Artificial Intelligence

BY

DR. ANUPAM GHOSH

DATE: .9TH SEPT, 2025

Email: anupam.ghosh@rediffmail.com

https://www.nsec.ac.in/fps/faculty.php?id=138

Google Scholar
https://scholar.google.com/citations?user=uejpl3kAAAAJ&hl=en

LinkedIn Profile:
https://www.linkedin.com/in/anupam-ghosh-1504273b/

Research Profile:
https://www.researchgate.net/profile/Anupam_Ghosh14

What is Artificial Intelligence?

- The power of a machine to copy intelligent human behaviour

Artificial Intelligence Tests: Turing Test

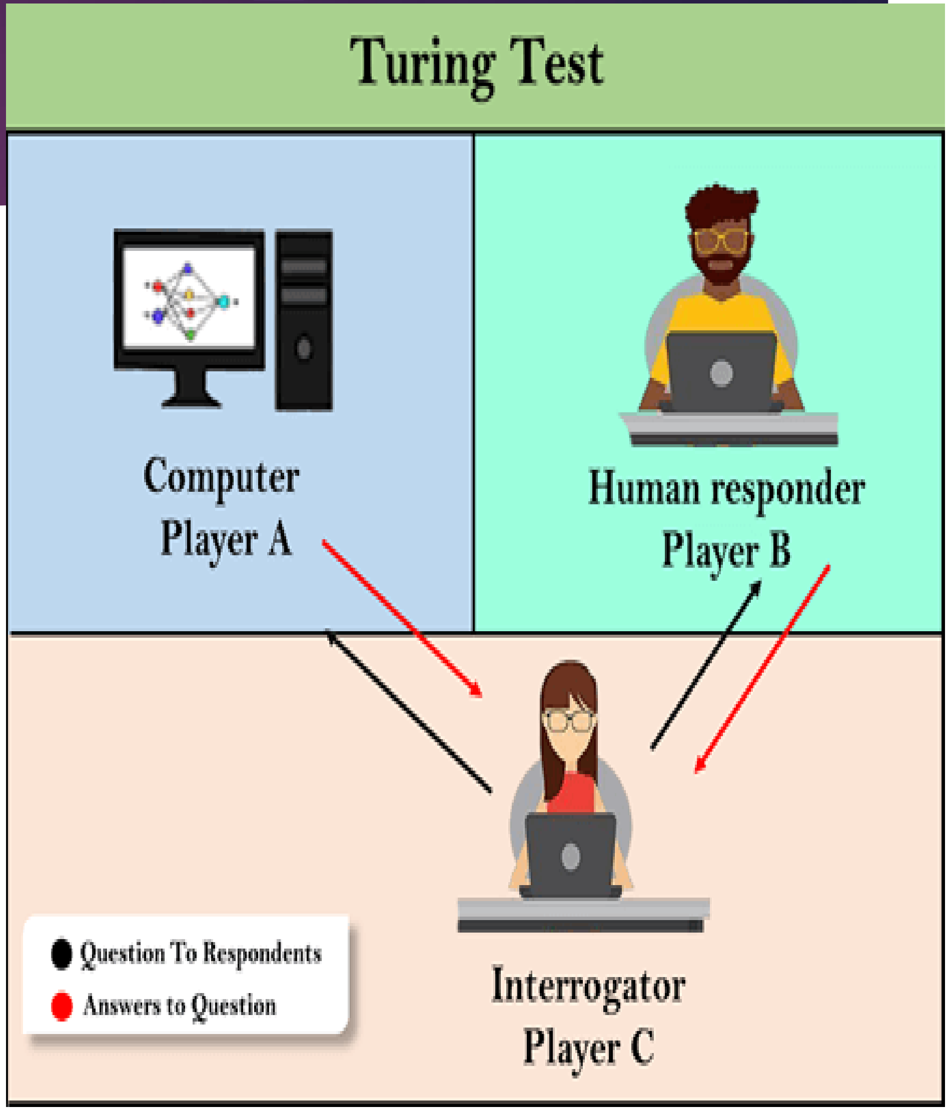
- Developed by Alan Turing
- Involves an interpreter, a human, and a computer.
- The computer and human have separate conversations with the interpreter.
- If the interpreter can't guess which is the computer or if the interpreter gets it wrong then the computer has Artificial Intelligence.

Turing Test in AI

- In 1950, Alan Turing introduced a test to check whether a machine can think like a human or not, this test is known as the Turing Test. In this test, Turing proposed that the computer can be said to be an intelligent if it can mimic human response under specific conditions.
- Consider, Player A is a computer, Player B is human, and Player C is an interrogator. Interrogator is aware that one of them is machine, but he needs to identify this on the basis of questions and their responses.

The questions and answers can be like:

- **Interrogator:** Are you a computer?
- **Player A (Computer):** No
- **Interrogator:** Multiply two large numbers such as (256896489*456725896)
- **Player A:** Long pause and give the wrong answer.
- In this game, if an interrogator would not be able to identify which is a machine and which is human, then the computer passes the test successfully, and the machine is said to be intelligent and can think like a human.



Features required for a machine to pass the Turing test:

- **Natural language processing:** NLP is required to communicate with Interrogator in general human language like English.
- **Knowledge representation:** To store and retrieve information during the test.
- **Automated reasoning:** To use the previously stored information for answering the questions.
- **Machine learning:** To adapt new changes and can detect generalized patterns.
- **Vision (For total Turing test):** To recognize the interrogator actions and other objects during a test.
- **Motor Control (For total Turing test):** To act upon objects if requested.

Intelligence - some definitions

- Intelligence:- ability to adapt oneself adequately to relatively new situations in life. (R. Pintner).
- Intelligence:- having learned or the ability to learn to adjust oneself to the environment. (Colvin)
- Intelligence:- the ability to carry out abstract thinking. (Terman)
- Intelligence:- innate general cognitive ability (Burt)
- Intelligence:- appropriate and adaptable behaviour in given circumstances. (Psihologija, group of authors, SK, Zagreb, 1992)
- Intelligence:- manifests itself only relative to specific social and cultural contexts. (J. Weizenbaum, 1975)

- A branch of computer science: Mathematical Sciences -> Computer Science -> Artificial Intelligence
- The branches of Artificial Intelligence (according to Association of Computing Machinery, ACM):

(1) General AI (cognitive modelling, philosophical foundations)

(2) Expert systems and applications

(3) Automated programming

(4) Deduction and theorem proving

(5) Formalisms and methods for knowledge representation

(6) Machine learning

(7) Understanding and processing of natural and artificial languages

(8) Problem solving, control methods, and state space search

(9) Robotics

(10) Computer vision, pattern recognition, and scene analysis

(11) Distributed artificial intelligence

Goals of AI

- To Create Expert Systems: The systems which exhibit intelligent behaviour, learn, demonstrate, explain, and advice its users

- To Implement Human Intelligence in Machines: Creating systems that understand, think, learn, and behave like humans

Applications of AI

- **Gaming:** AI plays crucial role in strategic games such as chess, poker, tic-tac-toe, etc., where machine can think of large number of possible positions based on heuristic knowledge.
- **Natural Language Processing**: It is possible to interact with the computer that understands natural language spoken by humans.
- **Expert Systems**: There are some applications which integrate machine, software, and special information to impart reasoning and advising. They provide explanation and advice to the users.
- **Vision Systems**: These systems understand, interpret, and comprehend visual input on the computer. For example,

  o Doctors use clinical expert system to diagnose the patient.

  o Police use computer software that can recognize the face of criminal with the stored portrait made by forensic artist.

- **Speech Recognition**: Some intelligent systems are capable of hearing and comprehending the language in terms of sentences and their meanings while a human talks to it. It can handle different accents, slang words, noise in the background, change in human's noise due to cold, etc.
- **Handwriting Recognition**: The handwriting recognition software reads the text written on paper by a pen or on screen by a stylus. It can recognize the shapes of the letters and convert it into editable text.
- **Intelligent Robots**: Robots are able to perform the tasks given by a human. They have sensors to detect physical data from the real world such as light, heat, temperature, movement, sound, bump, and pressure. They have efficient processors, multiple sensors and huge memory, to exhibit intelligence. In addition, they are capable of learning from their mistakes and they can adapt to the new environment.

Types of Intelligence

- **Linguistic intelligence**: The ability to speak, recognize, and use mechanisms of phonology (speech sounds), syntax (grammar), and semantics (meaning). Ex: Narrators, Orators
- **Musical intelligence**: The ability to create, communicate with, and understand meanings made of sound, understanding of pitch, rhythm. Ex: Musicians, Singers, Composers
- **Logical-mathematical intelligence**: The ability of use and understand relationships in the absence of action or objects. Understanding complex and abstract ideas. Ex: Mathematicians, Scientists
- **Spatial intelligence**: The ability to perceive visual or spatial information, change it, and re-create visual images without reference to the objects, construct 3D images, and to move and rotate them. Ex: Map readers, Astronauts, Physicists
- **Bodily-Kinesthetic intelligence**: The ability to use complete or part of the body to solve problems or fashion products, control over fine and coarse motor skills, and manipulate the objects. Ex: Players, Dancers
- **Intra-personal intelligence**: The ability to distinguish among one's own feelings, intentions, and motivations.
- **Interpersonal intelligence**: The ability to recognize and make distinctions among other people's feelings, beliefs, and intentions. Ex: Mass Communicators, Interviewers

What’s involved in Intelligence?

