



Machine Learning

Lesson 5

Clustering Techniques and Gaussian Mixture Model,

Lecturer: Dr. Msagha J Mbogholi, PhD

Flashback from Lesson 4

- Logistic regression is the type of regression used in scenarios where the outcome or dependent variable is binary (0 or 1) or discrete categorical (yes/no).
- With logistic regression we are interested in a dichotomous outcome, hence the need to convert the data points based on outcome to probabilities between 0 and 1. The sigmoid function allows us to achieve this.
- Multinomial logistic regression can be used in scenarios where classification is in multiclass, for example when a car can be classified as either a salon(sedan), pickup or station wagon.
- The support vector machine (SVM) is a supervised learning algorithm that is used for both classification and regression problems by introducing a margin that better helps in the process.
- Support vectors are used to determine the margin used by SVM.
- The hyperplane of SVM is simply the boundary that is defined as being the best among the many possible planes that could be used.
- SVM can be classified as being either linear SVM or non-linear SVM.
- Soft margins allow for generalization of the data and avoids overfitting which hard margins do. It is therefore important to fit the margins in such a way that it is not too generalized and not too overfit.

Content

- Clustering
- Gaussian Mixture Model



Part 1

Clustering

1.1 Introduction

- In lesson 4 the concept of supervised learning were introduced.
- Recall that it was stated that in supervised learning , The training data you feed to the algorithm includes the desired solutions, called labels.
- The two types of supervised learning that were covered are examples of classification and regression methods .
- Clustering has been studied for more that 40 years in many disciplines due to its broad applications(Basaraner et al.,2015)
- As a result many clustering techniques have been reported in literature.
- A broad taxonomy of existing clustering methods is shown in Fig. 1.
- It is not possible to cover all the techniques in this course. We emphasize on Model Based clustering techniques using Gaussian Mixture Model.
- In this lesson we will cover the most popular algorithm in this type of technique known as clustering using Gaussian Mixture Models (GMM).

1.1 Introduction (cont'd)

- Clustering is the task of grouping a set of objects such that similar objects end up in the same group and dissimilar objects are separated into different groups(Ben-David,2012)
- Clustering is one of the most widely used techniques for exploratory data analysis. Across all disciplines, from social sciences to biology to computer science, people try to get a first intuition about their data by identifying meaningful groups among the data points(Ben-David,2012).
- For example, computational biologists cluster genes on the basis of similarities in their expression in different experiments; retailers cluster customers, on the basis of their customer profiles, for the purpose of targeted marketing; and astronomers cluster stars on the basis of their special proximity(Ben-David,2012)
- The problem of data clustering has been widely studied in the data mining and machine learning literature because of its numerous applications to summarization, learning, segmentation, and target marketing (Aggarwal et al., 2014)
- In the absence of specific labeled information, clustering can be considered a concise model of the data which can be interpreted in the sense of either a summary or a generative model(Aggarwal et al., 2014)

1.2. Clustering Methods

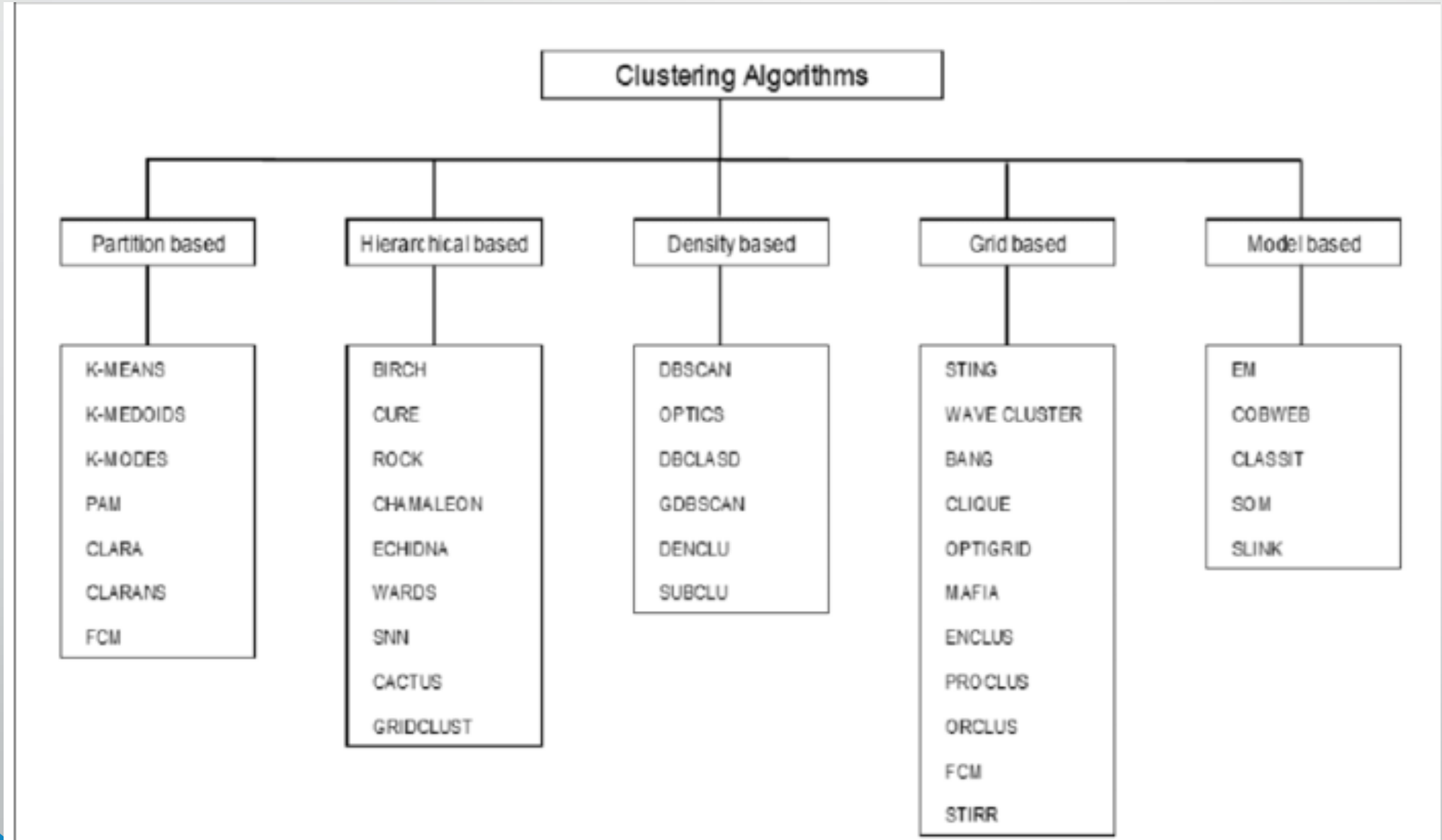


Fig 1 Taxonomy of cluster methods (Fahad et al., 2014)

