Deep Learning

Basic Maths for Deep Learning

Puneet Kumar Jain

CSE Department

National Institute of Technology Rourkela

References

- Mathematics for Deep Learning

https://www.d2l.ai/chapter_appendix-mathematics-for-deep-learning/index.html

- Essential Mathematics for Machine Learning

(By Prof. Sanjeev Kumar, Prof. S. K. Gupta | IIT Roorkee | NPTEL)

https://onlinecourses.nptel.ac.in/noc21_ma38/preview

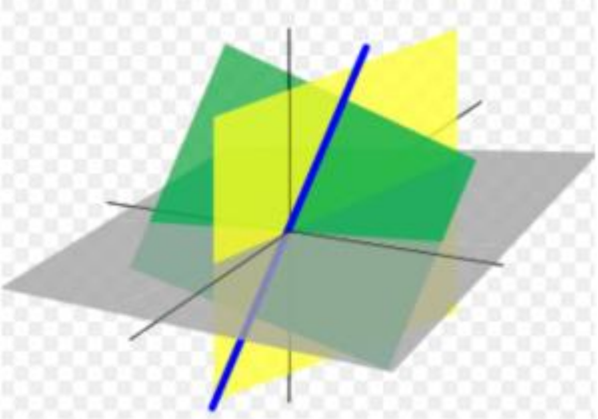
1. Linear Algebra

- Linear algebra is the branch of mathematics concerning linear equations such as

$$a_1x_1 + a_2x_2 \ldots\ldots + a_nx_n = b$$

  - By default, Vectors are denoted as Column Vectors.
  - In vector notation we say $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x}=b$

$$[a_1, a_2, \ldots a_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = b$$

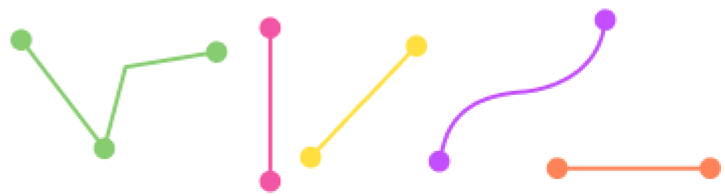


  - Called a linear transformation of $\boldsymbol{x}$

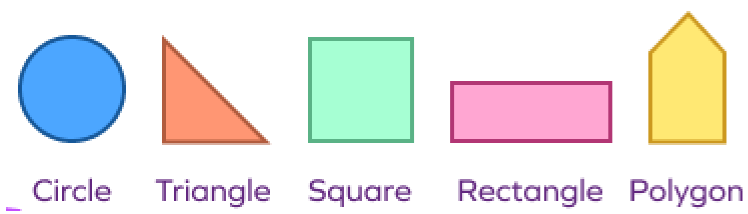
- **Linear algebra is fundamental to geometry, for defining objects such as lines, planes, rotations.**

2. Why Linear Algebra?

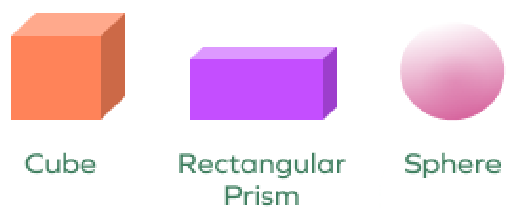
- Linear Algebra provides us with the mathematical tool to understand lower dimensions (2-D/3-D) and generalise for higher dimensions (n-D).
- 0-Dimensional : . (dot)
- 1-Dimensional



- 2-Dimensional

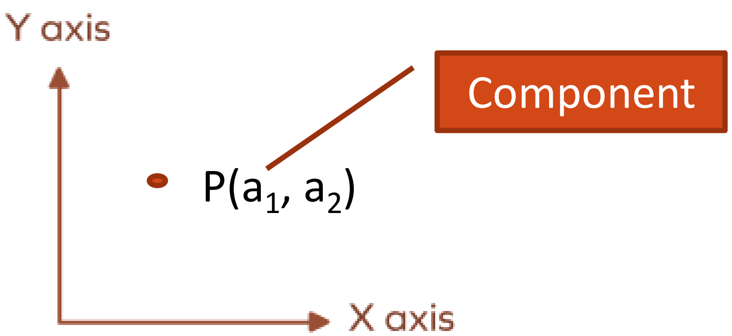


- 3-Dimensional

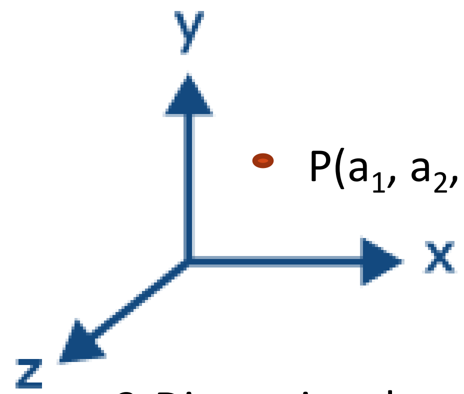


- n-Dimensional: Hypersphere, Hyperplane, Hypercube,...

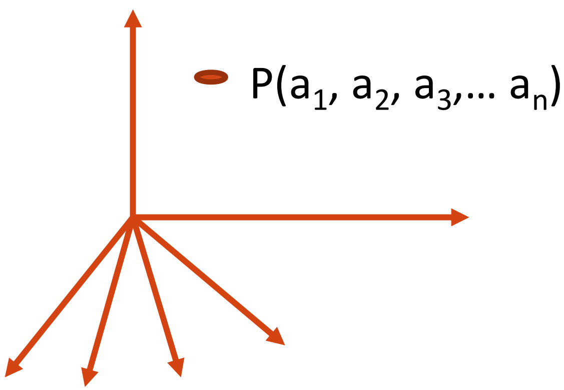
3. Point (Vector)



2-Dimensional Space

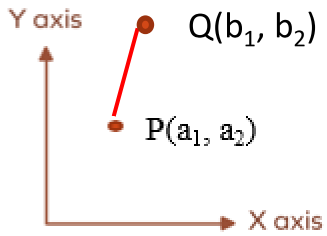


3-Dimensional Space



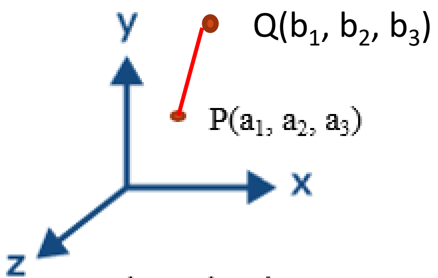
n-Dimensional Space

4. Distance between two Points



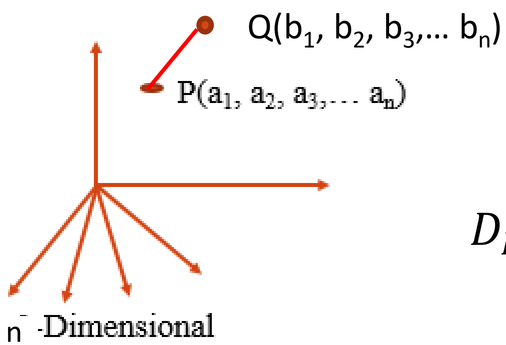
2-Dimensional
Space

$$D_{PQ} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$



3-Dimensional
Space

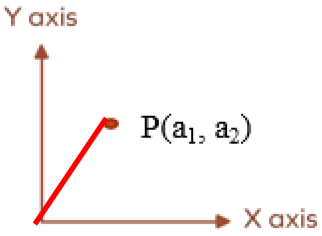
$$D_{PQ} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$$



n -Dimensional
Space

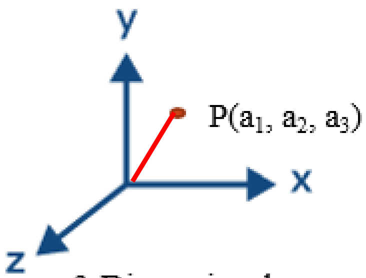
$$D_{PQ} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

4. Distance of a Point from Origin



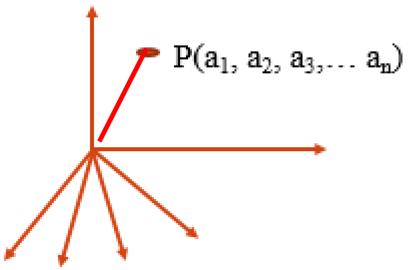
2-Dimensional Space

$$D = |P| = \sqrt{(a_1 - 0)^2 + (a_2 - 0)^2} = \sqrt{a_1^2 + a_2^2}$$



3-Dimensional Space

$$D = |P| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$



n -Dimensional Space

$$D = |P| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2}$$

5. Vector Operations

- $a = [a_1, a_2, \ldots, a_n]$
- $b = [b_1, b_2, \ldots, b_n]$

- Addition: $a + b = [a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n]$

- Subtraction: $a - b = [a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n]$