- Ability to interact with the real world:

  To perceive, understand, and act

  e.g., speech recognition and understanding and synthesis, image understanding, ability to take actions, have an effect

- Reasoning and Planning:

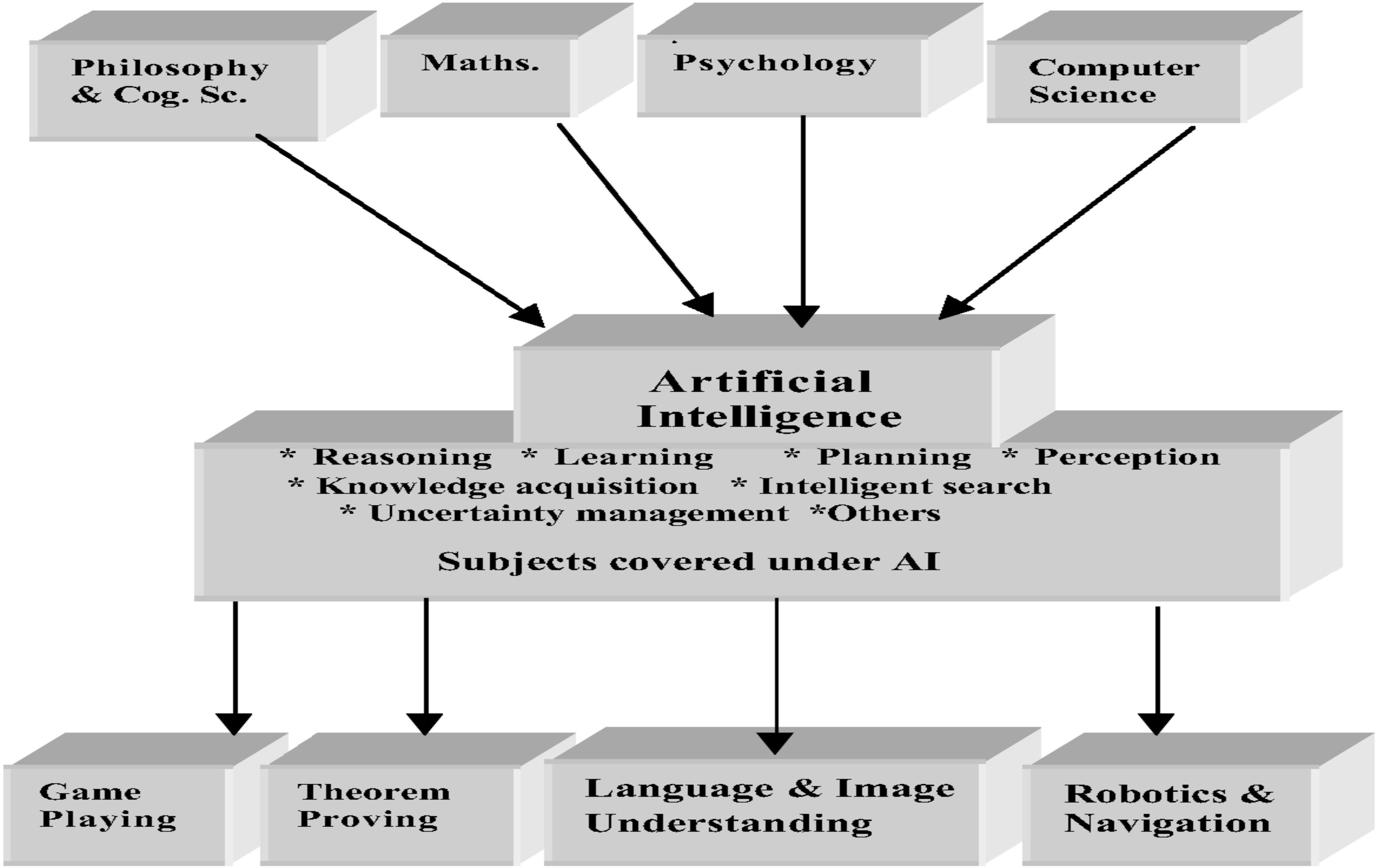
  Modelling the external world, given input solving new problems, planning, and making decisions, ability to deal with unexpected problems, uncertainties

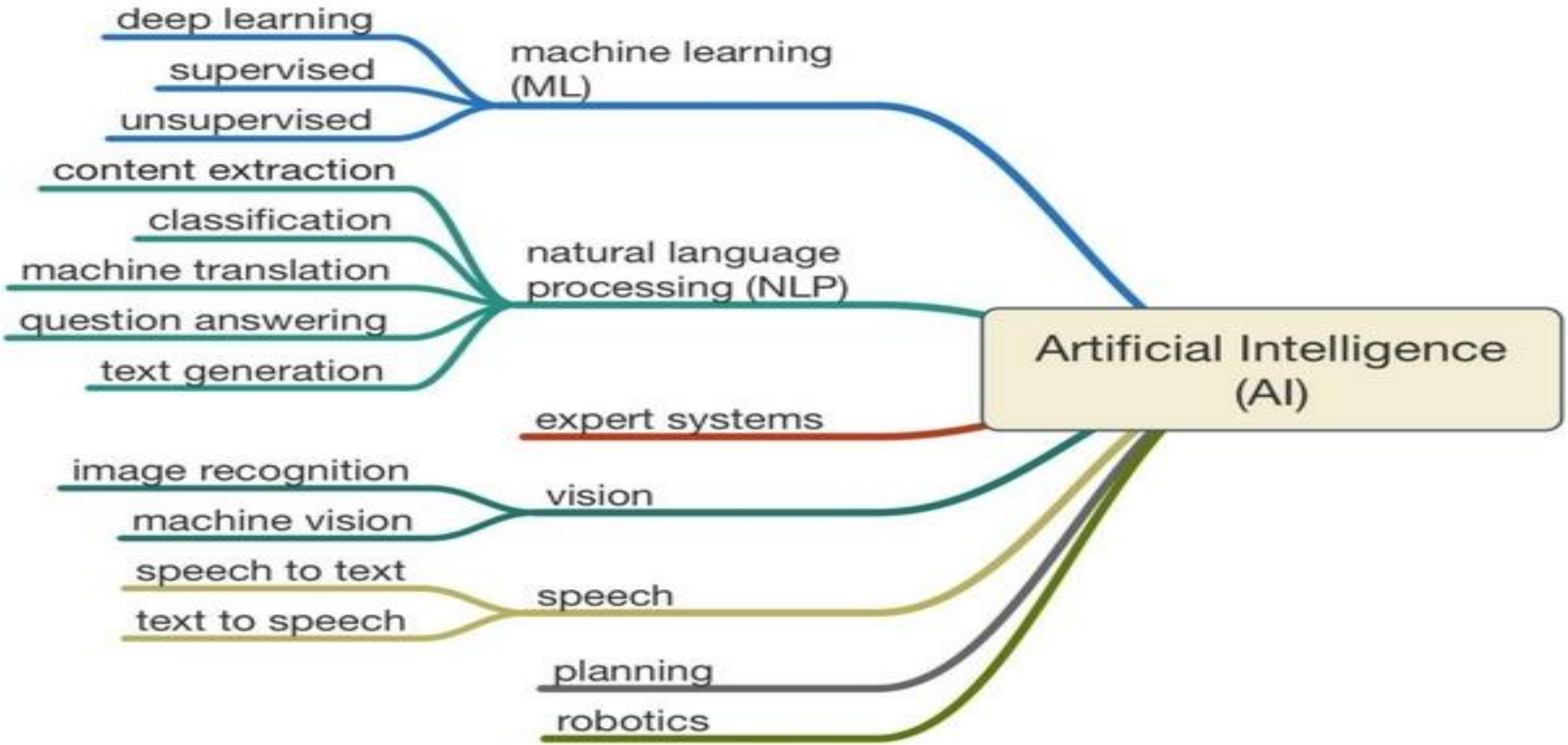
- Learning and Adaptation:

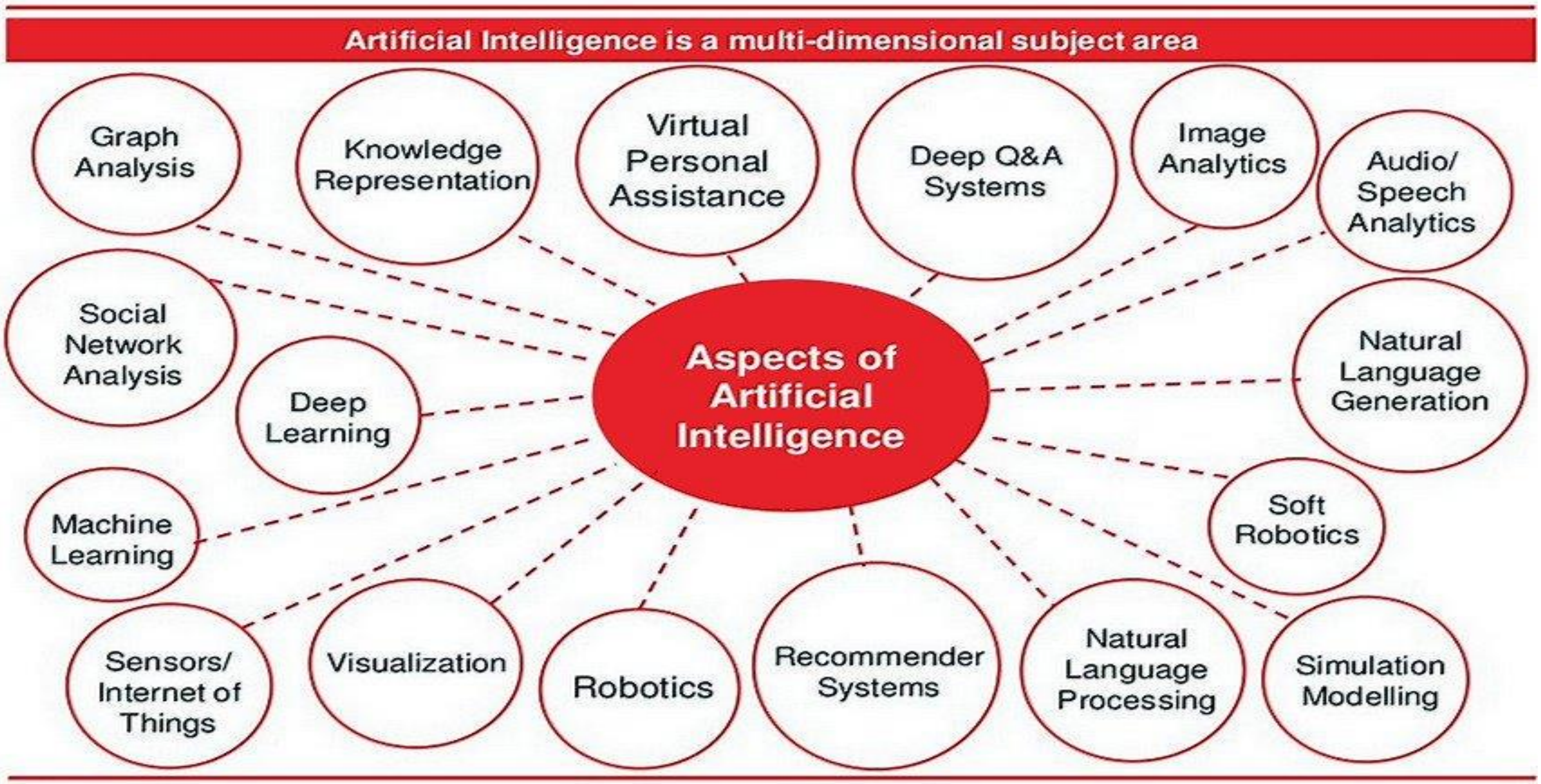
  we are continuously learning and adapting our internal models are always being “updated”

  e.g., a baby learning to categorize and recognize animals

PARENT DISCIPLINES OF AI




Artificial Intelligence (AI)
machine learning (ML)
deep learning
supervised
unsupervised
natural language processing (NLP)
content extraction
classification
machine translation
question answering
text generation
expert systems
vision
image recognition
machine vision
speech
speech to text
text to speech
planning
robotics


Artificial Intelligence is a multi-dimensional subject area
Aspects of Artificial Intelligence
Graph Analysis
Knowledge Representation
Virtual Personal Assistance
Deep Q&A Systems
Image Analytics
Audio/ Speech Analytics
Social Network Analysis
Deep Learning
Natural Language Generation
Machine Learning
Soft Robotics
Sensors/ Internet of Things
Visualization
Robotics
Recommender Systems
Natural Language Processing
Simulation Modelling

Example: Water Jug Problem

- Consider the following problem: A Water Jug Problem: You are given two jugs, a 4-gallon one and a 3-gallon one, a pump which has unlimited water which you can use to fill the jug, and the ground on which water may be poured. Neither jug has any measuring markings on it. How can you get exactly 2 gallons of water in the 4-gallon jug?

Solution: Phase-I

- State Representation and Initial State { we will represent a state of the problem as a tuple (x, y) where x represents the amount of water in the 4-gallon jug and y represents the amount of water in the 3-gallon jug. Note 0 <= x <= 4, and 0 <= y<= 3. Our initial state: (0,0)
- Goal Predicate state = (2,y) where 0 <= y <= 3.

Solution: Phase-II:--Operators

1. Fill 4-gal jug: (x,y) -> (4,y)
2. Fill 3-gal jug: (x,y) -> (x,3)
3. Empty 4-gal jug on ground: (x,y) -> (0,y)
4. Empty 3-gal jug on ground: (x,y) ->(x,0)
5. Pour water from 3-gal jug to fill 4-gal jug : (x,y) -> (4, y - (4 - x))

6. Pour water from 4-gal jug to fill 3-gal jug : (x,y) -> (x-(3-y), 3)
7. Pour all of water from 3-gal jug into 4-gal jug: (x,y) ->(x+y, 0)
8. Pour all of water from 4-gal jug into 3-gal jug: (x,y) ->(0, x+y)

Solution: Phase-III:--Rules

Gals in 4-gal jug	Gals in 3-gal jug	Rule Applied
0	0	
4	0	1. Fill 4
1	3	6. Pour 4 into 3 to fill
1	0	4. Empty 3
0	1	8. Pour all of 4 into 3
4	1	1. Fill 4
2	3	6. Pour into 3

Learning

BY

DR. ANUPAM GHOSH

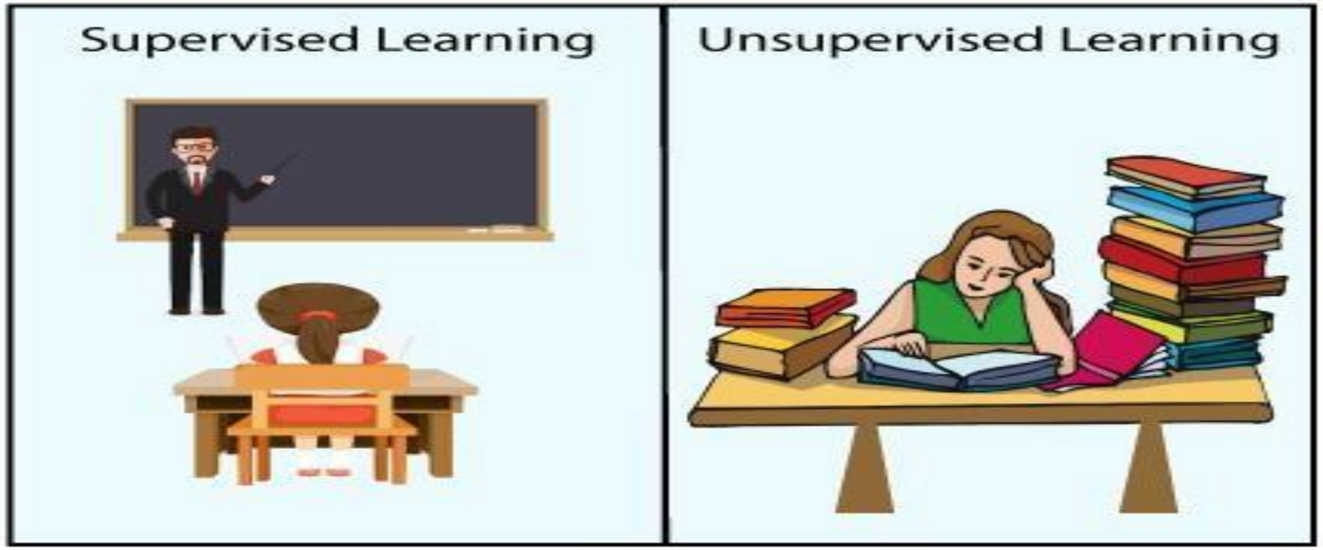
DATE: .9TH SEPT, 2025

Supervised vs. Unsupervised

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
...				
fruit n	...	...	...	...

Unsupervised learning:
Learning a model from unlabeled data.

Supervised learning:
Learning a model from labeled data.



Styles of Learning

Supervised	Unsupervised	Semi-Supervised	Reinforcement
• Data has **known labels** or output	• Labels or output unknown • Focus on **finding patterns and gaining insight** from the data	• Labels or output known for **a subset of data** • A blend of supervised and unsupervised learning	• Focus on **making decisions** based on previous experience • Policy-making with feedback
• Insurance underwriting • Fraud detection	• Customer clustering • Association rule mining	• Medical predictions (where tests and expert diagnoses are expensive, and only part of the population receives them)	• Game AI • Complex decision problems • Reward systems

sas

Clustering

- Clustering: Intuitively, finding clusters of points in the given data such that similar points lie in the same cluster
- Can be formalized using distance metrics in several ways
  - Group points into *k* sets (for a given *k*) such that the average distance of points from the centroid of their assigned group is minimized
    - Centroid: point defined by taking average of coordinates in each dimension.
  - Another metric: minimize average distance between every pair of points in a cluster
- Has been studied extensively in statistics, but on small data sets
  - Data mining systems aim at clustering techniques that can handle very large data sets
  - E.g., the Birch clustering algorithm (more shortly)

Clustering

- What is clustering?

