

Unit V

Unsupervised Learning I

Supervised Learning

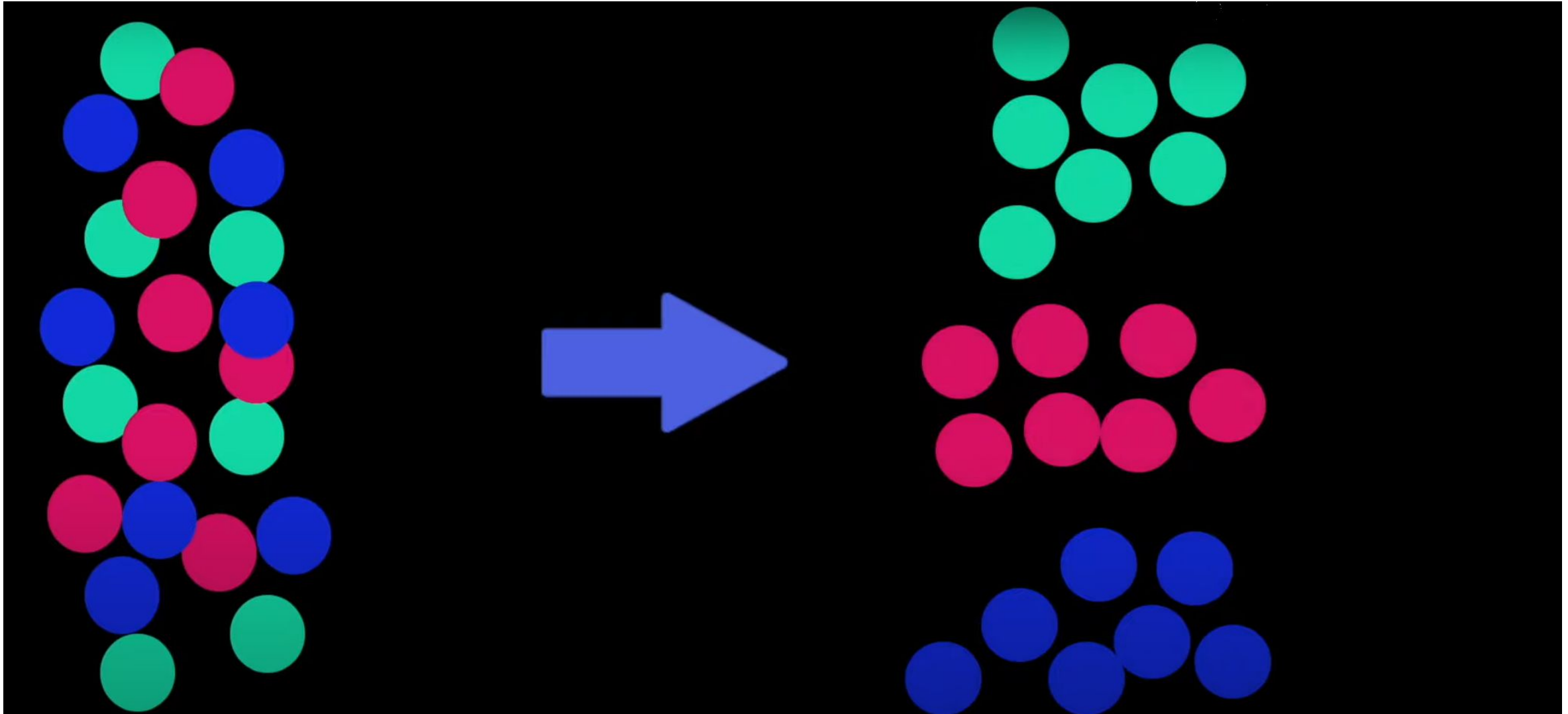
- Computer learns by making use of labeled data

Unsupervised Learning

- Computer learns by making use of unlabeled data

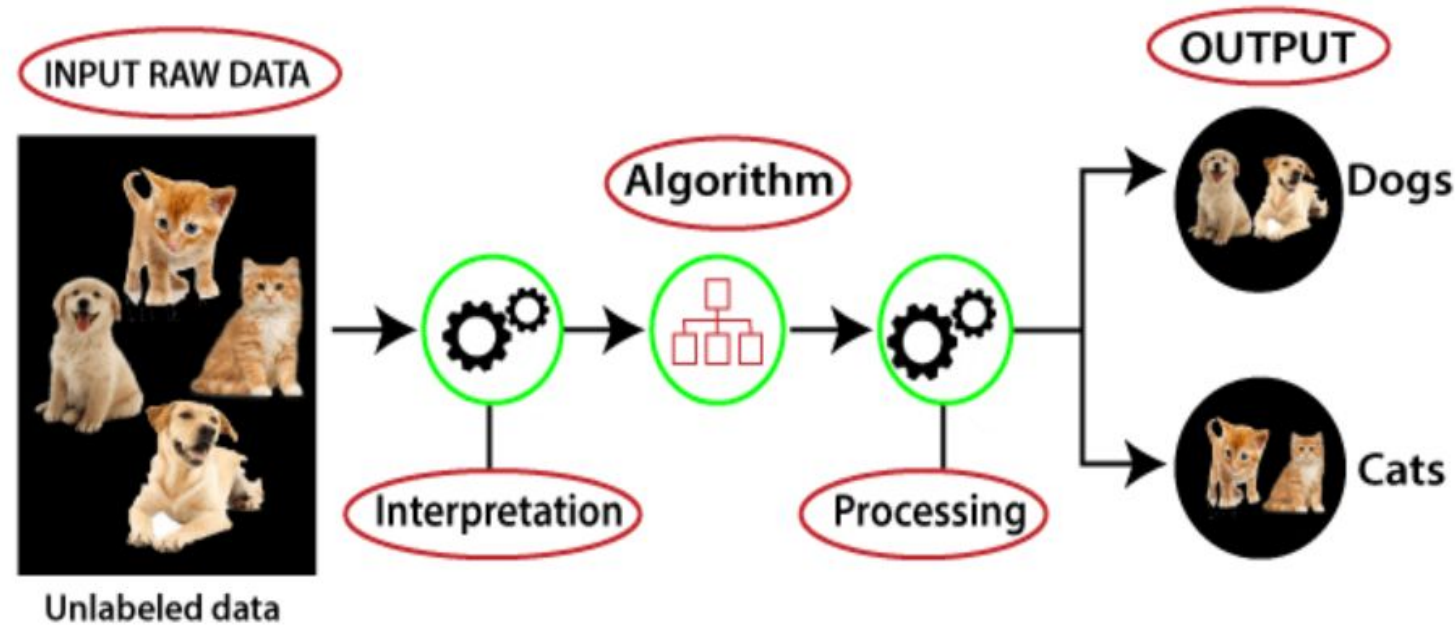
Features	Target
5.1 3.5 1.4 0.2	?
6.4 3.1 4.5 1.2	?
.....
.....
4.9 3.0 1.4 0.2	?

