



APSSDC

Andhra Pradesh State Skill Development Corporation



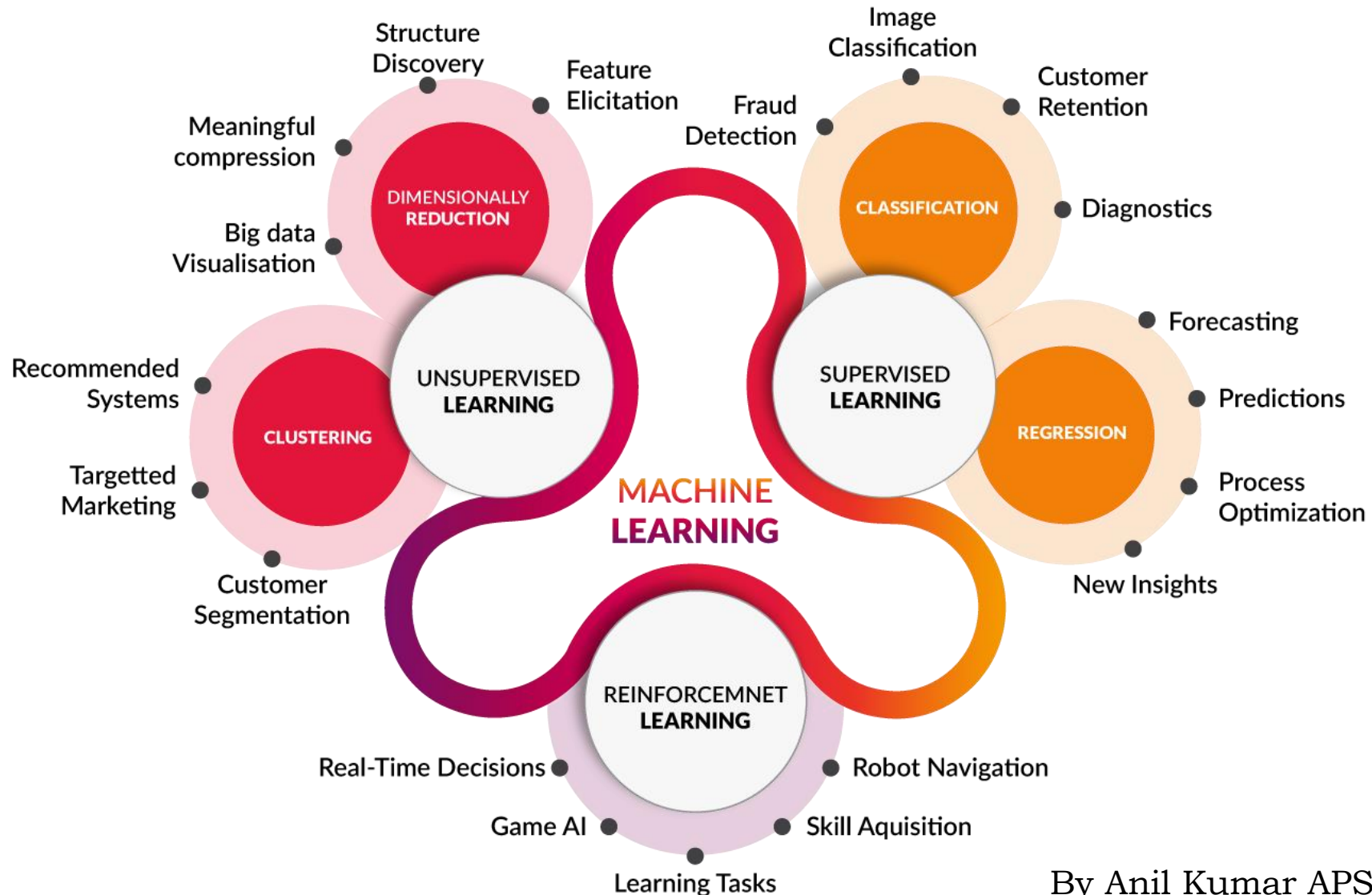
DIMENSIONALITY REDUCTION

DAY 9 AGENDA

Dimensionality
reduction

Principal Component
Analysis (PCA)