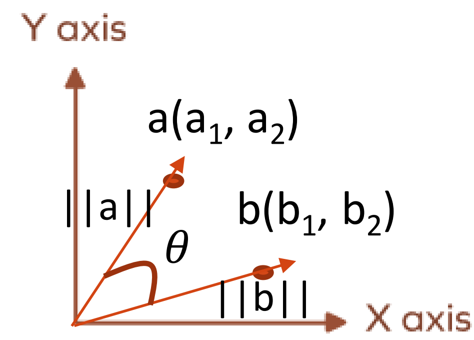
- Multiplication:
  - Dot Product: $a.b = [a_1 b_1 + a_2 b_2 + \cdots, a_n b_n]$

$$\mathrm{a.b} = [a_1, a_2, \ldots a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a^T b$$

  - Cross Product (Not much used in Machine Learning)

5. Vector Operations

- Dot Product: (Visualize in 2-D)
  - $a.b = \|a\|\|b\|\cos(\theta)$ *[***Geometry Definition***]*
  - *Where* $\|a\| = \sqrt{a_1^2 + a_2^2}$*=distance of a from Origin*
  - $a.b = a_1 b_1 + a_2 b_2$ **[Algebra Definition]**
  - The angle between two vectors=$\theta = \cos^{-1}\left(\frac{a.b}{\|a\|\|b\|}\right)$



$$\theta = \cos^{-1}\left(\frac{a_1 b_1 + a_2 b_2}{\|a\|\|b\|}\right)$$

When $\theta = 90^o$ ➔ $\cos(90) = 0$ ➔ *a.b=0*

$$\overset{\vec{a}}{\begin{bmatrix} 1 \\ 7 \\ 8 \end{bmatrix}} \cdot \begin{bmatrix} 3 & \overset{\vec{b}}{4} & 7 \end{bmatrix} = 1 \cdot 3 + 7 \cdot 4 + 8 \cdot 7 = 3 + 28 + 56 = 87$$

Hadamard Product (Element-wise Multiplication)

- It is named after French Mathematician Jacques Hadamard.

$$\vec{g} \circ \vec{h} \circ \vec{m}$$

- The order of matrices/vectors to be multiplied should be the same, and the resulting matrix will also be of the same order.

$$\overset{G}{\begin{bmatrix} 3 & 5 & 7 \\ 4 & 9 & 8 \end{bmatrix}} \circ \overset{H}{\begin{bmatrix} 1 & 6 & 3 \\ 0 & 2 & 9 \end{bmatrix}} = \overset{N}{\begin{bmatrix} 3\times1 & 5\times6 & 7\times3 \\ 4\times0 & 9\times2 & 8\times9 \end{bmatrix}}$$

- Hadamard Product is used in LSTM (Long Short-Term Memory) cells of Recurrent Neural Networks (RNNs).

5. Vector Operations

- Dot Product: (In n-D)
  - $a.b = \|a\|\|b\|\cos(\theta)$
  - *Where* $\|a\| = \sqrt{\sum_{i=1}^{n} a_i^2}$*=distance of a from Origin*
  - $a.b = a_1b_1 + a_2b_2 + \cdots + a_nb_n$=$\sum_{i=1}^{n} a_i\, b_i$
  - The angle between two vectors=$\theta = \cos^{-1}\left(\frac{a.b}{\|a\|\|b\|}\right)$

$$\theta = \cos^{-1}\left(\frac{\sum_{i=1}^{n} a_i b_i}{\|a\|\|b\|}\right)$$

When $\theta = 90^o$ ➔ $\cos(90) = 0$ ➔ *a.b=0*

$$a.a = a_1a_1 + a_2a_2 + \cdots + a_na_n = \sum_{i=1}^{n} a_i^2 = \|a\|^2$$

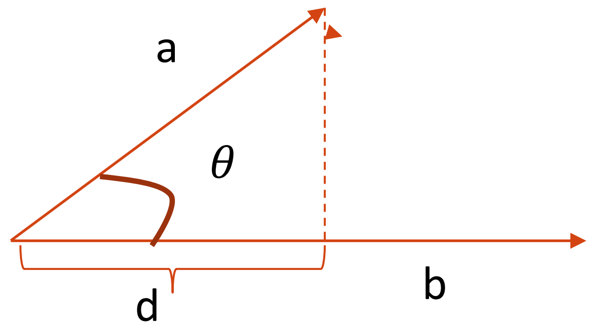
Dot product between two same vectors =(distance from Origin)^2

6. Projection

- $\cos(\theta) = \frac{p}{h} = \frac{d}{\|a\|}$

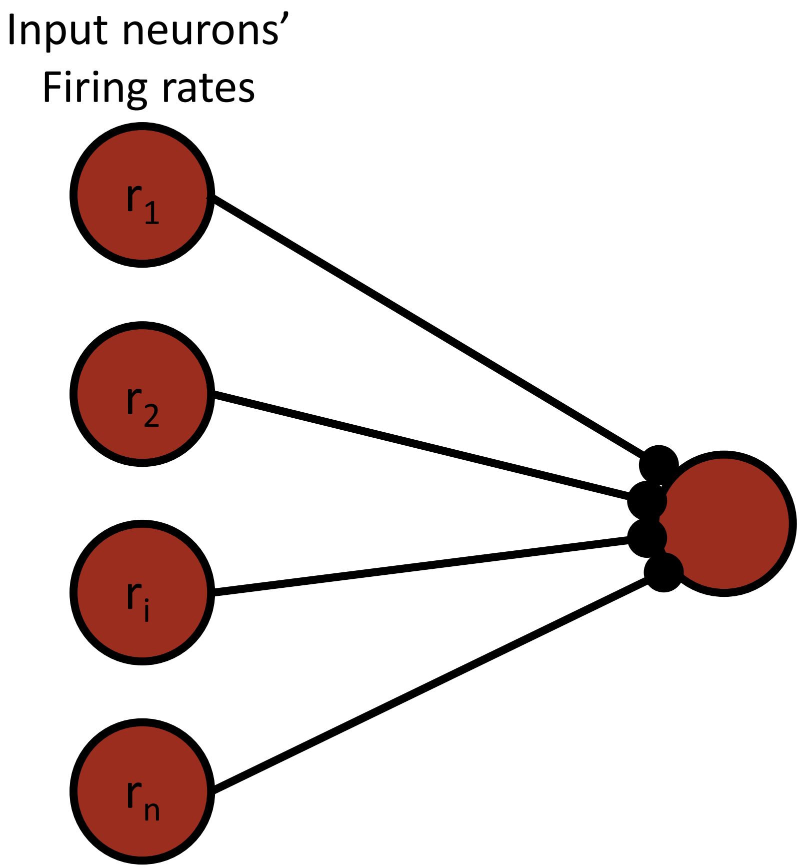
- *Projection of a on b i.e.* $d = \|a\|\cos(\theta)$

- $d = \frac{a.b}{\|b\|} = \frac{\|a\|\|b\|\cos(\theta)}{\|b\|} = \|a\|\cos(\theta)$



