Introduction and Linear Algebra

Libin Jiao

Dalian University of Technology

September 25, 2022

Outline

1. Introduction
2. Linear Algebra

Outline

1 Introduction

2 Linear Algebra

Introduction

Introduction

- MA2041 Mathematical Foundations of AI and Machine Learning
- Module Mark $=$ Coursework $\times$ 30%+Exam $\times$ 70%

What is artificial intelligence

- Artificial intelligence (AI), also known as machine intelligence, is a branch of computer science that focuses on building and managing technology that can learn to autonomously make decisions and carry out actions on behalf of a human being.
- Artificial intelligence is a constellation of many different technologies working together to enable machines to sense, comprehend, act, and learn with human-like levels of intelligence.

What is artificial intelligence

- Artificial intelligence (AI), also known as machine intelligence, is a branch of computer science that focuses on building and managing technology that can learn to autonomously make decisions and carry out actions on behalf of a human being.
- Artificial intelligence is a constellation of many different technologies working together to enable machines to sense, comprehend, act, and learn with human-like levels of intelligence.

What is Machine Learning?

- Arthur Samuel (1959): Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed
- Machine learning is the science of getting machines to learn and act in a similar way to humans while also autonomously learning from real-world interactions and sets of training data that we feed them
- Machine learning focuses on the development of computer programs that can access data and use it learn for themselves

What is Machine Learning?

- Arthur Samuel (1959): Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed
- Machine learning is the science of getting machines to learn and act in a similar way to humans while also autonomously learning from real-world interactions and sets of training data that we feed them
- Machine learning focuses on the development of computer programs that can access data and use it learn for themselves

What is Machine Learning?

- Arthur Samuel (1959): Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed
- Machine learning is the science of getting machines to learn and act in a similar way to humans while also autonomously learning from real-world interactions and sets of training data that we feed them
- Machine learning focuses on the development of computer programs that can access data and use it learn for themselves

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

What is Machine Learning?

Learning = Improving with experience at some task

- From [Simon 83] Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time
- From [Nilsson 96]A machine learns whenever it changes its structure, program, or data in such a manner that its expected future performance improves

What is Machine Learning?

Learning = Improving with experience at some task

- From [Simon 83] Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time
- From [Nilsson 96]A machine learns whenever it changes its structure, program, or data in such a manner that its expected future performance improves

What is Machine Learning?

Learning = Improving with experience at some task

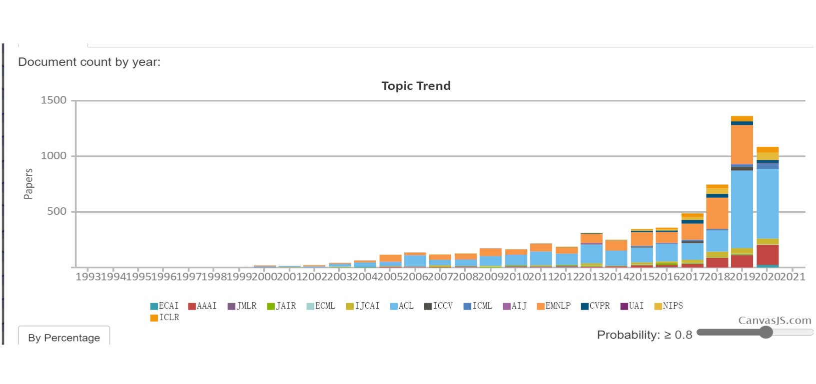
- From [Simon 83] Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time
- From [Nilsson 96]A machine learns whenever it changes its structure, program, or data in such a manner that its expected future performance improves

Machine Learning (and AI) is Very Hot

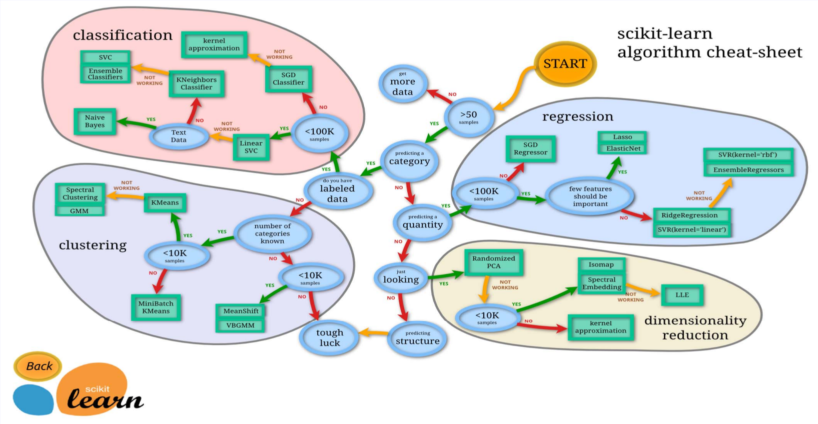
Countries and companies invest heavily in ML and AI.

