*) Vectors & Matrices

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

↓

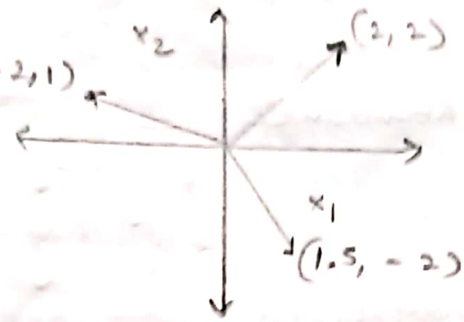
collection of the coordinates of a point in space.

↓

geometrically, ray connecting origin to the point.

↓

• vector quantified using magnitude and direction



vector representation

↓

magnitude: $\sqrt{(x_1)^2 + (x_2)^2}$

↓

root of the sum of the squares of the coordinates of the point represented by the vector

$$\sqrt{\sum_{i=1}^{N} x_i^2}$$

* ↓

L2 Norm of the vector

↓

euclidean norm

↓

possible to have same mag, but absolutely different directions.

↓

same direction and different magnitude, for ex. points on the line, $y = x$.

↓

Euclidean space: space in any finite no. of dimensions, in which points are designated by coordinates. (One for each dimension)

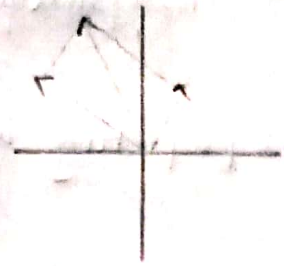
↓

distance computed using the euclidean distance formula

(•) adding 2 vectors (even subtraction)

↓

point wise addition giving another vector



↓

same for n dimensions

↓

geometrically, subtraction would be addition of the reverse of the vector to be subtracted.

(•) multiplying 2 vectors (dot product) → elementwise, resultant : scalar

$$\begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} = -2 + 1.5$$
$$= -0.5$$

$u = [u_1 \ u_2] \in \mathbb{R}^2$ → vector belong to 2 coordinate system.

$v = [v_1 \ v_2]$

↓

R^n for n coordinate system

dot product $(u, v) = u \cdot v = u^T v$

↓

transpose

$$= \sum_{i=1}^{n} u_i v_i$$

vector in direction of x-axis

↑

(•) unit vector : any vector of magnitude 1, ex. (1,0), or (0,1)

↓

vector in the same direction as a given direction, but whose magnitude = 1.

↓

dividing the vector by its own magnitude

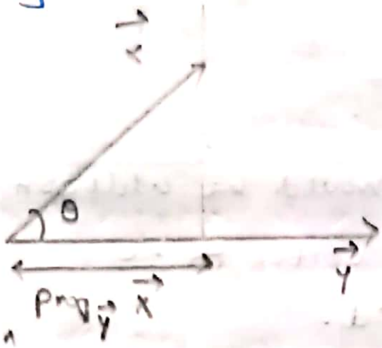
↓

$$u(2, 1.5) = \frac{(2, 1.5)}{2.5} = \frac{1}{2.5} \begin{bmatrix} 2 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$$

Scanned by CamScanner

similarly extended to n dimensions.

(•) projecting a vector onto another



$\vec{x}\cdot\vec{y} = xy\cos\theta$

$\frac{\vec{x}\cdot\vec{y}}{|y|} = x\cos\theta \cdot \hat{y}$

$= \frac{\vec{x}\cdot\vec{y}}{|y|} \cdot \frac{\vec{y}}{|y|}$

$= \left(\frac{\vec{x}\cdot\vec{y}}{\|y\|^2}\right)\vec{y}$

$proj_{\vec{y}}\vec{x}$

↓

same direction as $\vec{y}$ ✓

↓

Scaled down or scaled up version of $\vec{y}$

projection of $\vec{x}$ on $\vec{y}$.

$$proj_{\vec{y}}\vec{x} = \left(\frac{\vec{x}\cdot\vec{y}}{\|\vec{y}\|^2}\right)\vec{y}$$