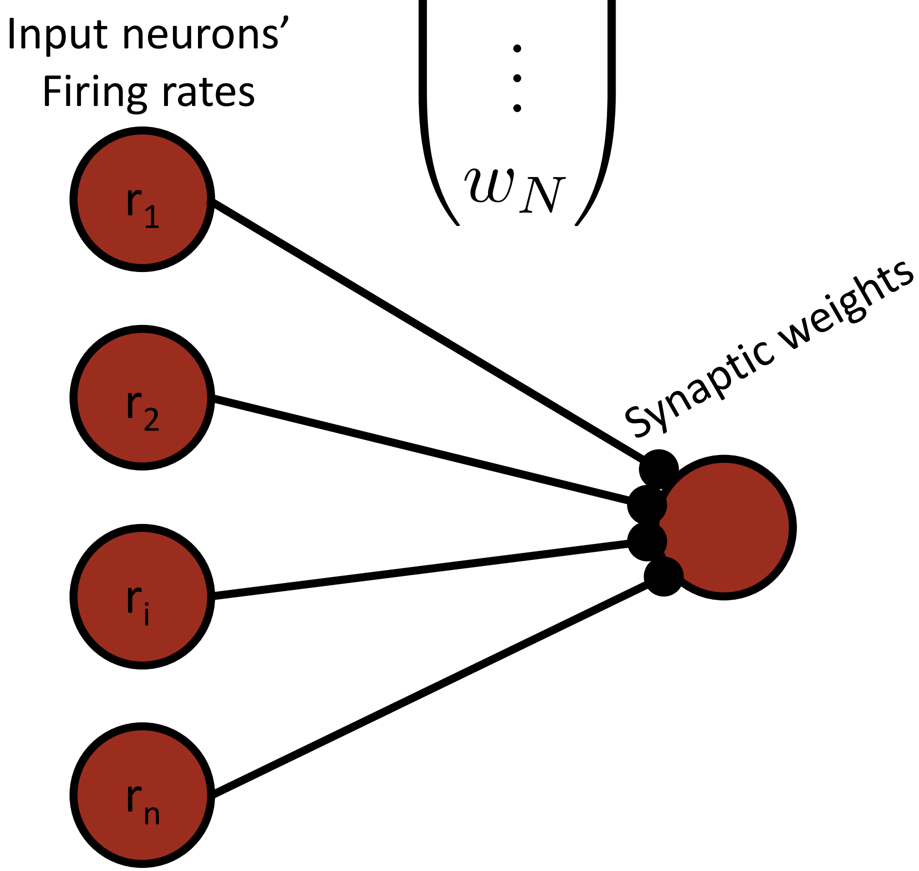
Example: linear feed-forward network

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_N \end{pmatrix}$$



Example: linear feed-forward network

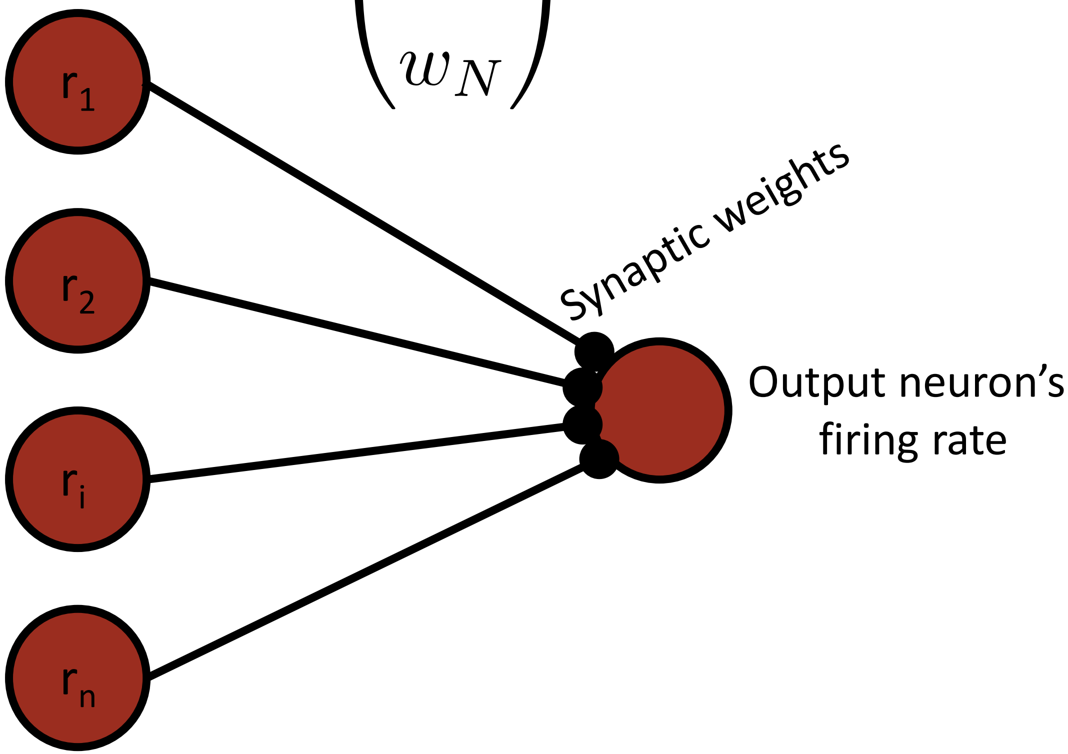
$$\begin{pmatrix} r_1 & r_2 & \cdots & r_N \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$$



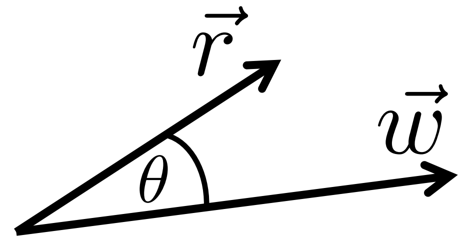
Example: linear feed-forward network

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_N \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} = r_1 w_1 + r_2 w_2 + \cdots + r_N w_N$$

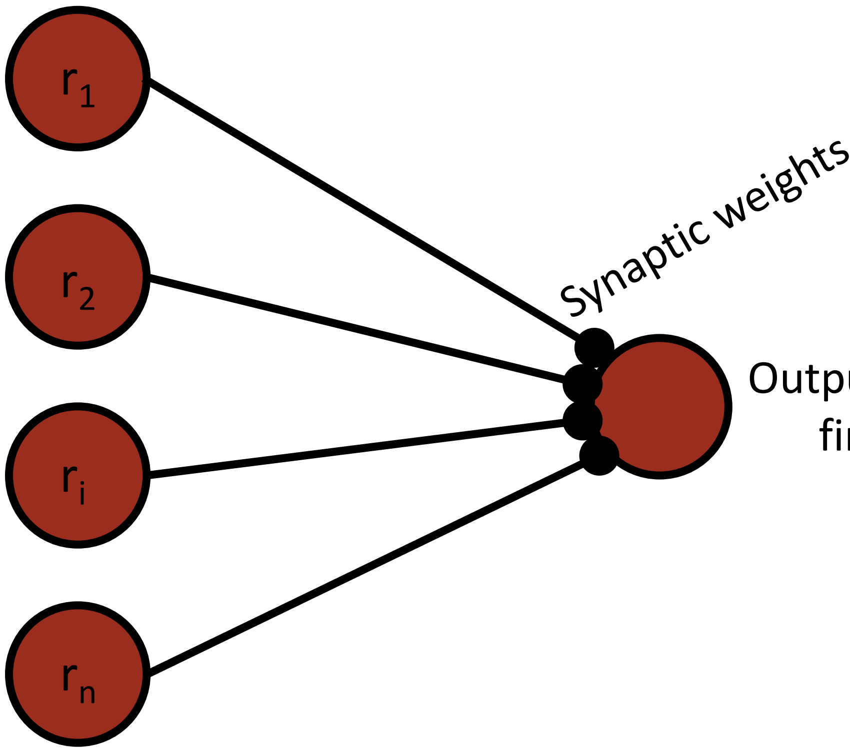
Input neurons' Firing rates



Example: linear feed-forward network



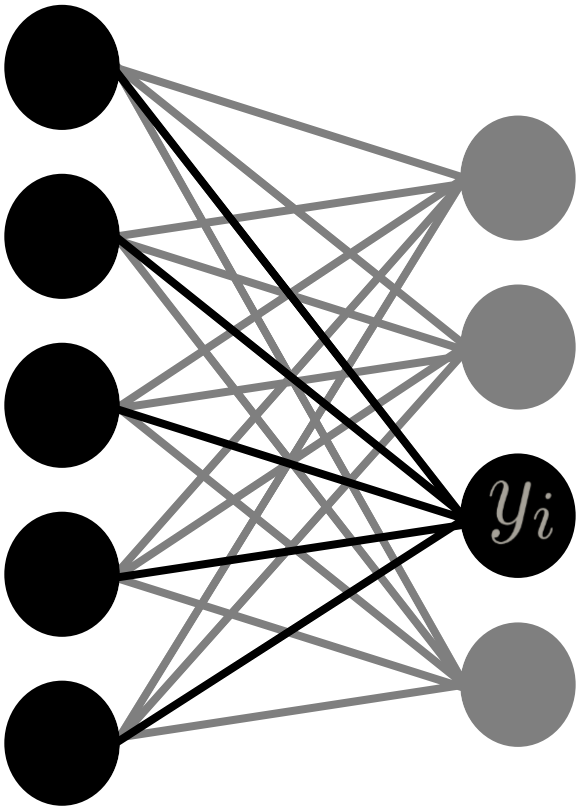
$$\vec{r} \cdot \vec{w} = |\vec{r}||\vec{w}| \cos(\theta)$$



- Insight: for a given input (L2) magnitude, the response is maximized when the input is parallel to the weight vector
- Receptive fields also can be thought of this way

Example: 2-layer linear network: **inner product point of view**

- *What is the response of cell y_i of the second layer?*

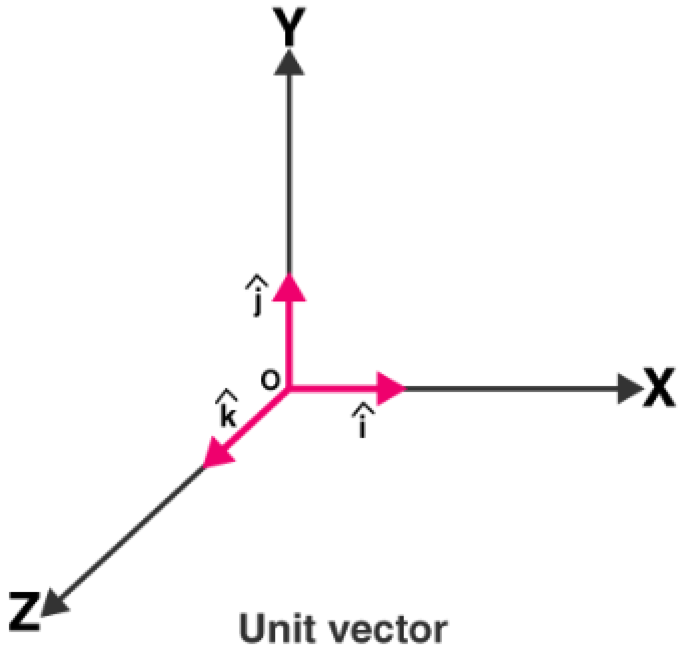


$$y_i = \sum_{j=1}^{N} W_{ij} x_j$$

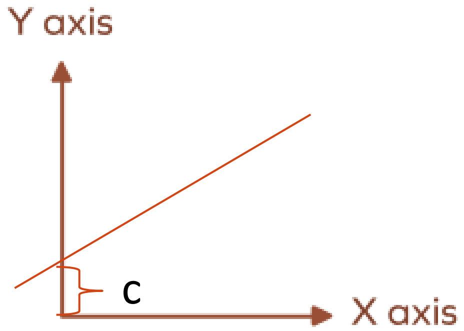
- *The response is the dot product of the i^{th} row of W with the vector x*