Machine Learning (and AI) is Very Hot

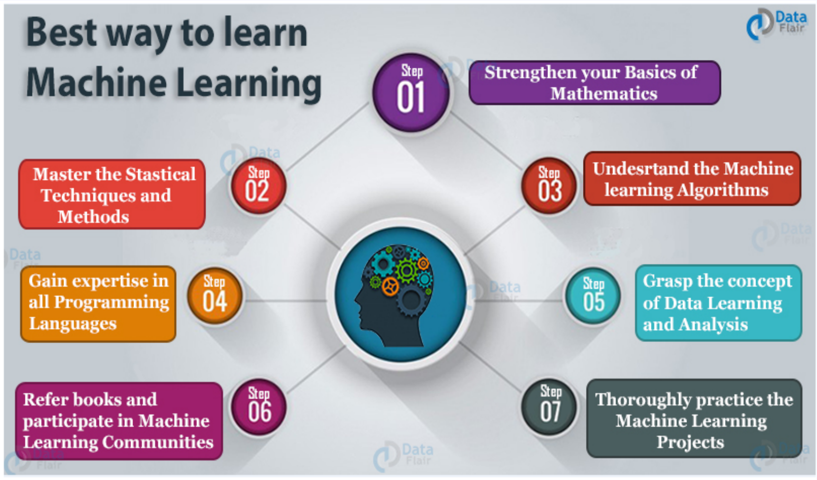
The number of research papers on AI and Machine Learning has been increasing sharply in the past few years.



http://aipano.cse.ust.hk/


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
<100K samples
Linear SVC
Text Data
Naive Bayes
KNeighbors Classifier
SVC
Ensemble Classifiers
SGD Classifier
kernel approximation
classification
number of categories known
<10K samples
KMeans
Spectral Clustering
GMM
MiniBatch KMeans
MeanShift
VBGMM
clustering
tough luck
predicting a quantity
regression
SGD Regressor
few features should be important
Lasso
ElasticNet
RidgeRegression
SVR(kernel='linear')
SVR(kernel='rbf')
EnsembleRegressors
just looking
predicting structure
Randomized PCA
Isomap
Spectral Embedding
LLE
kernel approximation
dimensionality reduction
YES
NO
NOT WORKING
Back
scikit learn

While machine learning has seen many success stories, and software is readily available to design and train rich and flexible machine learning systems, we believe that the mathematical foundations of machine learning are important in order to understand fundamental principles upon which more complicated machine learning systems are built.


Best way to learn
Machine Learning
Data Flair
Step 01
Strengthen your Basics of Mathematics
Step 02
Master the Stastical Techniques and Methods
Step 03
Undesrtand the Machine learning Algorithms
Step 04
Gain expertise in all Programming Languages
Step 05
Grasp the concept of Data Learning and Analysis
Step 06
Refer books and participate in Machine Learning Communities
Step 07
Thoroughly practice the Machine Learning Projects

Outline

1 Introduction

2 Linear Algebra

Linear Algebra

Systems of Linear Equations

Example

A company produces products $N_1, \dots, N_n$ for which resources $R_1, \dots, R_m$ are required. To produce a unit of product N_j, a_{ij} units of resource R_i are needed, where $i = 1, \dots, m$ and $j = 1, \dots, n$.

The objective is to find an optimal production plan, i.e., a plan of how many units x_j of product N_j should be produced if a total of b_i units of resource R_i are available and (ideally) no resources are left over.

Systems of Linear Equations

Example

A company produces products $N_1, \ldots, N_n$ for which resources $R_1, \ldots, R_m$ are required. To produce a unit of product N_j, a_{ij} units of resource R_i are needed, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

The objective is to find an optimal production plan, i.e., a plan of how many units x_j of product N_j should be produced if a total of b_i units of resource R_i are available and (ideally) no resources are left over.

If we produce $x_1, \ldots, x_n$ units of the corresponding products, we need a total of

$$a_{i1}x_1 + \cdots + a_{in}x_n$$

many units of resource R_i. An optimal production plan $(x_1, \ldots, x_n) \in \mathbb{R}^n$, therefore, has to satisfy the following system of equations:

$$\begin{gathered} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{gathered}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$.

If we produce $x_1, \ldots, x_n$ units of the corresponding products, we need a total of

$$a_{i1}x_1 + \cdots + a_{in}x_n$$

many units of resource R_i. An optimal production plan $(x_1, \ldots, x_n) \in \mathbb{R}^n$, therefore, has to satisfy the following system of equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$.

