Cluster Analysis

MA2233 - Team 5 Project

Table of Contents

What is Clustering or Cluster Analysis

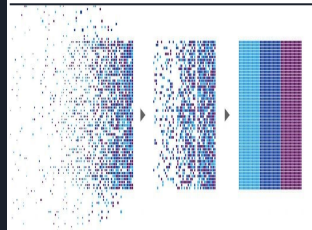
- Clustering is the task of grouping a set of objects in such a way that objects in the same group (cluster) are more similar to each other than to those in other groups (clusters).
- Cluster Analysis can be achieved by various algorithms that differ significantly in their understanding of what constitutes as a cluster and how to efficiently find them.
- Besides the term clustering, there are a number of terms with similar meanings, including *automatic classification, numerical taxonomy, botryology, typological analysis, and community detection. The subtle differences are often in the use of the results.*

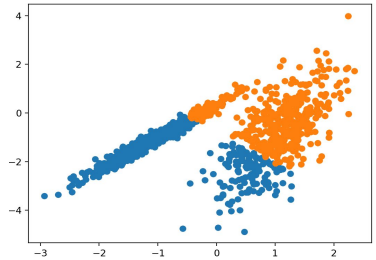
Clustering Algorithms

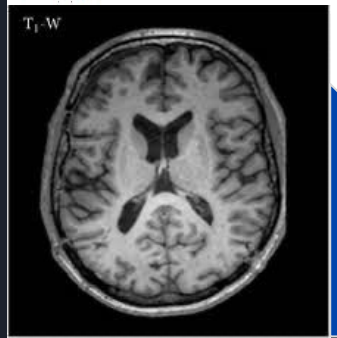
- Prototype-based
  - C-Means, Fuzzy C-Means
- Hierarchical
  - Single/Complete Linkage
- Distribution-based
  - Expectation Maximization
  - Gaussian Mixture Model
- Density-based
  - DBSCAN
  - DenCLUE
- Centroid Models
  - K-Means Algorithm
- MST-based

Real-life Applications

- Main Task in Exploratory Data Analysis
- Common Technique in Statistical Data Analysis
- Pattern Recognition
- Image Segmentation
- Information Retrieval
- Bioinformatics - Gene Expression Data
- Data Compression
- Computer Graphics
- Machine Learning







Problem Statement

01 Generate N points in 2-D that are from *k* number of clusters.

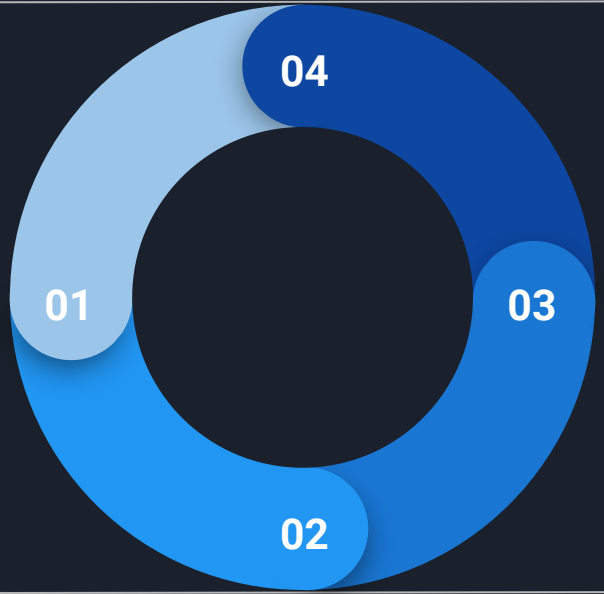
02 Find the MST of the above dataset using Euclidean distance.

03 Implement any 3 of the 5 algorithms for terminating the cluster.

04 A menu-driven MST-based clustering where the user can choose from the 3 algorithms.

Outlook of the Project

Random Points, Complete Graph, and Edges



Clustering based on different criteria

Kruskal's, Prim's and Breadth-First Search

Generating Minimum Spanning Tree

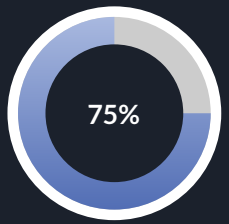
Problems faced in Implementation

- Understanding the termination criteria for clustering.
- Implementation of Minimum Spanning Tree.
- Clustering after formation of MST.
- Ensuring that the code works optimally with minimum Time & Space Complexities.

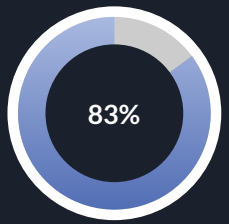
Persona 02

Berry Books

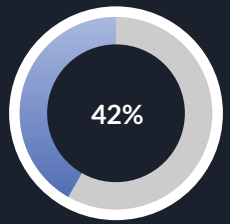
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eleifend a diam quis suscipit. Fusce venenatis nunc ut lectus convallis, sit amet egestas mi rutrum. Maecenas molestie ultricies euismod.



