Computational Linear Algebra

Prof. Phani Motamarri

Department of Computational and Data Sciences

Indian Institute of Science, Bangalore-560012

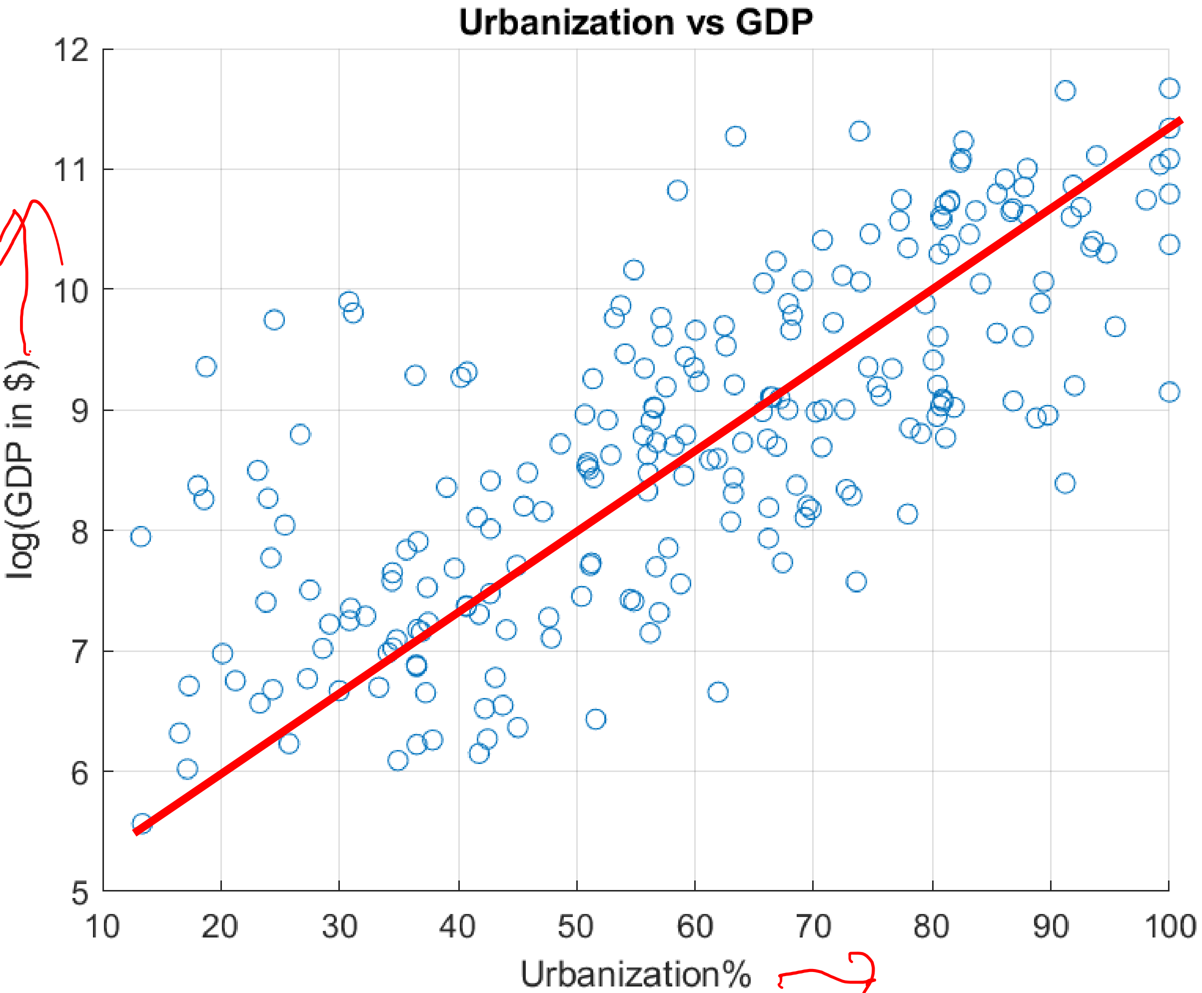
Computational Linear Algebra in Data Science

Motivation

$y = w_1 x + w_0$

$y_1^m = w_1 x_1 + w_0$

$y_2^m = w_1 x_2 + w_0$



URB(%)	GDP
18.59	8.26
23.98	8.27
30.95	7.35
34.47	7.65
36.50	6.87
40.71	7.38
42.71	7.48
51.37	9.26
54.08	9.47
56.75	8.73
58.52	10.82
60.31	9.24
.	.
..	..
...	...
70.79	9.00
73.85	11.31
77.22	10.57
80.80	10.58
84.07	10.05
87.64	9.61
89.74	8.96
93.36	10.35
100.00	11.67

Independent Data/Features:

Input data: % Urbanisation x

Dependent Data/Target: y

Output data: GDP

$$\varepsilon = \frac{1}{2}\sum_{i=1}^{i=N}(y_i - \underbrace{w^1 x_i - w^0}_{y_i^m})^2$$

Minimising the error

$$\frac{d\varepsilon}{dw^1} = \sum_{i=1}^{i=N}(y_i - w^1 x_i - w^0)x_i = 0$$

$$\frac{d\varepsilon}{dw^0} = \sum_{i=1}^{i=N} -(y_i - w^1 x_i - w^0) = 0$$

CDS Department of Computational and Data Sciences

Why Linear Algebra ?

Motivation

$w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$

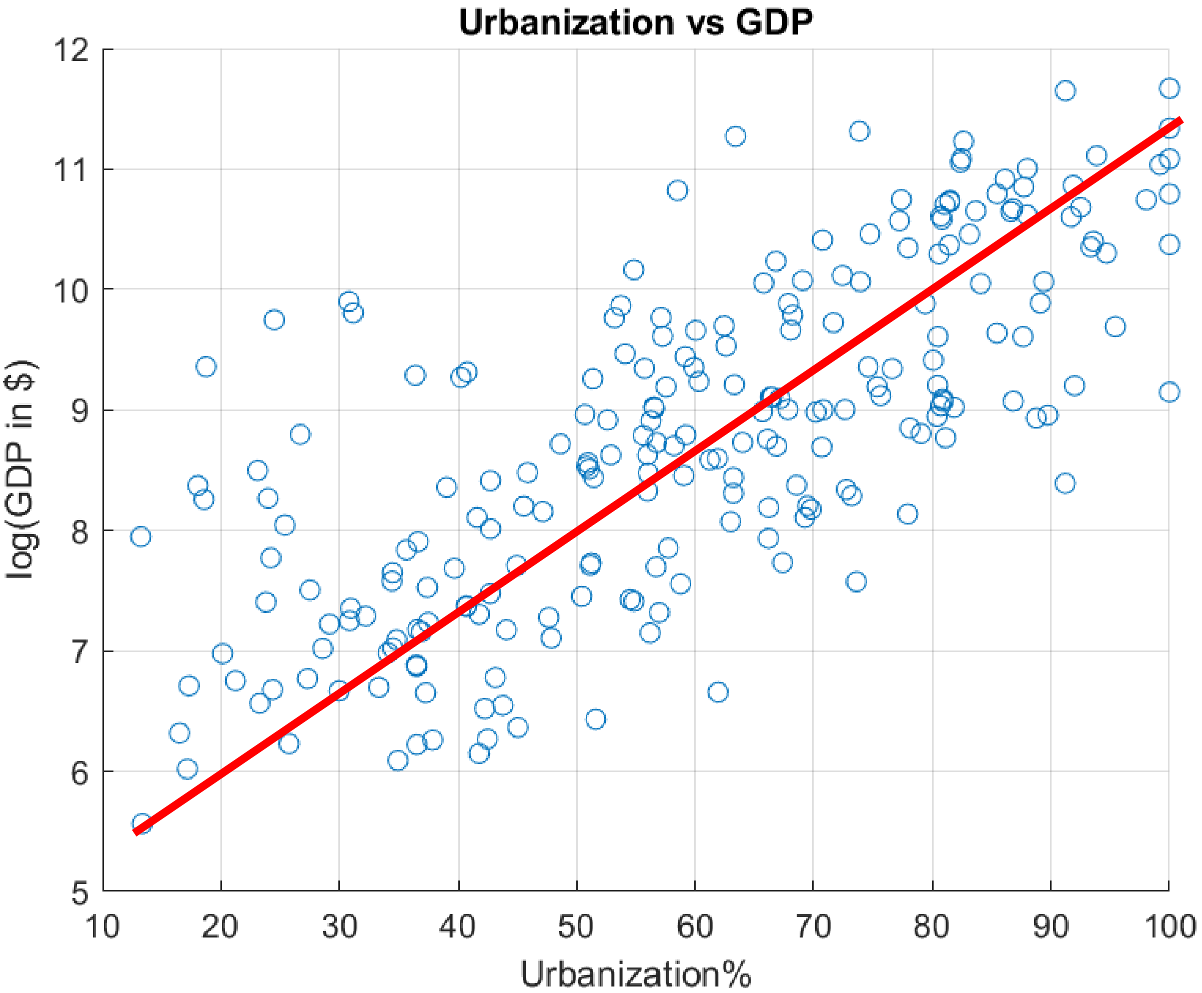
But

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix} \begin{bmatrix} w_1 \\ w_0 \end{bmatrix} \approx \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}$$

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix} \begin{bmatrix} w_1 \\ w_0 \end{bmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