Matrix

$$\boldsymbol{A} = (a_{ij})_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

- diagonal matrix: $\mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn}) =$

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix},$$

- identity matrix: $\mathrm{diag}(1, 1, \cdots, 1)$
- trace: $tr(\boldsymbol{A}) = \sum_j^n a_{jj}$

Matrix

$$A = (a_{ij})_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

- diagonal matrix: $\mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn})=$

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix},$$

- identity matrix: $\mathrm{diag}(1, 1, \cdots, 1)$
- trace: $tr(A) = \sum_j^n a_{jj}$

Matrix

$$\boldsymbol{A} = (a_{ij})_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

- diagonal matrix: $\mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn})$=

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix},$$

- identity matrix: $\mathrm{diag}(1, 1, \cdots, 1)$
- trace: $tr(\boldsymbol{A}) = \sum_j^n a_{jj}$

Matrix Addition/Subtraction

if C = A + B, then $(c_{ij}) = (a_{ij}) \pm (b_{ij})$

- commutative: $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$
- associative: $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$

Matrix Addition/Subtraction

if C = A + B, then $(c_{ij}) = (a_{ij}) \pm (b_{ij})$

- commutative: $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$
- associative: $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$

Multiply a Vector by a Matrix

$$\boldsymbol{Ax} = \boldsymbol{y}$$

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

write $\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}]$, then

$$y_i = \sum_{j=1}^{n} x_j \boldsymbol{a_j}$$

- $\boldsymbol{y}$ can be written as a weighted sum of $\boldsymbol{A}'s$ column vectors

Multiply a Vector by a Matrix

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

write $\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}]$, then

$$y_i = \sum_{j=1}^{n} x_j \boldsymbol{a_j}$$

- $\boldsymbol{y}$ can be written as a weighted sum of $\boldsymbol{A}'s$ column vectors

Multiply a Vector by a Matrix

$$\boldsymbol{Ax} = \boldsymbol{y}$$

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

write $\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}]$, then

$$y_i = \sum_{j=1}^{n} x_j \boldsymbol{a_j}$$

- $\boldsymbol{y}$ can be written as a weighted sum of $\boldsymbol{A}'s$ column vectors

Matrix Multiplication

if $\boldsymbol{C}_{m\times n} = \boldsymbol{A}_{m\times p} \times \boldsymbol{B}_{p\times n}$, then $(c_{ij}) = \sum_{k=1}^{p} a_{ik}b_{kj}$

- in general, non-commutative: $\boldsymbol{AB} \neq \boldsymbol{BA}$
- associative: $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- distributive: $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$

Matrix Multiplication

if $\boldsymbol{C}_{m\times n} = \boldsymbol{A}_{m\times p} \times \boldsymbol{B}_{p\times n}$, then $(c_{ij}) = \sum_{k=1}^{p} a_{ik} b_{kj}$

- in general, non-commutative: $\boldsymbol{AB} \neq \boldsymbol{BA}$
- associative: $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- distributive: $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$

Matrix Multiplication

if $\boldsymbol{C}_{m\times n} = \boldsymbol{A}_{m\times p} \times \boldsymbol{B}_{p\times n}$, then $(c_{ij}) = \sum_{k=1}^{p} a_{ik}b_{kj}$

- in general, non-commutative: $\boldsymbol{AB} \neq \boldsymbol{BA}$
- associative: $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- distributive: $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{A}\boldsymbol{B})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{A}\boldsymbol{B})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Transpose

- if $\boldsymbol{B} = \boldsymbol{A}^T$, then $\boldsymbol{B}_{ij} = \boldsymbol{A}_{ji}$
  - $\boldsymbol{A}^T$ is sometimes also denoted as $\boldsymbol{A}'$ or $\boldsymbol{A}^t$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$, $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- symmetric matrix: $a_{ij} = a_{ji}$ or $\boldsymbol{A} = \boldsymbol{A}^T$
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T\boldsymbol{A}$ is an identity matrix
- Matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $|\boldsymbol{A}| = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

Inverse

$$\boldsymbol{A}^{-1} = \frac{(\mathrm{cof}(\boldsymbol{A}))^T}{|\boldsymbol{A}|}$$

- $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$;
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$;
- $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T = \boldsymbol{A}^{-T}$.

Solving Systems of Linear Equations–The Minus-1 Trick

In the following, we introduce a practical trick for reading out the solutions $\boldsymbol{x}$ of a homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{A} \in \mathbb{R}^{k \times n}, \boldsymbol{x} \in \mathbb{R}^n$.

To start, we assume that $\boldsymbol{A}$ is in reduced row-echelon form without any rows that just contain zeros, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \end{bmatrix}$$

where $*$ can be an arbitrary real number, with the constraints that the first nonzero entry per row must be 1 and all other entries in the corresponding column must be 0.

Solving Systems of Linear Equations–The Minus-1 Trick

In the following, we introduce a practical trick for reading out the solutions $\boldsymbol{x}$ of a homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{A} \in \mathbb{R}^{k \times n}, \boldsymbol{x} \in \mathbb{R}^n$.
To start, we assume that $\boldsymbol{A}$ is in reduced row-echelon form without any rows that just contain zeros, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \end{bmatrix}$$

where $*$ can be an arbitrary real number, with the constraints that the first nonzero entry per row must be 1 and all other entries in the corresponding column must be 0.