1.3. Clustering Methods

- This clustering method helps grouping valuable data into clusters and picks appropriate results based on different techniques.
- There are various types of clustering methods as shown in Figure 1 , they are
 - Hierarchical methods
 - Partitioning methods
 - Density-based
 - Model-based clustering
 - Grid-based model

1.3.1. Hierarchical methods

- In hierarchical clustering, the number of clusters does not have to be specified(Richard et al. , 2013).
- Instead, the hierarchical clustering creates a hierarchy of clusters(Richard et al. , 2013).
- While similar posts are grouped into one cluster, similar clusters are again grouped into one uber-cluster(Richard et al. , 2013).
- This is done recursively, until one cluster is left, which contains everything in this hierarchy, one can then choose the desired number of clusters(Richard et al. , 2013).
- However, this comes at the cost of lower efficiency.

1.3.2 Partitioning Methods

- Pedamkar (2021) describes this method as follows:
- “The main goal of partitioning is relocation. They relocate partitions by shifting from one cluster to another, which makes an initial partitioning. It divides ‘n’ data objects into ‘k’ numbers of clusters. This partitioning method is preferred more than a hierarchical model in pattern recognition.
- The following criteria are set to satisfy the techniques:
- Each cluster should have one object.
- Each data object belongs to a single cluster.
- The most commonly used Partitioning techniques are the K-mean Algorithm. They divide into ‘K’ clusters represented by centroids. Then, each cluster center is calculated as a mean of that cluster, and the R function visualizes the result.”

1.3.3 Density Model

- From elementary physics we know that the density of a substance is its mass per unit volume; the formula is $\text{density} = \text{mass}/\text{volume}$.
- This concept is applied in density based clustering techniques.
- The principle lies in finding areas where the density is high and classifying these areas as belonging to the same cluster.
- Thus we find areas in the data where they are concentrated together and separate them from areas where there is empty space or where the data points are scattered.
- The data points that are scattered and not belonging to any cluster are referred to as noise.
- DBSCAN is the most well known density based algorithm that uses this method.

1.3.4 Model Based Clustering

- Grid-based methods are a specific class of density-based methods in which the individual regions of the data space which are explored are formed into a grid-like structure which uses cells to formulate the grids.
- Grid-like structures are often particularly convenient because of greater ease in putting together the different cells in the post-processing phase.
- Such grid-like methods can also be used in the context of high-dimensional methods, since the lower dimensional grids define clusters on subsets of dimensions (Aggarwal et al., 2014).

1.3.5 Grid Model

- In this approach, the objects are considered to be space-driven by partitioning the space into a finite number of cells to form a grid. Then, with the help of the grid, the clustering technique is applied for faster processing which is typically dependent on cells, not on objects (Pedamkar, 2021).

The steps involved are:

- Creation of grid structure
- Cell density is calculated for each cell
- Applying a sorting mechanism to their densities.
- Searching cluster centers and traversal on neighbor cells to repeat the process.

1.4 Importance of clustering Methods

- Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.
- This clustering analysis has been used for model analysis, vector region of attraction.
- Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.
- Clustering quality depends on the methods and the identification of hidden patterns.
- They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.
- They are used in outlier detections to detect credit card fraudulence.
- (Pedamkar, 2021)



Part 2

Gaussian Mixture Model

3.1 Introduction

- McGonagle et al. (in brilliant.org) describes Gaussian mixture models as: “a probabilistic model for representing normally distributed subpopulations within an overall population. Mixture models in general don't require knowing which subpopulation a data point belongs to, allowing the model to learn the subpopulations automatically. As the assignment of a subpopulation is unknown, these compose unsupervised learning.”
- GMM can be used to find clusters in datasets where the clusters may not be clearly defined.
- Additionally GMM can be used to estimate the probability that a new datapoint belongs to each cluster.
- Gaussian mixture models are extensively utilized in mining data, recognition of patterns, machine learning, and statistical analysis

3.1 Introduction

- As described in the previous slide a Gaussian distribution is nothing but a normal distribution, that is to say a bell shaped distribution.
- It is within our interest that our dataset be normally distributed in order to get reliable results. Sometimes it is not evident when examining the data at face value since when it is examined as a whole it does not appear to be normally distributed.
- However, within the data we may find that there are several normal distributions where each sample comes from.
- Presume we have a dataset where there are n clusters. For purposes of this demonstration let us assume $n = 4$.
- Thus there are 4 clusters. We assume that each cluster has a normal distribution within it; so there are 4 normal distributions. We then analyze each data point to see in which of the 4 distributions it fits best.
- What the algorithm does is to check each data point and calculate its probability of being in each of the 4 clusters.
- The cluster where the probability is highest is the one to which the data point will be placed.

3.2 Working of the GMM

- The Gaussian Mixture Model (GMM) is an unsupervised learning model; this means that there is no labelled data.
- In order for it to determine where each data point will fall, GMM has to start from the unknown and move to the known
- An initial random parameter is chosen. From here the algorithm goes to work; it will try and find the best weights and averages to bring together the estimated normal distributions to meet the ones specified in the initial data (the initial parameters).
- Once this is found then it will know where to place each data point.
- Understandably the calculations used to place data points using GMM are not for the faint hearted, especially for new students of machine learning.
- However, we shall go through a simple example to demonstrate the use of GMM.

3.2 GMM Example Application

- This example is drawn from ADipLearn and posted in 2018. The original video can be found at <https://youtu.be/fVsmnZqrBUs> and the example posted is originally done by them. We adopt this example to explain GMM and how it works as its easy and simple to follow. I have also added my own explanations to make it even more simpler.
- All diagrams in this example, except where explicitly labelled, are from <https://youtu.be/fVsmnZqrBUs>