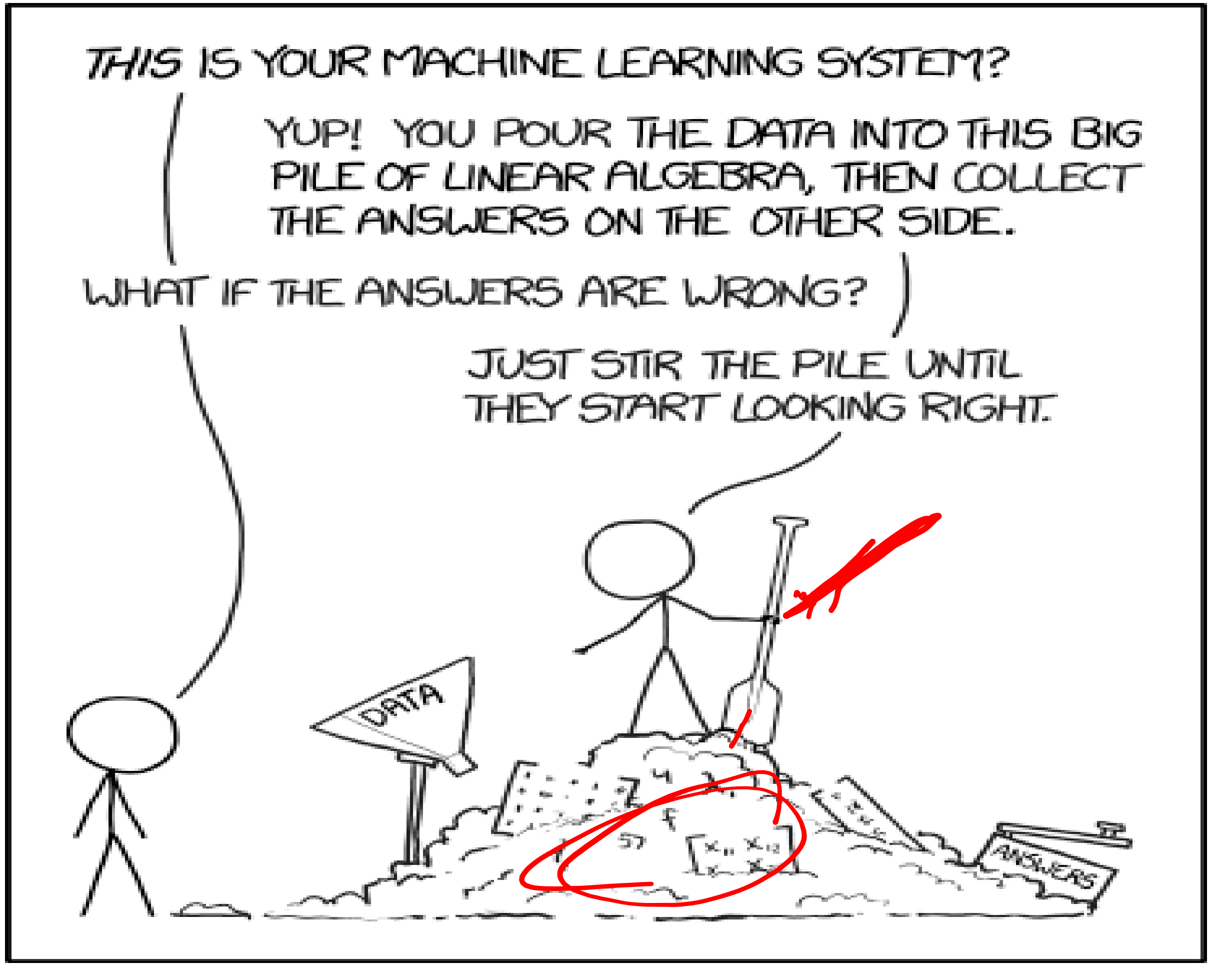
$$\boxed{\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{Xw})}$$

Minimize $\boldsymbol{\varepsilon}$



Computational Linear Algebra in Data Science

Why Linear Algebra ?



https://imgs.xkcd.com/comics/machine_learning.png

Course Outline:

1. Properties of Vectors, Matrices ,Tensors and the operations between them.
2. Algorithms in Computational Linear Algebra that are frequently encountered in Data Science/ML

Learning Objectives:

1. Understand foundational aspects of vectors, Matrices, operations between them and their relevance to Data Science.
2. Be able to understand the theory behind the many of the ML algorithms from a linear algebra perspective.

Vectors Matrices and Tensors

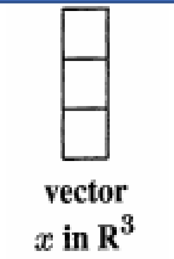
Vectors Matrices and Tensors

Vectors

An arrow that holds information of magnitude by its length and direction by the direction of the arrow ($\mathbb{R}^3$)

$$\begin{bmatrix} 5 \\ 4 \\ -2 \\ \dots \\ 1 \end{bmatrix}$$

A sequence of numbers whose order is important that can represent input or output features (N dimensional Array)



Class Notation: We will be representing vectors with bold lower-case alphabets.: $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$

$\mathbb{R}^3$

Eg:- $\begin{bmatrix} x \\ x \\ x \end{bmatrix}$ R G B

$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $n = 100$

GDP $n \geq 3$ 100

$y \in \mathbb{R}^{100}$

$y = \begin{bmatrix} B.P_1 \\ \vdots \\ B.P_{1000} \end{bmatrix}$

$y \in \mathbb{R}^{1000}$

$h = \begin{bmatrix} x & x \\ x & x \\ x & x \\ x & x \\ x & x \\ x & x \end{bmatrix}$

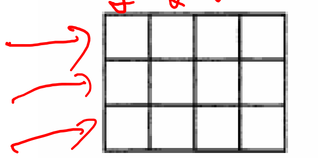
$h \in \mathbb{R}^{?}$

Vectors Matrices and Tensors

Matrices

$$\begin{bmatrix} 1 & 4 & 2.2 & 5 \\ 5 & 2 & 5 & -9 \\ -\cos(\alpha) & 0 & 0.5 & 1 \end{bmatrix} \; 3\times 4$$

- An organized structure of numbers that holds data of various labels
- Transforms vectors
- $N \times M$ dimensional array



matrix
A in $\mathbb{R}^{3 \times 4}$

Class Notation: We will be representing matrices with bold upper-case alphabets.: $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$

GDP

% urbanisation

Life expect

Education Index

$\mathbb{R}^n$

$m \to$ rows

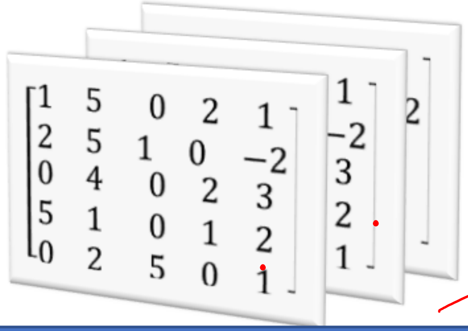
$n \to$ cols

$\mathbb{R}^{m \times n}$

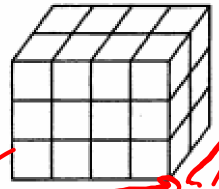
$\mathbb{R}^{3\times 4}$

Vectors Matrices and Tensors

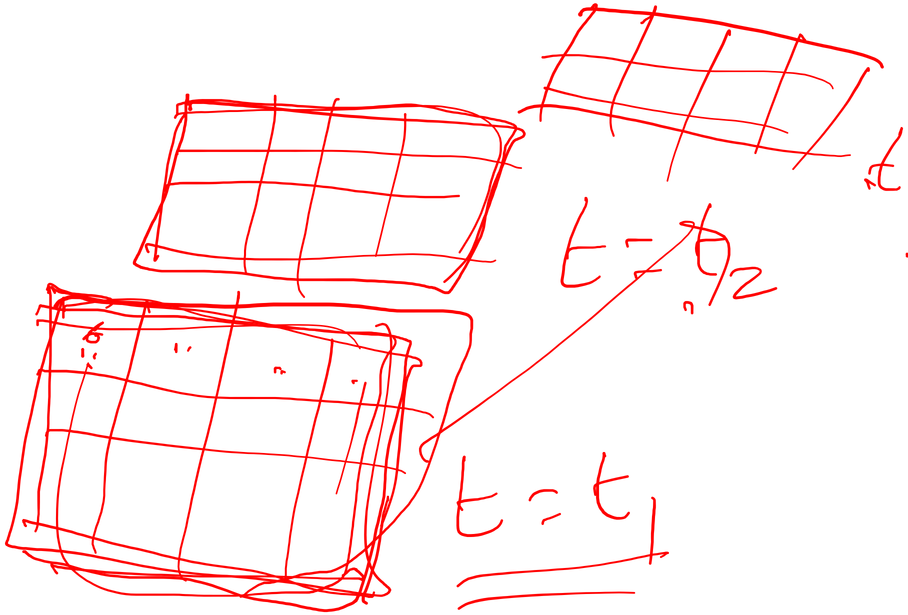
Tensors



- Higher dimension matrices
- Transforms lower dimension tensors
- $N \times M \times L$ dimensional array



tensor
T in $\mathbf{R}^{3 \times 4 \times 2}$

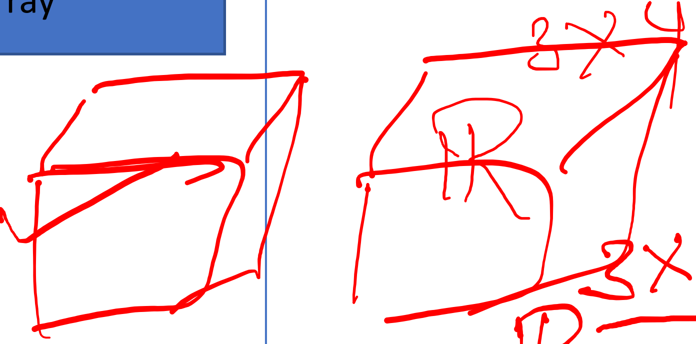


$t = t_3$

$t = t_2$

$t = t_1$

0
1



3×4

$\mathbb{R}$

$\to$ 2 slices

$3 \times 4 \times 2 \times 3$

$\mathbb{R}$

vector $\to$ 1st order

matrix $\to$ 2nd order

Tensor $\to$ 3rd order, 4th

Class Notation: We will be representing tensors with bold upper-case alphabets.: $\underline{\boldsymbol{A}}, \underline{\boldsymbol{B}}, \underline{\boldsymbol{C}}$