Solving Systems of Linear Equations–The Minus-1 Trick

In the following, we introduce a practical trick for reading out the solutions $\boldsymbol{x}$ of a homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{A} \in \mathbb{R}^{k\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$.
To start, we assume that $\boldsymbol{A}$ is in reduced row-echelon form without any rows that just contain zeros, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \end{bmatrix}$$

where $*$ can be an arbitrary real number, with the constraints that the first nonzero entry per row must be 1 and all other entries in the corresponding column must be 0.

The Minus-1 Trick

The columns $j_1, \ldots, j_k$ with the pivots (marked in bold) are the standard unit vectors $\boldsymbol{e}_1, \ldots, e_k \in \mathbb{R}^k$. We extend this matrix to an $n \times n$-matrix $\tilde{A}$ by adding $n - k$ rows of the form

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

so that the diagonal of the augmented matrix $\tilde{A}$ contains either 1 or −1. Then, the columns of $\tilde{A}$ that contain the −1 as pivots are solutions of the homogeneous equation system Ax = 0. To be more precise, these columns form a basis (Section 2.6.1) of the solution space of $\boldsymbol{Ax} = \boldsymbol{0}$, which we will later call the kernel or null space (see Section 2.7.3).

The Minus-1 Trick

The columns $j_1, \ldots, j_k$ with the pivots (marked in bold) are the standard unit vectors $\boldsymbol{e}_1, \ldots, e_k \in \mathbb{R}^k$. We extend this matrix to an $n \times n$-matrix $\tilde{A}$ by adding $n - k$ rows of the form

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

so that the diagonal of the augmented matrix $\tilde{A}$ contains either 1 or -1. Then, the columns of $\tilde{A}$ that contain the -1 as pivots are solutions of the homogeneous equation system Ax $=$ 0. To be more precise, these columns form a basis (Section 2.6.1) of the solution space of $\boldsymbol{Ax} = \boldsymbol{0}$, which we will later call the kernel or null space (see Section 2.7.3).

Vector Space

Definition

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$

$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   1. $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
   2. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x}$
4. Neutral element with respect to the outer operation: $\forall \boldsymbol{x} \in \mathcal{V} : 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Vector Space

Definition

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$

$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   1. $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
   2. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x}$
4. Neutral element with respect to the outer operation: $\forall \boldsymbol{x} \in \mathcal{V} : 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Vector Space

Definition

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$
$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   1. $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
   2. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x}$
4. Neutral element with respect to the outer operation: $\forall \boldsymbol{x} \in \mathcal{V} : 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Vector Space

Definition

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$
$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
    1. $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
    2. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x}$
4. Neutral element with respect to the outer operation: $\forall \boldsymbol{x} \in \mathcal{V} : 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Vector Space

Definition

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$
$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   1. $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
   2. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity (outer operation):
$\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x}$
4. Neutral element with respect to the outer operation:
$\forall \boldsymbol{x} \in \mathcal{V} : 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Vector Space

- The elements $\boldsymbol{x} \in V$ are called vectors. The neutral element of $(\mathcal{V}, +)$ is the zero vector $\mathbf{0} = [0, \ldots, 0]^\top$, and the inner operation $+$ is called vector addition.
- The elements $\lambda \in \mathbb{R}$ are called scalars and the outer operation $\cdot$ is a multiplication by scalars. Note that a scalar product is something different, and we will get to this in Section 3.2.

Vector Space

- The elements $\boldsymbol{x} \in V$ are called vectors. The neutral element of $(\mathcal{V}, +)$ is the zero vector $\mathbf{0} = [0, \ldots, 0]^\top$, and the inner operation $+$ is called vector addition.
- The elements $\lambda \in \mathbb{R}$ are called scalars and the outer operation $\cdot$ is a multiplication by scalars. Note that a scalar product is something different, and we will get to this in Section 3.2.

Vector Space

- The elements $\boldsymbol{x} \in V$ are called vectors. The neutral element of $(\mathcal{V}, +)$ is the zero vector $\mathbf{0} = [0, \ldots, 0]^\top$, and the inner operation $+$ is called vector addition.
- The elements $\lambda \in \mathbb{R}$ are called scalars and the outer operation $\cdot$ is a multiplication by scalars. Note that a scalar product is something different, and we will get to this in Section 3.2.

Vector Subspace

Definition

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}, \mathcal{U} \neq \emptyset$. Then $U = (\mathcal{U}, +, \cdot)$ is called vector subspace of V (or linear subspace) if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$. We write $U \subseteq V$ to denote a subspace U of V.