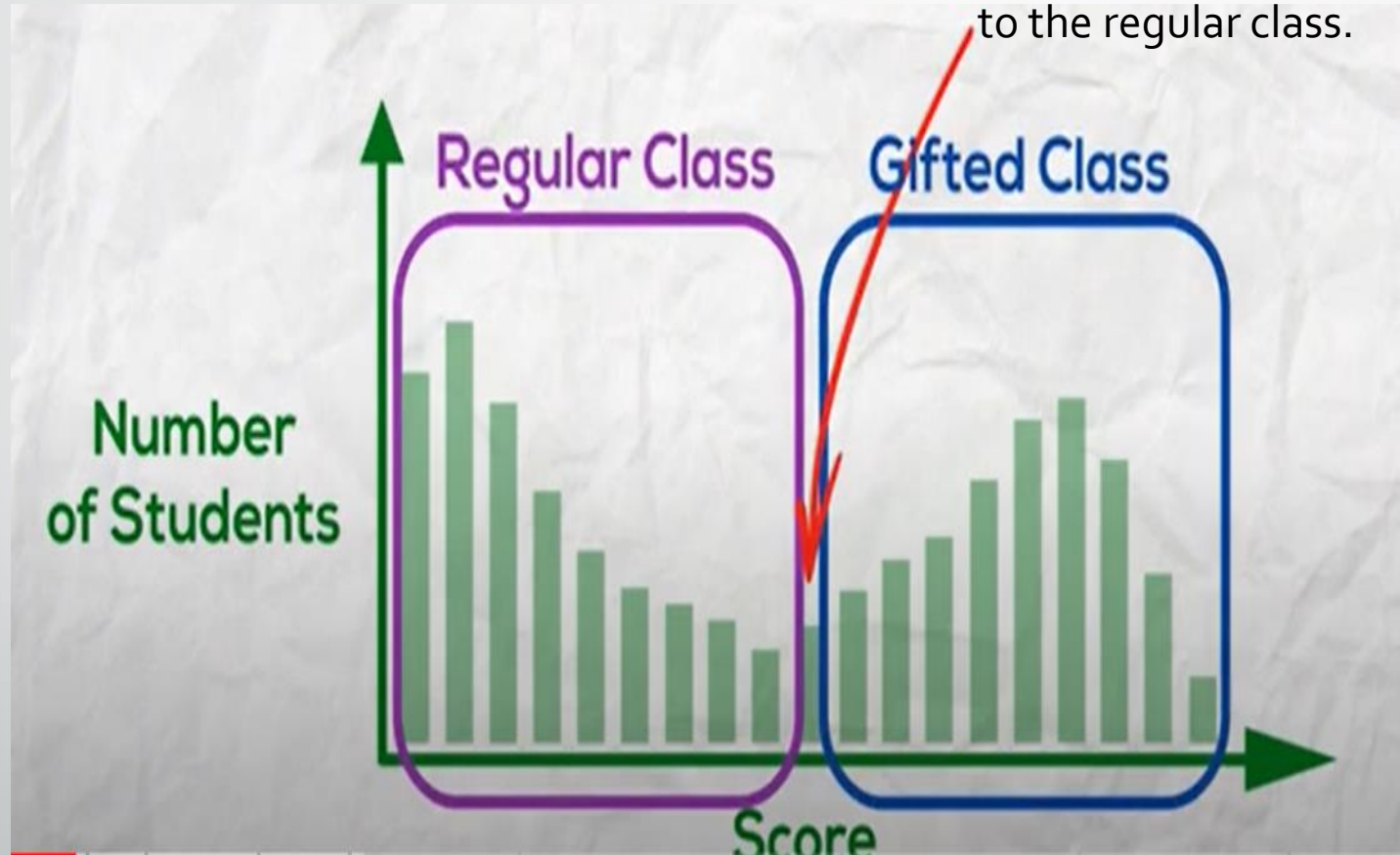
3.2 Gaussian Mixture Model-Example

Suppose we want to classify students as either regular class or gifted class.

- We can use their previous exam scores. Those who performed very well can be considered as gifted while the rest of the students were assigned to regular class.
- So we can use a histogram of last years exam scores to estimate the threshold to delineate the two classes.
- The histogram in Fig 2 creates two clusters of students, that is to say, those who will be classified as regular and those who will be classified as gifted.
- Based on our description of the GMM the n in our case is 2, i.e. we wish to classify new students as belonging to one of 2 clusters.

3.2 Histogram

We wish to find a threshold that will separate the two groups so that we can say students with x marks will belong to gifted class while the rest will belong to the regular class.



- Fig 2. Cluster histogram (ADipLearn, 2018)

3.2 Gaussian Mixture Model

- A Gaussian mixture model is parameterized by two types of values, the mixture component weights and the component means and variances/covariances as shown in Fig 3.
- The common notation for the symbol of summation in the fig 3 is (Σ)

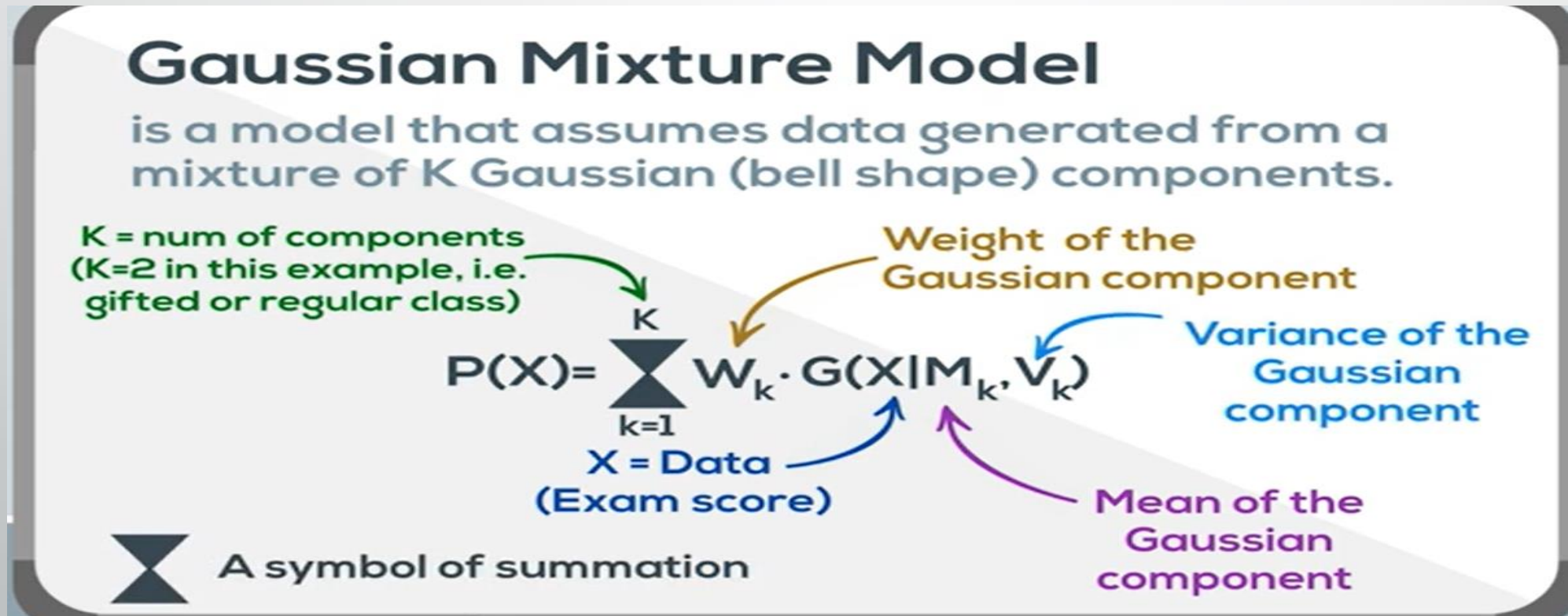


Fig 3 Gaussian Mixture Model probability formula (ADipLearn, 2018)