7. Unit Vector

- A vector is a quantity that has both magnitude, as well as direction.
- A vector that has a magnitude of 1 is a **unit vector**. It is also known as **Direction Vector**.
- Unit vector $\hat{a} = \dfrac{a}{\|a\|}$



8. Equation of a Line (2-D), Plane (3-D) & Hyperplane(n-D)

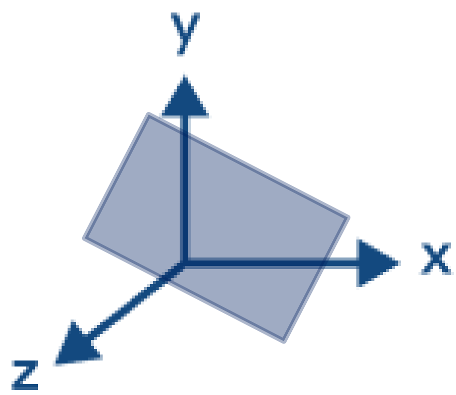


- 2-D:Line: $y = mx + c$

$$ax + by + c = 0$$

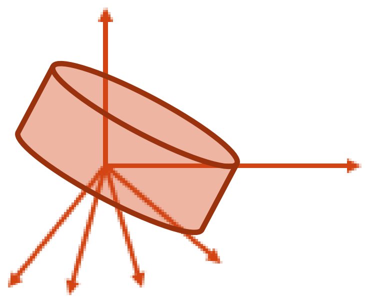
$$w_1 x_1 + w_2 x_2 + w_0 = 0$$

Note: *When $w_0 = 0$, the line passes through the origin.*



- 3-D: Plane:

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + w_0 = 0$$



- n-D:Hyperplane

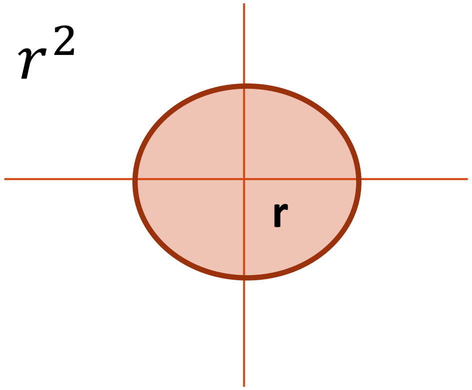
$$w_1 x_1 + \cdots + w_n x_n + w_0 = 0$$

$$w.x + w_0 = 0$$

When $w_0 = 0$, the hyperplane passes through origin.

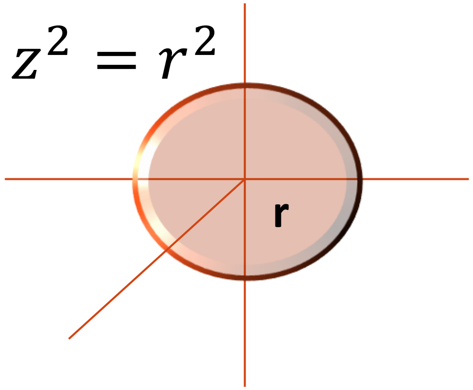
9. Equation of a Circle (2-D), Sphere (3-D) & Hypersphere(n-D)

- 2-D (Circle)
- Equation of a circle centred at origin: $x^2 + y^2 = r^2$
- A point $p(x_1,x_2)$ lies
  - Inside the circle if $x_1^2 + x_2^2 < r^2$
  - Outside the circle if $x_1^2 + x_2^2 > r^2$
  - On the circle if $x_1^2 + x_2^2 = r^2$



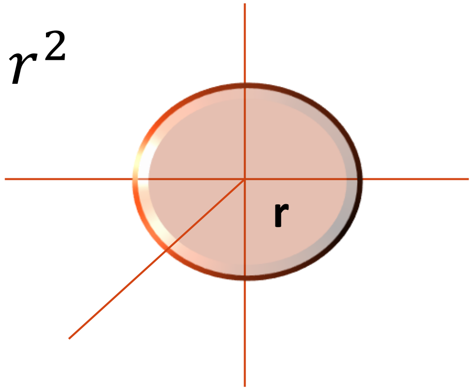
9. Equation of a Circle (2-D), Sphere (3-D) & Hypersphere(n-D)

- 3-D (Sphere)
- Equation of a circle centred at origin: $x^2 + y^2 + z^2 = r^2$
- A point p(x_1,x_2, x_3) lies
  - Inside the circle if $x_1^2 + x_2^2 + x_3^2 < r^2$
  - Outside the circle if $x_1^2 + x_2^2 + x_3^2 > r^2$
  - On the circle if $x_1^2 + x_2^2 + x_3^2 = r^2$



9. Equation of a Circle (2-D), Sphere (3-D) & Hypersphere(n-D)

- n-D (Hypersphere)
- Equation of a circle centred at origin: $\sum_{i=1}^{n} x_i^2 = r^2$
- A point $p(x_1, x_2, x_3 \ldots x_n)$ lies
  - Inside the circle if $\sum_{i=1}^{n} x_i^2 < r^2$
  - Outside the circle if $\sum_{i=1}^{n} x_i^2 > r^2$
  - On the circle if $\sum_{i=1}^{n} x_i^2 = r^2$



Other Structures

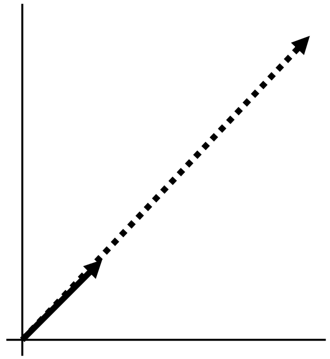
- Ellipse, Ellipsoid, Hyperellipsoid
- Square, Cube, Hypercube
- Rectangle, Hyperrectangle
- :

Eigenvectors & eigenvalues

Introduction to Linear Algebra
by Mark Goldman, and mily Mackevicius

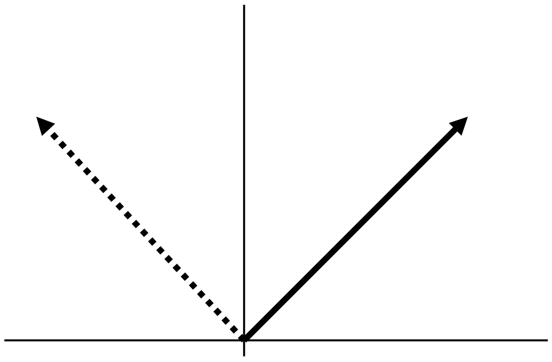
Matrices as linear transformations

$$\begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$



(stretching)

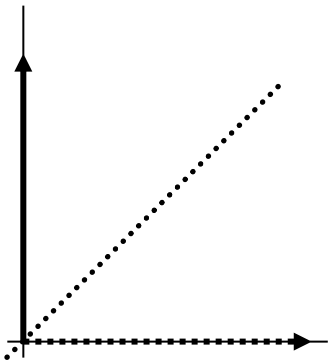
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



(rotation)

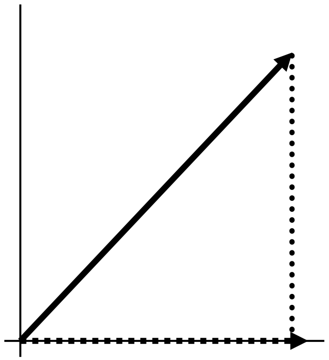
Matrices as linear transformations

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



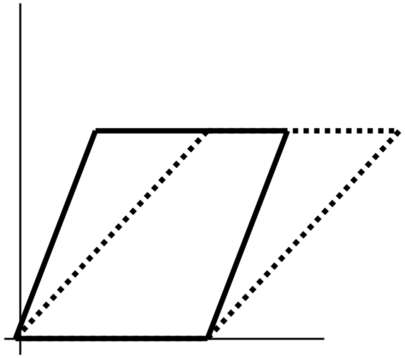
(reflection)

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$



(projection)

