CLUSTERING — EXAMPLE OF UNSUPERVISED LEARNING

DEFINITION : i) CLUSTERING IS A METHOD OF GROUPING DATA POINTS

ii) CLUSTERING IS A METHOD OF SEGMENTATION OF DATA POINTS

iii) CLUSTERING IS A METHOD OF CATEGORIZATION OF DATA POINTS

STEPS IN CLUSTERING PROCESS

1. FEATURE SELECTION
2. PROXIMITY MEASURE
3. CLUSTERING CRITERION
4. CLUSTERING ALGORITHM
5. VALIDATION OF RESULTS
6. INTERPRETATION OF RESULTS

Possible Clustering Applications

Clustering algorithms can be applied in many fields, for instance:

1. *Marketing*: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records;

2. *Biology*: classification of plants and animals given their features;

3. *Libraries*: book ordering;

4. *Insurance*: identifying groups of motor insurance policy holders with a high average claim cost; identifying frauds;

5. *City-planning*: identifying groups of houses according to their house type, value and geographical location;

6. *Earthquake studies*: clustering observed earthquake epicenters to identify dangerous zones;

7. *WWW*: document classification; clustering weblog data to discover groups of similar access patterns.