Vectors and Its Operations

Vector Space, Linear Independence, Basis vectors

$F =$ Real, Complex

- **Vector space:**

A vector space consists of a set $\boldsymbol{V}$ (elements of $\boldsymbol{V}$ are called vectors), a field F (elements of F are called scalars), and two operations

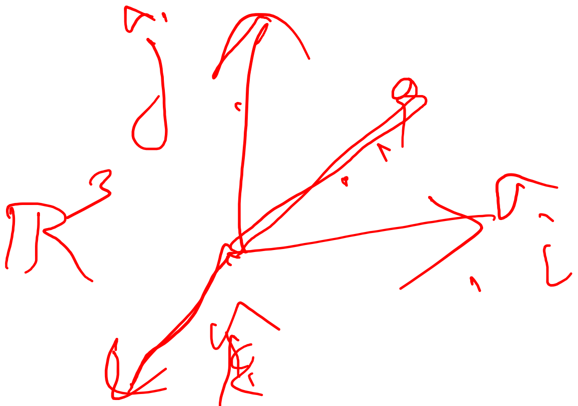
- An operation called vector addition takes two vectors $\boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{V}$, and produces a third vector, written $\boldsymbol{v} + \boldsymbol{w} \in \boldsymbol{V}$.
- An operation called scalar multiplication takes a scalar $c \in F$ and a vector $\boldsymbol{v} \in \boldsymbol{V}$, and produces a new vector, written $c\boldsymbol{v} \in \boldsymbol{V}$.

$\mathbb{R}^2$ $\qquad \underline{v}, \underline{w} \in V$ $\qquad v \in V \quad c\underline{v} \in V$

$\mathbb{R}^3 \rightarrow \underline{v} + \underline{w} = \underline{u}$

Vectors and Its Operations

Vector Space, Linear Independence, Basis vectors



- **Linear Independence and Dependence:**
  - If $\{\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots, \boldsymbol{v}_n\}$ are vectors in $\mathbb{R}^n$, then they are **linearly independent** if and only if

$$\sum_{i=1}^{i=N} \alpha^i \boldsymbol{v}_i = 0 \Rightarrow \alpha^i = 0$$

  - If $\boldsymbol{v}, \boldsymbol{w}$ are distinct vectors in $\mathbb{R}^2$ the linear combination represents any vector in the 2D-plane.

$\alpha_1 \hat{i} + \alpha_2 \hat{j} + \alpha_3 \hat{k} = \vec{v}$

- **Basis Vectors:**
  - Set of vectors in a given space, that are linearly independent which are used to represent any vector in that space. *E.g.*: $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are linearly independent and span the entire 3D space.

$V_1 \quad V_2 \quad V_3$

$\alpha_1 v_1 + \alpha_2 V_2 + \cdots \alpha_n V_n = 0$

$V_3 - V_2 = V_1$

$\alpha_n = 0$

Age ~~Date~~ Year of Birth ~~Date~~ of ... Education Per Capita

Current

Vectors and Its Operations

Dot Product/Inner Product

$$\boldsymbol{a}.\boldsymbol{b} = \boldsymbol{a}^T\boldsymbol{b}$$

e.g.: $(2 \quad 3 \quad 5)\begin{pmatrix} a \\ b \\ c \end{pmatrix} = 2a + 3b + 5c$

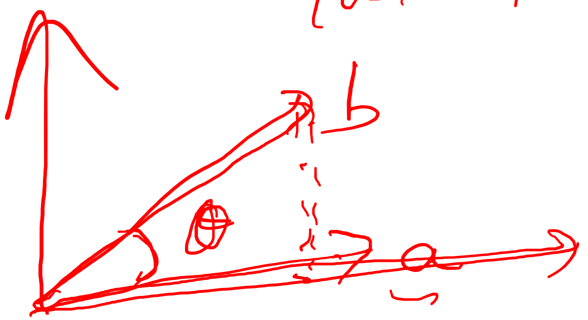
- **Properties of Inner/Dot Product**:
  - $\alpha(\boldsymbol{a}^T\boldsymbol{b}) = (\alpha\boldsymbol{a})^T\boldsymbol{b}$
  - $(\boldsymbol{a}+\boldsymbol{b})^T\boldsymbol{c} = \boldsymbol{a}^T\boldsymbol{c} + \boldsymbol{b}^T\boldsymbol{c}$
  - $\boldsymbol{a}^T\boldsymbol{a} \geq 0, equality\ holds\ if\ \boldsymbol{a} = \boldsymbol{0}$
- **Length of vector:**
  - $|\boldsymbol{a}| = \sqrt{\boldsymbol{a}^T\boldsymbol{a}}$
- **Angle between vectors:**
  - $\boxed{\cos(\theta) = \frac{\boldsymbol{a}.\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \frac{\sum_1^N a_i b_i}{|\boldsymbol{a}||\boldsymbol{b}|}}$

$|a| = \sqrt{a^T a}$
$= \sqrt{a \cdot a}$

$a = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}_{3\times 1}$ $\quad b = \begin{bmatrix} a \\ b \\ c \end{bmatrix}_{3\times 1}$

$\boxed{a \cdot b} = a^T b$
$= 1\times 3 \ 3\times 1 = \underline{1\times 1}$

$\cos\theta = \frac{a \cdot b}{|a|\,|b|}$



Vectors and Its Operations

Outer Product

$$\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^T$$

- $\begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} (a \quad b \quad c) = \begin{pmatrix} 2a & 2b & 2c \\ 3a & 3b & 3c \\ 5a & 5b & 5c \end{pmatrix}$
- The columns of $\boldsymbol{u}\boldsymbol{v}^T$ are multiples of $\boldsymbol{u}$. The rows of $\boldsymbol{u}\boldsymbol{v}^T$ are multiples of **v**
- **Only one linearly independent row**
- **These kind of matrices arise in Singular Value Decomposition.**

$a \cdot b = a^T b$

$a \otimes b = a b^T$

Vectors and Its Operations

Vector Norms

$\|x\|$

2 $\;$ 1

- Norm of a vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|$
- Norm is like absolute value, provides a metric for distance, closeness between vectors and often used in setting convergence criteria.
- A vector norm on $\mathbb{R}^n$ is a function that maps to real numbers. $f\colon\ \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies for $\boldsymbol{x}, \boldsymbol{y}\ \in\ \mathbb{R}^n\ and\ \alpha \in \mathbb{R}$:

$$\|\boldsymbol{x}\| \geq 0, \|\boldsymbol{x}\| = 0\ iff\ \boldsymbol{x} = 0$$
$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$
$$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$$

ε

$\|x\| \geq 0$

$\|\alpha x\| = |\alpha| \|x\|$

- Recall… goal of regression was to minimise vector $\boldsymbol{\varepsilon}$

Vectors and Its Operations

Vector Norms

- P-norms:
  - L1-Norm: $\|\boldsymbol{x}\|_1 = \sum |x_i|$
  - L2-Norm: $\|\boldsymbol{x}\|_2 = (\sum {x_i}^2)^{1/2}$
  - L-p Norm: $\|\boldsymbol{x}\|_p = (\sum {x_i}^p)^{1/p}$
  - L-inf Norm: $\|\boldsymbol{x}\|_\infty = \max(|x_i|)$

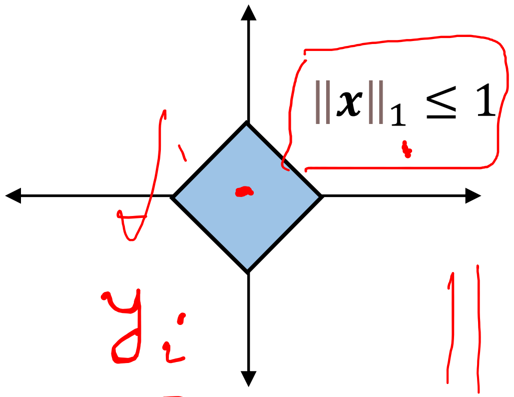
$x = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ \vdots \end{bmatrix}$

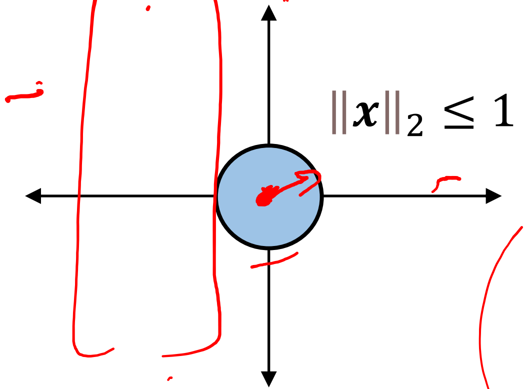
$\|\varepsilon\|_2$

$\sqrt{\varepsilon^T \varepsilon}$

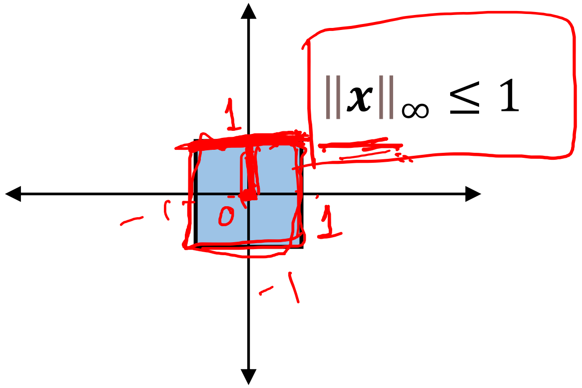
$y_m \quad y_i$

$\|x\|_2 \leq 1$





$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$



(x_1, x_2)