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + cy \\ y \end{pmatrix}$$

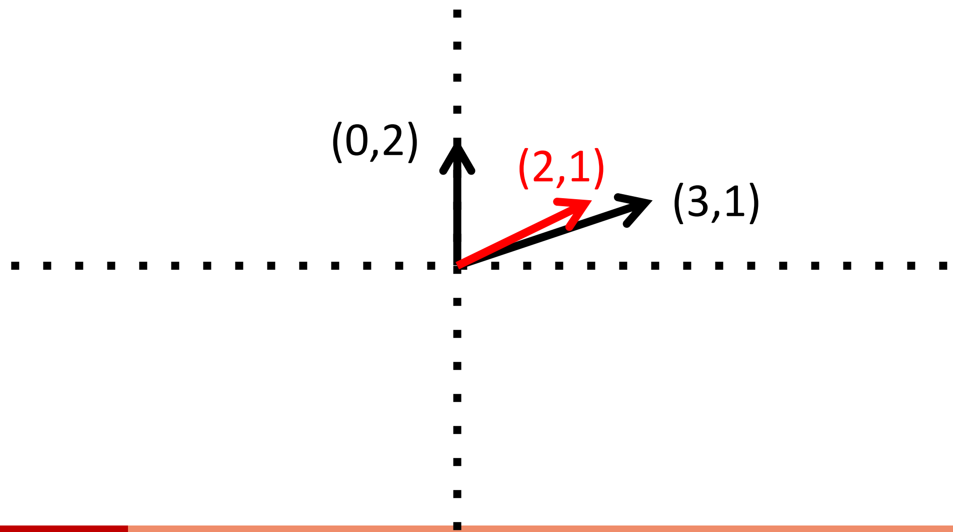


(shearing)

What do matrices do to vectors?

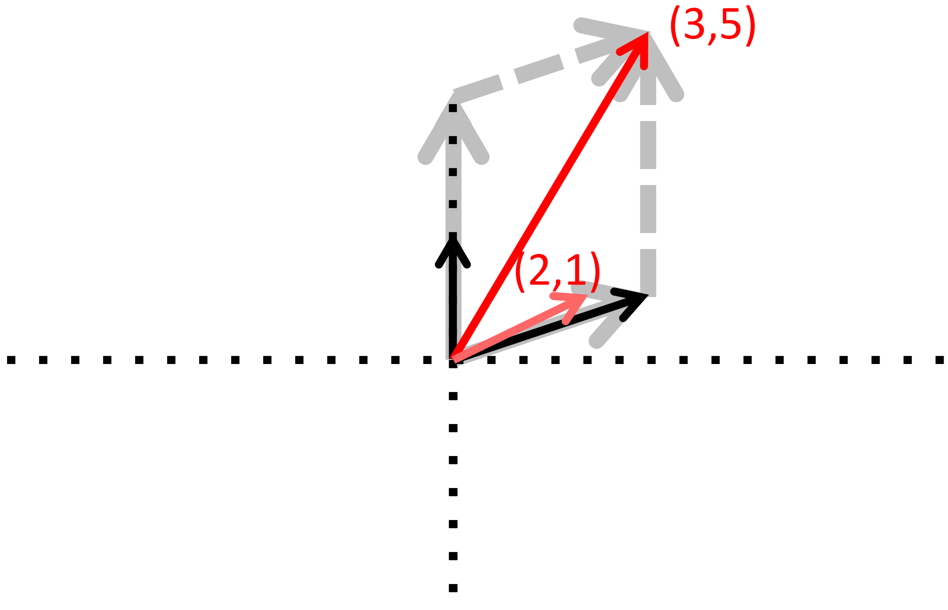
$\overleftrightarrow{M}$

$$\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} =$$



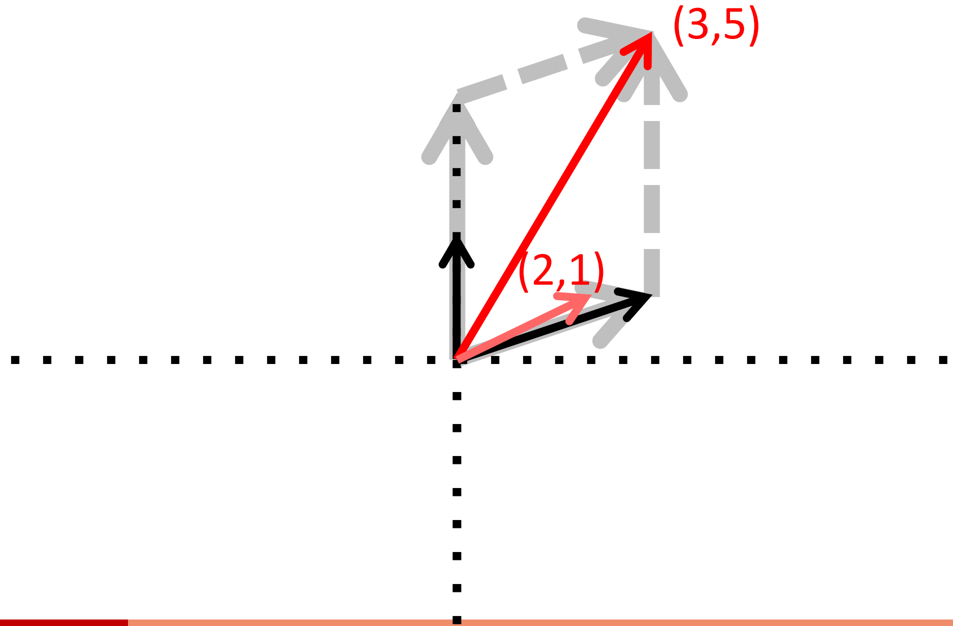
Recall

$$\overbrace{\begin{pmatrix}0 & 3\\ 2 & 1\end{pmatrix}}^{M}\begin{pmatrix}2\\ 1\end{pmatrix} = 2\begin{pmatrix}0\\ 2\end{pmatrix} + 1\begin{pmatrix}3\\ 1\end{pmatrix} = \begin{pmatrix}0\\ 4\end{pmatrix} + \begin{pmatrix}3\\ 1\end{pmatrix} = \begin{pmatrix}3\\ 5\end{pmatrix}$$



What do matrices do to vectors?

$$\overbrace{\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}}^{M} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 2 \end{pmatrix} + 1 \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$



- The new vector is:
  1) **rotated**
  2) **scaled**

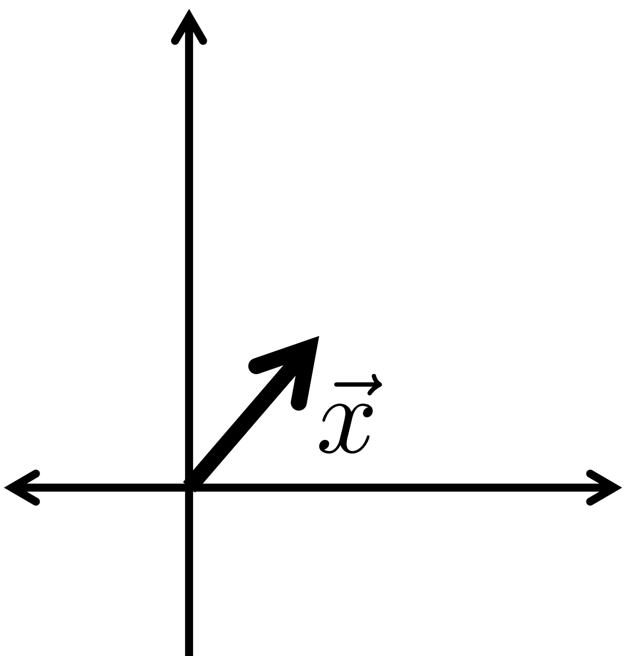
Are there any special vectors that **only get scaled**?

$$\overleftrightarrow{M}$$

$$\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}$$

Are there any special vectors that **only get scaled**?

