

REPORT FOR THE STUDY**Wonjun Lee¹ | Junseong Bang^{2,3}**¹Intelligence & Manufacturing Convergence Laboratory, ETRI, Daejun, Republic of Korea²Police Science & Public Safety ICT Research Center, ETRI, Daejun, Republic of Korea³Department of Computer Software, UST, Daejun, Republic of Korea**Abstract**

In this paper, we applied Gaussian Mixture Model to Mall Customer Segmentation data from kaggle. This Data set consist of basic information of customer and Spending Score. After Gaussian Mixture Model Silhouette Scores has a 0.41.

KEYWORDS:

Gaussian Mixture Model

1 | INTRODUCTION

Gaussian mixture model is one of the simplest unsupervised learning algorithms to solve the famous clustering problem.

In this paper we using Mall Customer Segmentation from kaggle. The dataset consists of basic information, such as the customer's ID, gender, and age, and annual income, Spending Score-assigned by the mall based on customer behavior and spanning nature.

This dataset is made up of 200 row, 5 columns.

2 | GAUSSIAN MIXTURE MODEL

A Gaussian mixture model is a stochastic model that assumes that all data points are generated from a mixture of Gaussian distributions.

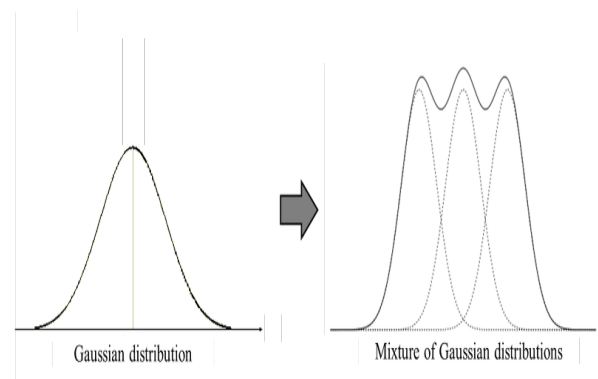
The Gaussian mixture model can generalize k-average clustering to integrate information about the covariance structure of data as well as the potential Gaussian center.

In probabilistic modeling, the probability distribution of all found clusters is estimated for each observation. Therefore, the Gaussian mixture model provides soft clustering different from the hard clustering provided by k-means, hierarchical, and density-based spatial clustering.

3 | METHODS

1. Load Dataset
2. Get lid of Useless Column

Abbreviations: ANA, anti-nuclear of antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

**FIGURE 1** Gaussian Mixture Model

3. Adjust Gaussian Mixture Model

4. Silhouette Score

4 | RESULT

After loading the data imported from Kaggle, the data was searched overall.

As a result, it was confirmed that the data had no duplicate values, and removing variables not necessary for learning.

We applied Gaussian Mixture Model to the data and applied silhouette score.

The Silhouette Score is used as an indicator of the number of clusters and is also used to evaluate the performance of clustering.

Silhouette score was used as an evaluation Metric and as a result, 0.42 was obtained

5 | CONCLUSTION

In this paper, we applied Gaussian Mixture Model to Mall Customer Segmentation data from kaggle. As a result of applying Silhouette score is 0.42.

In future studies, various clustering techniques will be applied to paint a way to enable more effective clustering.