Clustering

- Definition
  - Assignment of a set of observations into subsets so that observations in the same subset are similar in some sense
- Hard vs. Soft
  - Hard: same object can only belong to single cluster
  - Soft: same object can belong to different clusters
- Flat vs. Hierarchical
  - Flat: clusters are flat
  - Hierarchical: clusters form a tree
    - Agglomerative
    - Divisive

Similarity measures

- How to determine similarity between data points
  - using various distance metrics
- Let $\mathbf{x} = (x_1,\ldots,x_n)$ and $\mathbf{y} = (y_1,\ldots y_n)$ be n-dimensional vectors of data points of objects g_1 and g_2
  - g_1, g_2 can be two different genes in microarray data
  - n can be the number of samples

Summary of similarity measures

- Using different measures for clustering can yield different clusters
- Euclidean distance and correlation distance are the most common choices of similarity measure for microarray data
- Euclidean vs Correlation Example
  - g1 = (1,2,3,4,5)
  - g2 = (100,200,300,400,500)
  - g3 = (5,4,3,2,1)
  - Which genes are similar according to the two different measures?

K-MEANS CLUSTERING

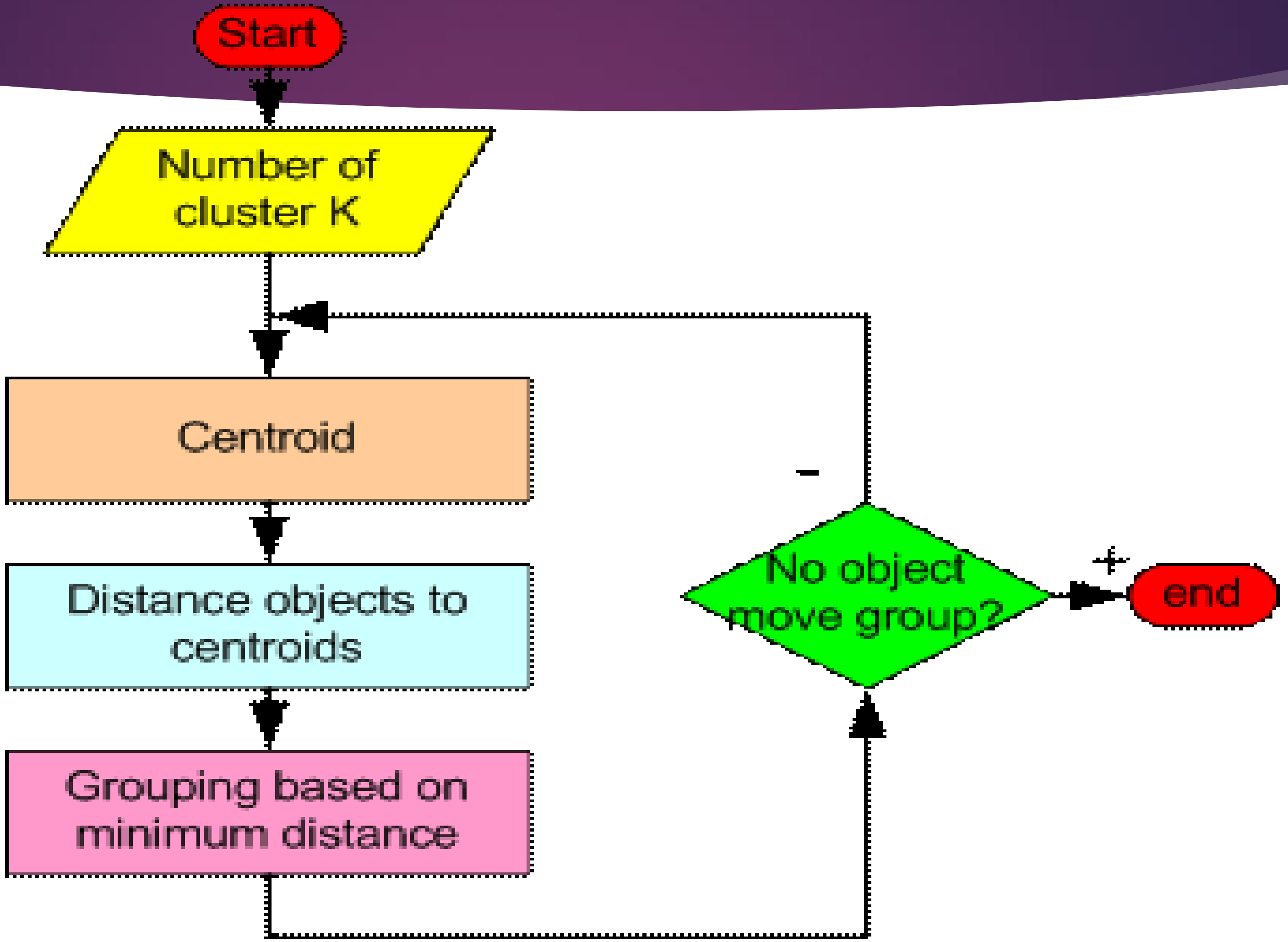
- The **k-means algorithm** is an algorithm to cluster *n* objects based on attributes into *k* partitions, where *k* < *n*.
- It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data.
- It assumes that the object attributes form a vector space.

- An algorithm for partitioning (or clustering) N data points into K disjoint subsets S_j containing data points so as to minimize the sum-of-squares criterion

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu_j|^2,$$

where x_n is a vector representing the the n^{th} data point and u_j is the geometric centroid of the data points in S_j.

How the K-Mean Clustering algorithm works?



- **Step 1:** Begin with a decision on the value of k = number of clusters .

- **Step 2**:

Put any initial partition that classifies the data into k clusters. You may assign the training samples randomly, or systematically as the following:

1.Take the first k training sample as single-element clusters

2. Assign each of the remaining (N-k) training sample to the cluster with the nearest centroid. After each assignment, recompute the centroid of the gaining cluster.

- **Step 3:** Take each sample in sequence and compute its distance from the centroid of each of the clusters.

If a sample is not currently in the cluster with the closest centroid, switch this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.

- **Step 4 .** Repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

A Simple example showing the implementation of k-means algorithm
(using K=2)

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Step 1:

Initialization: Randomly we choose following two centroids (k=2) for two clusters.

In this case the 2 centroid are: m1=(1.0,1.0) and m2=(5.0,7.0).

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

	Individual	Mean Vector
Group 1	1	(1.0, 1.0)
Group 2	4	(5.0, 7.0)

Step 2:

- Thus, we obtain two clusters containing:

  {1,2,3} and {4,5,6,7}.

- Their new centroids are:

$$m_1 = \left(\frac{1}{3}(1.0+1.5+3.0), \frac{1}{3}(1.0+2.0+4.0)\right) = (1.83, 2.33)$$

$$m_2 = \left(\frac{1}{4}(5.0+3.5+4.5+3.5), \frac{1}{4}(7.0+5.0+5.0+4.5)\right)$$

$$= (4.12, 5.38)$$

Individual	Centroid 1	Centroid 2
1	0	7.21
2 (1.5, 2.0)	1.12	6.10
3	3.61	3.61
4	7.21	0
5	4.72	2.5
6	5.31	2.06
7	4.30	2.92

$$d(m_1, 2) = \sqrt{|1.0-1.5|^2 + |1.0-2.0|^2} = 1.12$$

$$d(m_2, 2) = \sqrt{|5.0-1.5|^2 + |7.0-2.0|^2} = 6.10$$

Step 3:

- Now using these centroids we compute the Euclidean distance of each object, as shown in table.

- Therefore, the new clusters are:

  {1,2} and {**3**,4,5,6,7}

- Next centroids are: m1=(1.25,1.5) and m2 = (3.9,5.1)

Individual	Centroid 1	Centroid 2
1	1.57	5.38
2	0.47	4.28
3	2.04	1.78
4	5.84	1.84
5	3.15	0.73
6	3.78	0.54
7	2.74	1.08

- <u>Step 4</u> :