Vectors and Its Operations

$y = w_0 + w_1 x + w_2 x^2$

Overfitting, Norms & Regularization

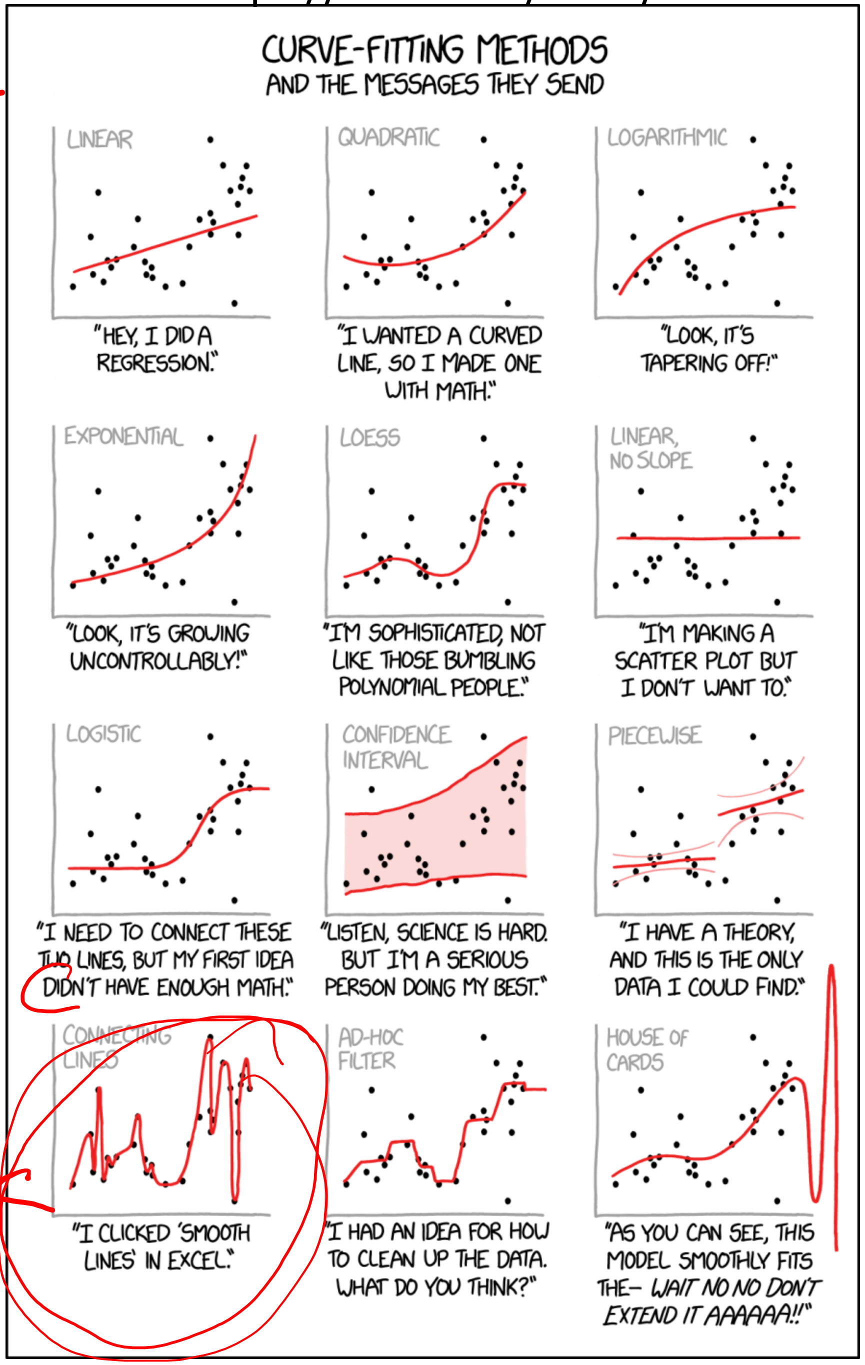
- Goal: Choose the one solution of the many that generalises well to unseen data.
- Overfitting: When the learning model explains the given data set but cannot generalise unseen data.
- Regularisation: Constraining parameters to reduce overfitting.
- Lasso Regression: (Least Absolute Shrinkage and Selection Operator) adds magnitudes of coefficients to the loss function. **Least important features magnitude reduces to 0**

$$L = ||(\boldsymbol{y} - \boldsymbol{w}^T \boldsymbol{X})||_2^2 + \lambda||\boldsymbol{w}||_1$$

- Ridge Regression: Square of magnitude of coefficient is added as penalty to Loss function

$$L = ||(\boldsymbol{y} - \boldsymbol{w}^T \boldsymbol{X})||_2^2 + \lambda||\boldsymbol{w}||_2$$

https://xkcd.com/2048/



Matrices and Its Operations

Common Types of Square Matrices

- Symmetric Matrices: $S_{ij} = S_{ji}$; $\boldsymbol{S} = \begin{bmatrix} -2 & 5 & 3 \\ 5 & 5 & 1 \\ 3 & 1 & 2 \end{bmatrix}$
- Orthogonal Matrices: $\boldsymbol{q}_i . \boldsymbol{q}_j = \delta_{ij}$; $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$; $\boldsymbol{Q} = \begin{bmatrix} -0.5145 & -0.7576 & 0.4016 \\ 0.8575 & -0.4546 & 0.2410 \\ 0 & 0.4684 & 0.8835 \end{bmatrix}$
- Triangular Matrices: $U_{i<j} = 0$; $\boldsymbol{U} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 2 \\ 0 & 0 & 5 \end{bmatrix}$
- Diagonal Matrices: $D_{i\neq j} = 0$; $\boldsymbol{D} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$;

$S_{ij} = S_{ji}$ $\boxed{S = S^T}$

$\delta_{ij} = 0 \quad i \neq j$

$= 1 \quad i = j$

$\cos 90$

0

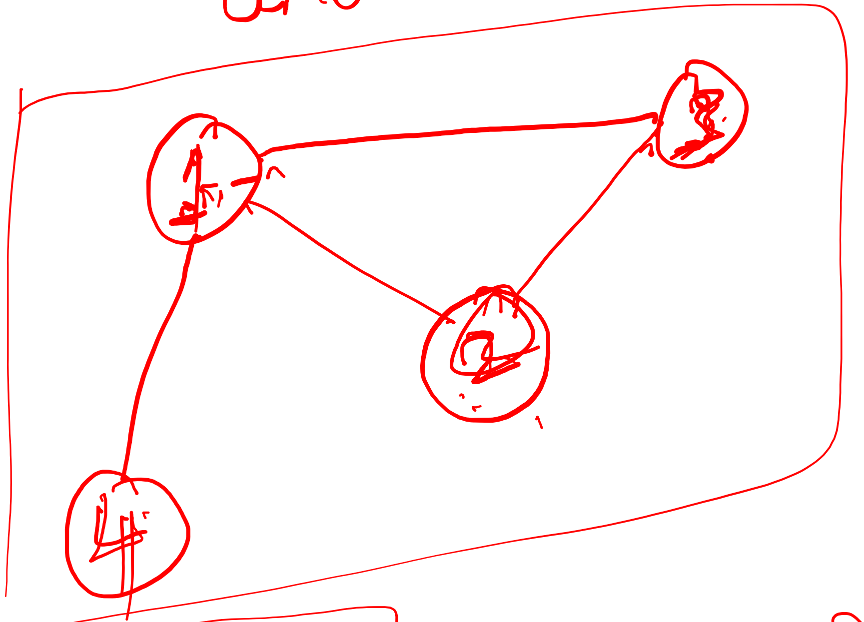
$a \cdot b = a^T b$

$\cos\theta = \dfrac{a \cdot b}{|a||b|}$

90

$a \cdot b = 0$

Undirected



$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$\boxed{A = A^T}$$

A_{ij} $i^{th} \rightarrow j^{th}$ node — 1,

2 3 — 0

Q $\boxed{Q^T Q = I}$

$$\boxed{\begin{array}{l} q_i \cdot q_j = 0 \\ q_i \cdot q_i = 1 \end{array}}$$

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

$$D = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} f(v_1, v_1) & f(v_1, v_2) & f(v_1, v_3) \\ f(v_2, v_1) & f(v_2, v_2) & f(v_2, v_3) \\ f(v_3, v_1) & f(v_3, v_2) & f(v_3, v_3) \end{bmatrix}$$

Matrices and Its Operations

Matrix Vector Multiplication

$$Ax = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_n] \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

m×1 m×1 m×1

$A_{m \times n}$ x $n \times 1$

$y = Ax$

Row Method

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

1. The result is a list of inner product: (dot product of row vector and column vector)

Column Method

A $y = Ax$

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

1. Linear Combination of columns of Matrix.
2. Vectorial approach
3. $\boldsymbol{b} = \sum x_i \boldsymbol{a}_i$

Matrices and Its Operations

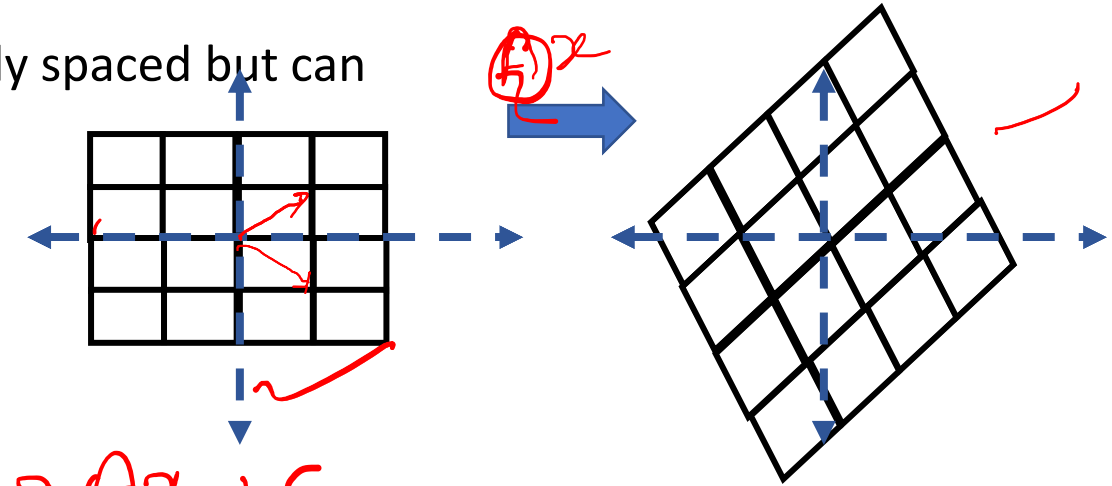
Linear Transformation

- **Transformation**: a function that takes in vector and outputs another vector

$T(x) = u$

- **Linear Transformation**:
  - Does not curve lines, equally spaced meshes are still equally spaced but can be rotated or stretched.

