5 Scalars and Vectors

Used in Physics, Maths and Computer Science — Data Science

- A Scalar is a single numerical value.
- A Scalar is a single numerical quantity and has no directions
- it represents a magnitude or quantity and has no directions

EXAMPLES

Car speed - 45 km/h
Temperature - 25°C
Examples of magnitude

APPLICATION IN DATA SCIENCE

DATASET

	f1	f2	f3
	–	–	–
	–	–	–
	–	–	–
	–	–	–
	–	–	–

Count of total records = 5

Average of feature f1 = — Scalar

SIMPLE LINEAR REGRESSION

slope ↲ → intercept

$y = mx + c$ → has scalar value

x: not considered a scalar

m: vector data point

A scalar - it has magnitude but no direction.

VECTORS

- A vector is a numerical value which has both magnitude and direction - in physics
- In Data Science is an ordered list of numbers it can represent a point in space or quantity with both magnitude and direction

EXAMPLE

- Speed of car is 45km/h AND is heading east

45 km/h → E

- Student marks dataset

IQ	Hours Studied	Pass/Fail
90	3 hrs	Fail
100	3 hrs	Pass

A vector representing person's IQ and the number of hours studied

[90, 3HRS]

magnitude, unit

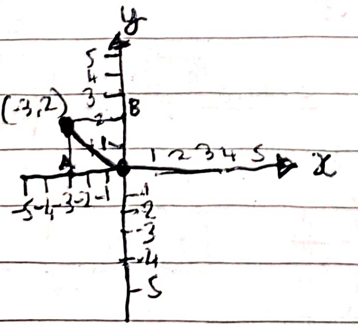
EXAMPLE

- A vector representing a person's weight over time [70, 72, 75, 73]

When you say direction it doesn't necessarily mean physical direction

A collection of values → 4 dimensions

$B = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$ (x, y)



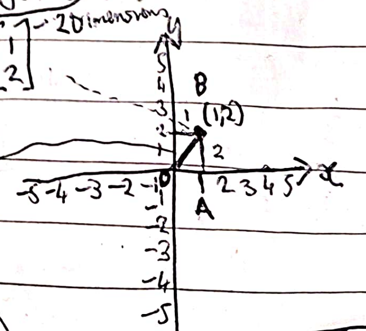
Hyp $\sqrt{(OA)^2 + (OB)^2}$

$= -3 + 2$

$= \sqrt{-1}$

Physics

$a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (x, y) - 2 Dimensions



Hyp $\sqrt{(OA)^2 + (OB)^2}$

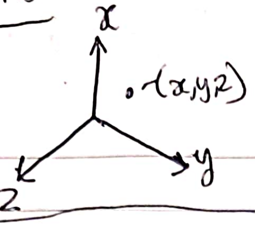
$= 1 + 4$

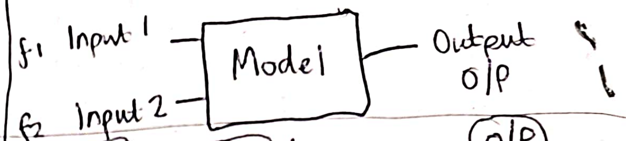
$\sqrt{5}$

(6)

3 dimensions

$$C = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

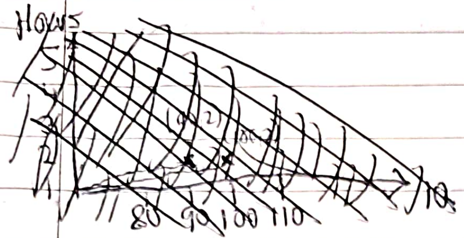


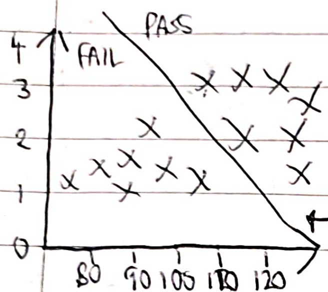


(f1)	(f2)	(O/P)
(IQ)	(No. Hours)	(Pass/Fail)

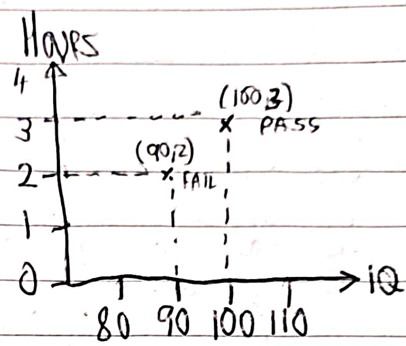
Vector $\begin{bmatrix} 90 & 2 \end{bmatrix}$, $\begin{bmatrix} 100 & 3 \end{bmatrix}$

