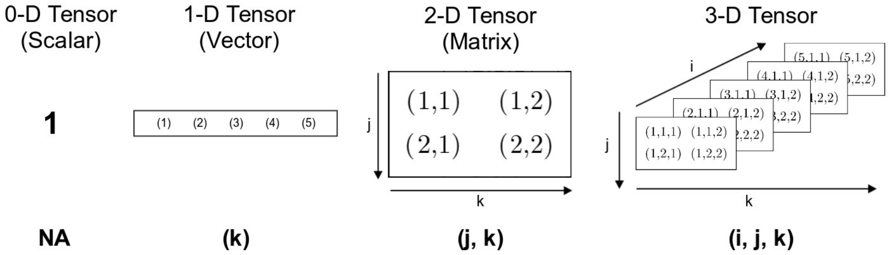
$$\boldsymbol{A}_{2\times 3} = \boldsymbol{A} = \begin{bmatrix} 5 & 9 & 1.2 \\ 3.5 & 5 & 6 \end{bmatrix} \text{ here, } A_{1,3} = 1.2$$

Matrices

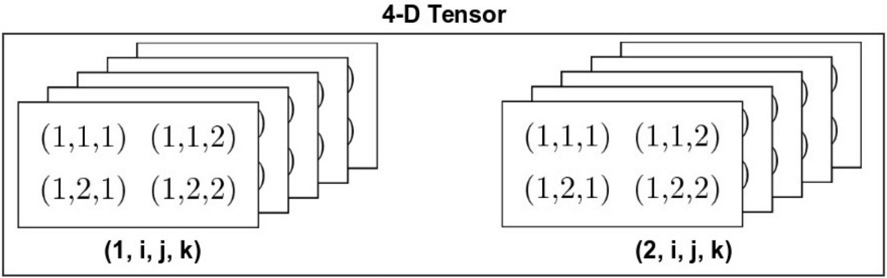
- *column-vector*: A matrix with m rows and 1 column.
- *row-vector*: A matrix with 1 row and n columns.
- *square matrix*: An $m \times n$ matrix where $m = n$
- *rectangular matrix*: An $m \times n$ matrix where $m \neq n$

Tensors

- A tensor is a mathematical object represented as a multidimensional array of numbers. It can be thought of as a generalization of the matrix to N-dimensions.



4D-Tensor



5D-Tensor

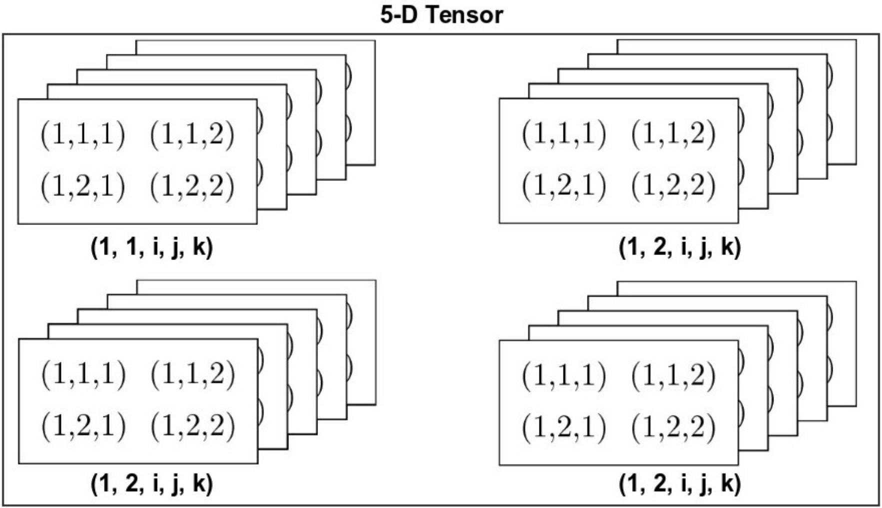
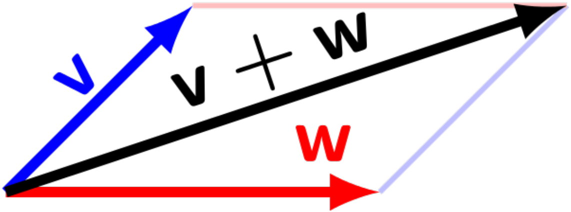

5-D Tensor
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 1, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)

Table of Contents

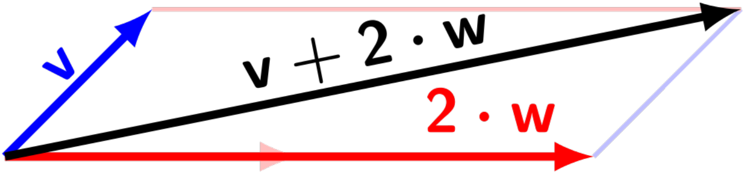
Vector addition

- Let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$,
- then $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$



