**Project report 10**

Gaussian Mixture Models: Bag of Words Representation

**NAME: DHIVAKAR.R**

**COURSE: AI and ML**

**Question:**

Using a gaussian mixture model, perform a simple clustering on the given 2D Dataset. Try to find the optimal number of clusters using python (you may use any module to implement this). Now implement the same from scratch using python and a dummy dataset generated using scikit learn dataset generating functions such as make blob.

**Prerequisites**

What things you need to install the software and how to install them:

Python 3.6 This setup requires that your machine has latest version of python. The following url <https://www.python.org/downloads/> can be referred to download python. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in how to run software section). To do that check this: <https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-asan-internal-or-externalcommand/> . Setting up PATH variable is optional as you can also run program without it and more instruction are given below on this topic.

Second and easier option is to download anaconda and use its anaconda prompt to run the commands. To install anaconda check this url <https://www.anaconda.com/download/> You will also need to download and install below 3 packages after you install either python or anaconda from the steps above Sklearn (scikit-learn) numpy scipy if you have chosen to install python 3.6

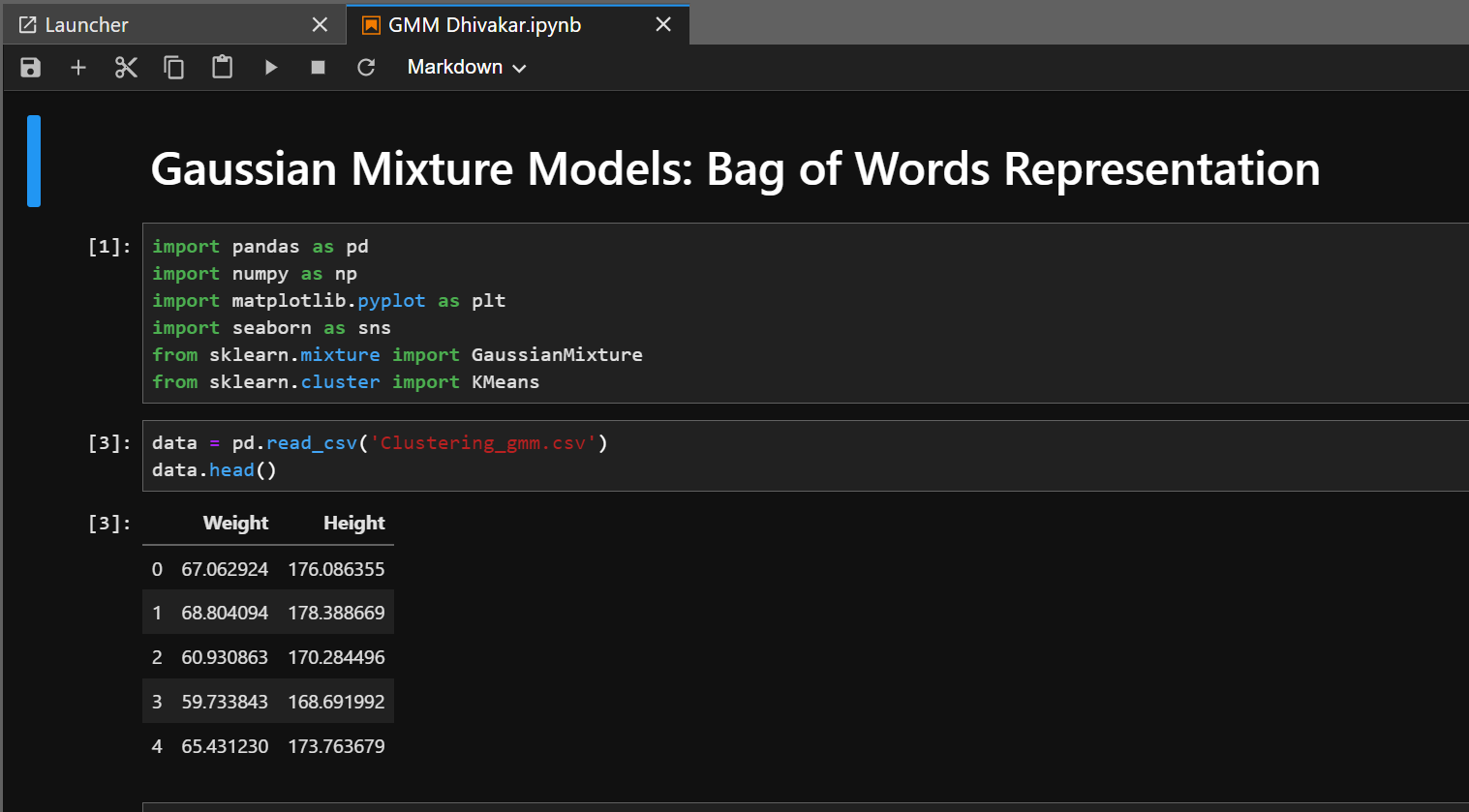
Dataset Link: Clustering\_GMM

<https://cdn.analyticsvidhya.com/wp-content/uploads/2019/10/Clustering_gmm.csv>

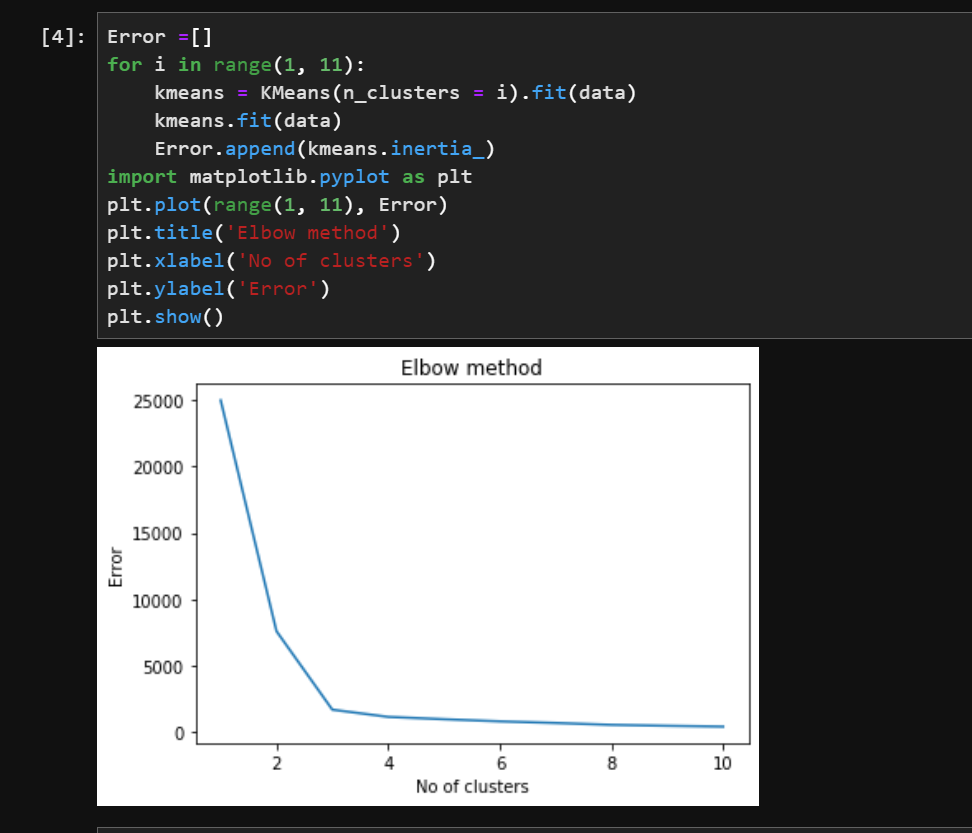
Method used :

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. It attempts to find a mixture of multi-dimensional Gaussian probability distributions that best model any input dataset allowing the model to learn automatically, i.e. in an unsupervised manner. The bag-of-words model is a way of representing text data when modelling text with machine learning algorithms which can be combined with GMM to get a useful model representation.