→ projection

↓

projection vector

Q. project $\vec{x} = [2, 2]$ on $\vec{y} = [0, 1]$

$$proj_{\vec{y}}\vec{x} = \frac{2}{1}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\2\end{bmatrix}$$

Q. project $\vec{x} = [1, 2, 5]$ on $\vec{y} = [2, 2, 1]$

$$proj_{\vec{y}}\vec{x} = \frac{11}{9}\cdot\begin{bmatrix}2\\2\\1\end{bmatrix}$$

$$= \left[\frac{22}{9}, \frac{22}{9}, \frac{11}{9}\right]$$

(•) angle b/w 2 vectors → $\vec{x}\cdot\vec{y} = xy\cos\theta$

$$\cos^{-1}\left(\frac{\vec{x}\cdot\vec{y}}{|x||y|}\right) = \theta$$

(•) orthogonal vectors → vectors that are ⊥ to each other.

$\vec{x}\cdot\vec{y} = 0$

(•) why do we care about vectors?

↓

objects can be numerically represented using vectors

↓

feature vectors (where dimensions have a meaning associated with them)

↓

image : vector of pixel values

↓

rows ← R^{1*7} → columns

↓

cosine similarity → if the angle between 2 vectors is less, than the vectors can be said to be very similar.

(•) matrices → collection of vectors

row vector

column vector ←

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

$A \in R^{3*3}$

↓

denoted by 9 values, spread across 3 rows and 3 columns.

3, $3*1$ vectors ← $(3*1)$

↓

Stacked together → R^{3*3}

(•) adding 2 matrices : element wise addition/subtraction (subtraction)

↓

dimension matching is important

(•) multiplying a matrix with a vector

$$\underset{2 \times 2}{\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}} * \underset{2 \times 1}{\begin{bmatrix} 5 \\ 6 \end{bmatrix}} = \underset{2 \times 1}{\begin{bmatrix} 23 \\ 34 \end{bmatrix}}$$

↓ geometric interpretation

what happens when a matrix hits a vector?

↓

✓ vector gets transformed into a new vector

$f(v) = Mv = u$

↓

again, can be extended to n dimensions

↓

no. of cols in matrix = no. of rows / elements in vector

↓

✓ if the no. of rows in matrix is more or less than the elements in the vector, than that may lead to increase or decrease in dimension (depending on col or row vector stacking)

(•) multiplying a matrix with a matrix

(cols in M1 = rows in M2)

↓

each row vector with each col vector between M1 and M2.

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

(*) alternate way of multiplying matrices?

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix} = \begin{bmatrix} a_{11} * b_{11} + a_{12} * b_{21} \\ a_{21} * b_{11} + a_{22} * b_{21} \end{bmatrix}$$

$$= b_{11} * \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{21} * \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$$

↓

multiplication of resp. vector element with column of matrix and then addition (vector)

↓

linear combination of inputs ← $y = mx_1 + nx_2$

$m \to b_{11}$, $x_1 \to \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$, $n \to b_{21}$, $x_2 \to \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$

↓

in this case, the o/p can be called the linear combination of the cols of the matrix, where the coefficients are the elements of the vector

↓

extending this to a matrix, each column in the 2nd matrix would represent the linear combination with that particular column vector

↓

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$b_{11} * \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{21} * \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$ → first col.

$b_{12} * \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{22} * \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$ → second col.

(•) why do we care about matrices ?

↓

one of the most common operation in DL

$W \cdot x + b \rightarrow R^{M*1}$ (final resultant vector dimension)

for one sample : $R^{m*n} \quad R^{n*1} \quad R^{m*1}$

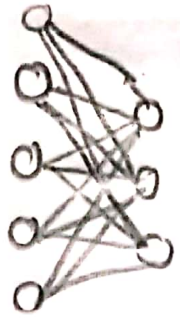
↓

for many samples

$R^{m*n} \cdot R^{n*s} + R^{m*1}$

↓

vector added for each sample



$(3*4)*(4*1)+(3*1)$

$(3*1)$

↓

activation for next layer

↓

$3*5$

↓

No. of samples in the dataset