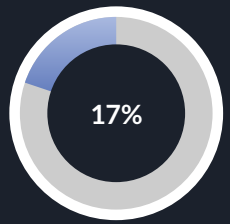
Lorem Ipsum



Lorem Ipsum



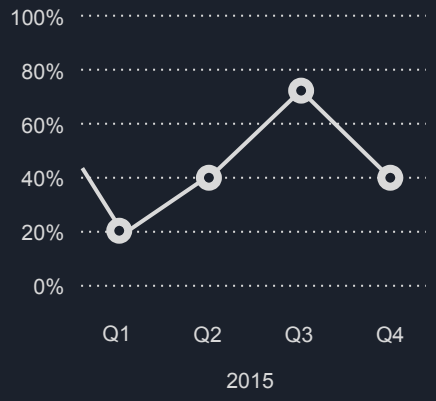
Lorem Ipsum

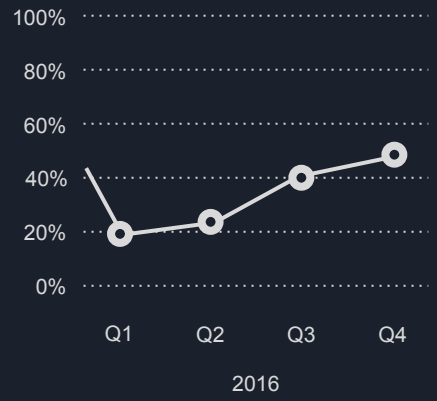


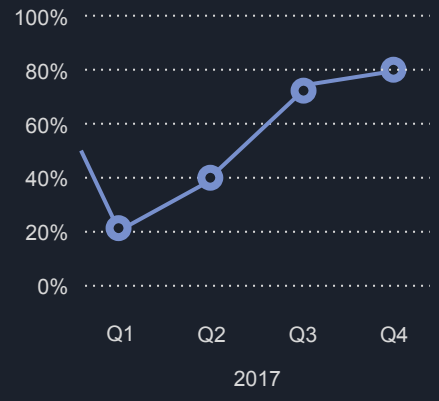
Lorem Ipsum

Market trends

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eleifend a diam quis suscipit. Fusce venenatis nunc ut lectus convallis, sit amet egestas mi rutrum. Maecenas molestie ultricies euismod. Morbi a rutrum nisl. Vestibulum laoreet enim id sem fermentum, sed aliquam arcu dictum. Donec ultrices diam sagittis nibh pellentesque eleifend.







Introducing: Lorem ipsum

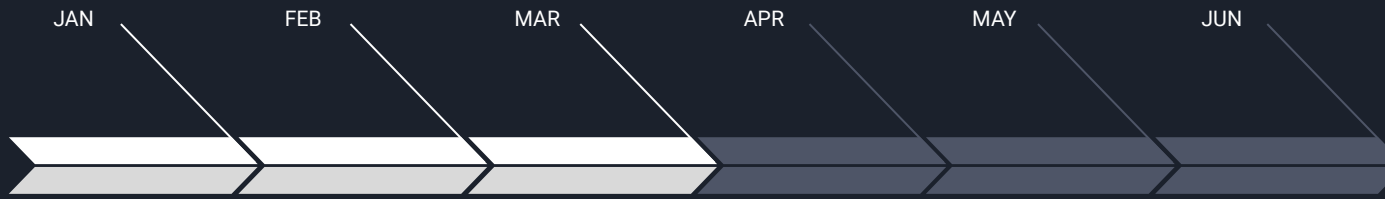
Showcase how your tools work across different devices

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eleifend a diam quis suscipit. Fusce venenatis nunc ut lectus convallis, sit amet egestas mi rutrum.

QUICK TIP

Try right clicking on a photo and using "Replace Image" to demonstrate your prototype.

Project timeline



Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing.

Sit Amet

Lorem ipsum dolor sit amet, consectetur adipiscing.

Donec Ultrices

Lorem ipsum dolor sit amet, consectetur adipiscing.

Litora

Lorem ipsum dolor sit amet, consectetur adipiscing.

Quis Diam

Lorem ipsum dolor sit amet, consectetur adipiscing.

Dolor Lorem

Lorem ipsum dolor sit amet, consectetur adipiscing.

MA2233 - Data Structures and Applications Lab
Group Project - 5

Course Instructor - Dr. Balasubramaniam Jayaram
jbala@math.iith.ac.in

Varunaditya Singhal
MA20BTECH11021

Anita Dash
MA20BTECH11001

Kethari Narasimha Vardhan
MA20BTECH11006

Thank you!