Fail = 0
Pass = 1

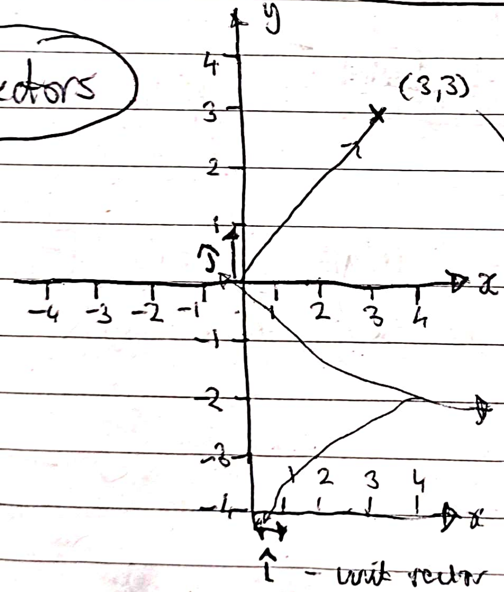




This line can be represented as $y = mx + c$



Unit Vectors



unit vector

$\hat{B}$ $\hat{V}$ $\hat{u}$ ← caret symbol or hat

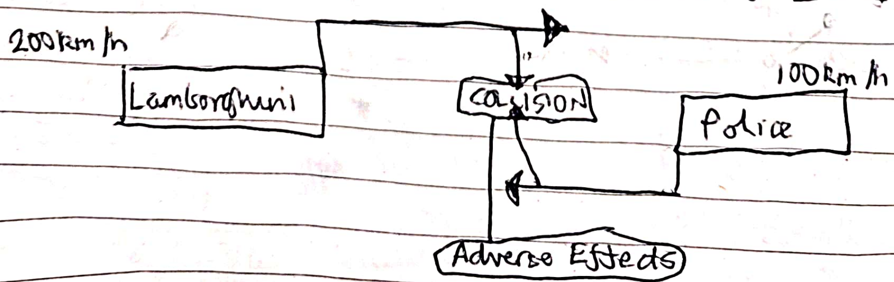
$\hat{u} = 1$ → unit vector

$(3,3) = 3\hat{i} + 3\hat{j}$

$\hat{i}$, $\hat{j}$ } unit vectors of x, y

= 1 magnitude

Example - Gaming Industry - Grand Theft Auto

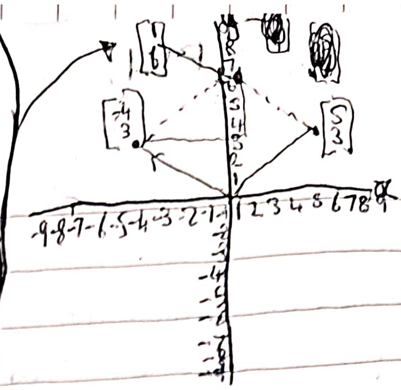


Addition of Vector

$P_1 = \begin{bmatrix} -4 \\ 3 \end{bmatrix} + P_2 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$

$\rightarrow P_1 + P_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 6 \end{bmatrix}$



$A \begin{bmatrix} x^1 \\ y^1 \\ z^1 \end{bmatrix} \quad B \begin{bmatrix} x^2 \\ y^2 \\ z^2 \end{bmatrix} \quad A+B \begin{bmatrix} x^1+x^2 \\ y^1+y^2 \\ z^1+z^2 \end{bmatrix} - \begin{bmatrix} x^3 \\ y^3 \\ z^3 \end{bmatrix}$

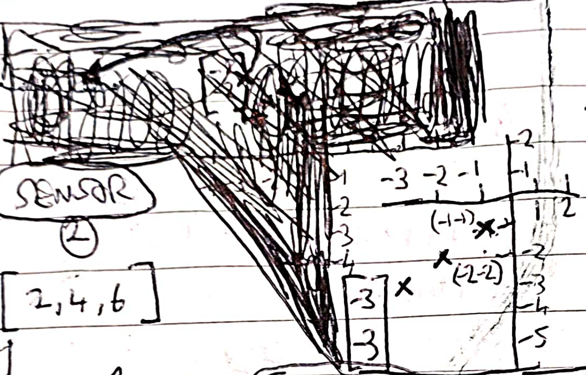
Example

$A \begin{bmatrix} -2 \\ -2 \end{bmatrix} + B \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$\begin{bmatrix} -3 \\ -3 \end{bmatrix}$

Example

Solving a usecase



SENSOR ①

SENSOR ②

[3,5,7] + [2,4,6]

Final Sensor

[5,9,13]

EDA and FE

EDA – Exploratory Data Analysis

FE – Feature Engineering

USED IN

= Use in this field of work – ① Data Aggregation Tasks

② Feature Engineering

③ Natural Language Process (NLP)