Vector Subspace

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\boldsymbol{x} \in \mathcal{V}$, and in particular for all $\boldsymbol{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element.

To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

1. $\mathcal{U} \neq \emptyset$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U :
   a. With respect to the outer operation:
   $$\forall \lambda \in \mathbb{R} \forall \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}.$$
   b. With respect to the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$.

Vector Subspace

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\boldsymbol{x} \in \mathcal{V}$, and in particular for all $\boldsymbol{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element.

To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

1. $\mathcal{U} \neq \emptyset$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U :
   a. With respect to the outer operation:
   $$\forall \lambda \in \mathbb{R} \forall \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}.$$
   b. With respect to the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$.

Vector Subspace

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\boldsymbol{x} \in \mathcal{V}$, and in particular for all $\boldsymbol{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element.

To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

1. $\mathcal{U} \neq \varnothing$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U :
   a. With respect to the outer operation:

$$\forall \lambda \in \mathbb{R} \forall \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}.$$

   b. With respect to the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$.

Vector Subspace

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\boldsymbol{x} \in \mathcal{V}$, and in particular for all $\boldsymbol{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element.

To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

1. $\mathcal{U} \neq \varnothing$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U :
   a. With respect to the outer operation:

$$\forall \lambda \in \mathbb{R} \forall \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}.$$

   b. With respect to the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$.

Vector Subspace

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\boldsymbol{x} \in \mathcal{V}$, and in particular for all $\boldsymbol{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element.

To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

1. $\mathcal{U} \neq \varnothing$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U :
   a. With respect to the outer operation:

$$\forall \lambda \in \mathbb{R} \forall \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}.$$

   b. With respect to the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$.

Linear Independence

Definition

Consider a vector space V and a finite number of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. Then, every $\boldsymbol{v} \in V$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in V$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$.

The $\mathbf{0}$-vector can always be written as the linear combination of k vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ because $\mathbf{0} = \sum_{i=1}^{k} 0 \boldsymbol{x}_i$ is always true. In the following, we are interested in non-trivial linear combinations of a set of vectors to represent $\mathbf{0}$, i.e., linear combinations of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, where not all coefficients λ_i are 0 .

Linear Independence

Definition

Consider a vector space V and a finite number of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. Then, every $\boldsymbol{v} \in V$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in V$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$.

The $\mathbf{0}$-vector can always be written as the linear combination of k vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ because $\mathbf{0} = \sum_{i=1}^{k} 0\boldsymbol{x}_i$ is always true. In the following, we are interested in non-trivial linear combinations of a set of vectors to represent $\mathbf{0}$, i.e., linear combinations of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, where not all coefficients λ_i are 0 .

Linear Independence

Definition

Consider a vector space V and a finite number of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. Then, every $\boldsymbol{v} \in V$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in V$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$.

The $\mathbf{0}$-vector can always be written as the linear combination of k vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ because $\mathbf{0} = \sum_{i=1}^{k} 0\boldsymbol{x}_i$ is always true. In the following, we are interested in non-trivial linear combinations of a set of vectors to represent $\mathbf{0}$, i.e., linear combinations of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, where not all coefficients λ_i are 0 .

Linear Independence

Definition

Let us consider a vector space V with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $x_1, \ldots, x_k$ are linearly dependent. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are linearly independent.

Linear independence is one of the most important concepts in linear algebra. Intuitively, a set of linearly independent vectors consists of vectors that have no redundancy, i.e., if we remove any of those vectors from the set, we will lose something.

Linear Independence

Definition

Let us consider a vector space V with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $x_1, \ldots, x_k$ are linearly dependent. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are linearly independent.

Linear independence is one of the most important concepts in linear algebra. Intuitively, a set of linearly independent vectors consists of vectors that have no redundancy, i.e., if we remove any of those vectors from the set, we will lose something.

Basis and Rank

Definition

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{x_1, \ldots, x_k\} \subseteq \mathcal{V}$. If every vector $v \in \mathcal{V}$ can be expressed as a linear combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \mathcal{A}$ is called a generating set of V. The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, we write $V = \text{span}[\mathcal{A}]$ or $V = \text{span}\,[x_1, \ldots, x_k]$

Generating sets are sets of vectors that span vector (sub)spaces, i.e., every vector can be represented as a linear combination of the vectors in the generating set. Now, we will be more specific and characterize the smallest generating set that spans a vector (sub)space.

Basis and Rank

Definition

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{x_1, \ldots, x_k\} \subseteq \mathcal{V}$. If every vector $v \in \mathcal{V}$ can be expressed as a linear combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \mathcal{A}$ is called a generating set of V. The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, we write $V = \text{span}[\mathcal{A}]$ or $V = \text{span}\,[x_1, \ldots, x_k]$

Generating sets are sets of vectors that span vector (sub)spaces, i.e., every vector can be represented as a linear combination of the vectors in the generating set. Now, we will be more specific and characterize the smallest generating set that spans a vector (sub)space.