$$\boldsymbol{x} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \longrightarrow \boxed{\boldsymbol{Ax}} \longrightarrow \boldsymbol{Ax} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$



$Ax = y$

$Qx = y$

$y = Ax + C$

$A(x + y) = Ax + Ay$

$A(\alpha x) = \alpha A x$

$\|x + y\| \leq \|x\| + \|y\|$

Matrices and Its Operations

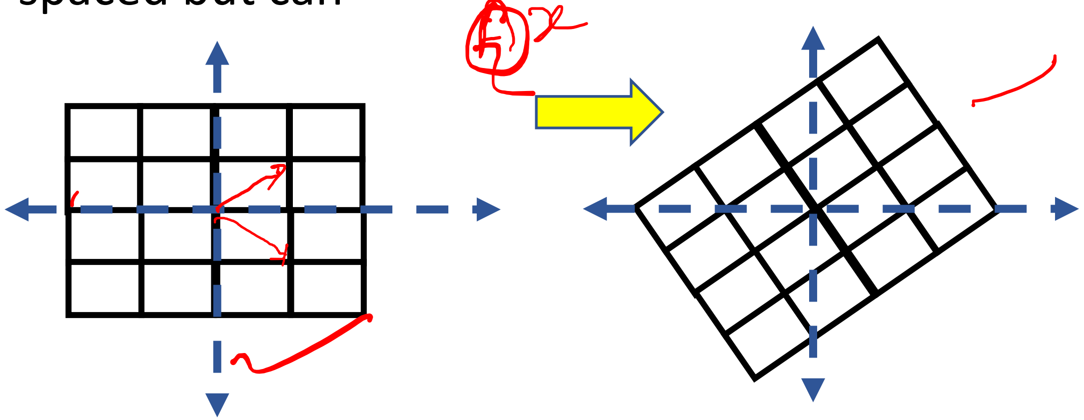
Linear Transformation

- **Transformation**: a function that takes in vector and outputs another vector

$T(v) = u$

- **Linear Transformation**:
  - Does not curve lines, equally spaced meshes are still equally spaced but can be rotated or stretched.

$$\boldsymbol{x} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \longrightarrow \boxed{\boldsymbol{Ax}} \longrightarrow \boldsymbol{Ax} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$



$Ax = y$

  - Action of orthogonal matrix on a vector rotates the vector

$Qx = y$

$y = Ax + c$

- **Matrix is a linear map between** $\mathbb{R}^m \rightarrow \mathbb{R}^n$
  - $\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{Ax} + \boldsymbol{Ay}$
  - $\boldsymbol{A}(\alpha \boldsymbol{x}) = \alpha \boldsymbol{Ax}$

$A(x+y) = Ax + Ay$

$A(\alpha x) = \alpha Ax$

$\|x+y\| \leq \|x\| + \|y\|$

Matrices and Its Operations

Matrix-Matrix Multiplication

$$\boldsymbol{XY} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_n][\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \cdots \quad \boldsymbol{y}_n]$$

Inner Product form

$$\underset{\boldsymbol{X}}{\begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 7 \end{bmatrix}} \underset{\boldsymbol{Y}}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} = \underset{\boldsymbol{Z}}{\begin{bmatrix} \boxed{2a+3c} & 2b+3d \\ 3a+4c & 3a+4b \\ 4a+7c & \boxed{4a+7d} \end{bmatrix}}$$

$$Z_{ij} = \sum_k x_{ik} y_{kj}$$

- $z_{ij} = \sum x_{ik} y_{kj}$

Linear Combination of columns

$$B = A\,C$$

$$[\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \boldsymbol{b}_3 \quad \cdots \quad \boldsymbol{b}_n] = \underbrace{[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_m]}_{A}[\boldsymbol{c}_1 \quad \boldsymbol{c}_2 \quad \boldsymbol{c}_3 \quad \cdots \quad \boldsymbol{c}_n]$$

Thus, the columns of $\boldsymbol{B}$ can be written as linear combinations of columns of $\boldsymbol{A}$

$b_1 =$

$$\boldsymbol{b}_j = \sum_{k=1}^{m} c_{kj}\boldsymbol{a}_k$$

Different interpretation…

Matrices and Its Operations

Matrix-Matrix Multiplication

Outer Product form

$$XY = [x_1 \quad x_2 \quad \cdots \quad x_n]\begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_n^T \end{bmatrix}$$

$$\boxed{XY = x_1 y_1{}^T + x_2 y_2{}^T + \ldots.\ x_n y_n{}^T}$$

- x_i are the vectors associated with the columns of the X matrix
- y_i are the vectors associated with the rows of the Y matrix
- Sum of rank 1 matrices formed by outer product of columns of X and rows of Y
- **Watch out during SVD…**

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

$$\begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 17 \end{bmatrix}$$

Different interpretation…

Matrices and Its Operations

Range, Rank, Null Space

$A x = 0$

$A x = 0$

$A x = y$

$A x = y$

- **Range of A :**
  - Range(A) is the set of all vectors that can be expressed as $\boldsymbol{Ax}$
  - Range(A) is the space spanned by columns of $\boldsymbol{A}$

$$[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_m]\begin{bmatrix}x_1\\ x_2\\ \cdots\\ x_m\end{bmatrix} = \sum_{i=1}^{i=m} x_i \boldsymbol{a}_i$$

$A =$

$n >> 3$ $\quad$ $n \times 3$

$(m \times n)$

$rank(A) = min(m, n)$

- **Rank of A :**
  - Column rank is the no. of linearly independent column vectors of $\boldsymbol{A}$ (dimension of columns space)
  - A matrix $\boldsymbol{A}$ with $m \geq n$ has a full $rank = n$, if and only if columns are linearly independent
- **Null space of A :**
  - null($\boldsymbol{A}$) is the set of all $\boldsymbol{x}$ that satisfy $\boxed{\boldsymbol{Ax} = \boldsymbol{0}}$
  - If null($\boldsymbol{A}$) ≠ 0, then $\boldsymbol{A}$ has linearly dependent columns.

Matrices and Its Operations

$x + 2y = 3$
$4x + 5y = 6$

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

$A \; x = b$

$y = w_0 + w_1 x$

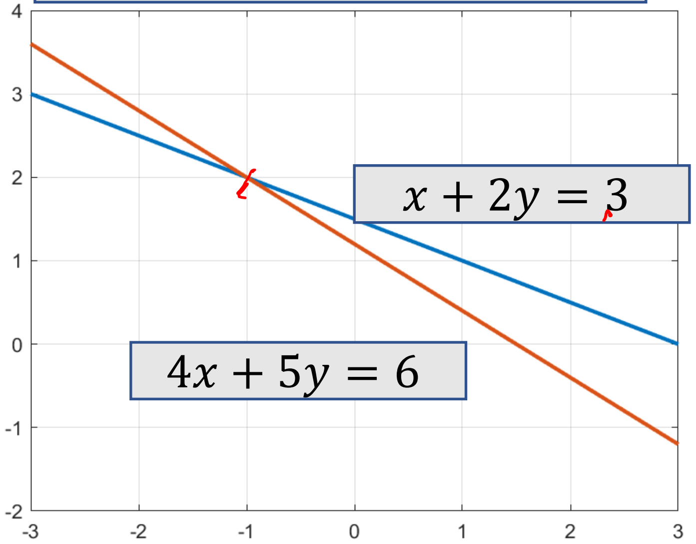
Range, Rank, Null Space

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$$
$$\boldsymbol{b} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$
$$\boldsymbol{a}_2 \neq c\boldsymbol{a}_1, c \in \mathbb{R}$$

Full Rank

$\boldsymbol{b} \in range(\boldsymbol{A})$

One Solution

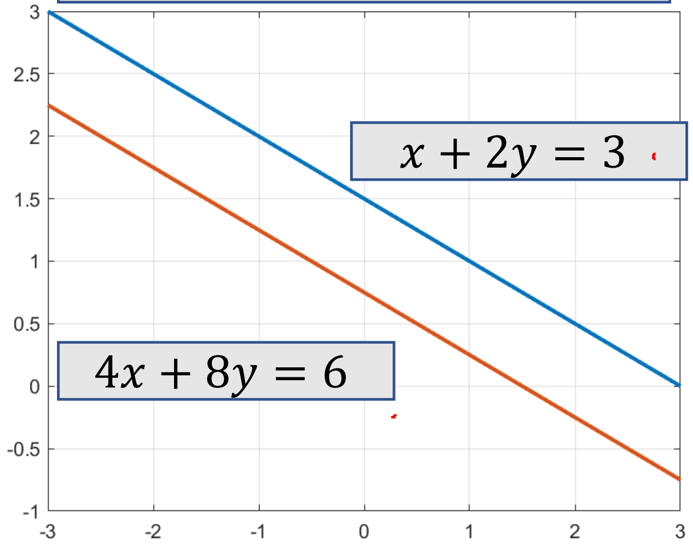


$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 4 & 8 \end{pmatrix}$$
$$\boldsymbol{b} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$
$$\boldsymbol{a}_2 = 2\boldsymbol{a}_1$$