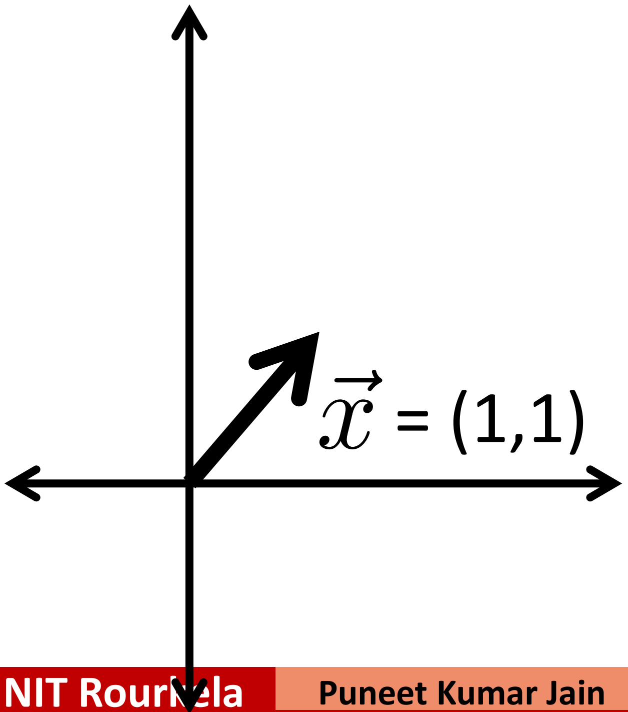
$$\overset{\overleftrightarrow{M}}{\begin{pmatrix}0 & 3\\ 2 & 1\end{pmatrix}}\begin{pmatrix}1\\ 1\end{pmatrix}$$



Try (1,1)

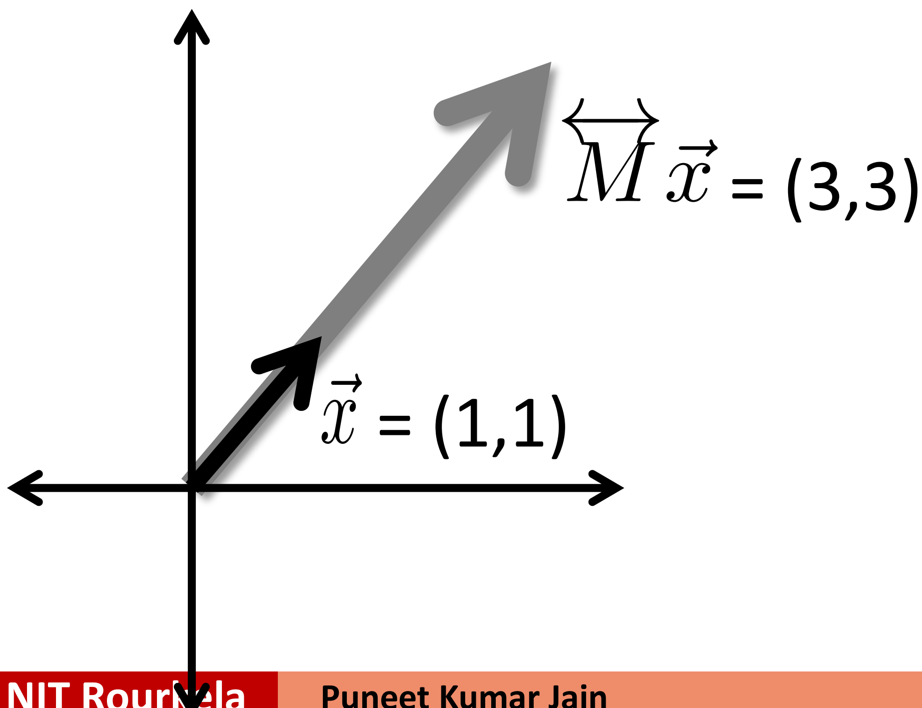
Are there any special vectors that **only get scaled**?

$$\overset{\overleftrightarrow{M}}{\begin{pmatrix}0 & 3\\ 2 & 1\end{pmatrix}}\begin{pmatrix}1\\ 1\end{pmatrix} = \begin{pmatrix}0\\ 2\end{pmatrix} + \begin{pmatrix}3\\ 1\end{pmatrix} = \begin{pmatrix}3\\ 3\end{pmatrix}$$



Are there any special vectors that **only get scaled**?

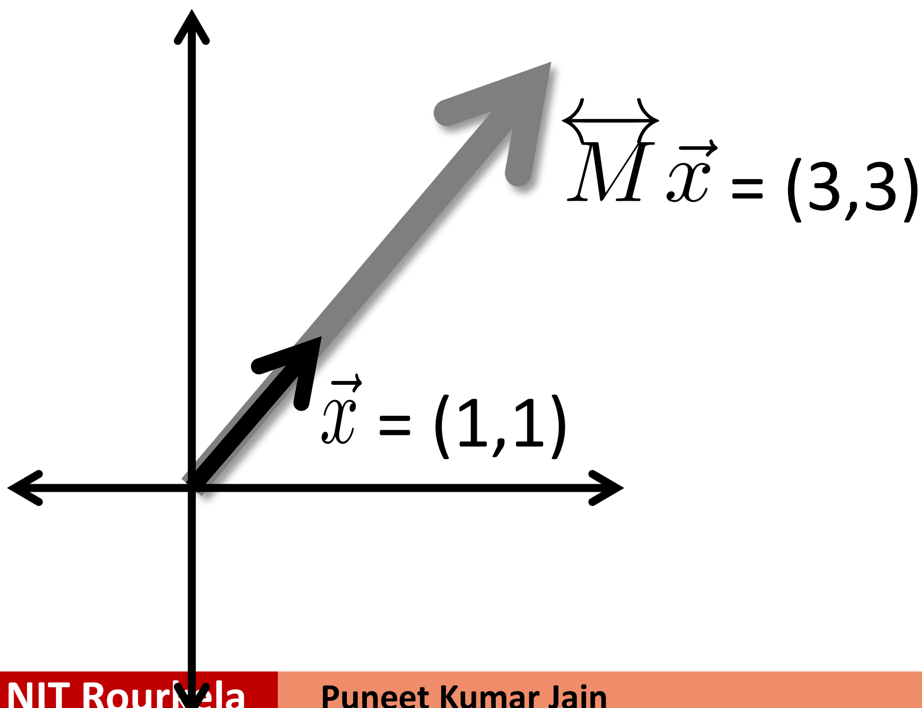
$$\overset{\overleftrightarrow{M}}{\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



Are there any special vectors that **only get scaled**?

$$\overleftrightarrow{M}$$

$$\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



- For this special vector, multiplying by M is like multiplying by a scalar.
- (1,1) is called an **eigenvector** of M
- 3 (the scaling factor) is called the **eigenvalue** associated with this eigenvector

• Vector space

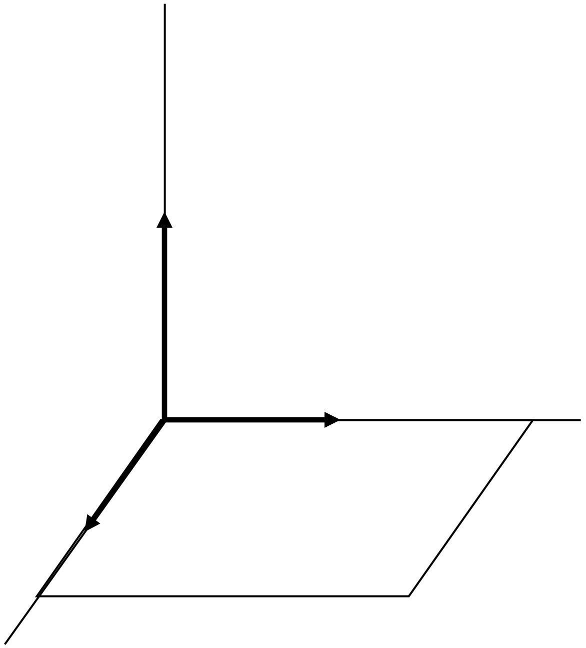
- 10-725 - Optimization
  1/16/08 Recitation
- Joseph Bradley

Vector spaces

- Formally, a *vector space* is a set of vectors which is closed under addition and multiplication by real numbers.
- A *subspace* is a subset of a vector space which is a vector space itself, e.g. the plane z=0 is a subspace of R^3 (It is essentially R^2.).
- We'll be looking at R^n and subspaces of R^n

Our notion of planes in R^3 may be extended to *hyperplanes* in R^n (of dimension n-1)

Note: subspaces must include the origin (zero vector).

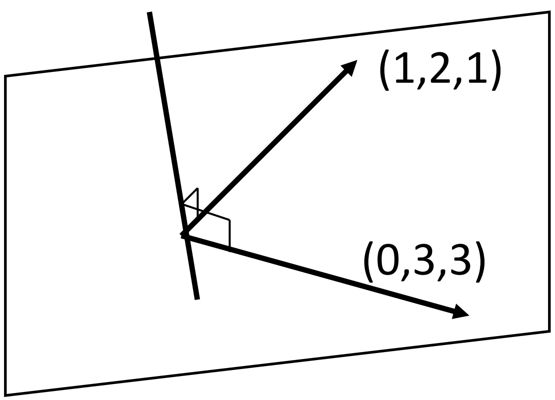


Linear system & subspaces

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

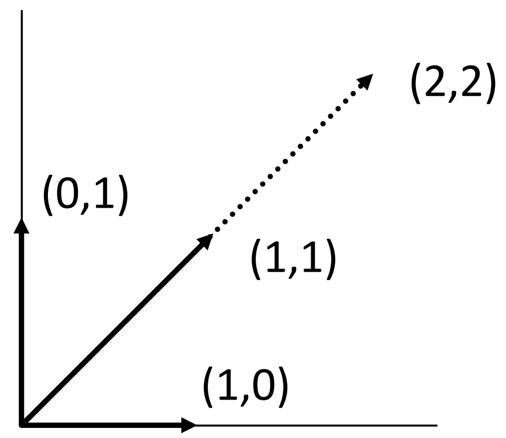
$$u\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + v\begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

