



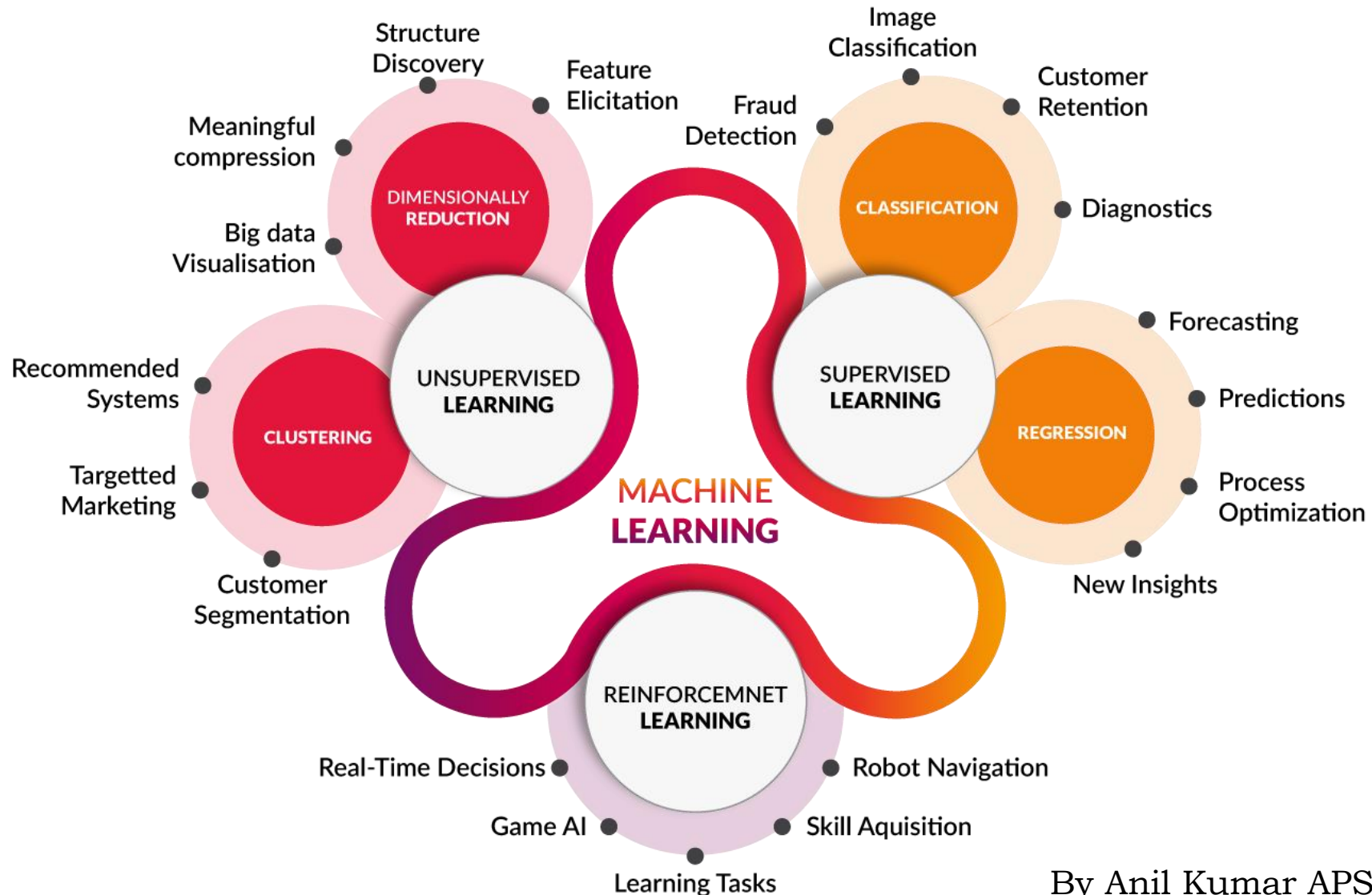
DIMENSIONALITY REDUCTION

DAY 9 AGENDA

Dimensionality
reduction

Principal Component
Analysis (PCA)

MACHINE LEARNING CATEGORIES



1d [1, 3, 3, 4, 5] – Vectors - Point

2d [[1,2]

[3,4]] → Matrix - Line

3d, Tensors –

180 * 180

1920 * 1080

4000 * 4000

STRUCTURE DISCOVERY

$$A + b + c + d + e = \text{out}$$

$$A + b = ab$$

$$C = 0$$

$$Ab + d + e = \text{out}$$

PRINCIPAL COMPONENT ANALYSIS

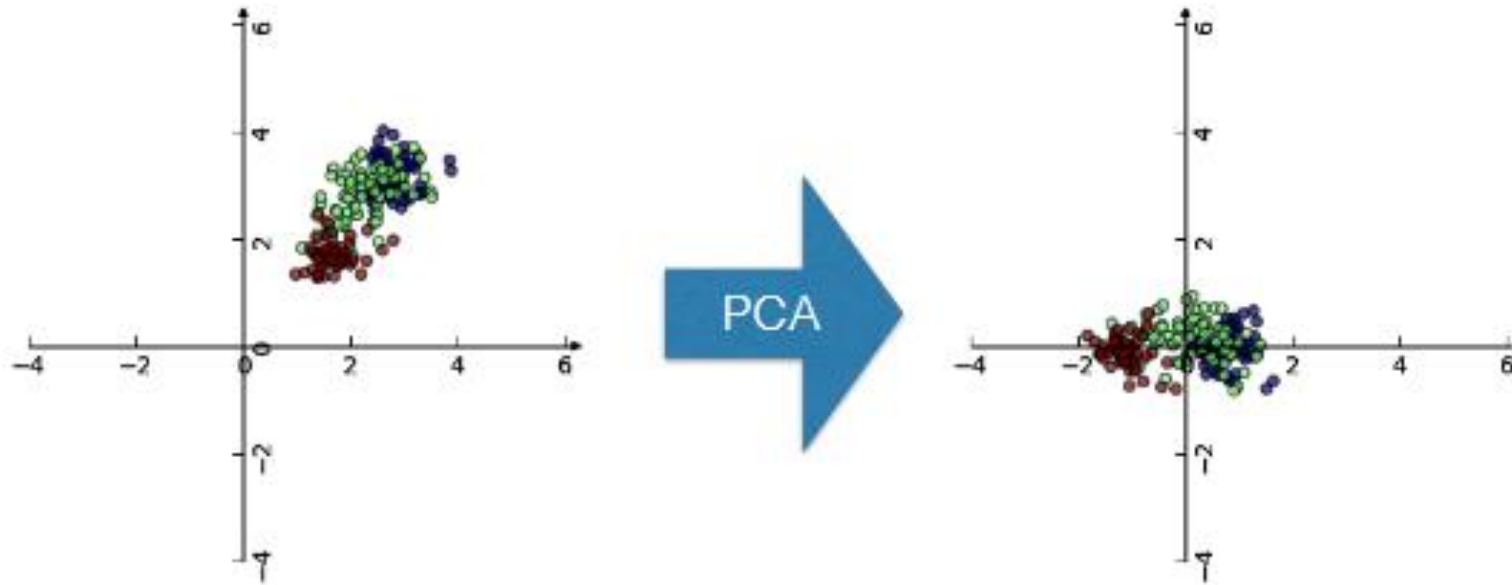
PCA = "Principal Component Analysis"

1. Fundamental dimension reduction technique
2. First step "decorrelation"
3. Second step reduces dimension

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

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

