DIMENSIONALITY REDUCTION

DAY 9 AGENDA

Dimensionality reduction

Principal Component Analysis (PCA)

By Anil Kumar APSSDC

WHAT IS DIMENSIONS

Dimension → measure in one direction

Line-1D, square-2D, cube-3D
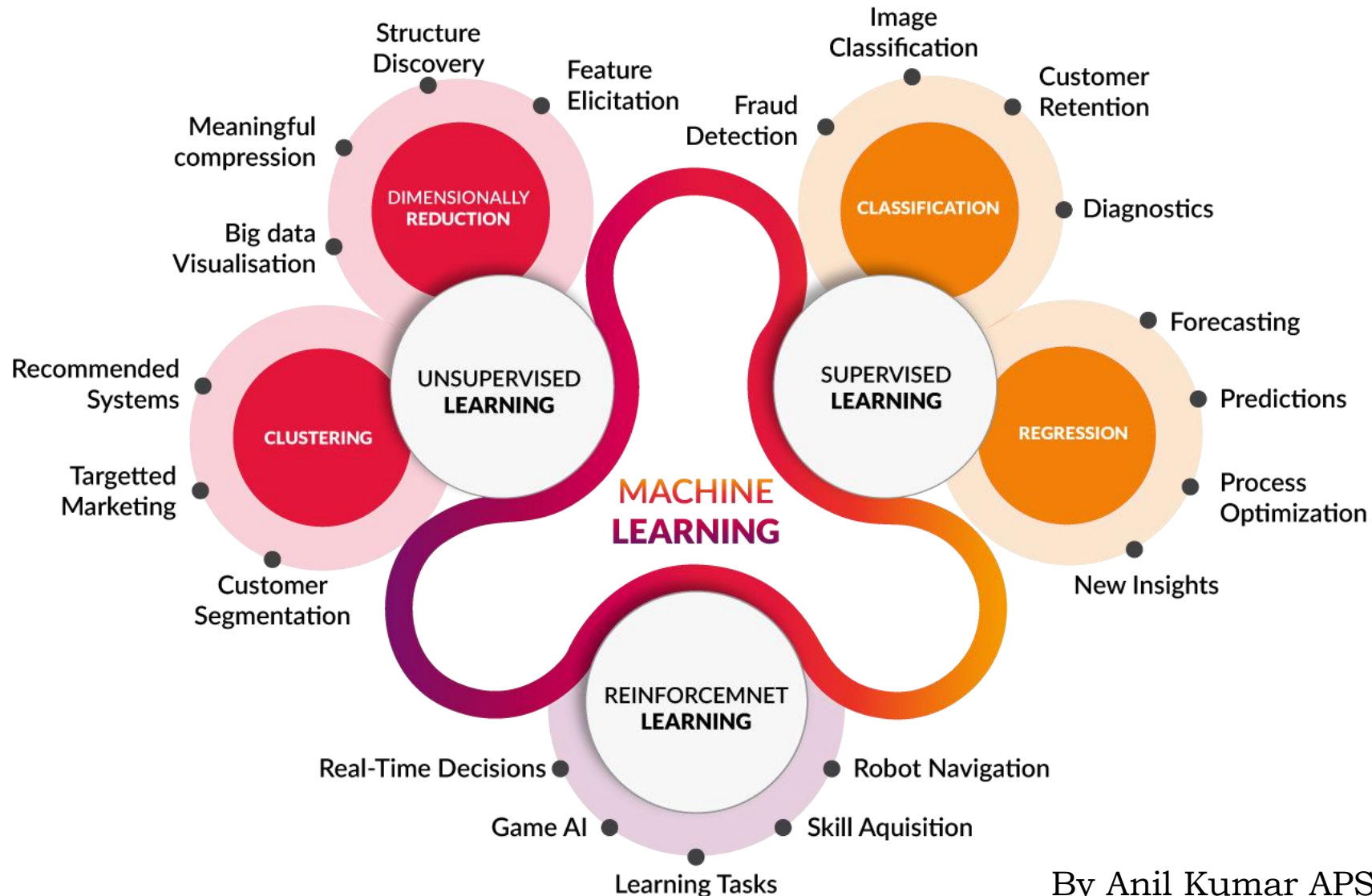
[1,2,3,4,5], → Vector