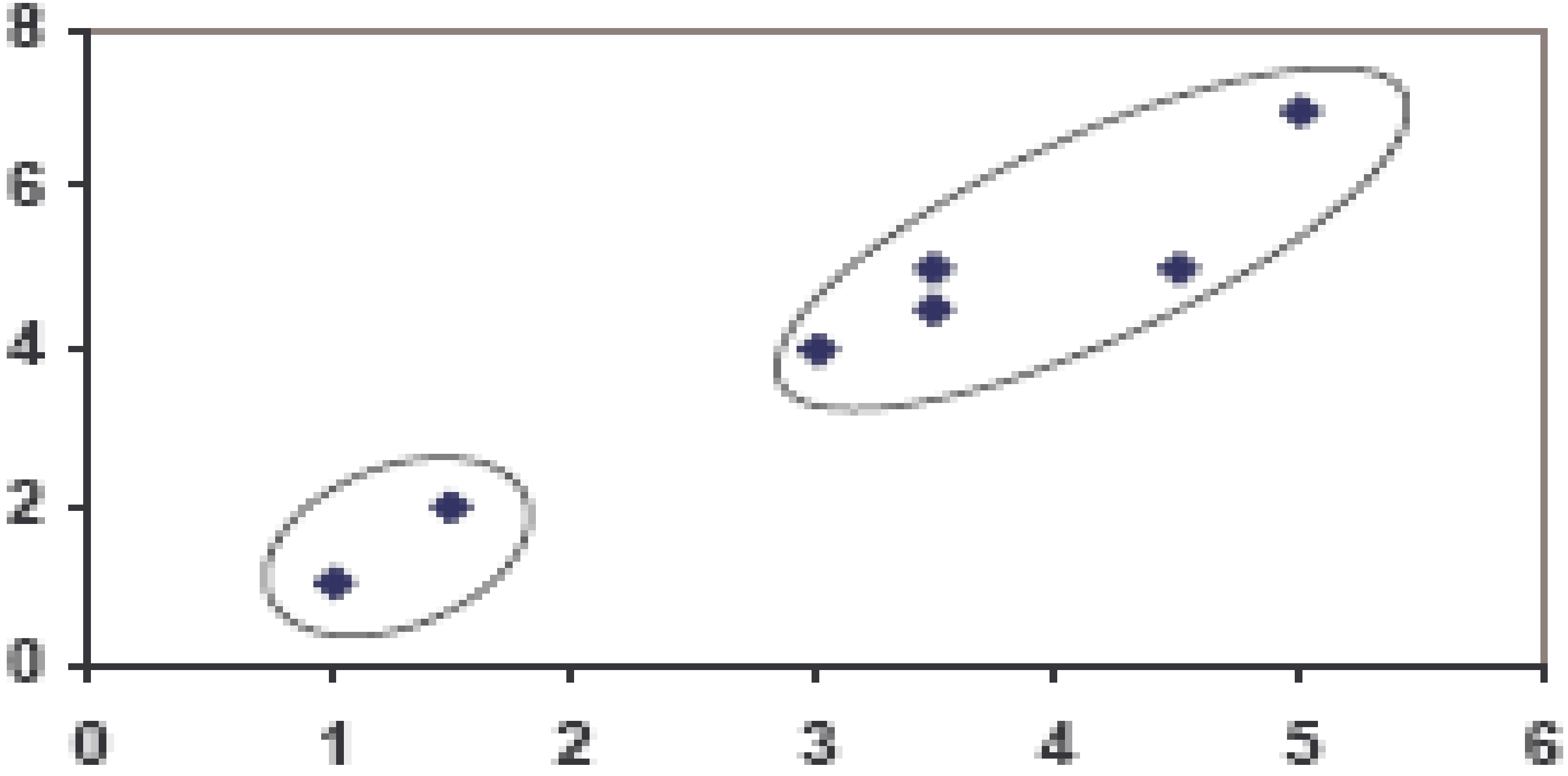
  The clusters obtained are:

  {1,2} and {3,4,5,6,7}

- Therefore, there is no change in the cluster.
- Thus, the algorithm comes to a halt here and final result consist of 2 clusters {1,2} and {3,4,5,6,7}.

Individual	Centroid 1	Centroid 2
1	0.56	5.02
2	0. 56	3.92
3	3.05	1.42
4	6.66	2.20
5	4.16	0.41
6	4.78	0.61
7	3.75	0.72

PLOT

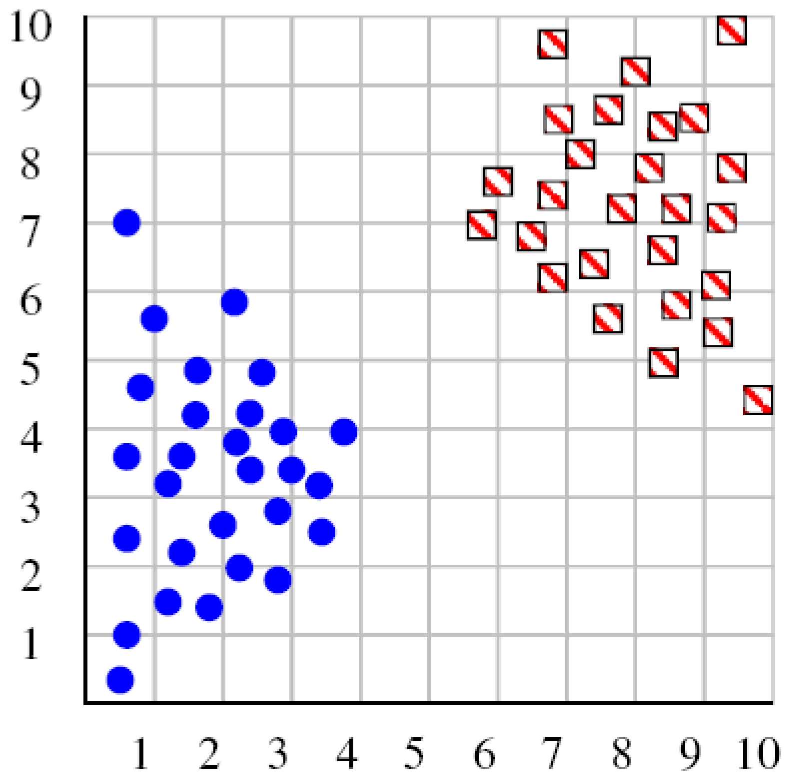


K-means: summary

- Algorithmically, very simple to implement
- *K*-means converges, but it finds a local minimum of the cost function
- Works only for numerical observations
- *K* is a user input; alternatively BIC (Bayesian information criterion) or MDL (minimum description length) can be used to estimate *K*
- Outliers can considerable trouble to *K*-means

