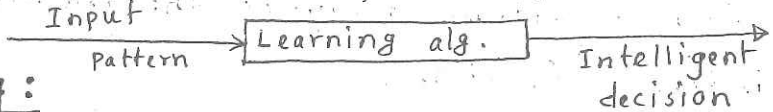
Machine Learning:

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that recognize complex patterns and make intelligent decisions based on input data.



Learning:

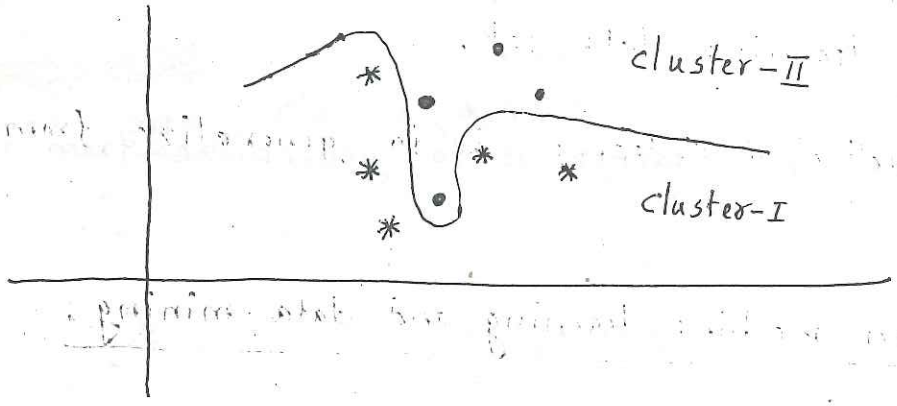
- Unsupervised learning
- Supervised learning

Unsupervised learning:

In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data.

characteristics

- Input data.
- No labeled data or teacher signal.



Approaches to unsupervised learning:

- K-mean clustering
- Modified K-mean clustering
- Hierarchical clustering
- Fuzzy C-Mean clustering

e) - Organizing map (SOM)

Supervised Learning:

Supervised learning is the machine learning task of inferring a function from labeled training data

Training data:
- Input object (a vector)
- Desired output value (supervisory signal)

A supervised learning algorithm analyzes the training data and produces an inferred function.

Inferred function is called
- classifier (if output is discrete)
- Regression function (if output is continuous)

Inferred function should predict the correct output value for any valid input object.

Generalization:

- Generalization in this context is the ability of an algorithm to perform accurately on new, unseen examples after having trained on a learning data set.
- The core objective of a learner is to generalize from its experience.

Difference between machine learning and data mining:

Machine learning:

Focuses on prediction based on known properties learned from the training data.

Data mining:

Discovery of unknown properties on the data.

Approaches to Supervised learning:

- Artificial neural network.
- Bayesian network.
- Support Vector machines.
- Decision tree learning.
- Hidden Markov Model.
- Association rule learning.

Unsupervised learning approaches:

K-Mean clustering:

Classify the following data into two clusters using the k-mean algorithm.

(3,4) (6,7) (4,5) (5,7) (2,6)

Ans:

(step1) Initial step: (Centroid Selection)

class-I: (3,4)

class-II: (4,5)

step 2, 3, 4:

Iteration1: Data	class-I (3,4)	class-II (4,5)	class
(3,4)	0	~~2.236~~ 1.414	class-I
(6,7)	4.242	2.828	class-II
(4,5)	~~2.236~~ 1.414	0	class-II
(5,7)	3.605	2.236	class-II
(2,6)	2.236	2.236	class-I

class-I : $(3,4)\ (2,6) \equiv \left(\frac{3+2}{2}, \frac{4+6}{2}\right)$

$\equiv (2.5, 5)$

class-II : $(6,7)\ (4,5)\ (5,7) \equiv \left(\frac{6+4+5}{3}, \frac{7+5+7}{3}\right)$

$\equiv (5, 6.33)$

Iteration 2:

Data	class-I (2.5, 5)	class-II (5, 6.33)	class
(3,4)	1.12	3.07	class-I
(6,7)	4.03	1.208	class-II
(4,5)	1.5	1.66	class-I
(5,7)	3.2	0.67	class-II
(2,6)	1.12	3.01	class-I

class-I : $(3,4)\ (4,5)\ (2,6) \equiv \left(\frac{3+4+2}{3}, \frac{4+5+6}{3}\right)$

$\equiv (3, 5)$

class-II : $(6,7)\ (5,7) \equiv \left(\frac{6+5}{2}, \frac{7+7}{2}\right)$

$\equiv (5.5, 7)$

Iteration 3 :

Data	class-I (3, 5)	class-II (5.5, 7)	class
(3,4)	1.0	3.91	class-I
(6,7)	3.61	0.50	class-II
(4,5)	1.0	2.5	class-I
(5,7)	2.83	~~0.25~~ 0.5	class-II
(2,6)	1.12	3.64	class-I

class-I : $(3,4)\ (4,5)\ (2,6) \equiv \left(\frac{3+4+2}{3}, \frac{4+5+6}{3}\right)$

$\equiv (3, 5)$

class-II : $(6,7)\ (5,7) \equiv \left(\frac{6+5}{2}, \frac{7+7}{2}\right)$

$\equiv (5.5, 7)$

Algorithm stop

Output:

Data	class
(3,4) (4,5) (2,6)	I
(6,7) (5,7)	II

K- Mean clustering algorithm:

step1: Place K points the space represented by the objects that are being clustered. These points represent initial group centroids.

step2: Assign each object to the group that has the closest centroid.

step3: When all objects have been assigned, recalculate the positions of the K centroids.

step4: Repeat Step 2 and 3 until the centroids no longer move.

K- Mean clustering algorithm:

step1: place K points the space represented by the objects that are being clustered. These points represent initial group centroids.

step2: Assign each object to the group that has the closest centroid.

step3: When all objects have been assigned, recalculate the positions of the K centroids.

step4: Repeat step 2 and 3 until the centroids no longer move.

Modified K- Means clustering algorithm:

step1: Select initial group centroids - cleverly.

step2: Same as K-mean

step3: Same as K-mean

step4: Same as K-mean

Example using the previous data:

step1: Initial centroid selection

Data	(3,4)	(6,7)	(4,5)	(5,7)	(2,6)
(3,4)	0	4.242	1.414	3.605	2.236
(6,7)		0	2.828	1	4.123
(4,5)			0	2.24	2.24
(5,7)				0	3.162
(2,6)					0

class-I : (3,4)

class-II : (6,7)

step 2, 3, 4 :

Iteration-I :

Data	class-I (3,4)	class-II (6,7)	class
(3,4)	0	4.242	class-I
(6,7)	4.242	0	class-II
(4,5)	1.414	2.828	class-I
(5,7)	3.605	1	class-II
(2,6)	2.236	4.123	class-I

— Fuzzy C-Mean Clustering

class-I: $(3,4)\ (4,5)\ (2,6) \equiv \left(\frac{3+4+2}{3}, \frac{4+5+6}{3}\right)$

$\equiv (3, 5)$

class-II: $(6,7)\ (5,7) \equiv \left(\frac{6+5}{2}, \frac{7+7}{2}\right)$

$\equiv (5.5, 7)$

Iteration-II:

Data	class-I (3, 5)	class-II (5.5, 7)	class
(3, 4)	1	3.91	class-I
(6, 7)	3.61	0.5	class-II
(4, 5)	1	2.5	class-I
(5, 7)	2.83	0.5	class-II
(2, 6)	1.414	3.64	class-I

class-I: $(3,4)\ (4,5)\ (2,6) \equiv \left(\frac{3+4+2}{3}, \frac{4+5+6}{3}\right)$

$\equiv (3, 5)$

class-II: $(6,7)\ (5,7) \equiv \left(\frac{6+5}{2}, \frac{7+7}{2}\right)$

$\equiv (5.5, 7)$

Algorithm stop

Data	class	centroid
(3,4) (4,5) (2,6)	class-I	(3, 5)
(6, 7) (5, 7)	class-II	(5.5, 7)