- Linear systems define certain subspaces
- Ax = b is solvable iff b may be written as a linear combination of the columns of A
- The set of possible vectors b forms a subspace called the *column space* of A



Linear independence and basis

- Vectors $v_1,\dots,v_k$ are linearly independent if $c_1v_1+\dots+c_kv_k = 0$ implies $c_1=\dots=c_k=0$



i.e. the nullspace is the origin

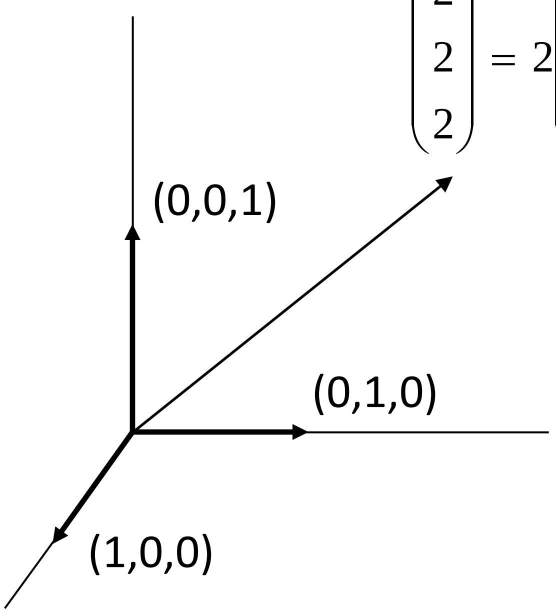
$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- If all vectors in a vector space may be expressed as linear combinations of $v_1,\dots,v_k$, then $v_1,\dots,v_k$ *span* the space.

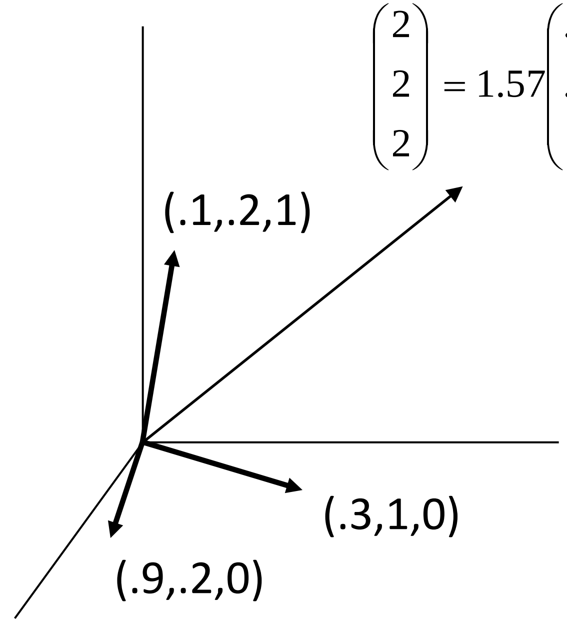
Linear independence and basis

- A *basis* is a set of linearly independent vectors which span the space.
- The *dimension* of a space is the # of "degrees of freedom" of the space; it is the number of vectors in any basis for the space.
- A basis is a maximal set of linearly independent vectors and a minimal set of spanning vectors.

$$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$



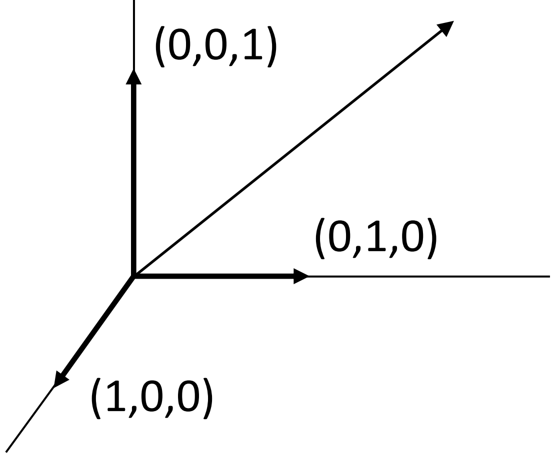
$$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 1.57\begin{pmatrix} .9 \\ .2 \\ 0 \end{pmatrix} + 1.29\begin{pmatrix} .3 \\ 1 \\ 0 \end{pmatrix} + 2\begin{pmatrix} .1 \\ .2 \\ 1 \end{pmatrix}$$



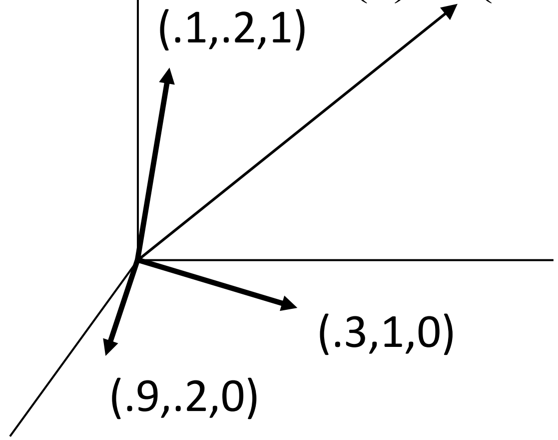
Basis transformations

We may write v=(2,2,2) in terms of an alternate basis:

$$\begin{pmatrix}2\\2\\2\end{pmatrix} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}2\\2\\2\end{pmatrix}$$

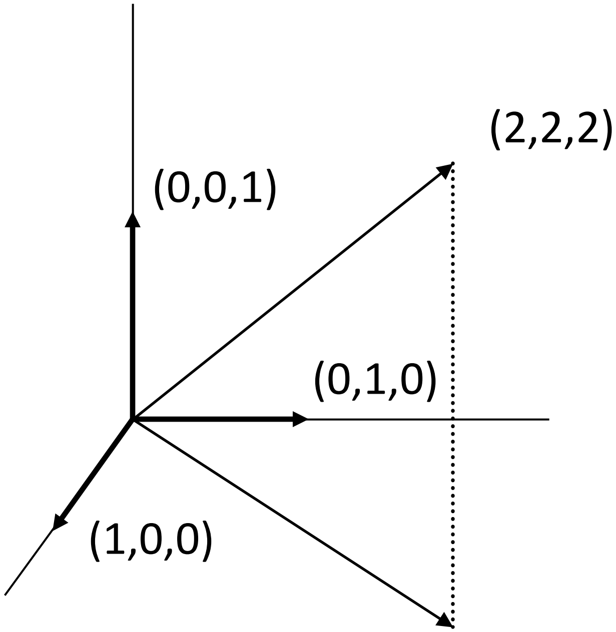


$$\begin{pmatrix}2\\2\\2\end{pmatrix} = \begin{pmatrix}.9&.3&.1\\.2&1&.2\\0&0&1\end{pmatrix}\begin{pmatrix}1.57\\1.29\\2\end{pmatrix}$$

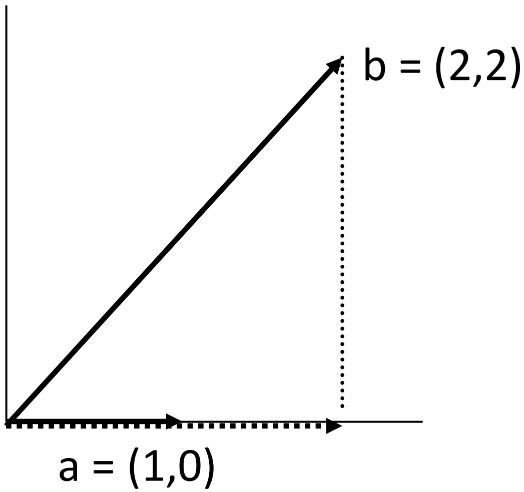


Components of (1.57,1.29,2) are projections of v onto new basis vectors, normalized so new v still has same length.