EXAMPLE - ECOMMERCE WEBSITE

Reviews	Sentiment
Product is Good	1
Product is bad	0

INPUTS → Model → OUTPUT

Text → convert → Vector

Using different Technics

OHE, TFIDF, BOW

WORD2VEC

change into numbers

① Called WORD EMBEDDING

① Data [0.2, 0.1, 0.4]

② Science [0.3, 0.7, 0.2]

③ Data Science = [0.5, 0.8, 0.6]

② IMAGE PROCESSING

Colour image [R, G, B]

– R – red channel = [255, 128, 0]

– G – green channel = [128, 255, 0]

– B blue channel = [64, 64, 255]

Convert to greyscale

255, 128, 64
128, 255, 64
0, 0, 255

[149, 149, 85]

MULTIPLICATION OF VECTORS

② Element wise Multiplication

THREE TYPES

① DOT PRODUCT / Inner Product – The dot product of two vectors results in a scalar and is calculated as the sum of the products of their corresponding components

$A = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ $B = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ $A \cdot B = 2 \times 4 + 3 \times 5 = 8 + 15 = 23$

$A \cdot B$ – dot product

T = Transpose

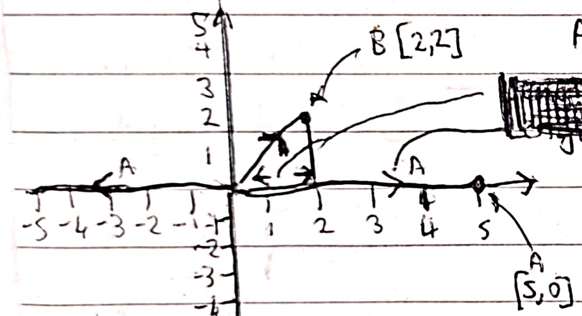
23 = scalar value

$A \cdot B^T = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 & 5 \end{bmatrix}$

$A \begin{bmatrix} 5 \\ 0 \end{bmatrix}$ $B = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ + \\ 0 \end{bmatrix} = 10$

LENGTH of A = 2
LENGTH of B = 5

$A \cdot B = 10$



$A = \begin{bmatrix} -5 \\ 0 \end{bmatrix}$ $B = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -10 \\ 0 \end{bmatrix} = -10$

Length of A –5
Length of B 2 = –10

$A \cdot B = 0$ = Project the vector to the origin

AVENGERS

$\frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{6}{3.872 \times 2.646} = 0.586$

MOVIE B

Step 1 – Dot product of $A \cdot B$

$A \cdot B = 1 \cdot 2 + 2 \cdot 0 + 0 \cdot 1 + 3 \cdot 1 + 1 \cdot 1 = 2 + 3 + 1 = 6$

Step 2 – $\|A\|$ $\|B\|$?

$\|A\| = \sqrt{1^2 + 2^2 + 0^2 + 3^2 + 1^2} = \sqrt{15} = 3.872$

$\|B\| = \sqrt{2^2 + 0^2 + 1^2 + 1^2 + 1^2} = \sqrt{7} = 2.646$

called EUCLEDIAN DISTANCE

$\cos\theta = \frac{A \cdot B}{\|A\| \ \|B\|}$

magnitudes

Recommendation System

Netflix – Action Movie → Recommendation of other movies.

AVENGERS = [1, 2, 0, 3, 1] (A)

Action, Drama, Comedy, Romance

RECOMMEND ? MOVIE B = [2, 0, 1, 1, 1]

APPLICATION OF DOT PRODUCT IN DATA SCIENCE

Gen AI App = RAG System (AI)

① Cosine Similarity

It is a measured used to determine how similar two vectors are. It calculates the cosine of the angle between two vectors providing similarity score that ranges from –1 (dissimilar) to +1 (completely similar)

APPLICATION OF DOT PRODUCTS IN DATA SCIENCE

COSINE SIMILARITY

$$\cos\theta = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

Magnitude

MOVIE A · PULP FICTION

MOVIE B ?

$[1,2,0,3,1]$ ← Characteristics → $[2,0,1,1,1]$

STEP 1 $A \cdot B = 1\cdot2 + 2\cdot0 + 0\cdot1, 3\cdot1, 1\cdot1 = \boxed{6}$

STEP 2 $\|A\| \cdot \|B\| = \sqrt{1^2+2^2+0^2+3^2+1^2} = \sqrt{15} = 3.872$

$\sqrt{2^2+0^2+1^2+1^2+1^2} = \sqrt{7} = 2.646$

$$\frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{6}{3.872 \times 2.646} = +0.586$$

Calculates the cosine of the angle between two vectors providing similarity score that ranges from
-1 = dissimilar to
+1 = completely similar

Movie A is 58.6% similar to Movie B

Vector Databases / RAG SYSTEM