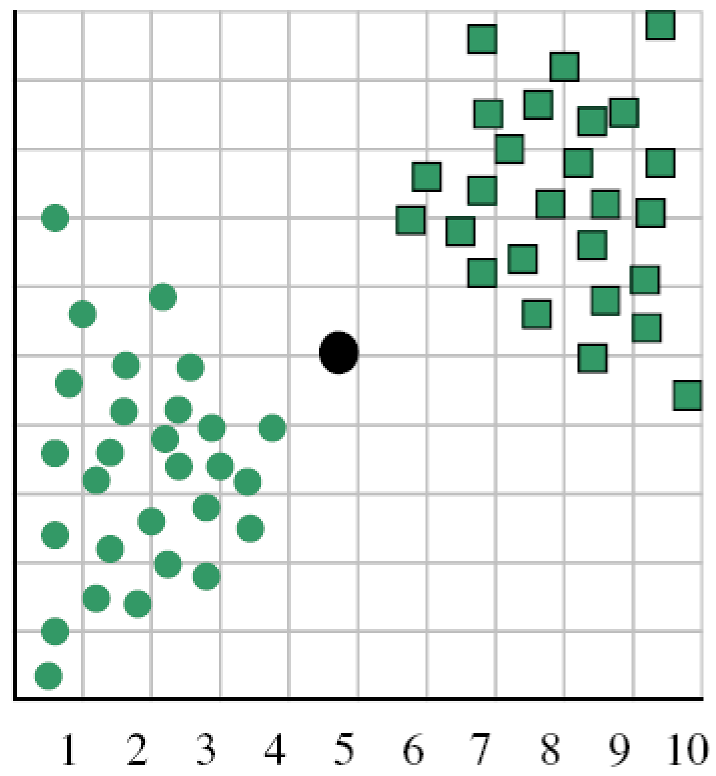
Deciding K

- Try a couple of K



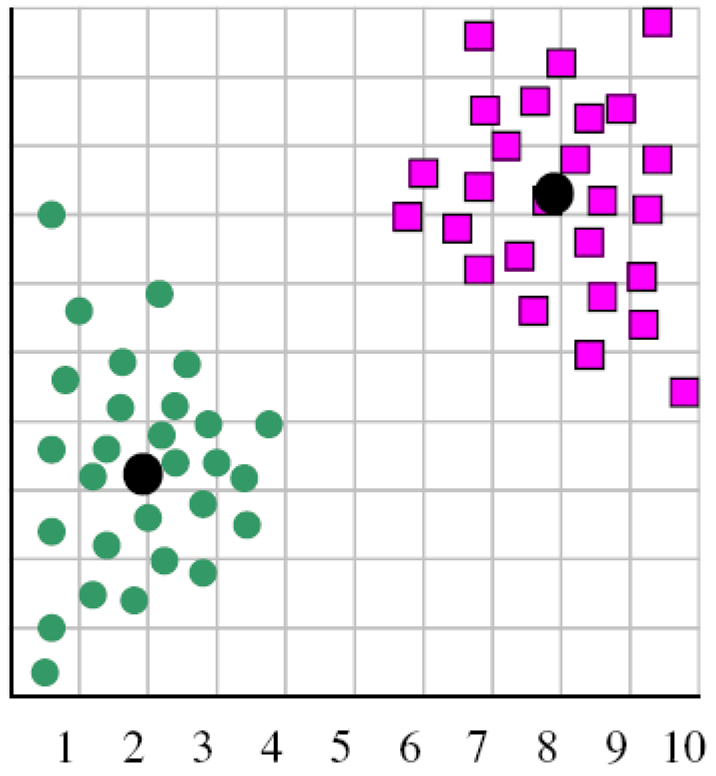
Deciding K

- When k = 1, the objective function is 873.0



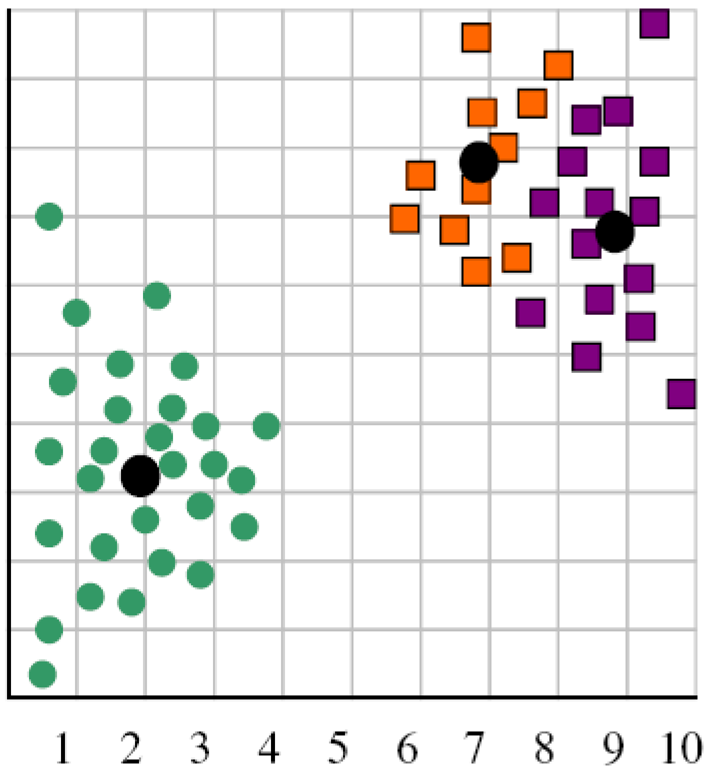
Deciding K

- When k = 2, the objective function is 173.1



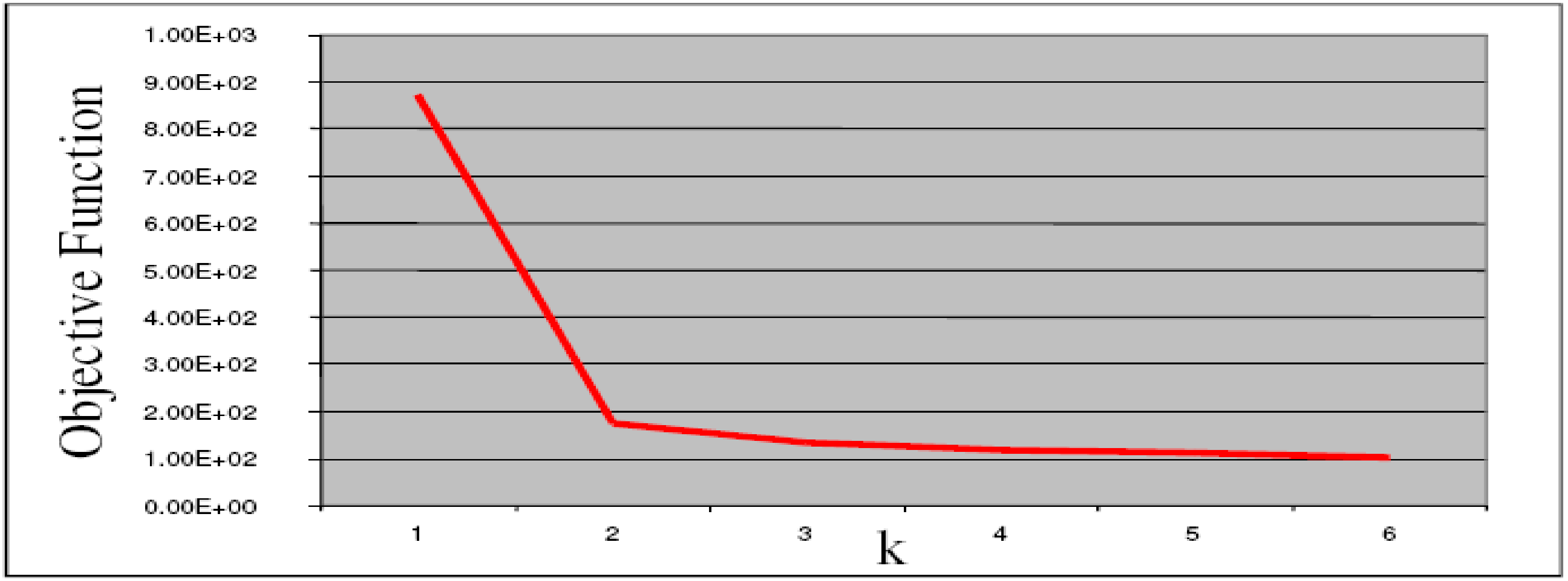
Deciding K

- When k = 3, the objective function is 133.6



Deciding K

- We can plot objective function values for k=1 to 6
- The abrupt change at k=2 is highly suggestive of two clusters
- "knee finding" or "elbow finding"
- Note that the results are not always as clear cut as in this toy example



Back

Weaknesses of K-Means Clustering

1. When the numbers of data are not so many, initial grouping will determine the cluster significantly.
2. The number of cluster, K, must be determined before hand. Its disadvantage is that it does not yield the same result with each run, since the resulting clusters depend on the initial random assignments.
3. We never know the real cluster, using the same data, because if it is inputted in a different order it may produce different cluster if the number of data is few.
4. It is sensitive to initial condition. Different initial condition may produce different result of cluster. The algorithm may be trapped in the *local optimum*.

Applications of K-Means Clustering

- It is relatively *efficient and fast*. It computes result at **O(tkn)**, where n is number of objects or points, k is number of clusters and t is number of iterations.
- k-means clustering can be applied to *machine learning or data mining*
- *Used on acoustic data in speech understanding to convert waveforms into one of k categories (known as Vector Quantization or Image Segmentation).*
- *Also used for choosing color palettes on old fashioned graphical display devices and Image Quantization.*

Geometric Distance Measure

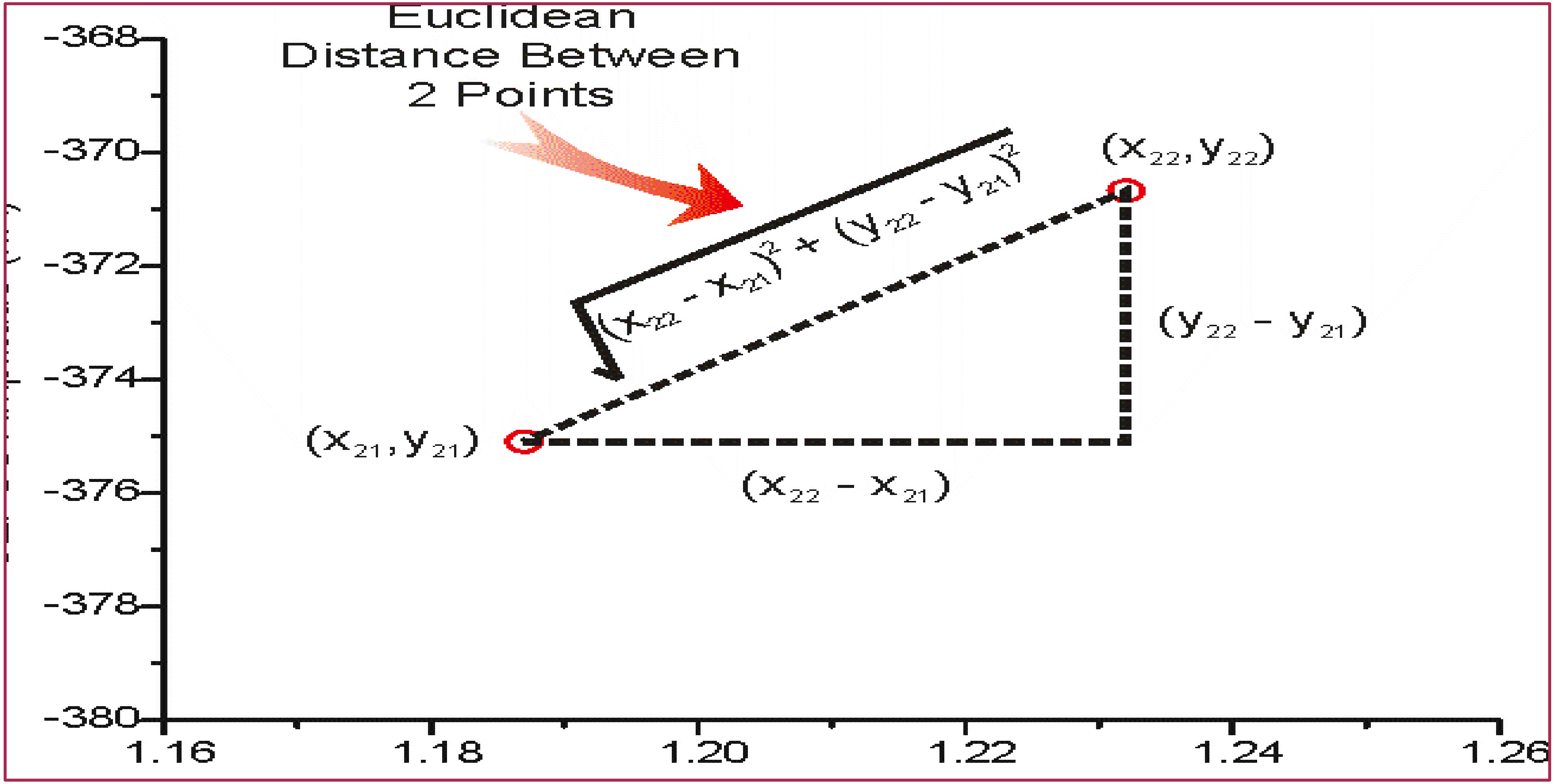
- Geometric distance metrics, primarily, tends to measure the similarity between two or more vectors solely based on the distance between two points in multi-dimensional space.

- The examples of such type of geometric distance measures are Minkowski distance, Euclidean distance and Manhattan distance.