If the Data is labeled , How the model is learning



Essentially, the model is trying to figure out some kind of structure in the data and will extract the useful information from it.

- **Example:** Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs.



Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given. Now, this unlabeled input data is fed to the machine learning model in order to train it. Firstly, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.

Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.

K-Means Clustering

- Suppose that the data mining task is to cluster points into three clusters, where the points are:

A1(2,10),A2(2,5),A3(8,4)

B1(5,8), B2(7,5),B3(6,4)

C1(1,2),C2(4,9)

- The distance function is Euclidean distance.
- Suppose initially we assign A1,B1 and C1 as the center of each cluster respectively. Considering this, we divide data points into clusters.
- Considering this divide the dataset into different clusters.
- If initial centroids are not given then select any of these particular points as the initial centroids and then continue applying the K-Means algorithm.

Data points	Points	(2,10)	(5,8)	(1,2)	Cluster
A1	(2,10)	0.00	3.61	8.06	1
A2	(2,5)	5.00	4.24	3.16	3
A3	(8,4)	8049	5.00	7.28	2
B1	(5,8)	3.61	0.00	7.21	2
B2	(7,5)	7.07	3.61	6.71	2
B3	(6,4)	7.21	4.21	5.39	2
C1	(1,2)	8.06	7.21	0.00	3
C2	(4,9)	2.24	1.41	7.62	2

A1 to A1 $d(p_1, p_2) = 0$

A2 to A1 = 5.00

New Centroids

A1 = (2,10)

B1 = $((8+5+7+6+4)/5 = 6$ $(4+8+5+4+9)/5 = 6$

B1 = (6,6)

C1 = $(2+1)/2 = 1.5$ $(5+2)/2 = 3.5$

C1 = (1.5, 3.5)

Data points	Points	(2,10)	(1.5,3.5)	(1,2)	Cluster	Cluster
A1	(2,10)	0.00	5.66	6.52	1	1
A2	(2,5)	5.00	4.12	1.58	3	3
A3	(8,4)	8.49	2.83	6.52	2	2
B1	(5,8)	3.61	2.24	5.70	2	2
B2	(7,5)	7.07	1.41	5.70	2	2
B3	(6,4)	7.21	2.00	4.53	2	2
C1	(1,2)	8.06	6.40	1.58	3	3
C2	(4,9)	2.24	3.61	6.04	2	1

A1 = (3,9.5)
B1 = (6.5,5.25)
C1 = (1.5,3.5)

Data points	Points	(2,10)	(1.5,3.5)	(1,2)	Cluster	Cluster
A1	(2,10)	0.00	5.66	6.52	1	1
A2	(2,5)	5.00	4.12	1.58	3	3
A3	(8,4)	8.49	2.83	6.52	2	2
B1	(5,8)	3.61	2.24	5.70	2	2
B2	(7,5)	7.07	1.41	5.70	2	2
B3	(6,4)	7.21	2.00	4.53	2	2
C1	(1,2)	8.06	6.40	1.58	3	3
C2	(4,9)	2.24	3.61	6.04	2	1

A1= (3.67,9)
B1 = (7,4.33)
C1 = (1.5,3.5)

Data points	Points	(3.67,9)	(7,4.33)	(1.5,3.5)	Cluster	Cluster	Cluster	Cluster
A1	(2,10)	1.94	7.56	6.52	1	1	1	1
A2	(2,5)	4.33	5.04	1.58	3	3	3	3
A3	(8,4)	6.62	1.05	6.52	2	2	2	2
B1	(5,8)	1.67	4.08	5.70	2	2	1	1
B2	(7,5)	5.21	0.67	5.70	2	2	2	2
B3	(6,4)	5.52	1.05	4.58	2	2	2	2
C1	(1,2)	7.49	6.44	1.58	3	3	3	3
C2	(4,9)	0.33	5.55	6.04	2	1	1	1

A1=(

- A1, B1, C2 belongs to cluster 1
- A3, B2, B3 Cluster 2
- A2, C1 Cluster 3