Basis

Definition

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{A} \subseteq \mathcal{V}$. A generating set $\mathcal{A}$ of V is called minimal if there exists no smaller set $\overline{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V}$ that spans V. Every linearly independent generating set of V is minimal and is called a basis of V.

Basis

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of V.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Basis

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of V.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Basis

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of V.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Basis

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \varnothing$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of V.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Basis

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of V.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Rank

Definition

The number of linearly independent columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the rank of $\boldsymbol{A}$ and is denoted by $\mathrm{rk}(\boldsymbol{A})$.

Rank

The rank of a matrix has some important properties:

- $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(\boldsymbol{A})$. Later we will call this subspace the image or range. A basis of U can be found by applying Gaussian elimination to A to identify the pivot columns.
- The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(\boldsymbol{A})$. A basis of W can be found by applying Gaussian elimination to $\boldsymbol{A}^\dagger$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is regular (invertible) if and only if $\text{rk}(\boldsymbol{A}) = n$.

Rank

The rank of a matrix has some important properties:

- $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}\left(\boldsymbol{A}^\top\right)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \mathrm{rk}(\boldsymbol{A})$. Later we will call this subspace the image or range. A basis of U can be found by applying Gaussian elimination to A to identify the pivot columns.
- The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \mathrm{rk}(\boldsymbol{A})$. A basis of W can be found by applying Gaussian elimination to $\boldsymbol{A}^\dagger$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is regular (invertible) if and only if $\mathrm{rk}(\boldsymbol{A}) = n$.

Rank

The rank of a matrix has some important properties:

- $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(\boldsymbol{A})$. Later we will call this subspace the image or range. A basis of U can be found by applying Gaussian elimination to A to identify the pivot columns.
- The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(\boldsymbol{A})$. A basis of W can be found by applying Gaussian elimination to $\boldsymbol{A}^\dagger$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is regular (invertible) if and only if $\text{rk}(\boldsymbol{A}) = n$.

Rank

The rank of a matrix has some important properties:

- $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(\boldsymbol{A})$. Later we will call this subspace the image or range. A basis of U can be found by applying Gaussian elimination to A to identify the pivot columns.
- The rows of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(\boldsymbol{A})$. A basis of W can be found by applying Gaussian elimination to $\boldsymbol{A}^\dagger$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ it holds that $\boldsymbol{A}$ is regular (invertible) if and only if $\text{rk}(\boldsymbol{A}) = n$.

Rank

The rank of a matrix has some important properties:

- $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}^\top)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \mathrm{rk}(\boldsymbol{A})$. Later we will call this subspace the image or range. A basis of U can be found by applying Gaussian elimination to A to identify the pivot columns.
- The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \mathrm{rk}(\boldsymbol{A})$. A basis of W can be found by applying Gaussian elimination to $\boldsymbol{A}^\dagger$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is regular (invertible) if and only if $\mathrm{rk}(\boldsymbol{A}) = n$.

properties...

- For all $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and all $b \in \mathbb{R}^m$ it holds that the linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be solved if and only if $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A} \mid \boldsymbol{b})$, where $\boldsymbol{A} \mid \boldsymbol{b}$ denotes the augmented system.
- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ the subspace of solutions for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ possesses dimension $n - \mathrm{rk}(\boldsymbol{A})$. Later, we will call this subspace the kernel or the null space.
- A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. This means that the rank of a full-rank matrix is the lesser of the number of rows and columns, i.e., $\mathrm{rk}(\boldsymbol{A}) = \min(m, n)$. A matrix is said to be rank deficient if it does not have full rank.

properties...

- For all $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and all $b \in \mathbb{R}^m$ it holds that the linear equation system $\boldsymbol{Ax} = \boldsymbol{b}$ can be solved if and only if $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A} \mid \boldsymbol{b})$, where $\boldsymbol{A} \mid \boldsymbol{b}$ denotes the augmented system.
- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ the subspace of solutions for $\boldsymbol{Ax} = \boldsymbol{0}$ possesses dimension $n - \mathrm{rk}(\boldsymbol{A})$. Later, we will call this subspace the kernel or the null space.
- A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. This means that the rank of a full-rank matrix is the lesser of the number of rows and columns, i.e., $\mathrm{rk}(\boldsymbol{A}) = \min(m, n)$. A matrix is said to be rank deficient if it does not have full rank.

properties...

