****

**Electrical and Computers Engineering Department**

**Clustering and Unsupervised Learning**

**Final Report**

Elad Sofer 312124662

Yehonatan Dahan 313441131

Table of Contents

[1 Introduction 4](#_Toc141032122)

[2 Task 0: Dataset Generation 4](#_Toc141032123)

[2.1 Bi-Gaussian 5](#_Toc141032124)

[2.2 Hash Sign 6](#_Toc141032125)

[3 Dataset 8](#_Toc141032126)

[3.1 Dataset collecting 8](#_Toc141032127)

[3.2 Data description 9](#_Toc141032128)

[3.2.1 IMU Relevant background 9](#_Toc141032129)

**List of figures**

Figure ‎2.1 – Sample of 2 Simple Clusters 4

Figure ‎2.2 – 3D visualiation of the clusters 5

Figure ‎2.3 – Sample generated in a bi-gaussian dataset 6

Figure ‎2.4 – 3D view of the bi-baussian dataset 6

Figure ‎2.5 - 7

Figure ‎2.6 - 7

Figure ‎3.1: Shimmer3 IMU 9

# Introduction

Add

# Task 0: Dataset Generation

This task focuses on the process of creating a 2-D dataset for testing clustering algorithms. The dataset was created in MATLAB by using Gaussian. Gaussian mixtures were utilized as a powerful approach to generate synthetic data resembling real-world patterns and structures. The initial step involved determining the desired number of clusters to be included in the dataset. Each cluster was represented by a Gaussian distribution with its unique mean and variance parameters. An example of a “simple” dataset is shown below:



Figure ‎2.1 – Sample of 2 Simple Clusters

Another visualization of the clusters could be shown as a 3-dimensional plot of the created gaussians, representing the points density vs. the 2D location:

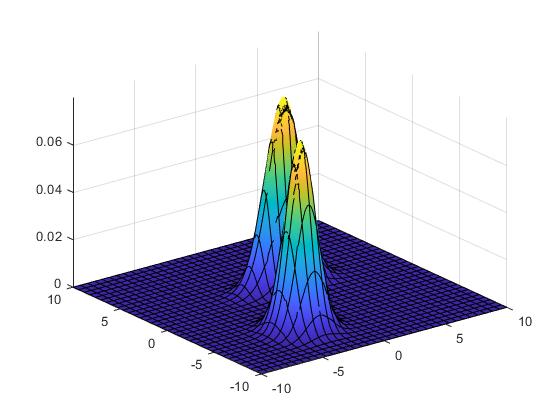


Figure ‎2.2 – 3D visualiation of the clusters

By manipulating the defined means and variances, clusters with different shapes and characteristics were created. The alteration of means resulted in the formation of clusters with distinct centers, while variations in variances led to clusters of varying sizes and shapes, such as elongated or round clusters. This allowed for the creation of diverse datasets, facilitating the evaluation and comparison of various clustering algorithms to gain a better understanding of their strengths and limitations.

Once the Gaussian mixture model was set up, the next step was to sample data points from the distribution. The sampling process was based on the specified means and variances of each cluster. For each data point, the algorithm selected a cluster based on the probabilities dictated by the Gaussian distributions. Consequently, the resulting dataset consisted of data points scattered across the different clusters, with each point assigned to a specific cluster based on the Gaussian mixture model.

Some Examples are shown below:

## Bi-Gaussian

Two gaussians were created, such that the first centered around (2,2) with variances (1,10) (around x and y axes respectively) and the second centered around (-5,-5) with variances (10,1):



Figure ‎2.3 – Sample generated in a bi-gaussian dataset

The 3-dimensional view is shown below:

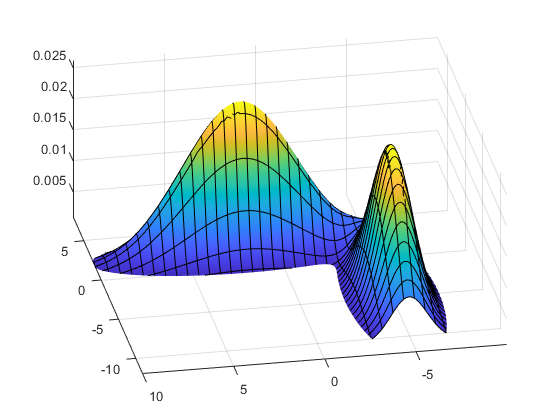


Figure ‎2.4 – 3D view of the bi-baussian dataset

## Hash Sign

Creating a "HASH (#)" symbol in a two-dimensional space using Gaussian distributions, was done by determining the mean and variance for each line in the "#" symbol. The mean of each distribution corresponds to the position of the line, while the variance determines the thickness or width of the line. By using this technique, we were able to create a smooth representation of the "#" symbol in a two-dimensional space as shown below:



Figure ‎2.5 – Hash symbol dataset

And in 3-dimensional space:

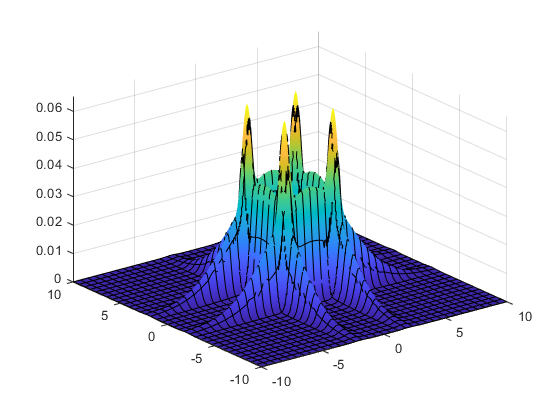


Figure ‎2.6 – 3D view of Hash symbol dataset

As could be seen in the 3-dimensional view, the points joints of the Hash symbol (the point where the lines meet) are the densest and could be mistakenly thought of as the centers (this point will be further discussed in the next chapters).

## Asymmetric Hash sign

In a similar way to the previous section, a Hash sign was created, although in this dataset the sigh was created uneven, to “break” the symmetry:



Figure ‎2.7 -

A 3D view of the dataset:

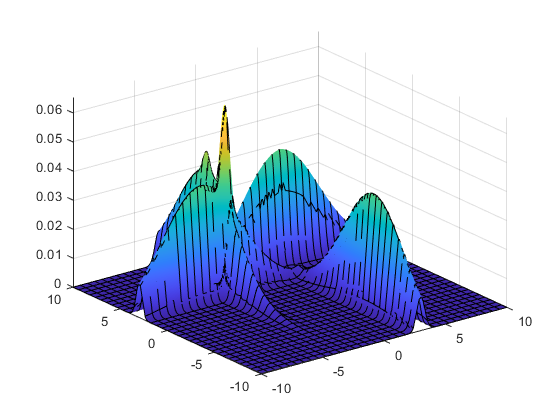


Figure ‎2.8 -

As can be seen, there is a dense point at the intersection of the 2 lines, and another dominant point (lower density) next to it.

# Task 1: Maximum likelihood estimator

# Task 2: Unsupervised Optimal Fuzzy Clustering (UOFC)

## Euclidian distance

## Exponential distance

# Task 3: Agglomerative hierarchical clustring