[[1,2,3,4],

[4,3,2,1]], -> matrix

3d data → Tensors → images (RGB), videos,

180 * 180 → 180rows and 180columns

1920 * 1080 → 1920 rows, 1080 columns →

By Anil Kumar APSSDC

MACHINE LEARNING CATEGORIES

A + b + c + d = out

5 + 1 + 0 + 8 = 14

A*b = a → 5, 1 → 5 → ab→ e → e + c + d = 14

C = 0

E + d → output

DIMENSIONALITY REDUCTION METHODS

The various methods used for dimensionality reduction include:

1. Principal Component Analysis (PCA)

2. Linear Discriminant Analysis (LDA)

3. Generalized Discriminant Analysis (GDA)

By Anil Kumar APSSDC

PRINCIPAL COMPONENT ANALYSIS

PCA = "Principal Component Analysis"

1. Fundamental dimension reduction technique

2. First step "decorrelation"

3. Second step reduces dimension

By Anil Kumar APSSDC

DIMENSION REDUCTION

1. More efficient storage and computation

2. Remove less-informative "noise" features

3. which cause problems for prediction tasks, e.g. classification, regression

4. Represents same data, using less features

5. Important part of machine-learning pipelines

6. Can be performed using PCA

By Anil Kumar APSSDC

PCA ALIGNS DATA WITH AXES
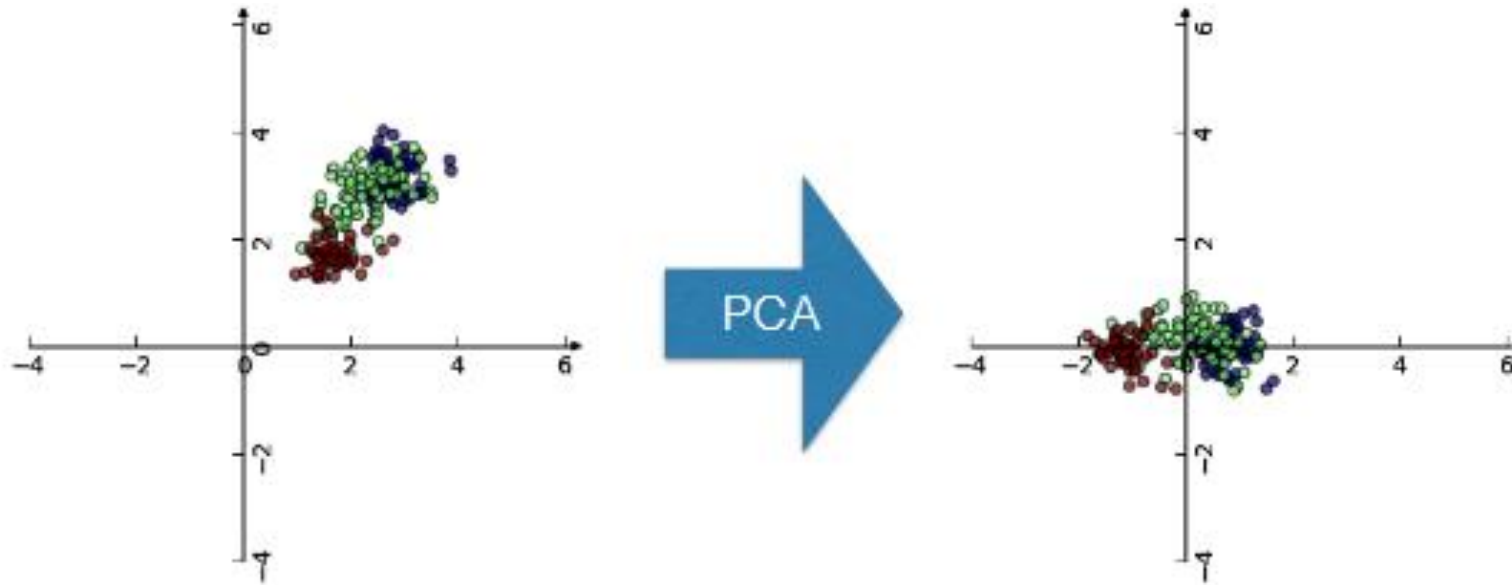
PRINCIPAL COMPONENTS

"Principal components" = directions of variance

PCA aligns principal components with the axes

Available as components_attribute of PCA object

Each row defines displacement from mean

The Principal Components