Projections

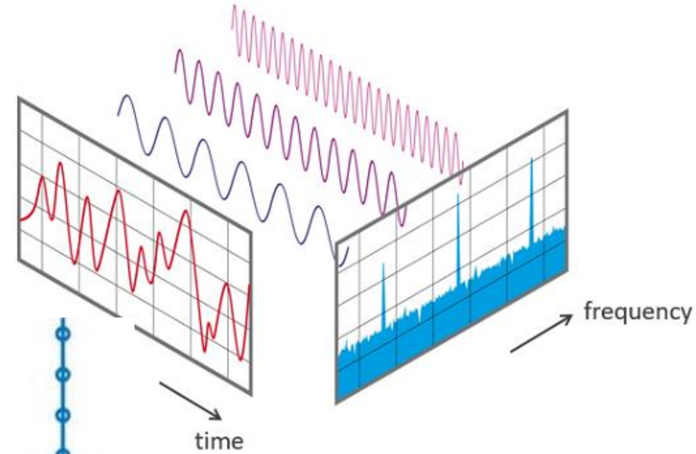


$$\begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$



$$c = \frac{a^T b}{a^T a} a = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

DFT as change of basis



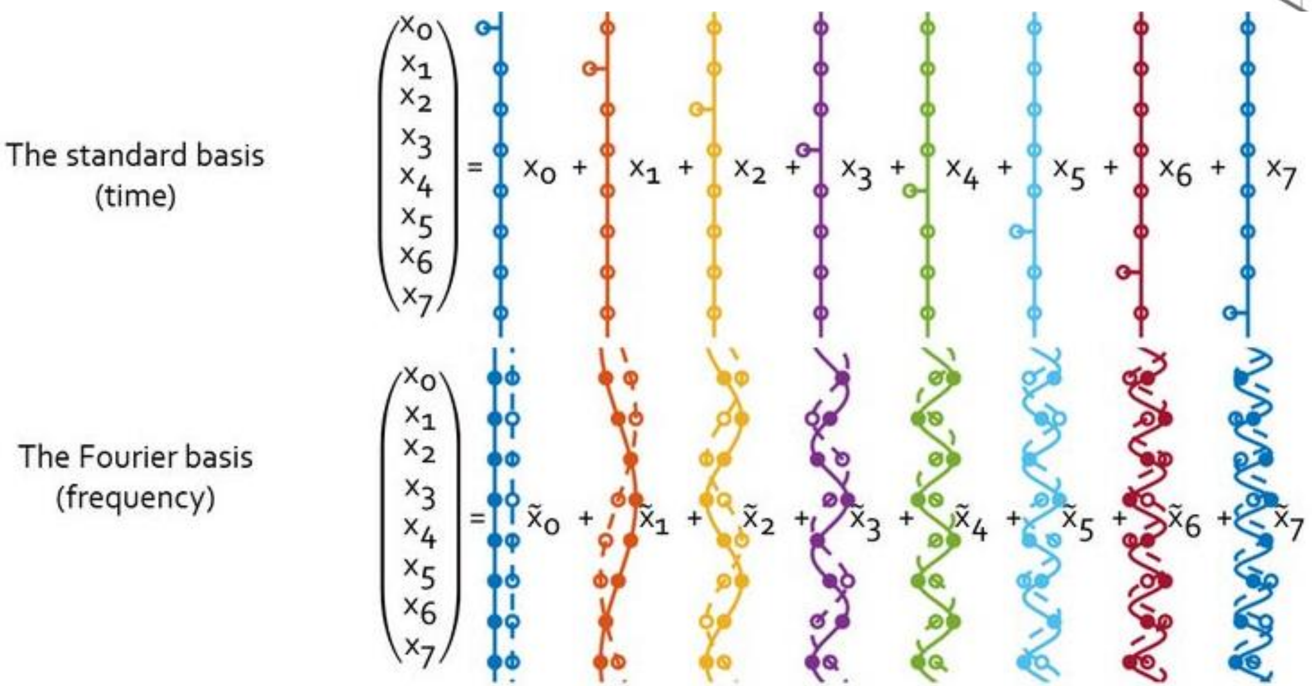


Image reference:
https://x.com/neuro_gal/status/1389539070129344514

About subspaces

- The *rank* of A is the dimension of the column space of A.
- It also equals the dimension of the *row space* of A (the subspace of vectors which may be written as linear combinations of the rows of A).

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix}$$

(1,3) = (2,3) – (1,0)

Only 2 linearly independent rows, so rank = 2.

About subspaces

<u>Fundamental Theorem of Linear Algebra</u>:

If A is m x n with rank r,

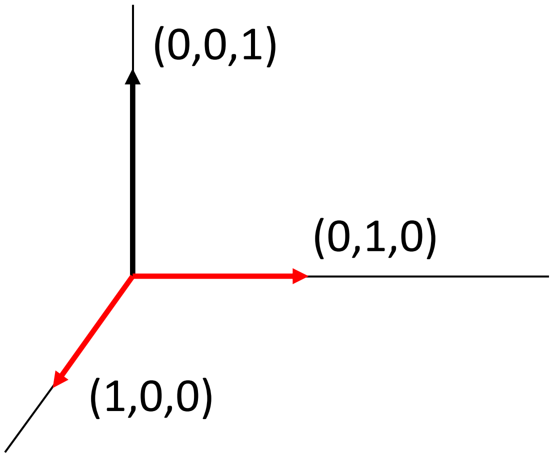
Column space(A) has dimension r

Nullspace(A) has dimension n-r (= *nullity* of A)

Row space(A) = Column space(A^T) has dimension r

Left nullspace(A) = Nullspace(A^T) has dimension m - r

<u>Rank-Nullity Theorem</u>: rank + nullity = n



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

m = 3

n = 2

r = 2

Matrix inversion

- To solve Ax=b, we can write a closed-form solution if we can find a matrix A^{-1}

  s.t. $AA^{-1} = A^{-1}A = I$ (identity matrix)

- Then Ax=b iff $x=A^{-1}b$:

  $x = Ix = A^{-1}Ax = A^{-1}b$

- A is *non-singular* iff A^{-1} exists iff Ax=b has a unique solution.

- Note: If A^{-1}, B^{-1} exist, then $(AB)^{-1} = B^{-1}A^{-1}$,

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

Non-square matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{pmatrix}$$

m = 3
n = 2
r = 2
System Ax=b may not have a solution (x has 2 variables but 3 constraints).

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{pmatrix}$$

m = 2
n = 3
r = 2
System Ax=b is underdetermined (x has 3 variables and 2 constraints).

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Differentiation

Derivatives

$\Rightarrow$ Derivative $\rightarrow$

Let's assume

$$y = x^2$$

$$\frac{dy}{dx} = 2x.$$

$$\frac{d^2y}{dx^2} = 2.$$

Partial derivative

$$y = \theta_1^2 + \theta_2^2$$

('dau') — $\frac{\partial y}{\partial w} = \begin{bmatrix} 2\theta_1 \\ \\ 2\theta_2 \end{bmatrix}$ $\longleftarrow$ $\nabla_\theta L(\theta)$: vector of partial derivative with respect to θ_i

$\frac{\partial y}{\partial b} =$ $\longleftarrow$ This is called gradient

$\downarrow$ Second order derivative.
(gradient of gradient)

$$\nabla_\theta^2 L(\theta) = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_1^2} & \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 L}{\partial \theta_2^2} \end{bmatrix}$$ $\longleftarrow$ Hessian matrix.

Derivatives

if $y = uv$ | u and v are function of x and differentiable

$\because u = x^2 \quad v = 3x$

then $\frac{\partial y}{\partial x} = \frac{\partial u}{\partial x} \cdot v + u \frac{\partial v}{\partial x}$

(A) Product rule

(B) Quotient rule

if $y = \frac{u}{v} = u \cdot v^{-1}$

$$\frac{\partial y}{\partial x} = \frac{\partial u}{\partial x} \cdot v^{-1} + u \frac{\partial (v^{-1})}{\partial x} \longrightarrow \frac{-1}{v^2} \frac{\partial v}{\partial x}.$$

$$= \frac{\partial u}{\partial x} v^{-1} + \left(-\frac{u}{v^2}\right) \frac{\partial v}{\partial x}$$

$$= \frac{\frac{\partial u}{\partial x} v - u \frac{\partial v}{\partial x}}{v^2}$$

© Chain rule:

$$y = z^2 \quad \text{and} \quad z = 3x$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x}$$

another example:

$$z = e^{xy} \qquad x = 2u + v \qquad y = \frac{u}{v}$$

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial u}$$

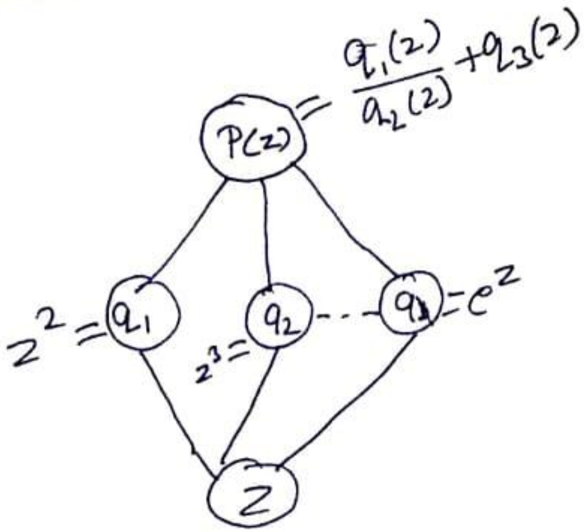
$$\frac{\partial z}{\partial x} = y e^{xy}$$

$$\frac{\partial z}{\partial y} = x e^{xy}$$

$$\frac{\partial x}{\partial u} = 2$$

$$\frac{\partial y}{\partial u} = \frac{1}{v}$$

generalization



$$\frac{dP(z)}{dz} = \frac{\partial P(z)}{\partial q_1(z)} \cdot \frac{\partial q_1(z)}{\partial z} + \frac{\partial P(z)}{\partial q_2(z)} \cdot \frac{\partial q_2(z)}{\partial z} \cdots$$

$$= \sum_{i=1}^{n} \frac{\partial P(z)}{\partial q_i(z)} \cdot \frac{\partial q_i(z)}{\partial z}$$

End of Topic