

Experiment-2.1

Aim of the Experiment:

Implementation of Unsupervised Learning Algorithm.

Theory:

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets without human intervention, in contrast to supervised learning where labels are provided along with the data. The most common unsupervised learning method is cluster analysis, which applies clustering methods to explore data and find hidden patterns or groupings in data.

Unsupervised learning algorithms are machine learning algorithms that analyze and cluster unlabeled datasets. They can identify patterns in data without human intervention.

Benefits of Unsupervised Learning Algorithms:

- 1) Complex processing: They are better suited for complex processing tasks, such as organizing large datasets into clusters.
- 2) Identify patterns: They can identify previously undetected patterns in data and can help identify features useful for categorizing data.
- 3) Discover structure: They discover the underlying structure of a dataset using only input features.
- 4) Require little human supervision: They require little human supervision and prep work.

K-means Clustering:

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

K-means Clustering Algorithm:

- 1) Select value of K and Randomly select initial cluster centers from data points.
- 2) Calculate the distances between each data point and each cluster center.
- 3) Assign each data point to the cluster with the nearest center.
- 4) Update the cluster centers based on the mean of the data points assigned to each cluster.
- 5) Repeat steps 2 to 4 until convergence, i.e., until the cluster centers no longer change between iterations.

Code for Experiment :

```
% load data
my_data = load('dataset.txt');
data_size = size(my_data);
num = data_size(1);
data = my_data(:, 1:2);
fprintf("Data Size:");
disp(data_size);

% normalize data
epsilon = 0.01;
data_mean = mean(data);
data_mean = repmat(data_mean, [num,1]);
data_var = var(data);
data_var = repmat(data_var, [num,1]);
data = (data - data_mean)./sqrt(data_var + epsilon);

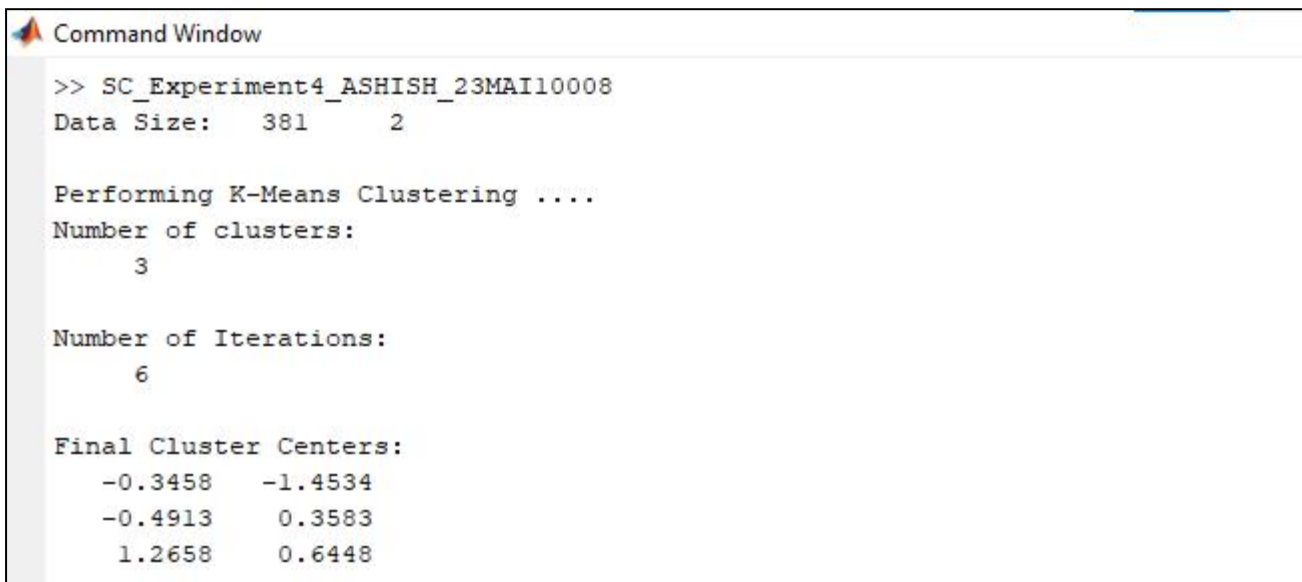
% perform k_means algorithm
disp("Performing K-Means Clustering ....")
clusters = 3;
disp("Number of clusters:")
disp(clusters)
[cluster_label, step, final_centers] = k_means(data, clusters, num);

disp('Number of Iterations:');
disp(step);
disp('Final Cluster Centers:');
disp(final_centers);

% Plot the data points with colors representing their clusters
figure;
gscatter(data(:,1), data(:,2), cluster_label);
hold on;
plot(final_centers(:,1), final_centers(:,2), 'kx', 'MarkerSize', 10, 'LineWidth', 2);
title('K-Means Clustering Result');
xlabel('Feature 1');
ylabel('Feature 2');
legend('Cluster 1', 'Cluster 2', 'Cluster 3', 'Cluster Centers');
hold off;
```

```
function [label, step, center] = k_means(data, clusters, num)
    % Initialize initial clusters
    index = randperm(num, clusters);
    dis = zeros(num, clusters);
    label = zeros(num, 1);
    center = data(index, :);
    step = 0;
    % while loop runs until convergence
    while(1)
        pre_center = center;
        % calculate distance between data points and cluster centers
        for i = 1:num
            for j = 1:clusters
                dis(i, j) = norm(data(i,:) - center(j, :));
            end
        end
        % construct new clusters
        for i = 1:num
            label(i) = find(dis(i,:) == min(dis(i,:)));
        end
        % attain new centers
        for i = 1:clusters
            one_cluster = data(label == i, :);
            center(i, :) = mean(one_cluster);
        end
        if (center == pre_center)
            break;
        end
        step = step + 1;
    end
end
```