3.3 The Gaussian Function

The Gaussian function:

In matrix form: n refers to the number of feature sets. In this example, $n = 1$ (coz we use the exam score only)

$$G(X|M, V) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} [X-M]^T V^{-1} [X-M] \right]$$

$\pi = 3.1416$

V = variance of exam score in a component

M = mean matrix of exam score in a component

Probability of X (i.e. the exam score) belonging to a certain component

Figure 4 The Gaussian function (ADipLearn, 2018)

Steps of Gaussian Mixture Model

- Steps of Gaussian Mixture Model
- Step 1: Split the histogram into two equal portions. Each portion refers to either of the two clusters (remember $n = 2$)
- Step 2: initialize the mean and variance in each component and assign the weight depending on the number of clusters. (in this case since are 2 clusters then initial weight will be $\frac{1}{2} = 0.5$; if there were 4 clusters it would be $\frac{1}{4} = 0.25$, and so on).
- Step 3: calculate the new mean, variance and weight for each component (use known formulas for mean, variance and weight)

Repeat step 3 until the computed parameters are close to their previous value, that is, vary the weight until the computed parameters are acceptably close the initial values.

3.4 Example

- For purposes of this example, assume user data as depicted in table 1, on the scores of students. From the table the total number of students = 259

Number of students	Score
10	40
14	45
19	50
28	55
22	60
19	65
16	70
12	75
30	80
34	85
24	90
19	95
12	100

Table 1. students scores (Source : Mbogho, 2019)

3.3 Example (cont'd)

- The same data presented graphically as shown in Fig 5.
- This presents a different view of the data and enables us see the distribution of marks.

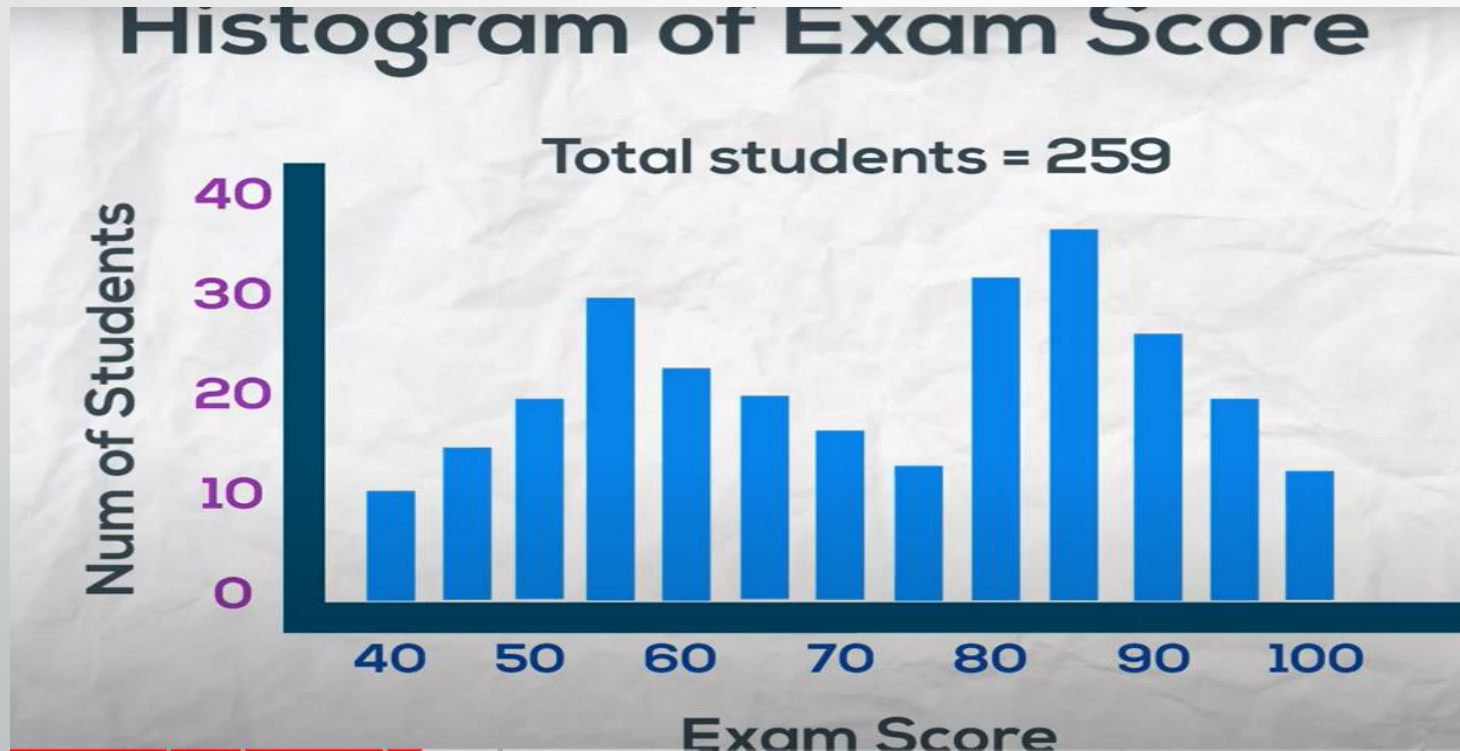


Fig 5. Histogram of exam score (ADipLearn, 2018)

3.3 Example –step 1(cont'd)

- Step 1: get the mean of all the points and split the histogram
- In this instance we shall be computing the weighted mean, whose formula is given by:

- $$\text{Weight} = \frac{\sum_{i=1}^n (\text{Weight}_i \times X_i)}{\sum_{i=1}^n (\text{Weight}_i)}$$

- Applying this to our data yields:

- $$\frac{(40 \times 10) + (45 \times 14) + (50 \times 19) + (55 \times 28) + \dots + (95 \times 19) + (100 \times 12)}{10 + 14 + 19 + 28 + 22 + 19 + 16 + 12 + 30 + 34 + 24 + 19 + 12} = 71.62$$

$$10 + 14 + 19 + 28 + 22 + 19 + 16 + 12 + 30 + 34 + 24 + 19 + 12$$

Example (cont'd)

- The mean we have computed is now assumed to be the threshold which will separate the two clusters, as shown in fig 6.