Not Full Rank

$\boldsymbol{b}$ *is not in* $range(\boldsymbol{A})$

Parallel: No solution

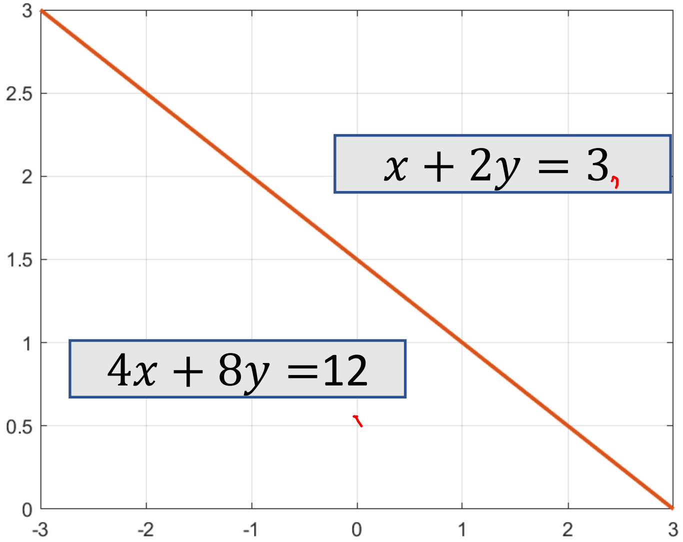


$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 4 & 8 \end{pmatrix}$$
$$\boldsymbol{b} = \begin{pmatrix} 3 \\ 12 \end{pmatrix}$$
$$\boldsymbol{a}_2 = 2\boldsymbol{a}_1$$

Not Full Rank

$\boldsymbol{b} \in range(\boldsymbol{A})$

Whole line of solution



Matrices and Its Operations

Matrix Inversion → Full rank properties and matrix-matrix multiplication

$A_{m \times m}$

A^{-1}

$\boxed{AA^{-1} = I}$

- Non-singular/invertible matrix is a full rank square matrix $m \times m$
- The m columns of such a matrix are linearly independent
- Implies, we can write any vector in $\mathbb{R}^m$ as a unique linear combination of column vectors of $\boldsymbol{A}$. We can represent the canonical basis vector $\boldsymbol{e_j}$ as unique linear combination of $\boldsymbol{A}$
- $\boldsymbol{e_j} = \sum z_{ij} \boldsymbol{a_j}$
- $\boldsymbol{I} = \boldsymbol{AZ} \quad \boldsymbol{Z} = \boldsymbol{A}^{-1}$

$\mathbb{R}^m$

$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_m \\ | & | & & | \end{bmatrix}_{m \times m}$

$$\boldsymbol{e}_j = \begin{array}{cl} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \boxed{1} \\ \vdots \\ 0 \end{bmatrix} & \begin{array}{l} \\ \\ \\ \\ \leftarrow j^{th} \text{ location} \\ \\ \\ \end{array} \end{array}$$

Matrices and Its Operations

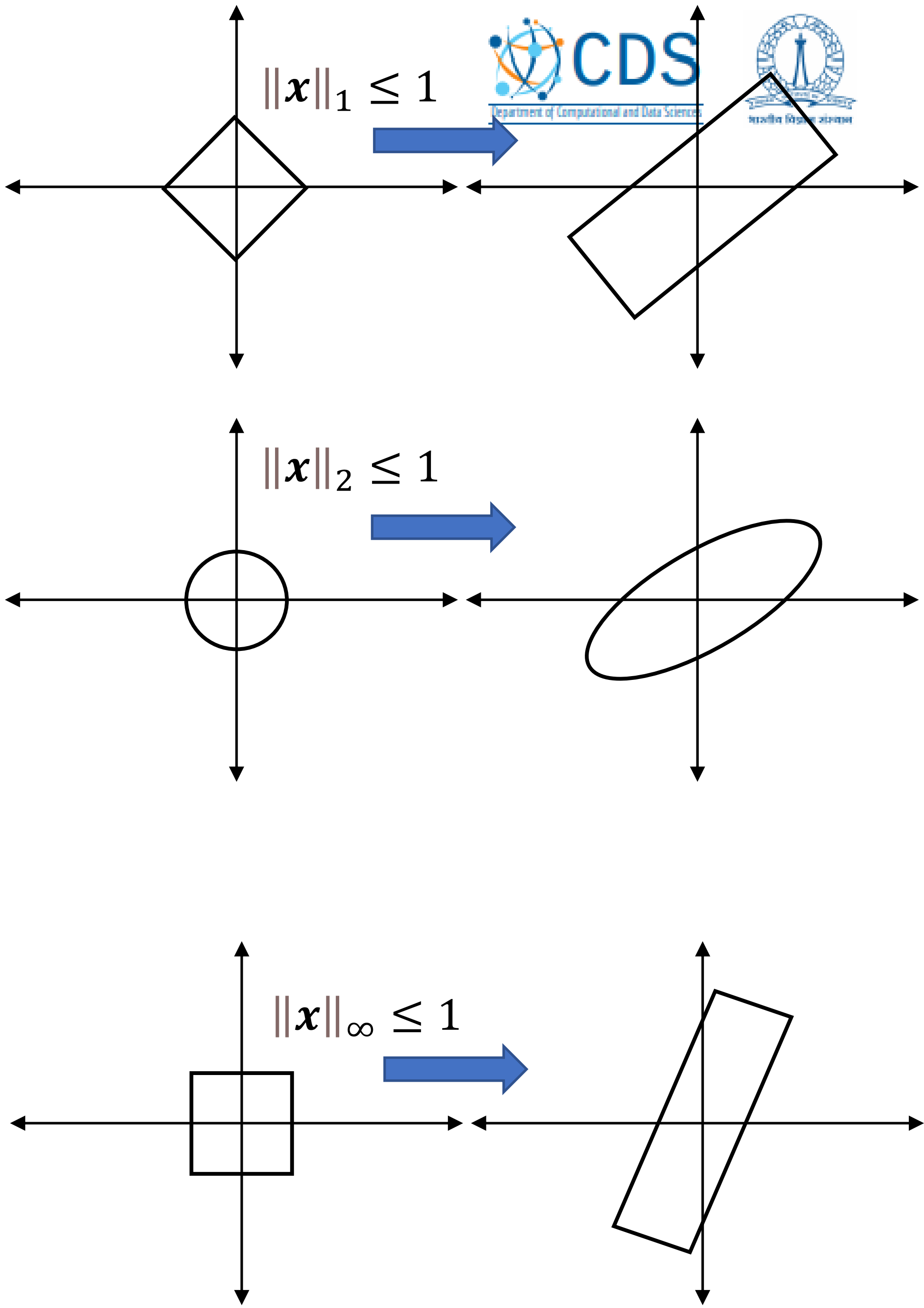
Matrix Norm

$$\mathbf{A} = \begin{pmatrix} 10 & 2 \\ 1 & 11 \end{pmatrix}$$

- An $m \times n$ matrix can be viewed as a vector in mn dimensional space. And the norm can be calculated similar to a mn vector norm
- Induced Matrix Norms: Given vector norms $\|.\|_{m}$ and $\|.\|_{n}$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the induced matrix norm $\|\boldsymbol{A}\|_{(m,n)}$ is the smallest no. K, that satisfies for all $\boldsymbol{x} \in \mathbb{R}^{n}$:

$$\|\boldsymbol{Ax}\|_{m} \leq K\|\boldsymbol{x}\|_{n}$$

$$\|\boldsymbol{A}\|_{(m,n)} = \max\left(\frac{\|\boldsymbol{Ax}\|_{m}}{\|\boldsymbol{x}\|_{n}}\right)$$



Matrices and Its Operations

Matrix Norm

- **Frobenius Norm:**

$$\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

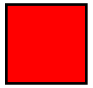
- Matrix norms are invariant under orthogonal matrix transformation

$$\|\boldsymbol{QA}\|_F = \|\boldsymbol{A}\|_F$$

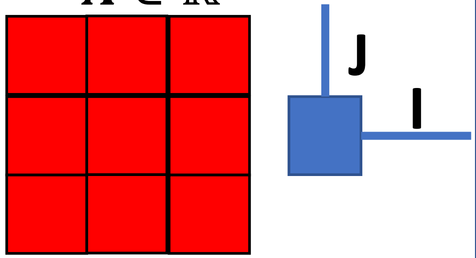
Tensors and Its Operations

Motivation

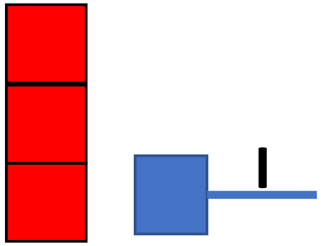
Scalar
Order 0 Tensor
$a \in \mathbb{R}$



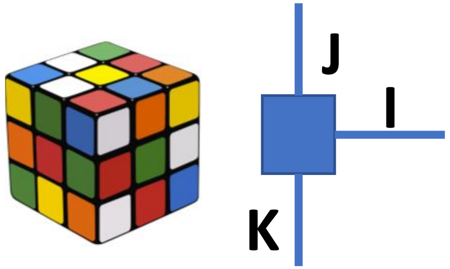
Matrix
Order 2 Tensor
$\boldsymbol{A} \in \mathbb{R}^{n \times m}$