MACHINE LEARNING CATEGORIES



DIMENSION

Direction of VIEW

No. of coordinates

Number of views

Space

Shape of the array

1d – line, scale, if our data is behaving in one direction $[1,2,3,4,5]$ - Vector

2d - if our data is behaving in two direction – $[[1,2],[3,4]]$ - Matrix

3d - if our data is behaving in three direction , cube, volume -

3d,4d,5d,.....nd – Tensors - Tensorflow

FEATURE EXTRACTION

Correlation Matrix

RM, LSTAT \rightarrow MEDV

RM & LSTAT \rightarrow High correlation \rightarrow Internal correlation

Random Forest

D1 – RM - 3

D2 – RM

D3 – LSTAT - 1

D4 - RM

D5 – ZN - 1

D6 – NOX - 2

D7 - NOX

STRUCTURE DISCOVERY

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

$$A + b = ab$$

$$C = 0$$

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

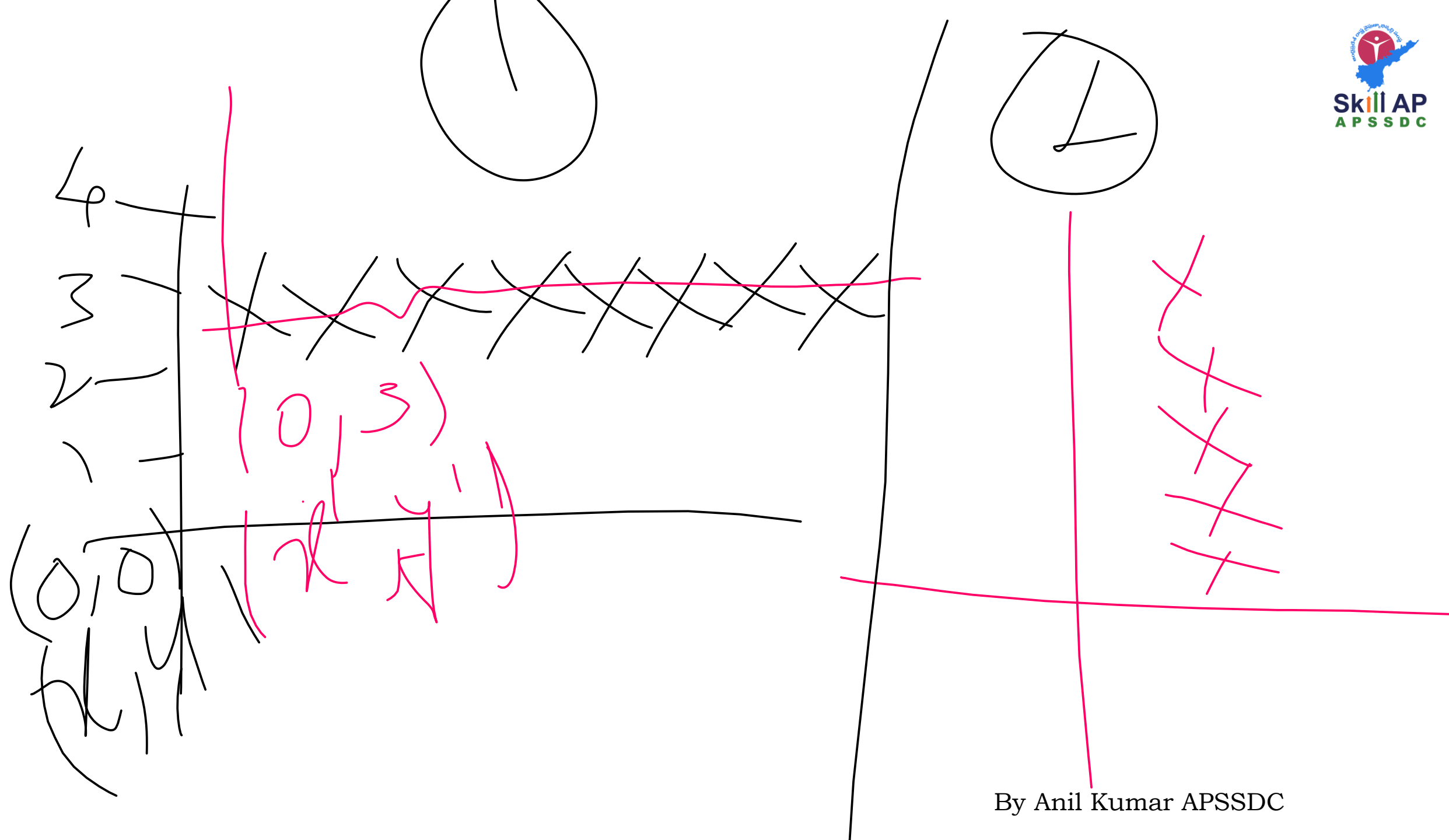
DIMENSION REDUCTION

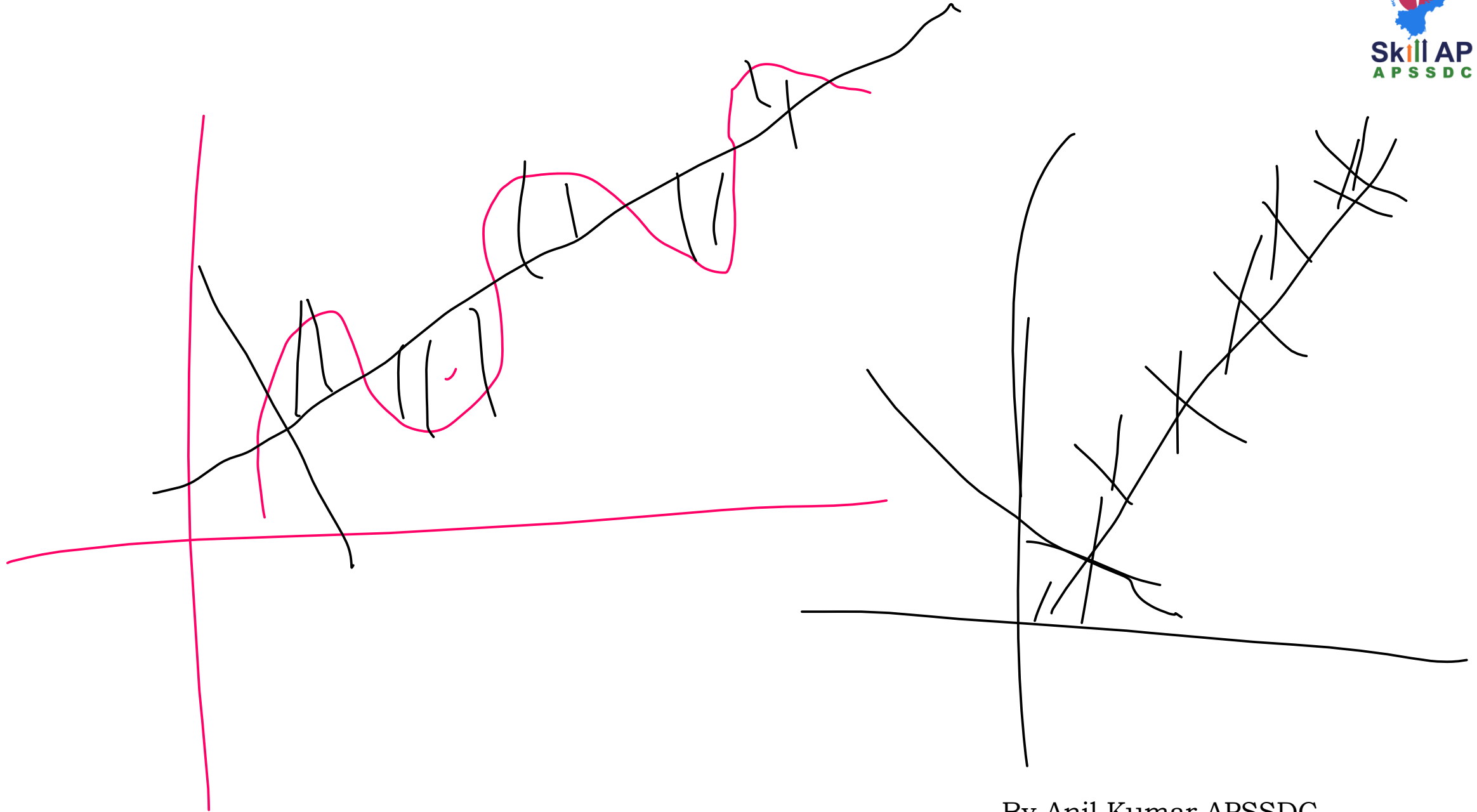
1. More efficient storage and computation
2. Remove less-informative "noise" features
3. Which cause problems for prediction tasks, e.g. classification, regression
4. Represents same data, using less features
5. Important part of machine-learning pipelines
6. Can be performed using PCA – Principle Component Analysis