Scalar Multiplication

- Let $\mathbf{w} = (w_1, w_2, \dots, w_n)$, and c a scalar,
- then $c \cdot \mathbf{w} = (c \cdot w_1, c \cdot w_2, \dots, c \cdot w_n)$



Matrix Operations

- *Matrix Addition*: Two matrices can only be added together if they have the same shape. Given two matrices $\boldsymbol{A}, \boldsymbol{B}$ of the same shape, then $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, where $C_{i,j} = A_{i,j} + B_{i,j}$

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

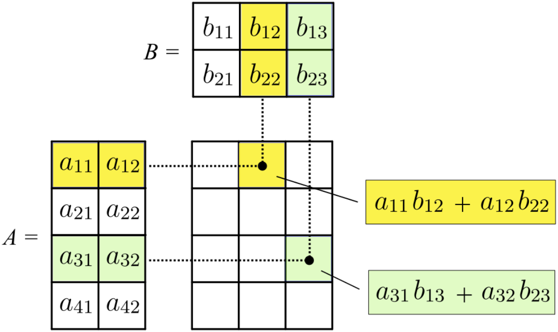
$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

- *Scalar Multiplication and Addition*: Given a matrix $\boldsymbol{A}$ and scalars c, d, the matrix $\boldsymbol{B} = c \cdot \boldsymbol{A} + d$ is given by $B_{i,j} = c \cdot A_{i,j} + d$

Matrix Operations

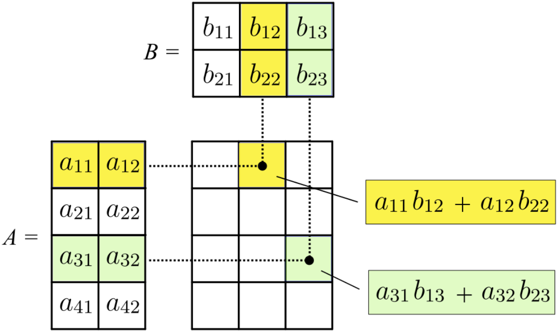
- *Matrix Multiplication*: Two matrices can only be multiplied together if the number of columns of the first matrix match the number of rows in the second. More concretely, given two matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{p\times q}$, the product $\boldsymbol{C} = \boldsymbol{AB}$ is only defined when $n = p$. The product is defined as:

$$C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$



- *Matrix Multiplication*: Two matrices can only be multiplied together if the number of columns of the first matrix match the number of rows in the second. More concretely, given two matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{p\times q}$, the product $\boldsymbol{C} = \boldsymbol{AB}$ is only defined when $n = p$. The product is defined as:

$$C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$



Transpose Operation

- Transpose of a matrix (or a vector represented as a matrix) is the mirror image of the matrix across its main diagonal.
- Easier way to visualize: rows become columns and columns become rows.
- Transpose of a matrix $\boldsymbol{A}$ is written as $\boldsymbol{A}^\top$

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

Table of Contents

More about Spaces

- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)
- Length (magnitude, size) of a vector
- Angle between two vector

More about Spaces

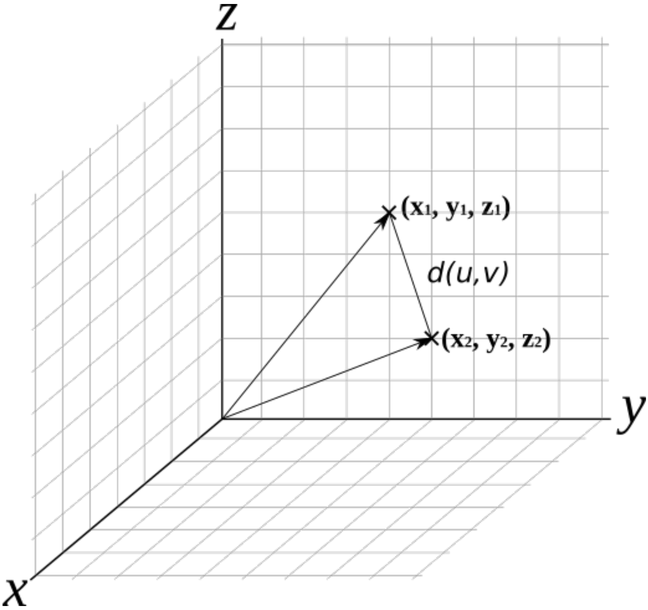
- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)... Metric space
- Length (magnitude, size) of a vector... Normed Space
- Angle between two vector... Inner Product Space