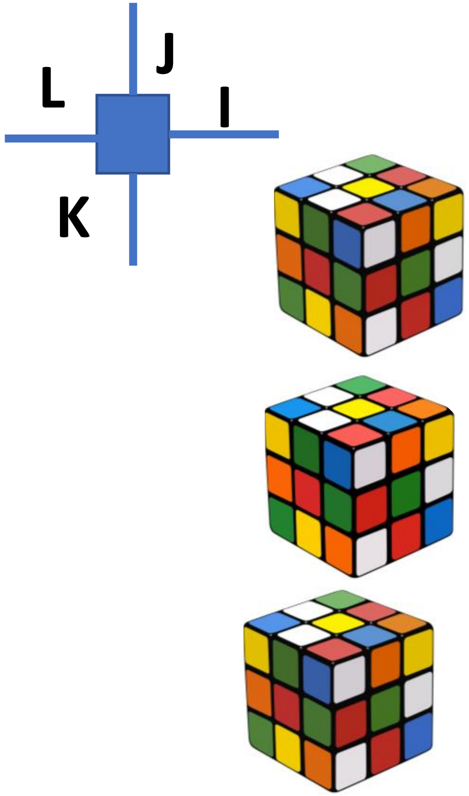
Vector
Order 1 Tensor
$\boldsymbol{a} \in \mathbb{R}^{n}$



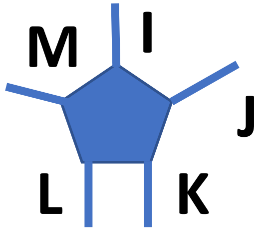
Order 3 Tensor
$\underline{\boldsymbol{A}} \in \mathbb{R}^{n \times m \times l}$

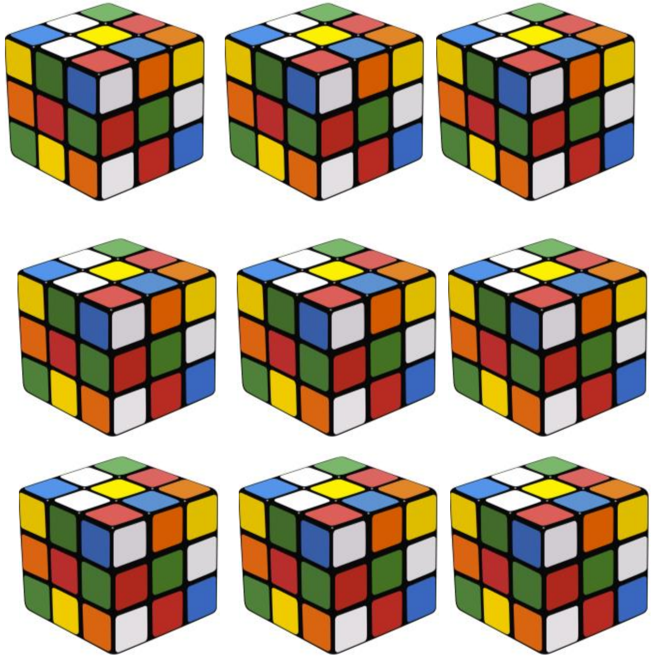


Order 4 Tensor
$\underline{\boldsymbol{A}} \in \mathbb{R}^{n \times m \times l \times p}$



Order 5 Tensor
$\underline{\boldsymbol{A}} \in \mathbb{R}^{n \times m \times l \times p \times q}$



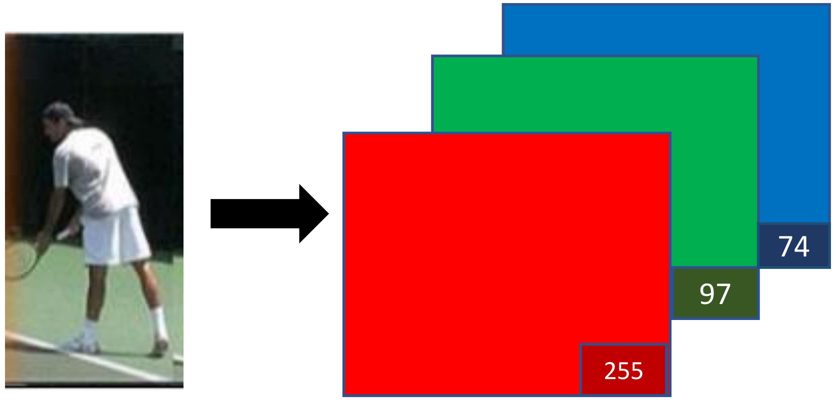


Tensors and Its Operations

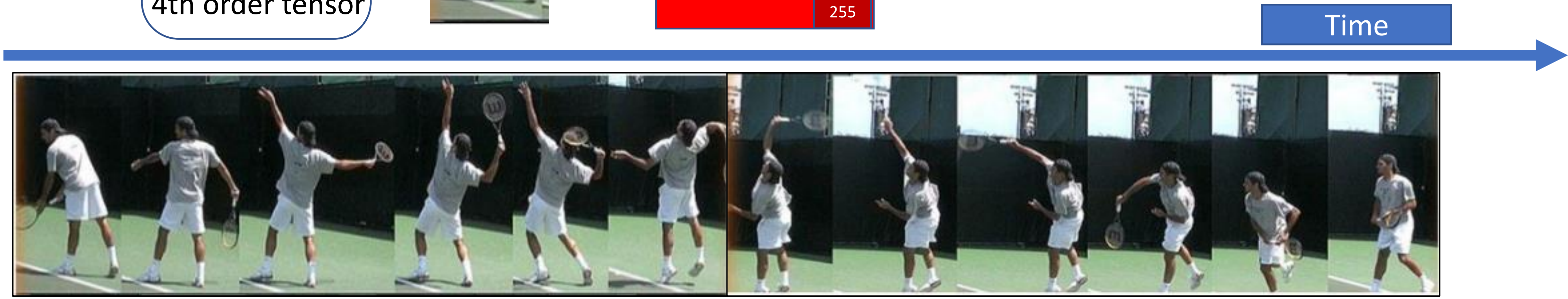
Motivation

- Mostly used in image processing, Time-series analysis where the input data have parameters that vary in more than 2 dimensions.

3[rd] order tensor



4th order tensor



Tensors and Its Operations

Addition and Subtraction

- **Tensor Sum:**
  - Sum of tensors is given by $\underline{\boldsymbol{C}} = \underline{\boldsymbol{A}} + \underline{\boldsymbol{B}}$, where $\underline{\boldsymbol{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{C}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$ (element-wise addition)

$$c_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N} + b_{i_1 i_2 \ldots i_N}$$

- **Tensor Subtraction:**
  - Sum of tensors is given by $\underline{\boldsymbol{C}} = \underline{\boldsymbol{A}} - \underline{\boldsymbol{B}}$, where $\underline{\boldsymbol{A}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{C}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$ (element-wise subtraction)

$$c_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N} - b_{i_1 i_2 \ldots i_N}$$

Let $\underline{\boldsymbol{A}} \in \mathbb{R}^{2 \times 2 \times 2}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{2 \times 2 \times 2}$, $\underline{\boldsymbol{C}} \in \mathbb{R}^{2 \times 2 \times 2}$

$$\left[\begin{pmatrix} 5 & 2 \\ -2 & 9 \end{pmatrix} \ \begin{pmatrix} 1 & 0 \\ 6 & -3 \end{pmatrix}\right] + \left[\begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \ \begin{pmatrix} 1 & 0 \\ 6 & -4 \end{pmatrix}\right] = \left[\begin{pmatrix} 8 & 3 \\ 0 & 10 \end{pmatrix} \ \begin{pmatrix} 2 & 0 \\ 12 & -7 \end{pmatrix}\right]$$

- **Tensor Quantitative Product:**
  - Given for two same order-same size tensors c $= \underline{\boldsymbol{A}} \blacksquare \underline{\boldsymbol{B}}$, where $\underline{\boldsymbol{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$

$$c = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \dots i_N} b_{i_1 i_2 \dots i_N}$$

$$\boxed{\left\|\underline{\boldsymbol{A}}\right\|_F = (\underline{\boldsymbol{A}} \blacksquare \underline{\boldsymbol{A}})^{1/2}}$$

Let $\underline{\boldsymbol{A}} \in \mathbb{R}^{2\times 2}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{2\times 2}$,

$$c = \sum_{i=1}^{2} \sum_{j=1}^{2} A_{ij} B_{ij} = \sum_{i=1}^{2} (A_{i1} B_{i1} + A_{i2} B_{i2}) = A_{11} B_{11} + A_{12} B_{12} + A_{21} B_{21} + A_{22} B_{22}$$

$$\begin{bmatrix} 1 & 0 \\ 5 & -4 \end{bmatrix} \blacksquare \begin{bmatrix} 2 & 6 \\ -3 & 2 \end{bmatrix} = -21$$

Tensors and Its Operations

Tensor Product

- **Tensor Quantitative Product:**

Let $\underline{\boldsymbol{A}} \in \mathbb{R}^{2\times2\times2}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{2\times2\times2}$,

$$c = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} A_{ijk}B_{ijk} = \sum_{i=1}^{2}\sum_{j=1}^{2}(A_{ij1}B_{ij1} + A_{ij2}B_{ij2})$$

$$= \sum_{i=1}^{2} A_{i11}B_{i11} + A_{i12}B_{i12} + A_{i21}B_{i21} + A_{i22}B_{i22}$$

$$= A_{111}B_{111} + A_{112}B_{112} + A_{121}B_{121} + A_{122}B_{122} + A_{211}B_{211} + A_{212}B_{212} + A_{221}B_{221} + A_{222}B_{222}$$

$$\begin{bmatrix} \begin{pmatrix} 5 & 2 \\ -2 & 9 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 6 & -3 \end{pmatrix} \end{bmatrix} \blacksquare \begin{bmatrix} \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 6 & -4 \end{pmatrix} \end{bmatrix} = 15 + 2 - 4 + 9 + 1 + 0 + 36 + 12 = 71$$