- Minkowski distance is the general form of Euclidean and Manhattan distance. Mathematically, it can be represented as the following:

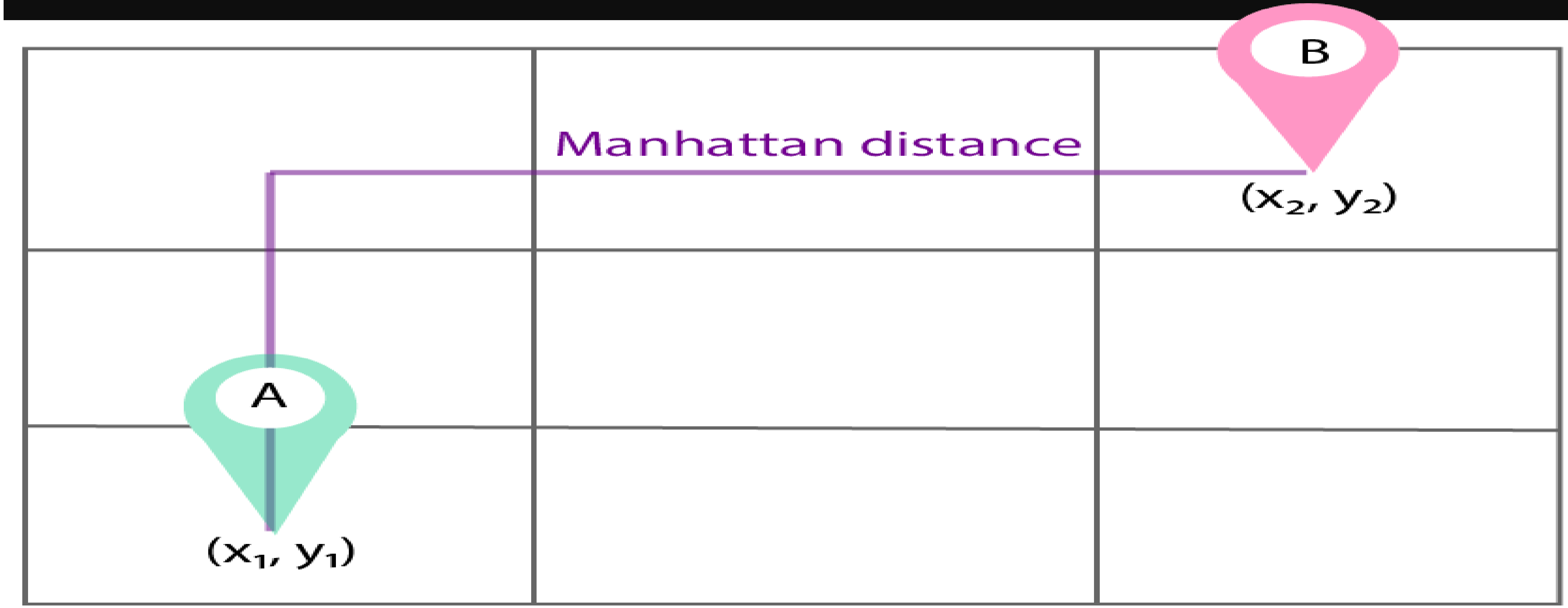
$$D(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{l=1}^{d} |x_{il} - x_{jl}|^{1/p} \right)^{p}$$

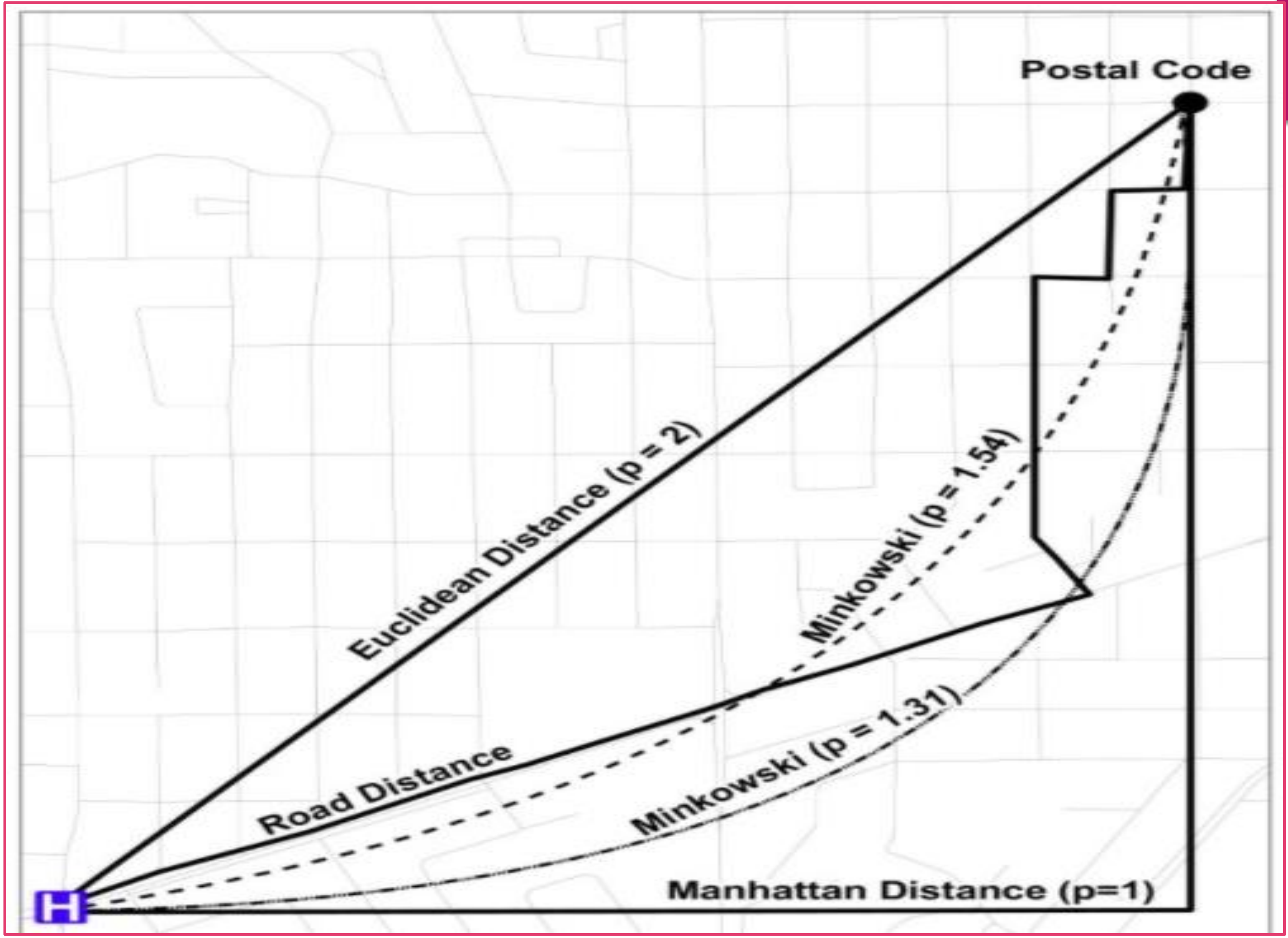
$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$



Manhattan Distance

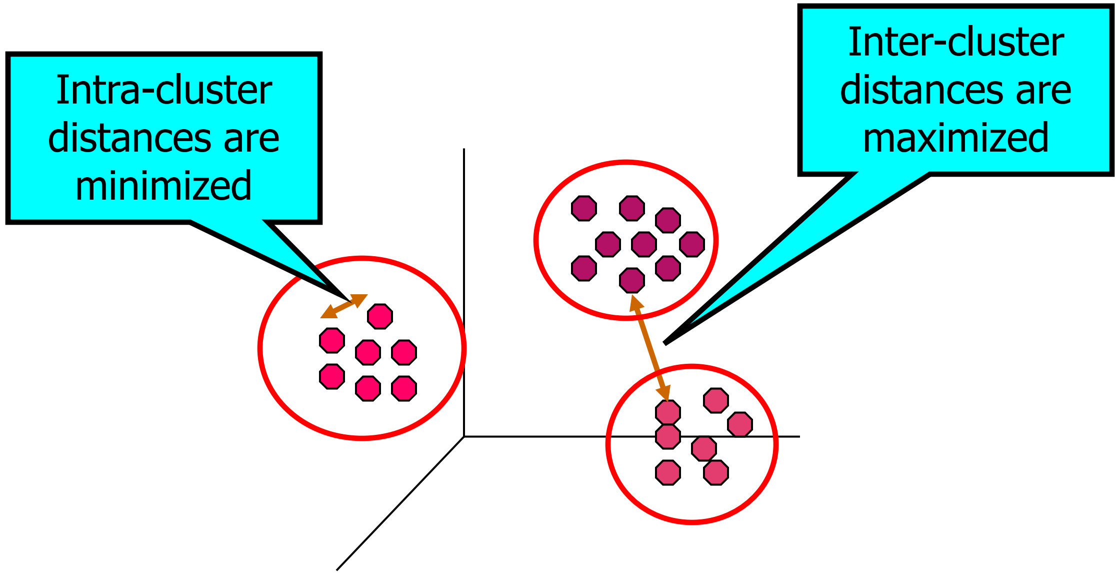
- If you want to find Manhattan distance between two different points (x1, y1) and (x2, y2) such as the following, it would look like the following:
- Manhattan distance = (x2 – x1) + (y2 – y1)