- For all $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and all $b \in \mathbb{R}^m$ it holds that the linear equation system $\boldsymbol{Ax} = \boldsymbol{b}$ can be solved if and only if $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A} \mid \boldsymbol{b})$, where $\boldsymbol{A} \mid \boldsymbol{b}$ denotes the augmented system.
- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ the subspace of solutions for $\boldsymbol{Ax} = \boldsymbol{0}$ possesses dimension $n - \mathrm{rk}(\boldsymbol{A})$. Later, we will call this subspace the kernel or the null space.
- A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. This means that the rank of a full-rank matrix is the lesser of the number of rows and columns, i.e., $\mathrm{rk}(\boldsymbol{A}) = \min(m, n)$. A matrix is said to be rank deficient if it does not have full rank.

Rank

Example

$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ $\boldsymbol{A}$ has two linearly independent rows/columns so that $\mathrm{rk}(\boldsymbol{A}) = 2$.

Rank

Example

$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Here, we see that the number of linearly independent rows and columns is 2 , such that $\mathrm{rk}(\boldsymbol{A}) = 2$.

Rank

Example

$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \sim \dots \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Here, we see that the number of linearly independent rows and columns is 2 , such that $\text{rk}(\boldsymbol{A}) = 2$.

Rank

Example

$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \sim \dots \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Here, we see that the number of linearly independent rows and columns is 2 , such that $\mathrm{rk}(\boldsymbol{A}) = 2$.

Linear Mappings

Definition

For vector spaces V, W, a mapping $\Phi : V \to W$ is called a linear mapping (or vector space homomorphism/linear transformation) if

$$\forall \boldsymbol{x}, \boldsymbol{y} \in V \, \forall \lambda, \psi \in \mathbf{R} : \Phi(\lambda \boldsymbol{x} + \psi \boldsymbol{y}) = \lambda \Phi(\boldsymbol{x}) + \psi \Phi(\boldsymbol{y})$$

It turns out that we can represent linear mappings as matrices (Section 2.7.1 Definition 2.19).

Linear Mappings

Definition

For vector spaces V, W, a mapping $\Phi : V \to W$ is called a linear mapping (or vector space homomorphism/linear transformation) if

$$\forall \boldsymbol{x}, \boldsymbol{y} \in V \,\forall \lambda, \psi \in \mathbf{R} : \Phi(\lambda \boldsymbol{x} + \psi \boldsymbol{y}) = \lambda \Phi(\boldsymbol{x}) + \psi \Phi(\boldsymbol{y})$$

It turns out that we can represent linear mappings as matrices (Section 2.7.1 Definition 2.19).

Linear Mappings

Definition

(Injective, Surjective, Bijective). Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}, \mathcal{W}$ can be arbitrary sets. Then Φ is called:

- Injective if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y}$.
- Surjective if $\Phi(\mathcal{V}) = \mathcal{W}$.
- Bijective if it is injective and surjective.

Linear Mappings

Definition

(Injective, Surjective, Bijective). Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}, \mathcal{W}$ can be arbitrary sets. Then Φ is called:

- Injective if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \Longrightarrow \boldsymbol{x} = \boldsymbol{y}$.
- Surjective if $\Phi(\mathcal{V}) = \mathcal{W}$.
- Bijective if it is injective and surjective.

Linear Mappings

Definition

(Injective, Surjective, Bijective). Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}, \mathcal{W}$ can be arbitrary sets. Then Φ is called:

- Injective if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \Longrightarrow \boldsymbol{x} = \boldsymbol{y}$.
- Surjective if $\Phi(\mathcal{V}) = \mathcal{W}$.
- Bijective if it is injective and surjective.

Linear Mappings

Definition

(Injective, Surjective, Bijective). Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}, \mathcal{W}$ can be arbitrary sets. Then Φ is called:

- Injective if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y}$.
- Surjective if $\Phi(\mathcal{V}) = \mathcal{W}$.
- Bijective if it is injective and surjective.

Affine Subspaces

In the following, we will have a closer look at spaces that are offset from the origin, i.e., spaces that are no longer vector subspaces. Moreover, we will briefly discuss properties of mappings between these affine spaces, which resemble linear mappings.

Remark. In the machine learning literature, the distinction between linear and affine is sometimes not clear so that we can find references to affine spaces/mappings as linear spaces/mappings.

Affine Subspaces

In the following, we will have a closer look at spaces that are offset from the origin, i.e., spaces that are no longer vector subspaces. Moreover, we will briefly discuss properties of mappings between these affine spaces, which resemble linear mappings.

Remark. In the machine learning literature, the distinction between linear and affine is sometimes not clear so that we can find references to affine spaces/mappings as linear spaces/mappings.

Affine Subspace

Definition

Let V be a vector space, $\boldsymbol{x}_0 \in V$ and $U \subseteq V$ a subspace, then the subset

$$\begin{aligned} L = \boldsymbol{x}_0 + U &:= \{\boldsymbol{x}_0 + \boldsymbol{u} : \boldsymbol{u} \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called affine subspace or linear manifold of V.

U is called direction or direction space, and $\boldsymbol{x}_0$ is called support point. In Chapter 12, we refer to such a subspace as a hyperplane.
Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not a (linear) subspace (vector subspace) of V for $x_0 \notin U$.

