

School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET301
Year- 2022
Date- 16-11-2022

Type- Core
Course Name-AIML
Semester- Odd
Batch- V Sem

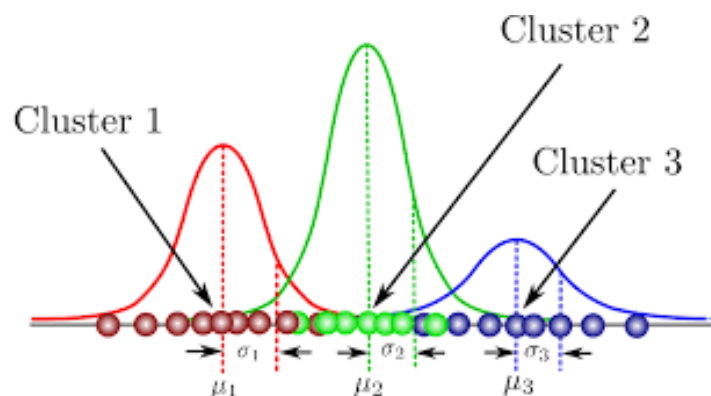
Lab Assignment No. 11_2.2

Exp. No.	Name	CO-1	CO-2	CO-3
11.2.2	Gaussian mixture models clustering		✓	✓

Objective: To Implement Gaussian Mixture Models (GMM) clustering (Unsupervised Learning).

Description

Gaussian Mixture Models (GMMs) assume that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Hence, a Gaussian Mixture Model tends to group the data points belonging to a single distribution together. Let's say we have three Gaussian distributions (more on that in the next section) – GD1, GD2, and GD3. These have a certain mean (μ_1 , μ_2 , μ_3) and variance (σ_1 , σ_2 , σ_3) value respectively. For a given set of data points, our GMM would identify the probability of each data point belonging to each of these distributions.



Algorithm

- Initialize the mean μ_k , the covariance matrix Σ_k and the mixing coefficients π_k by some random values (or other values).
- Compute the C_k values for all k .
- Again Estimate all the parameters using the current C_k

values.

- Compute log-likelihood function.
- Put some convergence criterion
- If the log-likelihood value converges to some value (or if all the parameters converge to some values) then stop, else return to Step 2.

Download Use the following link to download the dataset (5)

<https://archive.ics.uci.edu/ml/machine-learning-databases/wine/>

About Dataset:

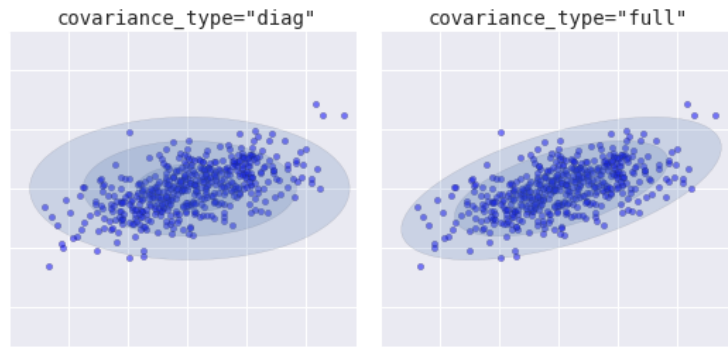
These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

The attributes are

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

Apply following Steps to perform the clustering

1. Import the necessary Libraries
2. Read the dataset
3. Check the presence of missing values. Handle it if present
4. Handle the categorical data if present
5. Selecting the feature i.e., Identify the Independent variables and perform the extraction. (Hint: Remove the Target Column as it is Unsupervised Learning Problem i.e., 1st column quality of wine)
6. Select the n_component value (5)
7. Training the GMM algorithm on the training dataset (Hint: from sklearn.mixture import GMM) (15)
8. Perform the prediction and identify to which cluster data belongs (5)
9. Training the GMM algorithm on the training dataset by setting (15)
covariance_type="diag"
covariance_type="full"



10. Perform the prediction and identify to which cluster data belongs (5)

11. Visualizing the clusters (Hint: Use `.scatter()` function of matplotlib lib) (20)

Extra: Try for Scratch Implementation of GMM algorithm

Suggested Platform: Python: Jupyter Notebook/Azure Notebook/Google Colab.