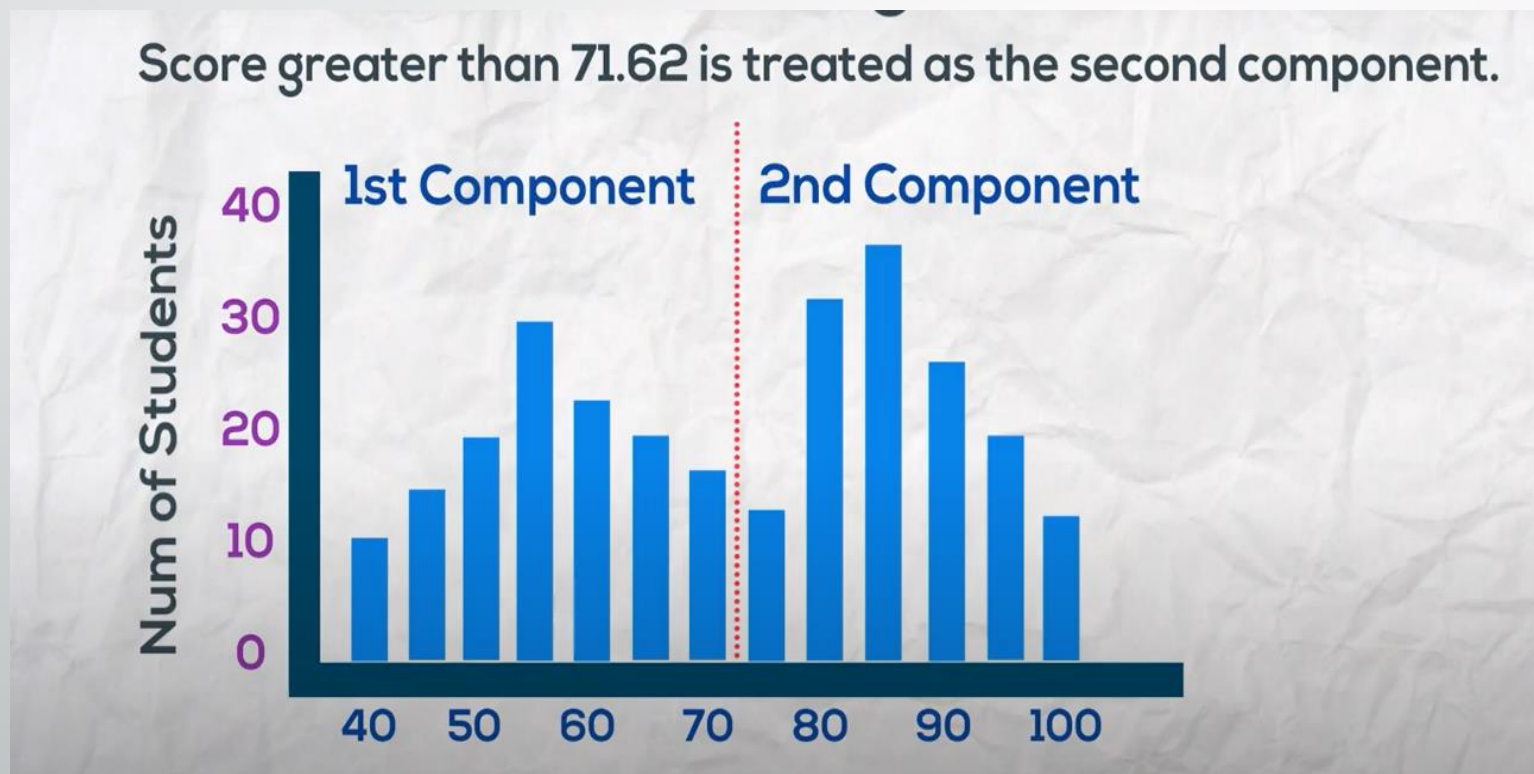


Fig 6. Cluster histogram (ADipLearn, 2018)

3.3 Example-step 2 (cont'd)

- Step 2 : Initialization
- First component (regular learner): we now compute the mean of the first group; this is the group whose marks are less than 71.62:

$$\text{Mean} = \frac{(40*10)+(45*15)+(50*19)+(55*28)+(60*22)+(65*19)+(70*16)}{10+14+19+28+22+19+16} = 56.21$$

The variance tells us the largest variance of any data point.

$$\text{Variance (V1)} = \frac{10(40-56.21)^2 + 14*(45-56.21)^2 + \dots + 16*(70-56.21)^2}{(10 + 14 + \dots + 16) - 1} = 78.64$$

So let us call the mean of the first component (regular) M_1 and the corresponding variance V_1 : $M_1=56.21$ and $V_1=78.64$

3.3 Example-step 2 (cont'd)

Step 2 : Initialization

- Second component (gifted):
- Mean = $\frac{75*15+80*30+85*34+90*24+95*19+100*12}{12+30+34+24+19+12} = 86.68$
- Variance = $\frac{12(75-86.88)^2 + 30*(80-86.88)^2 + \dots + 12*(100-86.88)^2}{(12+30+\dots+12)-1} = 52.16$

So using similar notation as for the first component we get:

- $M_2 = 86.68$
- $V_2 = 52.16$

3.3 Example-step 2 (cont'd)

- Step 3: calculate the new mean, variance and weight for each component (use equations)

Repeat step 3 until the computed parameters are close to their previous value

Using equation (1) – compute new weights using the obtained mean, variance and number of students.

