For the CS293 Project, I have exercised option 3, with Professor's approval and have taken up a project on **clustering algorithms**

Clustering Algorithms

Overview

I have implemented four major clustering algorithms:

- K-means clustering
- Gaussian Bayesian Clustering
- Density Based Clustering
- Hierarchical Clustering

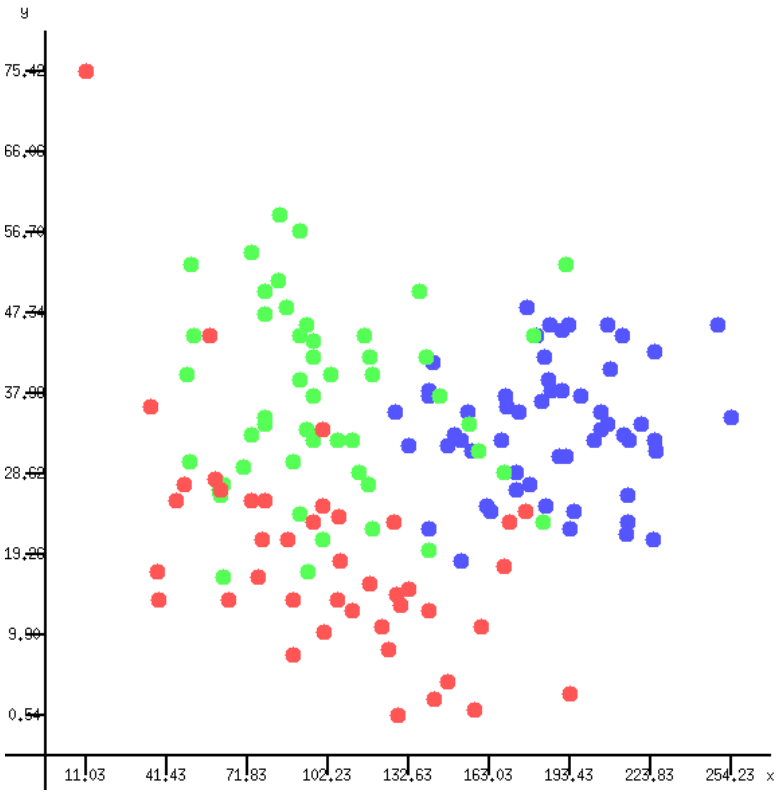
I have also implemented a plotter using simplecpp as a renderer for these algorithms. The plotter can render scatter plots for all of the data, showing different colors for different clusters.

Structure

```
.
├── include
│   ├── cluster
│   │   ├── densityBasedEstimator.hpp
│   │   ├── gaussianEstimator.hpp
│   │   ├── hierarchicalEstimator.hpp
│   │   └── kMeansEstimator.hpp
│   ├── plotter
│   │   └── plotter.hpp
│   └── util
│       ├── edge.hpp
│       ├── matrix2.hpp
│       ├── point.hpp
│       ├── range_point.hpp
│       ├── union_point.hpp
│       └── vec2.hpp
├── src
    ├── cluster
    │   ├── densityBasedEstimator.cpp
    │   ├── gaussianEstimator.cpp
    │   ├── hierarchicalEstimator.cpp
    │   └── kMeansEstimator.cpp
    ├── plotter
    │   └── plotter.cpp
    └── util
        ├── edge.cpp
        ├── matrix2.cpp
        ├── point.cpp
        ├── range_point.cpp
        ├── union_point.cpp
        └── vec2.cpp
```

Example

I have created a sample of the clustering algorithms at work using the Wine Clustering Dataset from Kaggle, which clusters different beverages based on their characteristics. The K-means clustering algorithm works the best for this dataset.



Use of Datastructures and Algorithms

The following Datastructures were used:

- Hash map : `std::unordered_map<int, int> colors`
- Red-black Tree : `std::multimap<int, int> clusters`
- Priority Queue : `std::priority_queue<edge, std::vector<edge>, std::function<bool(edge &, edge &)>> edge_list;`
- Union finding structure : `class union_point`
- Graph : `class range_point`

The following algorithms were used:

- K-means clustering
- Gaussian clustering
- Breadth first search