Result/Output :



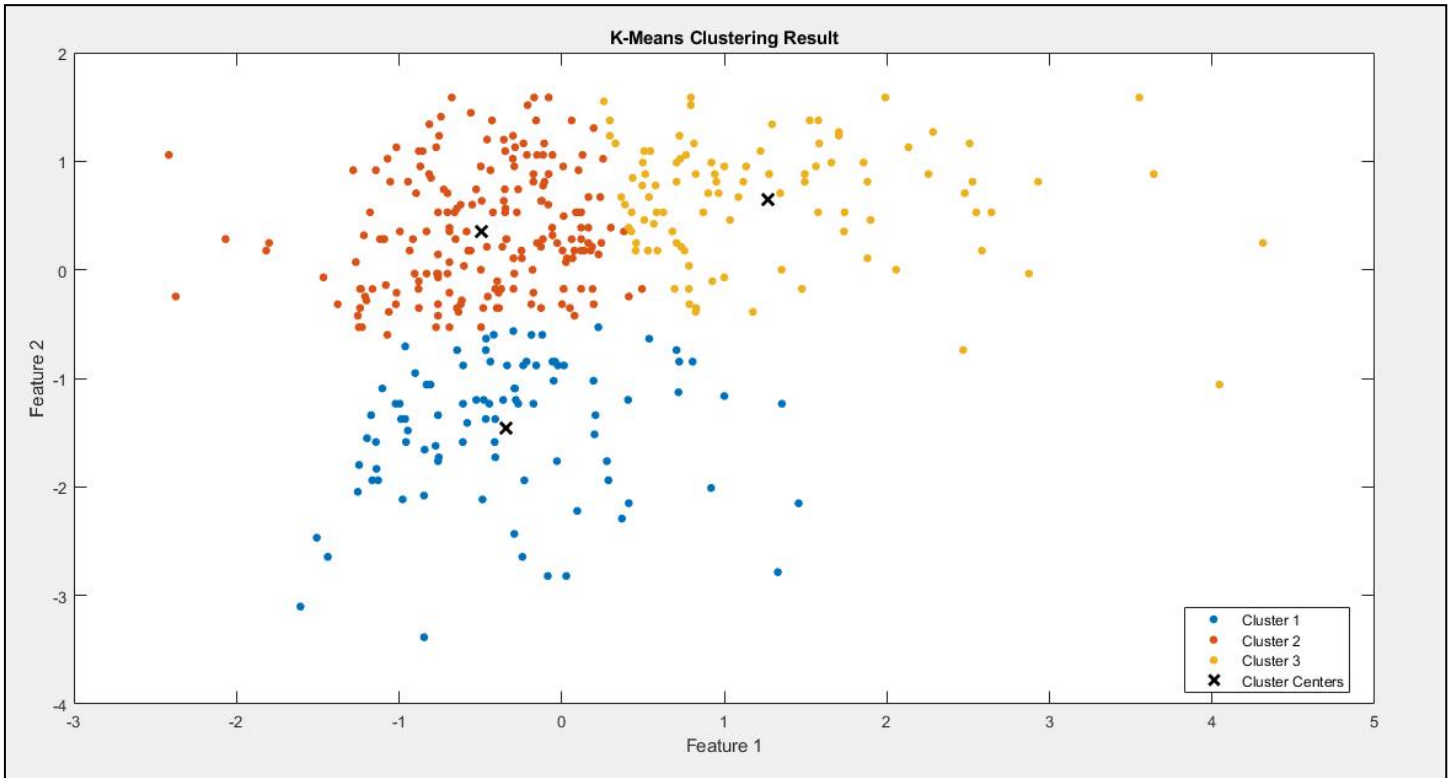
The screenshot shows a MATLAB Command Window with the following text:

```
>> SC_Experiment4_ASHISH_23MAI10008
Data Size:   381      2

Performing K-Means Clustering ....
Number of clusters:
      3

Number of Iterations:
      6

Final Cluster Centers:
    -0.3458    -1.4534
    -0.4913     0.3583
     1.2658     0.6448
```



Learning outcomes :

1. Learnt about basic concepts of Unsupervised learning algorithm.
2. Learnt about K-Means Clustering Algorithm.
3. Learnt about how centroid gets updated after each iteration.
4. Learnt about difference between Supervised and Unsupervised learning.
5. Learnt about difference between Classification and Clustering.