PRINCIPAL COMPONENT ANALYSIS

PCA = "Principal Component Analysis"

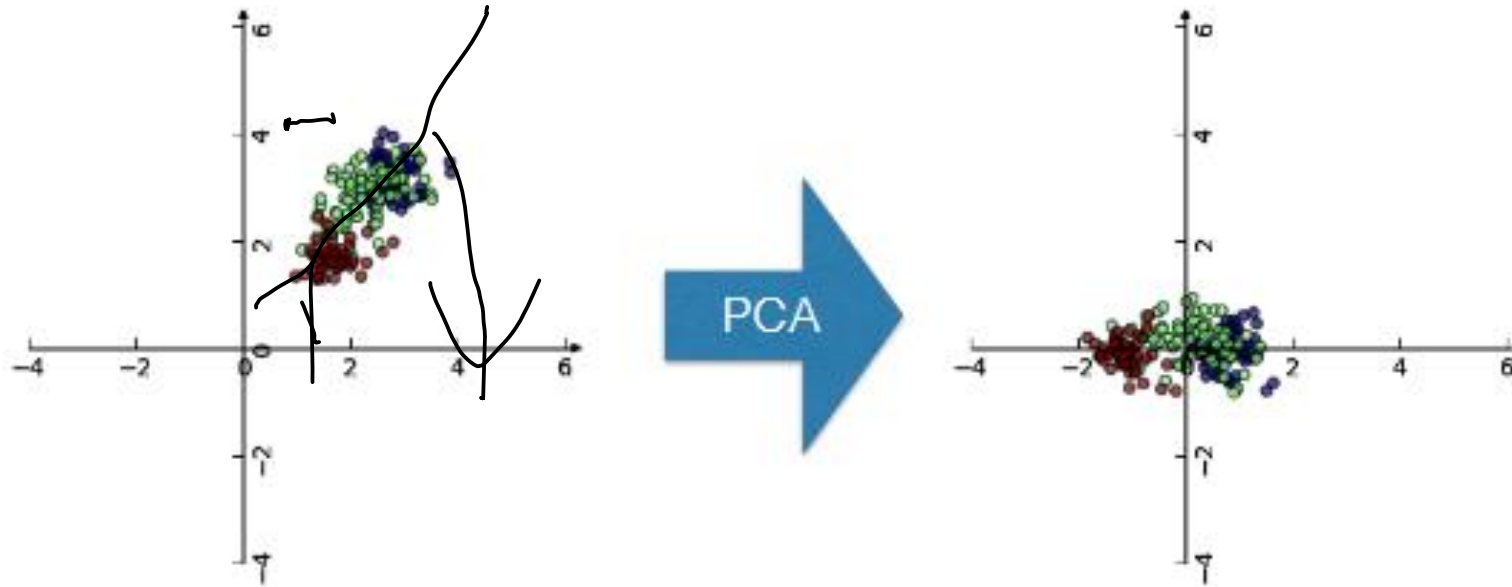
1. Fundamental dimension reduction technique
2. First step "decorrelation"
3. Second step reduces dimension





By Anil Kumar APSSDC

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