Substitute the following in the equation

N which is the number of students=259

W (weights) – 0.5 since we had only 2 clusters

K=2 (because we have 2 clusters)

M- $M_1= 56.21$ or $M_2 = 86.66$

X = exam scores (40,45...100)

V- $V_1=78.64$, $V_2=52.16$

3.3 Example-step 3a (cont'd)

- Use the equation shown in figure 3 to computer new weights
- we solve the numerator first using the equation in figure 2 to get $G(X|M_1,V_1)$
- Substitute $X=40,45,50,55,60,65,70,75,80,85,90,95,100$, $M- M_1= 56.21$ and $V_1=78.64$
- We get $G(X| M_1,V_1)= 0.0006,0.020,0.035,0.045,0.041,0.020,0.019,0.005,0,0,0,0$
- Substituting them in the equation in figure 7 we get the new weight as shown in figure 8

$$W_{k,new} = \frac{1}{N} \frac{\sum_{n=1}^N W_k \cdot G(X_n | M_k, V_k)}{\sum_{k=1}^K W_k \cdot G(X_n | M_k, V_k)}$$

Figure 7 New weight Formula (ADipLearn, 2018)

$$\begin{aligned}
&= \left[\frac{0.5 \times 0.008}{0.5 \times (0.008 + 0)} \times 10 + \frac{0.5 \times 0.002}{0.5 \times (0.002 + 0)} \times 14 + \frac{0.5 \times 0.035}{0.5 \times (0.035 + 0)} \times 19 + \frac{0.5 \times 0.045}{0.5 \times (0.045 + 0)} \times 28 + \frac{0.5 \times 0.041}{0.5 \times (0.041 + 0)} \times 22 + \right. \\
&\quad \frac{0.5 \times 0.028}{0.5 \times (0.0028 + 0.001)} \times 19 + \frac{0.5 \times 0.013}{0.5 \times (0.013 + 0.004)} \times 16 + \frac{0.5 \times 0.005}{0.5 \times (0.005 + 0.015)} \times 12 + \frac{0.5 \times 0.001}{0.5 \times (0.001 + 0.036)} \times 30 + \frac{0.5 \times 0}{0.5 \times (0 + 0.054)} \times 34 + \\
&\quad \left. \frac{0.5 \times 0}{0.5 \times (0 + 0.05)} \times 24 + \frac{0.5 \times 0}{0.5 \times (0 + 0.028)} \times 19 + \frac{0.5 \times 0}{0.5 \times (0 + 0.01)} \times 12 \right] \times \frac{1}{259} \\
&= 0.494
\end{aligned}$$

- **Figure 8** New weight value (ADipLearn, 2018)

3.3 Example-step 3 (cont'd)

Use the equation shown in figure 3 we compute new weight 2

We solve the denominator f using the equation in figure 2 to get $G(X|M_2, V_2)$

Substitute $X=40,45,50,55,60,65,70,75,80,85,90,95,100$, $M- M_1= 86.66$
and $V_1=56.16$

We get $G(X| M_2, V_2)= 0,0,0,0,0,0.001,0.004,0.016,0.036,0.054,0.05,0.028,0.01$

Substituting them in the equation in figure 3 we get the new weight as shown in figure

$$\begin{aligned} &= \left[\frac{0.5 \times 0}{0.5 \times (0.008 + 0)} \times 10 + \frac{0.5 \times 0}{0.5 \times (0.002 + 0)} \times 14 + \frac{0.5 \times 0}{0.5 \times (0.035 + 0)} \times 19 + \frac{0.5 \times 0}{0.5 \times (0.045 + 0)} \times 28 + \frac{0.5 \times 0}{0.5 \times (0.041 + 0)} \times 22 + \right. \\ &\quad \frac{0.5 \times 0.001}{0.5 \times (0.0028 + 0.001)} \times 19 + \frac{0.5 \times 0.004}{0.5 \times (0.013 + 0.004)} \times 16 + \frac{0.5 \times 0.015}{0.5 \times (0.005 + 0.015)} \times 12 + \frac{0.5 \times 0.036}{0.5 \times (0.001 + 0.036)} \times 30 + \frac{0.5 \times 0.054}{0.5 \times (0 + 0.054)} \times 34 + \\ &\quad \left. \frac{0.5 \times 0.05}{0.5 \times (0 + 0.05)} \times 24 + \frac{0.5 \times 0.028}{0.5 \times (0 + 0.028)} \times 19 + \frac{0.5 \times 0.01}{0.5 \times (0 + 0.01)} \times 12 \right] \times \frac{1}{259} \\ &= 0.506 \end{aligned}$$

Figure 9 Weight 2 calculation (ADipLearn, 2018)

3.3 Example-step 3 (cont'd)

- Before we conduct iteration:
- $W_1=0.5$
- $W_2=0.5$
- After step 3a the new weights are:
- $W_1 = 0.494$
- $W_2 = 0.506$
- Now lets work on the mean : M_1 and M_2

Notice that this formula is same as calculating the weight:

Unsupervised Learning: Gaussian Mixture Model (1D GMM)

Step 3b: Iteration (Compute new mean)

$$M_{k,new} = \frac{\sum_{n=1}^N \frac{W_k \cdot G(X_n | M_k, V_k)}{\sum_{k=1}^K W_k \cdot G(X_n | M_k, V_k)} X_n}{\sum_{n=1}^N \frac{W_k \cdot G(X_n | M_k, V_k)}{\sum_{k=1}^K W_k \cdot G(X_n | M_k, V_k)}}$$

5:33 / 9:40 • Step 3b: Iteration (Compute new mean) [Scroll for details](#)

- Fig 10 Formula for new mean (ADipLearn, 2018)

3.3 Example-step 3b (cont'd)

Using equation(1) we calculate new means

The same procedure is used as the one in calculating weights

We get $G(X| M,V) = 0.0006, 0.020, 0.035, 0.045, 0.041, 0.020, 0.019, 0.005, 0, 0, 0, 0$

$$\begin{aligned} &= \frac{0.5 \times 0.008}{0.5 \times (0.008 + 0)} \times 40 \times 10 + \frac{0.5 \times 0.002}{0.5 \times (0.002 + 0)} \times 45 \times 14 + \frac{0.5 \times 0.035}{0.5 \times (0.035 + 0)} \times 50 \times 19 + \frac{0.5 \times 0.045}{0.5 \times (0.045 + 0)} \times 55 \times 28 + \\ &\frac{0.5 \times 0.041}{0.5 \times (0.041 + 0)} \times 60 \times 22 + \frac{0.5 \times 0.028}{0.5 \times (0.0028 + 0.001)} \times 65 \times 19 + \frac{0.5 \times 0.013}{0.5 \times (0.013 + 0.004)} \times 70 \times 16 + \frac{0.5 \times 0.005}{0.5 \times (0.005 + 0.015)} \times 75 \times 12 + \\ &\frac{0.5 \times 0.001}{0.5 \times (0.001 + 0.036)} \times 80 \times 30 + \frac{0.5 \times 0}{0.5 \times (0 + 0.054)} \times 85 \times 34 + \frac{0.5 \times 0}{0.5 \times (0 + 0.05)} \times 90 \times 24 + \frac{0.5 \times 0}{0.5 \times (0 + 0.028)} \times 95 \times 19 + \\ &\frac{0.5 \times 0}{0.5 \times (0 + 0.01)} \times 100 \times 12 = \mathbf{7728.26} \quad (\text{This is the mean of the numerator}) \end{aligned}$$