Tensors and Its Operations

Tensor Product

- **Right Kronecker Product:**
  - Given by $\underline{\boldsymbol{C}} = \underline{\boldsymbol{A}} \otimes_R \underline{\boldsymbol{B}}$, where $\underline{\boldsymbol{A}} \in \mathbb{R}^{I_3 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$,

$\underline{\boldsymbol{C}} \in \mathbb{R}^{J_1 I_1 \times J_2 I_2 \times \cdots \times J_N I_N}$(note same order tensors as input)

$$\overline{i_N j_N} = j_N + (i_N - 1)J_N$$

$$c_{\overline{i_1 i_2 \ldots i_N j_1 j_2 \ldots j_N}} = a_{i_1 i_2 \ldots i_N} b_{j_1 j_2 \ldots j_N}$$

$$\underline{\boldsymbol{C}} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \otimes_R \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

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

Tensors and Its Operations

Hadamard Product

- Also called Tensor Element Product
- Defined for two same-order same-size tensors, $\underline{\boldsymbol{C}} = \underline{\boldsymbol{A}} \circledast \underline{\boldsymbol{B}}$, where $\underline{\boldsymbol{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$

$$c_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N} b_{i_1 i_2 \ldots i_N}$$

Let $\underline{\boldsymbol{A}} \in \mathbb{R}^{2\times2\times2}$, $\underline{\boldsymbol{B}} \in \mathbb{R}^{2\times2\times2}$, $\underline{\boldsymbol{C}} \in \mathbb{R}^{2\times2\times2}$

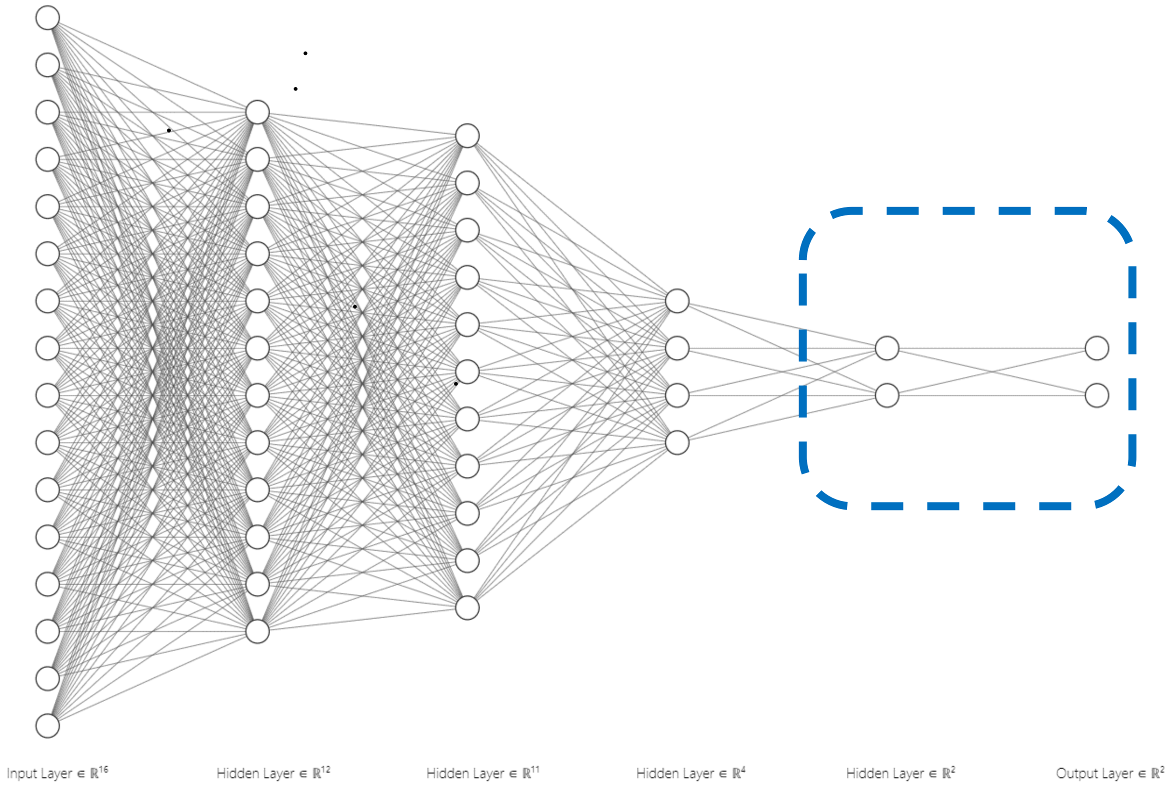
$$\left[\begin{pmatrix} 5 & 2 \\ -2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 6 & -3 \end{pmatrix}\right] \circledast \left[\begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 6 & -4 \end{pmatrix}\right] = \left[\begin{pmatrix} 15 & 2 \\ -4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 36 & 12 \end{pmatrix}\right]$$

Tensors and Its Operations

Hadamard Product

<u>Minimising Loss function in Neural Network:</u>

- Hadamard product and Right Kronecker Product is used for optimised implementation of backpropagation in Neural Networks
- We would like to see the effect of layer $N-1$ on output layer N
- Recall the logistic regression that can be considered as a single neuron…



Tensors and Its Operations

Hadamard Product

Minimising Loss function in Neural Network:

$$\boldsymbol{x} = (x_1 \quad x_2)^T$$
$$\boldsymbol{y} = (y_1 \quad y_2)^T$$

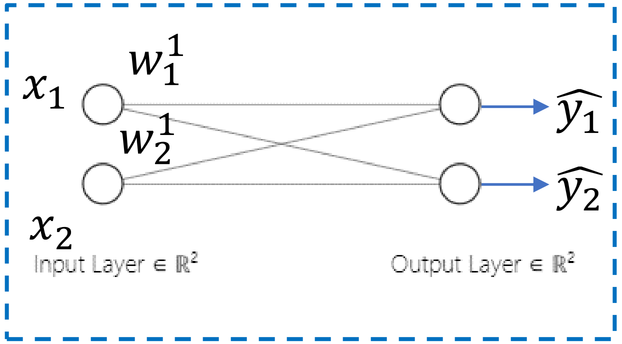
Input Data

Output Data

$$W = \begin{bmatrix} w_1^1 & w_2^1 \\ w_1^2 & w_2^2 \end{bmatrix}$$

Weights

$$(\widehat{y_1} \quad \widehat{y_2})^T = \left[\frac{1}{1 + e^{-\sum_{i=1}^{i=2} w_1^i x_i}} \quad \frac{1}{1 + e^{-\sum_{i=1}^{i=2} w_2^i x_i}}\right]^T$$



$$L = \frac{(\boldsymbol{y} - \widehat{\boldsymbol{y}})^T (\boldsymbol{y} - \widehat{\boldsymbol{y}})}{2}$$

Aim: To minimize L we need to evaluate $\frac{\partial L}{\partial \boldsymbol{W}}$

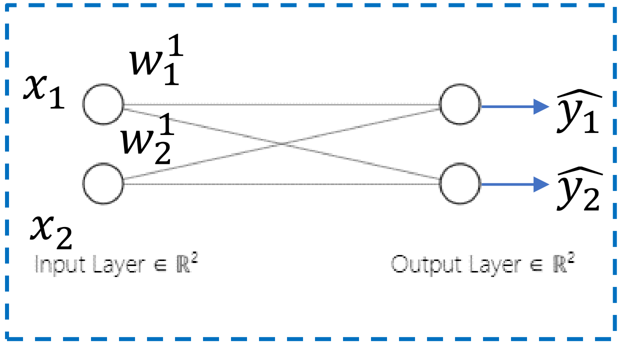
Tensors and Its Operations

Hadamard Product

Minimising Loss function in Neural Network:

$$\frac{\partial L}{\partial w_j^1} = (y_j - \widehat{y_j})x_1\widehat{y_j}(1 - \widehat{y_j})$$

$$\frac{\partial L}{\partial w_j^2} = (y_j - \widehat{y_j})x_2\widehat{y_j}(1 - \widehat{y_j})$$



$$\frac{\partial L}{\partial \boldsymbol{W}} = \begin{bmatrix} \frac{\partial L}{\partial w_1^1} & \frac{\partial L}{\partial w_2^1} \\ \frac{\partial L}{\partial w_1^2} & \frac{\partial L}{\partial w_2^2} \end{bmatrix}$$

Tensors and Its Operations

Hadamard Product

Minimising Loss function in Neural Network:

$$\frac{\partial L}{\partial \boldsymbol{W}} = (x_1 \quad x_2) \otimes_R \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \end{bmatrix} \circledast \begin{bmatrix} 1 - \widehat{y_1} \\ 1 - \widehat{y_2} \end{bmatrix} \circledast \begin{bmatrix} y_1 - \widehat{y_1} \\ y_2 - \widehat{y_2} \end{bmatrix}$$

$$\frac{\partial L}{\partial \boldsymbol{W}} = (x_1 \quad x_2) \otimes_R \begin{bmatrix} \dfrac{1}{1 + e^{-\sum_{i=1}^{i=2} w_1^i x_i}} \\ \dfrac{1}{1 + e^{-\sum_{i=1}^{i=2} w_2^i x_i}} \end{bmatrix} \circledast \begin{bmatrix} 1 - \dfrac{1}{1 + e^{-\sum_{i=1}^{i=2} w_1^i x_i}} \\ 1 - \dfrac{1}{1 + e^{-\sum_{i=1}^{i=2} w_2^i x_i}} \end{bmatrix} \circledast \begin{bmatrix} y_1 - \widehat{y_1} \\ y_2 - \widehat{y_2} \end{bmatrix}$$