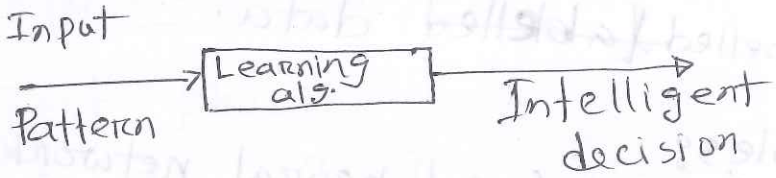
Theory

Lecture Note: 1

What do you understand by Machine Learning? Difference between supervised learning and unsupervised learning.

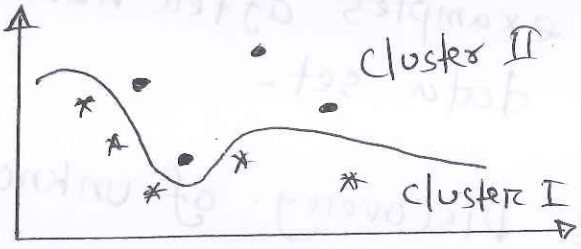
Machine learning is a branch of artificial intelligence. It is a scientific discipline with the design and development of algorithms that recognize complex patterns and make intelligence decisions based on input data.



Unsupervised Learning: In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabelled data.

Characteristics:

- Input data
- No output / No teacher / or No labelled data.



Examples:

- k-means clustering
- Modified k-means clustering
- Hierarchical clustering
- Fuzzy-c-mean clustering.

Supervised learning: supervised learning is the machine learning task of inferring a function from labelled training data.

characteristics:
- Input data
- Output data/ Teacher data or/ labelled ~~labelled~~ data.

Examples:
- Artificial neural network.
- Bayesian network
- Support Vector machine
- Hidden Markov model
- Decision tree learning

Generalization: Generalization is the ability of an algorithm to perform accurately on new, unseen examples after having trained on a learning data set.

Data mining: Discovery of unknown properties on the data.

Difference between close test and open test.

Close test: During the ~~training~~ test phase if known input data and output data are apply in the test phase, that test is called close test.

open test: During the test phase if unknown input data and output data are apply in the test phase that test is called open test.

Training phase:

AND

A	B	e
1	1	1
0	0	0

Test phase

0 1 ?? } open test.
1 0 ??

Test phase

0 0 0 } close test.

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

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

Soln:

(Step 1) Initial step: (Centroid Selection)

Let, Class - I : (3,4)
Class - II : (4,5)

step 2, 3, 4 :

Iteration 1:

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

$\therefore$ 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	1.204	Class-II
(4,5)	1.5	1.66	Class-I
(5,7)	3.20	0.67	Class-II
(2,6)	1.12	3.01	Class-I

$\therefore$ 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	0.5	Class-II
(2,6)	1.414	3.64	Class-I

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

$\equiv (3,5)$

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	Centroid
(3,4) (4,5) (2,6)	I	(3,5)
(6,7) (5,7)	II	(5.5,7)

K-Means clustering Algorithm:

step 1: Place K points into the space, represented by the objects that are being clustered. These points represent initial group centroids.

step 2: Assign each object to the group that has the closest centroid.

step 3: When all objects have been assigned, recalculate the positions of the K centroids.

step 4: Repeat step 2 and 3 until the centroids no longer move.

* Classify this data (2,7) using the previous experiment.

Ans:

Class - I: (Distance) $d_1 = \text{dist}\{(2,7),(3,5)\}$

$$= \sqrt{(2-3)^2 + (7-5)^2}$$

$$= \sqrt{1+2^2}$$

$$= \sqrt{5}$$

$$= 2.236$$

Class - II: (Distance) $d_2 = \text{dist}\{(2,7),(5.5,7)\}$

$$= \sqrt{(2-5.5)^2 + (7-7)^2}$$

$$= \sqrt{(-3.5)^2 + 0}$$

$$= \sqrt{12.25}$$

$$= 3.5$$

Here 2.236 is smaller than 3.5

So the given data (2,7) is in class-I.

* Classify the following Data into two clusters using the Modified k-means algorithm. (3,4) (6,7) (4,5) (5,7) (2,6)

Soln:

Step 1: 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)	4.242	0	2.828	1.0	4.123
(4,5)	1.414	2.828	0	2.236	2.236
(5,7)	3.605	1.0	2.236	0	3.162
(2,6)	2.236	4.123	2.236	3.162	0

∴ Class-I : (3,4)
Class-II : (6,7)

step 2,3,4 : Iteration-1 :

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	1.0	Class-II
(2,6)	2.236	4.123	Class-I

∴ Class I: (3,4) (4,5) (2,6) $\equiv \left(\frac{3+4+2}{3}, \frac{4+5+}{3}\right.$

$\equiv (3,5)$

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

$\equiv (5.5, 7)$

Iteration 2:

Data	Class-I (3,5)	Class-II (5.5,7)	Class
(3,4)	1.0	3.91	Class-I
(6,7)	3.61	0.5	Class-II
(4,5)	1.0	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.

Output:

Data	Class	Centroid
(3,4) (4,5) (2,6)	Class-I	(3,5)
(6,7) (5,7)	Class-II	(5.5,7)

* Modified k-Means clustering algorithm:

Step1: Select initial group centroids cleverly.

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.