Figure 11 New means calculation (numerator) (ADipLearn, 2018)

3.3 Example-step 3b (cont'd)

- Using equation(1) we calculate new means
- The same procedure is used as the one in calculating weights
- We get $G(X| M_1, V_1) =$
 $0, 0, 0, 0, 0, 0.001, 0.004, 0.016, 0.036, 0.054, 0.05, 0.028, 0.01$

$$\begin{aligned}
 &= \frac{0.5 \times 0.008}{0.5 \times (0.008 + 0)} \times 10 + \frac{0.5 \times 0.002}{0.5 \times (0.002 + 0)} \times 14 + \frac{0.5 \times 0.035}{0.5 \times (0.035 + 0)} \times 19 + \frac{0.5 \times 0.045}{0.5 \times (0.045 + 0)} \times 28 + \frac{0.5 \times 0.041}{0.5 \times (0.041 + 0)} \times 22 + \\
 &\quad \frac{0.5 \times 0.028}{0.5 \times (0.0028 + 0.001)} \times 19 + \frac{0.5 \times 0.013}{0.5 \times (0.013 + 0.004)} \times 16 + \frac{0.5 \times 0.005}{0.5 \times (0.005 + 0.015)} \times 12 + \frac{0.5 \times 0.001}{0.5 \times (0.001 + 0.036)} \times 30 + \frac{0.5 \times 0}{0.5 \times (0 + 0.054)} \times 34 + \\
 &\quad \frac{0.5 \times 0}{0.5 \times (0 + 0.05)} \times 24 + \frac{0.5 \times 0}{0.5 \times (0 + 0.028)} \times 19 + \frac{0.5 \times 0}{0.5 \times (0 + 0.01)} \times 12 \\
 &= \mathbf{128.06} \quad (\text{This is the denominator})
 \end{aligned}$$

Fig 12 New means calculation (denominator) (ADipLearn, 2018)

3.3 Example-step 3 (cont'd)

- We divide the numerator by denominator and get the new Mean 1
- $M_1 \text{ new} = 7228.28/128.06 = 55.45$
- We skip the computation of $M_{2, \text{new}}$ since it can be done in a similar way

3.3 Example-results from step 3a and 3b (cont'd)

Before we conduct iteration:

$$W_1 = 0.4 \quad M_1 = 56.21 \quad V_1 = 78.64$$

$$W_2 = 0.5 \quad M_2 = 86.68 \quad V_2 = 52.16$$

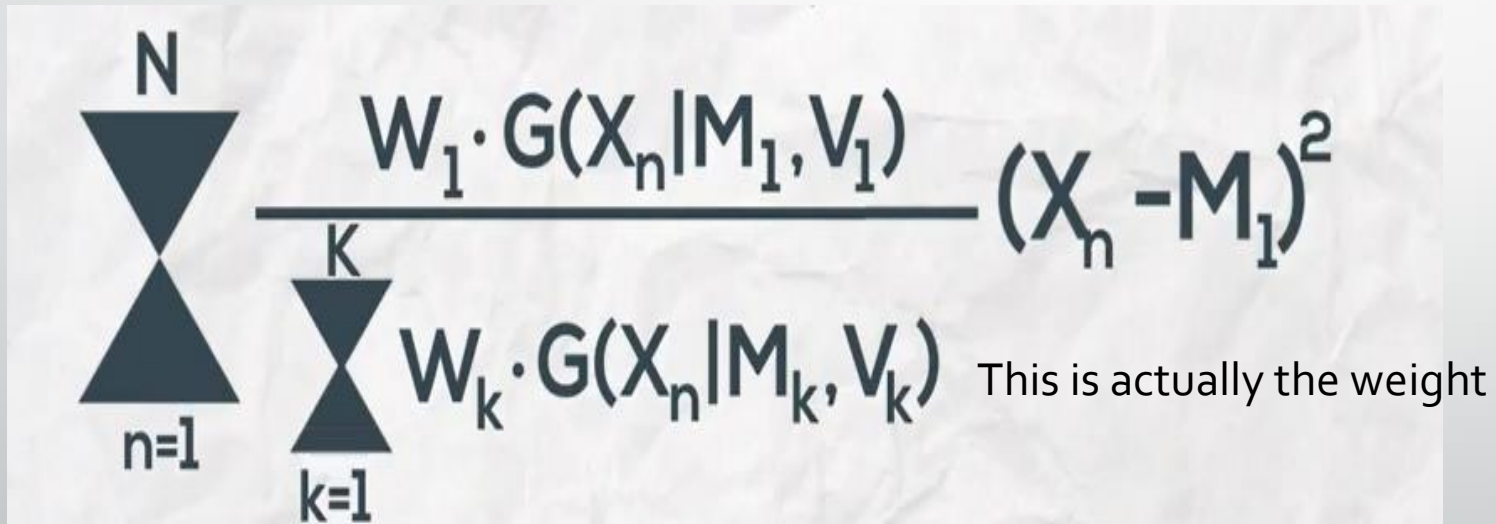
After step 3a, 3b and 3c

$$W_1 = 0.494 \quad M_1 = 56.45 \quad V_1 = 85.95$$

$$W_2 = 0.506 \quad M_2 = 86.46 \quad V_2 = 57.64$$

3.3 Example-step 3c (cont'd)

- We compute new variance using the equation shown in figure 13 and then output for the numerator is shown in Figure 14 (since the denominator is actually the weight).

The diagram shows a mathematical formula for variance. The numerator is a sum from n=1 to N of W_1 * G(X_n | M_1, V_1) * (X_n - M_1)^2. The denominator is a sum from k=1 to K of W_k * G(X_n | M_k, V_k). A note next to the denominator says "This is actually the weight".
$$\sum_{n=1}^N \frac{W_1 \cdot G(X_n | M_1, V_1)}{\sum_{k=1}^K W_k \cdot G(X_n | M_k, V_k)} (X_n - M_1)^2$$