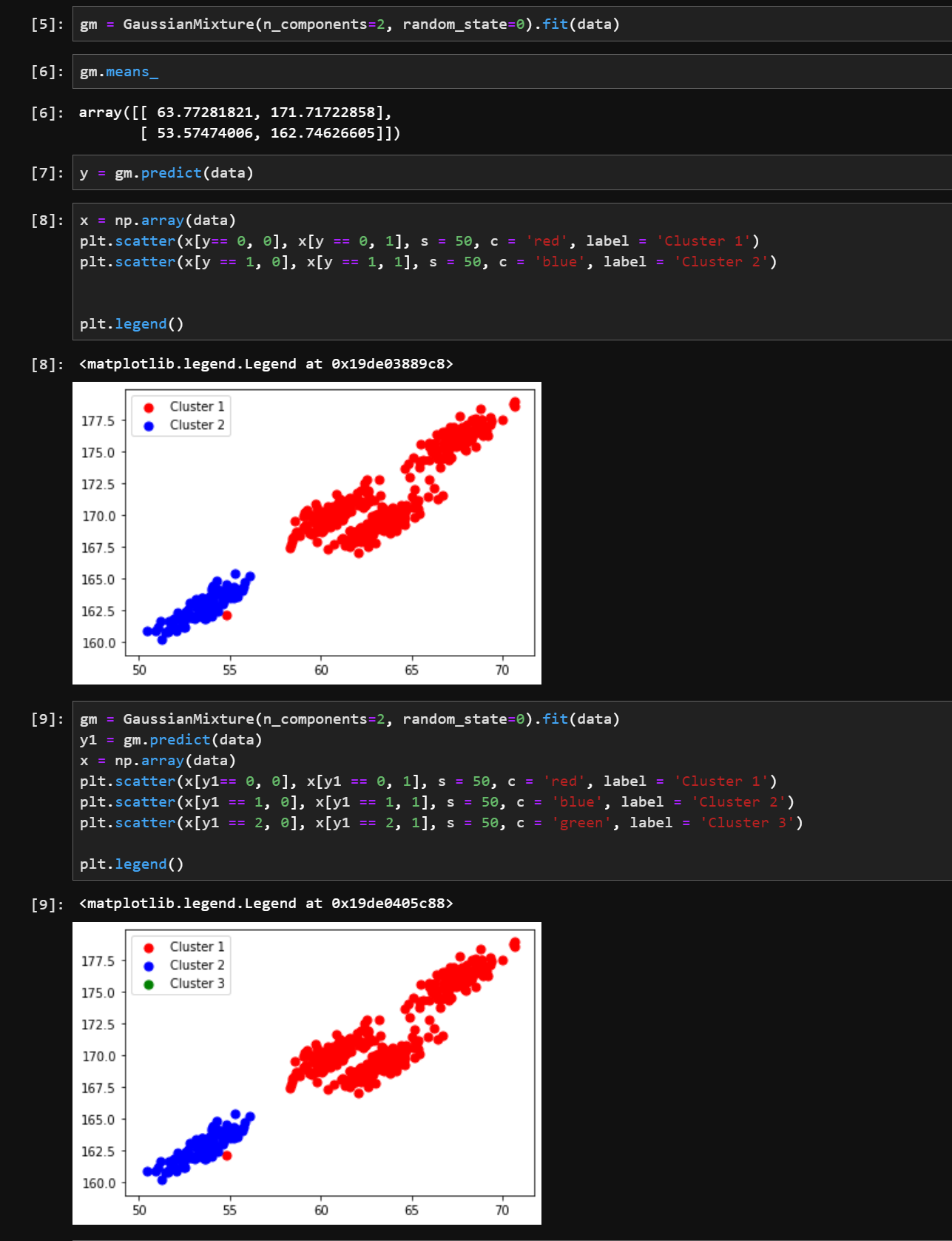
**Load all the required libraries**

****

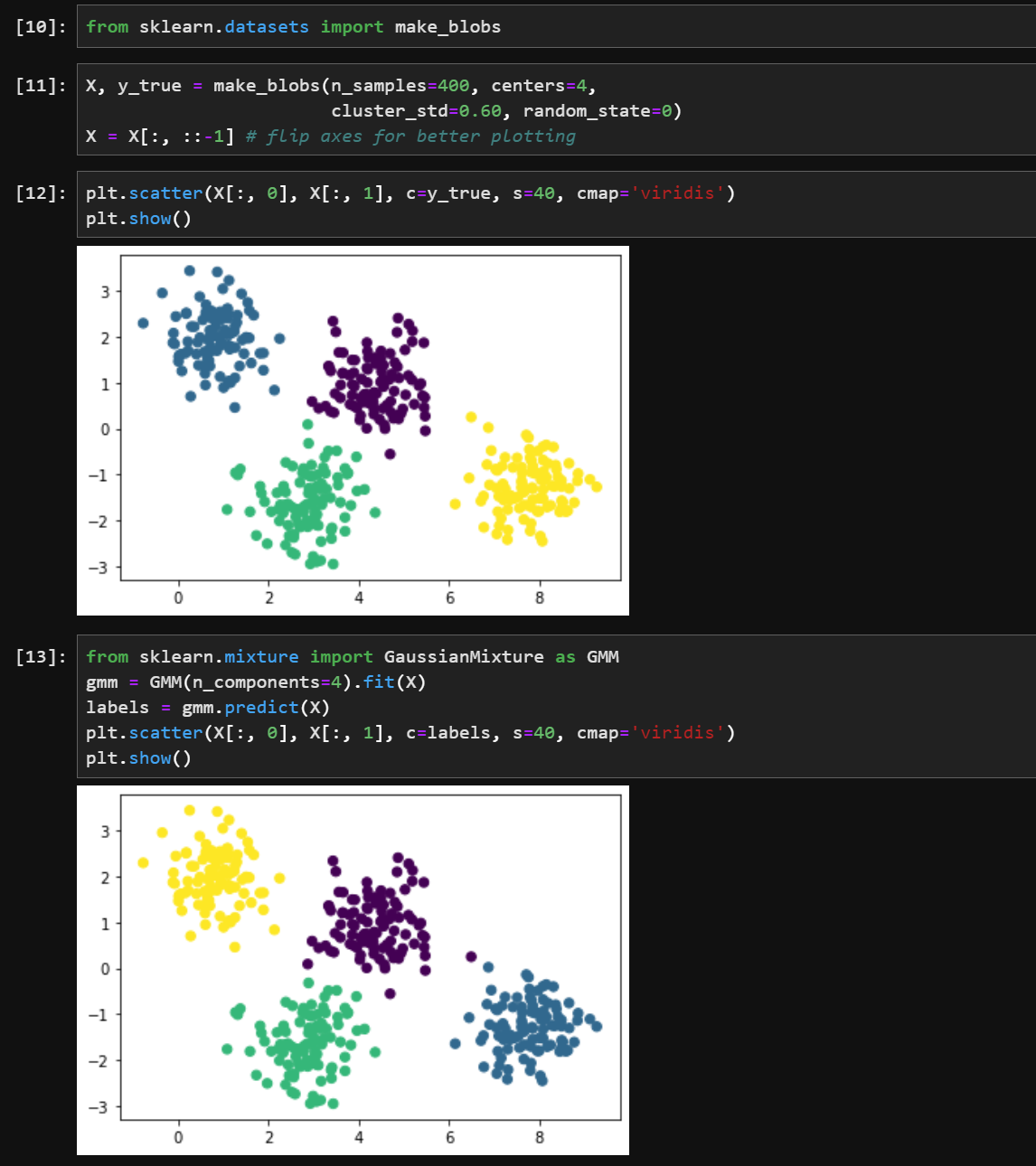
**Elbow Method:**

****

**Model Building Using GMM on Clustering Gmm dataset**

****

**Importing Libraries and Making dummy dataset**

****