Metric Spaces

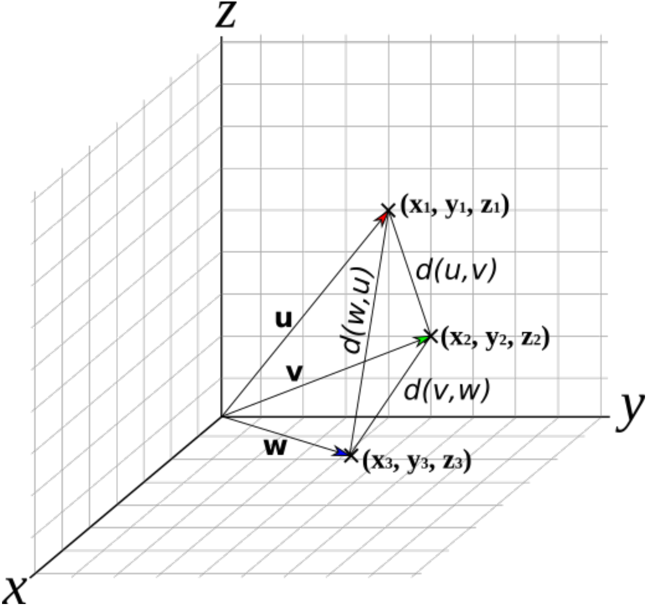
- A vector space with a *distance function* giving the distance between any two points is called a metric space. [1]
- For a function, $d(\cdot, \cdot)$, to be a distance function, it is needs to satisfy the following properties.
- *Identity of Indiscernibles*: $d(\mathbf{u}, \mathbf{v}) = 0 \Rightarrow \mathbf{u} = \mathbf{v}$
- *Symmetry*: $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
- *Triangle Inequality*: $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

[1] This definition is correct but not complete, to flush out the actual details we need to go more into pure math, which is not required for us.

Metric Spaces: Visualization



Metric Spaces: Visualization



Examples of Metric Spaces

- $\mathbb{R}^n$ with euclidean distance. If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$,

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}$$

- Manhattan distance or taxi-cab distance or L_1-distance. If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$,

$$d(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} |u_i - v_i|$$

- Wasserstein metric. A metric that gives a measure of distance between two probability distributions. Useful in Deep learning. Formula? Too brutal!

More about Spaces

- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)... Metric space ✓
- Length (magnitude, size) of a vector... Normed Space
- Angle between two vector... Inner Product Space

Normed Space

- A *Norm* is a function that gives the length (magnitude, size) of a vector. The notation $||\cdot||$ is often used for a norm.
- For a function, $f(\cdot) = ||\cdot||$, to be a norm, it has to satisfy the following properties.
- *Triangle Inequality*: $||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
- *Absolute Homogeneity*: $||a\mathbf{u}|| = a||\mathbf{u}||$
- *Positive Definite*: If $||\mathbf{u}|| = 0 \Rightarrow \mathbf{u} = 0$

Examples of Normed spaces

- $\mathbb{R}^n$ with ℓ_1-*norm* (Taxicab norm):

$$||\mathbf{u}|| = ||\mathbf{u}||_1 = \sum_{i=1}^{n} |u_i|$$

- $\mathbb{R}^n$ with ℓ^2-*norm* (Euclidean norm):

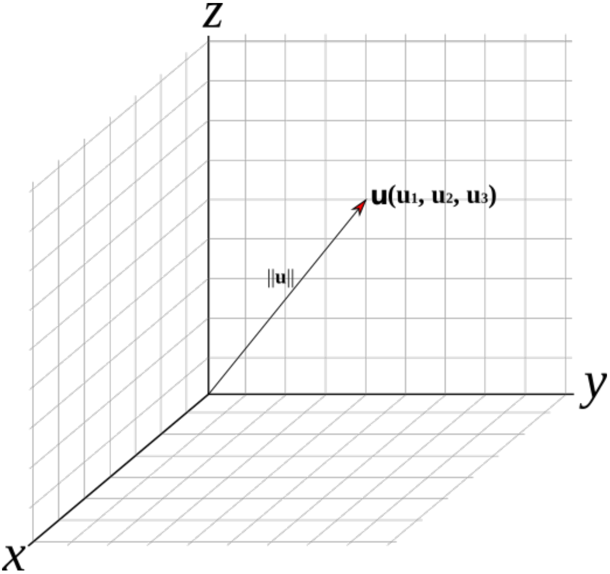
$$||\mathbf{u}|| = ||\mathbf{u}||_2 = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$$

- $\mathbb{R}^n$ with *max-norm* (Infinity norm):

$$||\mathbf{u}|| = ||\mathbf{u}||_\infty = \max(|u_1|, |u_2|, \ldots, |u_n|)$$

Visualization: Normed Spaces

Do you a see a connection between metric spaces and normed spaces?



Norm induces a Metric

- A Norm induces a Metric in the space.
- If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ and $||\mathbf{u}||$, $||\mathbf{v}||$ are their respective norms, then the norm $||\cdot||$ induces the following metric:

$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

Visualization: Norm and Metrics