Affine Subspace

Definition

Let V be a vector space, $\boldsymbol{x}_0 \in V$ and $U \subseteq V$ a subspace, then the subset

$$\begin{aligned} L = \boldsymbol{x}_0 + U &:= \{\boldsymbol{x}_0 + \boldsymbol{u} : \boldsymbol{u} \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called affine subspace or linear manifold of V.

U is called direction or direction space, and $\boldsymbol{x}_0$ is called support point. In Chapter 12, we refer to such a subspace as a hyperplane.

Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not a (linear) subspace (vector subspace) of V for $x_0 \notin U$.

Affine Subspace

Definition

Let V be a vector space, $\boldsymbol{x}_0 \in V$ and $U \subseteq V$ a subspace, then the subset

$$\begin{aligned} L = \boldsymbol{x}_0 + U &:= \{\boldsymbol{x}_0 + \boldsymbol{u} : \boldsymbol{u} \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called affine subspace or linear manifold of V.

U is called direction or direction space, and $\boldsymbol{x}_0$ is called support point. In Chapter 12, we refer to such a subspace as a hyperplane.

Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not a (linear) subspace (vector subspace) of V for $x_0 \notin U$.

Affine Subspace

Definition

Let V be a vector space, $\boldsymbol{x}_0 \in V$ and $U \subseteq V$ a subspace, then the subset

$$\begin{aligned} L = \boldsymbol{x}_0 + U &:= \{\boldsymbol{x}_0 + \boldsymbol{u} : \boldsymbol{u} \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called affine subspace or linear manifold of V.

U is called direction or direction space, and $\boldsymbol{x}_0$ is called support point. In Chapter 12, we refer to such a subspace as a hyperplane.

Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not a (linear) subspace (vector subspace) of V for $x_0 \notin U$.

Affine Subspace

Definition

Let V be a vector space, $\boldsymbol{x}_0 \in V$ and $U \subseteq V$ a subspace, then the subset

$$L = \boldsymbol{x}_0 + U := \{\boldsymbol{x}_0 + \boldsymbol{u} : \boldsymbol{u} \in U\}$$
$$= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{u}\} \subseteq V$$

is called affine subspace or linear manifold of V.

U is called direction or direction space, and $\boldsymbol{x}_0$ is called support point. In Chapter 12, we refer to such a subspace as a hyperplane.
Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not a (linear) subspace (vector subspace) of V for $x_0 \notin U$.

Affine Subspace

Examples of affine subspaces are points, lines, and planes in $\mathbb{R}^3$, which do not (necessarily) go through the origin. Remark. Consider two affine subspaces $L = \boldsymbol{x}_0 + U$ and $\tilde{L} = \tilde{\boldsymbol{x}}_0 + \tilde{U}$ of a vector space V. Then, $L \subseteq \tilde{L}$ if and only if $U \subseteq \tilde{U}$ and $\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_0 \in \tilde{U}$

Affine Subspace

Affine subspaces are often described by parameters: Consider a k-dimensional affine space $L = x_0 + U$ of V. If $(b_1, \dots, b_k)$ is an ordered basis of U, then every element $\boldsymbol{x} \in L$ can be uniquely described as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_k \boldsymbol{b}_k$$

where $\lambda_1, \dots, \lambda_k \in \mathbb{R}$. This representation is called parametric equation of L with directional vectors $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ and parameters $\lambda_1, \dots, \lambda_k$.

Affine Subspace

Affine subspaces are often described by parameters: Consider a k-dimensional affine space $L = x_0 + U$ of V. If $(b_1, \ldots, b_k)$ is an ordered basis of U, then every element $\boldsymbol{x} \in L$ can be uniquely described as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_k \boldsymbol{b}_k$$

where $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. This representation is called parametric equation of L with directional vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ and parameters $\lambda_1, \ldots, \lambda_k$.

Affine Subspace

Affine subspaces are often described by parameters: Consider a k-dimensional affine space $L = x_0 + U$ of V. If $(b_1, \ldots, b_k)$ is an ordered basis of U, then every element $\boldsymbol{x} \in L$ can be uniquely described as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_k \boldsymbol{b}_k$$

where $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. This representation is called parametric equation of L with directional vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ and parameters $\lambda_1, \ldots, \lambda_k$.

Affine Subspace

Example

One-dimensional affine subspaces are called lines and can be written as $\boldsymbol{y} = \boldsymbol{x}_0 + \lambda \boldsymbol{x}_1$, where $\lambda \in \mathbb{R}$, where $U = \operatorname{span}[\boldsymbol{x}_1] \subseteq \mathbb{R}^n$ is a one-dimensional subspace of $\mathbb{R}^n$. This means that a line is defined by a support point x_0 and a vector x_1 that defines the direction.

Thanks for your attention!