This is actually the weight

Fig 13 Variance formula (ADipLearn, 2018)

$$\begin{aligned}
&= \frac{0.5 \times 0.008}{0.5 \times (0.008 + 0)} \times (40 - 56.21)^2 \times 10 + \frac{0.5 \times 0.002}{0.5 \times (0.002 + 0)} \times (45 - 56.21)^2 \times 14 + \frac{0.5 \times 0.035}{0.5 \times (0.035 + 0)} \times (50 - 56.21)^2 \times 19 + \frac{0.5 \times 0.045}{0.5 \times (0.045 + 0)} \times (55 - 56.21)^2 \times 28 + \\
&\frac{0.5 \times 0.041}{0.5 \times (0.041 + 0)} \times (60 - 56.21)^2 \times 22 + \frac{0.5 \times 0.028}{0.5 \times (0.028 + 0.001)} \times (65 - 56.21)^2 \times 19 + \frac{0.5 \times 0.013}{0.5 \times (0.013 + 0.004)} \times (70 - 56.21)^2 \times 16 + \frac{0.5 \times 0.005}{0.5 \times (0.005 + 0.015)} \times (75 - 56.21)^2 \times 12 + \\
&\frac{0.5 \times 0.001}{0.5 \times (0.001 + 0.036)} \times (80 - 56.21)^2 \times 30 + \frac{0.5 \times 0}{0.5 \times (0 + 0.054)} \times (85 - 56.21)^2 \times 34 + \frac{0.5 \times 0}{0.5 \times (0 + 0.05)} \times (90 - 56.21)^2 \times 24 + \frac{0.5 \times 0}{0.5 \times (0 + 0.028)} \times (95 - 56.21)^2 \times 19 + \\
&\frac{0.5 \times 0}{0.5 \times (0 + 0.01)} \times (100 - 56.21)^2 \times 12 = \mathbf{11006.97}
\end{aligned}$$

- **Figure 14** Variance value (numerator) (ADipLearn, 2018)

3.3 Example-step 3c (cont'd)

- We then compute the denominator the someway and obtain 128.
- We then divide the numerator by the denominator $1106/128= 86$.
- We compute the second variance the same way

3.3 Example-results from step 3a,3b,3c(cont'd)

- We can keep doing the iteration , replace the six new parameters in the equations , until there is no significant change
- After step 3a,3b and 3c
- $W_1 = 0.494$ $M_1 = 56.45$ $V_1 = 85.95$
- $W_2 = 0.506$ $M_2 = 86.46$ $V_2 = 57.64$
- We can keep doing the iterations , replace the new six parameters in the equations until there is no significant change

3.4 Example-step 4 (cont'd)

- The threshold can be found when the following requirements fulfills
- i.e. the data point which is located in both the first Gaussian component and the second Gaussian component using equation in Figure 15 and 16

$$W_1 \cdot G(X|M_1, V_1) = W_2 \cdot G(X|M_2, V_2)$$

Figure 15 *Threshold Formula* (ADipLearn, 2018)

Threshold Formula

Therefore,

$$\frac{W_1}{(2\pi)^{\frac{1}{2}}|V_1|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}[X-M_1]^T V_1^{-1}[X-M_1]\right] =$$

$$\frac{W_2}{(2\pi)^{\frac{1}{2}}|V_2|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}[X-M_2]^T V_2^{-1}[X-M_2]\right]$$

$$\frac{W_1 |V_2|^{\frac{1}{2}}}{W_2 |V_1|^{\frac{1}{2}}} = \frac{\exp\left[-\frac{1}{2}[X-M_2]^T V_2^{-1}[X-M_2]\right]}{\exp\left[-\frac{1}{2}[X-M_1]^T V_1^{-1}[X-M_1]\right]}$$

- **Figure 16** Threshold Formula (ADipLearn, 2018)

3.4 Example-step 4 (cont'd)

Take natural log on both sides:

$$\log \left[\frac{w_1 |v_2|^{\frac{1}{2}}}{w_2 |v_1|^{\frac{1}{2}}} \right] = -\frac{[x-m_2]^2}{2v_2} + \frac{[x-m_1]^2}{2v_1}$$

$$2v_1v_2 \log \left[\frac{w_1 |v_2|^{\frac{1}{2}}}{w_2 |v_1|^{\frac{1}{2}}} \right] = -v_1[x-m_2]^2 + v_2[x-m_1]^2$$

$$[v_2-v_1]x^2 + 2[v_1m_2 - v_2m_1]x + v_2m_1^2 - v_1m_2^2 - 2v_1v_2 \log \left[\frac{w_1 |v_2|^{\frac{1}{2}}}{w_2 |v_1|^{\frac{1}{2}}} \right] = 0$$

Figure 17 Threshold formula (ADipLearn, 2018)

3.4 Example-step 4 (cont'd)

- Assume $a=[V_2-V_1] = 57.64-85.95=-28.31$
- $b=2[V_1M_2-V_2M_1] = 2(85.95*86.46-57.64)= 8354.918$
- c is the last part of the equation which works out to = -457866
- You must have noticed the equation is of the form of a quadratic equation, that is $ax^2 + bx + c = 0$, which is why we find a , b and c .

3.4 Example-step 4 (cont'd)

- Substituting the values of a,b, and c in the quadratic equations we get 72.72 as the final score

The solution of quadratic equation:

$$X = \frac{-B \pm [B^2 - 4AC]^{\frac{1}{2}}}{2A}$$

With

A = -28.31, B = 8354.918, C = -457866

The solution would be:

X = 72.72 or ~~X = 222.4~~
Out of exam score range!

Figure 18 The final score (ADipLearn, 2018)

3.3 Example-Final Results (cont'd)

- Any student with exam score less than 73. he or she would be most likely from the regular class.
- Any student with exam score greater than or equal to 73 would be most likely form the gifted cluster.

Summary

- The clustering method helps grouping valuable data into clusters and picks appropriate results based on different techniques.
- Clustering is used when data is not labelled; it is an unsupervised machine learning technique.
- Clustering methods can be largely grouped as being partition based, hierarchical based, density based, grid based or model based .
- GMM can be used to find clusters in datasets where the clusters may not be clearly defined.
- Gaussian mixture models take into account the variance, mean and weights of the data

References

- A.D.L. [ADipLearn]. (2018, December 2). *Unsupervised Learning: Gaussian Mixture Model (1D GMM)* [Video]. YouTube. <https://www.youtube.com/watch?v=fVsmnZqrBUs&feature=youtu.be>
- Aggarwal, C. (2014). *Data clustering Algorithms and Applications*. CRC Press.
- Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
- Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, 2(3), 267-279.
- McGonagle, J., Pilling, G., Dobre, A., Tembo, V., Kurmukov, A., Chumbley, A., Ross, E., & Khim, J. (n.d.). *Gaussian mixture model*. Brilliant Math & Science Wiki. Retrieved April 20, 2022, from <https://brilliant.org/wiki/gaussian-mixture-model/>
- Mbogho, A. W. (2019). *Introduction 3. Machine Learning*.
- Pedamkar, P. (2021, October 18). *Clustering methods: Importance and techniques of clustering methods*. EDUCBA. Retrieved April 20, 2022, from <https://www.educba.com/clustering-methods/>
- Richard et al., (2014). *Algorithm for clustering and data*. CRC Press.