Postal Code
Euclidean Distance (p = 2)
Minkowski (p = 1.54)
Road Distance
Minkowski (p = 1.31)
Manhattan Distance (p=1)
H

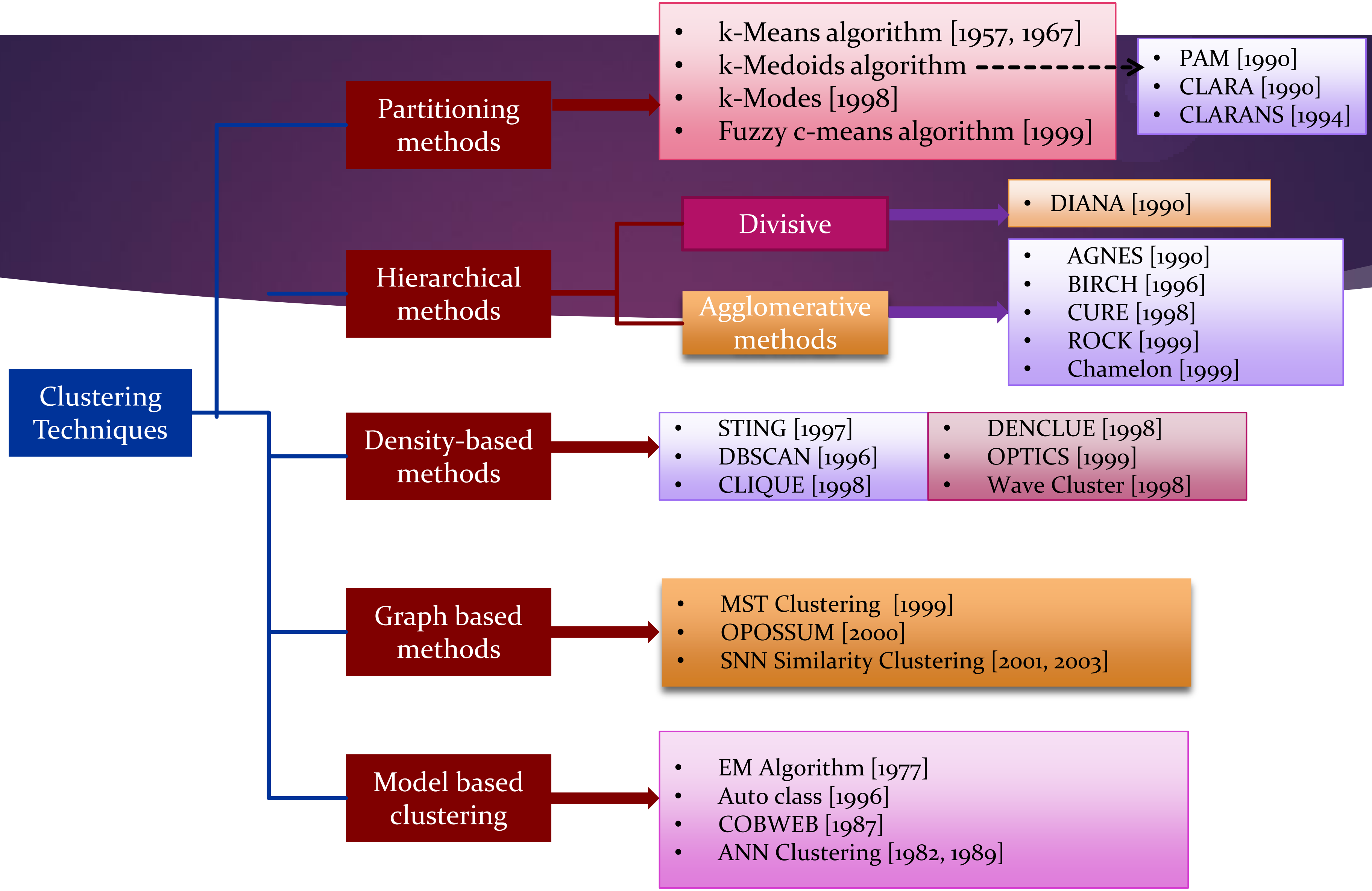
What is Cluster Analysis?

- Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



Quality: What Is Good Clustering?

- A <u>good clustering</u> method will produce high quality clusters with
  - high <u>intra-class</u> similarity
  - low <u>inter-class</u> similarity
- The <u>quality</u> of a clustering result depends on both the similarity measure used by the method and its implementation
- The <u>quality</u> of a clustering method is also measured by its ability to discover some or all of the <u>hidden</u> patterns



Clustering Techniques
Partitioning methods
k-Means algorithm [1957, 1967]
k-Medoids algorithm
k-Modes [1998]
Fuzzy c-means algorithm [1999]
PAM [1990]
CLARA [1990]
CLARANS [1994]
Hierarchical methods
Divisive
DIANA [1990]
Agglomerative methods
AGNES [1990]
BIRCH [1996]
CURE [1998]
ROCK [1999]
Chamelon [1999]
Density-based methods
STING [1997]
DBSCAN [1996]
CLIQUE [1998]
DENCLUE [1998]
OPTICS [1999]
Wave Cluster [1998]
Graph based methods
MST Clustering [1999]
OPOSSUM [2000]
SNN Similarity Clustering [2001, 2003]
Model based clustering
EM Algorithm [1977]
Auto class [1996]
COBWEB [1987]
ANN Clustering [1982, 1989]

K-MEDOIDS CLUSTERING

The mean in k-means clustering is sensitive to outliers. Since an object with an extremely high value may substantially distort the distribution of data.

Hence we move to k-medoids.

Instead of taking mean of cluster we take the most centrally located point in cluster as it’s center.

These are called medoids.

k-Medoids

- It is also called as Partitioning Around Medoid algorithm.
- A medoid can be defined as the point in the cluster, whose dissimilarities with all the other points in the cluster is minimum.
- The dissimilarity of the medoid(Ci) and object(Pi) is calculated by using $\boldsymbol{E = |Pi - Ci|}$

Algorithm:

1. Initialize: select k random points out of the n data points as the medoids.
2. Associate each data point to the closest medoid by using any common distance metric methods.
3. While the cost decreases:

For each medoid m, for each data o point which is not a medoid:

1. Swap m and o, associate each data point to the closest medoid, recompute the cost.
2. If the total cost is more than that in the previous step, undo the swap.

K-MEDOIDS - PAM ALGORITHM

PAM stands for **Partitioning Around Medoids**.

GOAL: To find Clusters that have minimum average dissimilarity between objects that belong to same cluster.

ALGORITHM:

1. Start with initial set of medoids.
2. Iteratively replace one of the medoids with a non-medoid if it reduces total sum of SSE of resulting cluster.

SSE is calculated as below:

$$SSE(X) = \sum_{i=1}^{K} \sum_{x \in C_i} dist^2(m_i, x)$$

Where k is number of clusters and x is a data point in cluster C_i and M_i is medoid of C_i

ADVANTAGES and DISADVANTAGES of PAM

Advantages:

PAM is more flexible as it can use any similarity measure.

PAM is more robust than k-means as it handles noise better.

Disadvantages:

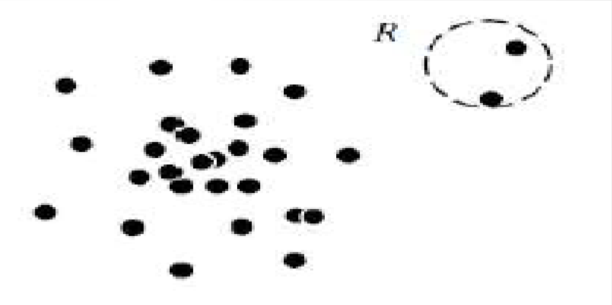
PAM algorithm for K-medoid clustering works well for dataset but cannot scale well for large data set due to high computational overhead.

PAM COMPLEXITY : $O(k(n-k)^2)$ this is because we compute distance of n-k points with each k point, to decide in which cluster it will fall and after this we try to replace each of the medoid with a non medoid and find it's distance with n-k points.

Outlier Detection in Clustering

- An **outlier** is a data object that deviates significantly from the rest of the objects, as if it were generated by a different mechanism.
- Outliers are referred as "abnormal" data.
- Outliers are different from noisy data. Noise is a random error or variance in a measured variable.
- Outliers are interesting because they are suspected of not being generated by the same mechanisms as the rest of the data.
- Outlier detection is also related to novelty detection in evolving data sets.

The objects in region R are outliers



Types of Outliers

- ✓ Global Outliers
- ✓ Contextual Outliers
- ✓ Collective Outliers

Outlier detection techniques.

Supervised, Semi-Supervised, and Unsupervised Methods

- Supervised methods model data normality and abnormality.
- In some application scenarios, objects labeled as "normal" or "outlier" are not available. Thus, an unsupervised learning method has to be used.
- In some cases where only a small set of the normal and/or outlier objects are labeled, but most of the data are unlabeled.

Statistical Methods, Proximity-Based Methods and Clustering-Based Methods

- Statistical methods (also known as model-based methods) make assumptions of data normality.
- The effectiveness of proximity-based methods relies heavily on the proximity (or distance) measure used.
- Clustering-based methods assume that the normal data objects belong to large and dense clusters, whereas outliers belong to small or sparse clusters, or do not belong to any clusters.

It will be Continued....