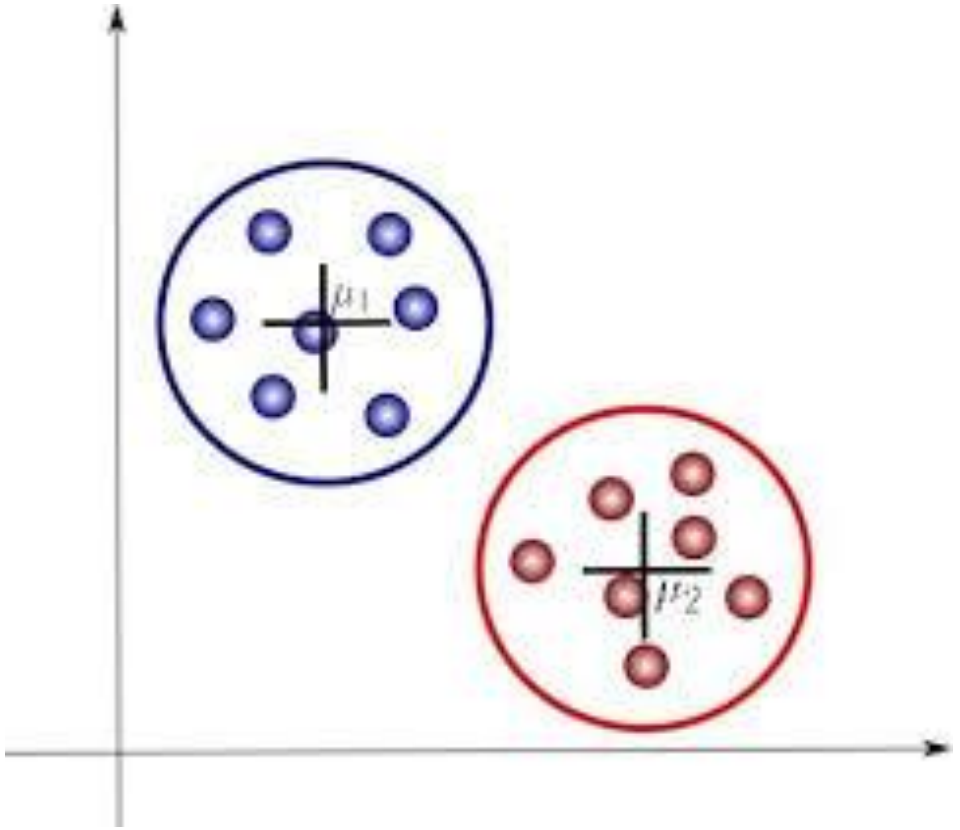
Gaussian Mixture Model (GMM):



What is a Gaussian Mixture Model (GMM)?

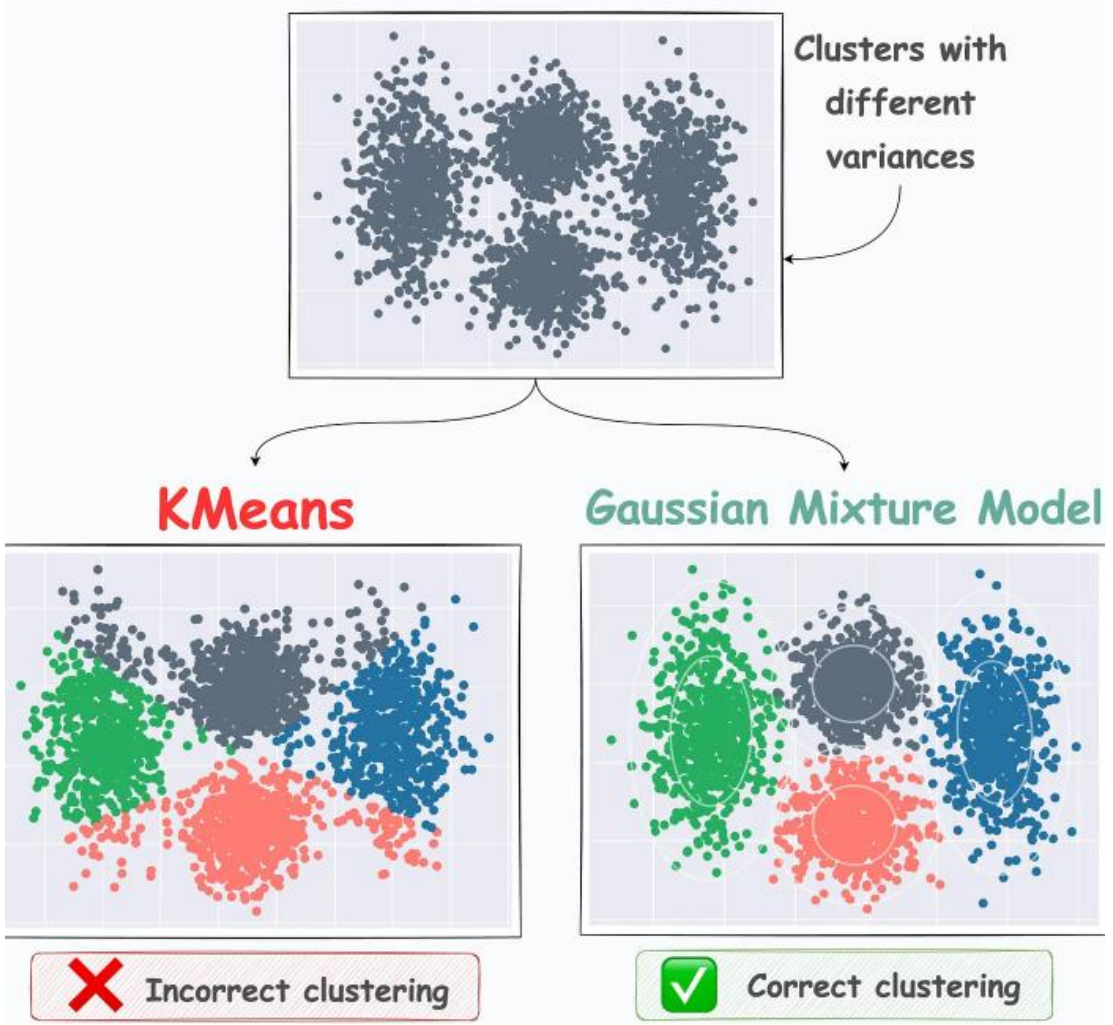
- A **Gaussian Mixture Model (GMM)** is a statistical model used for **clustering** data.
  It assumes that the data is generated from a combination of **multiple Gaussian (normal) distributions**.
- Unlike **K-Means**, where each data point belongs to only one cluster, **GMM performs soft clustering**, meaning each point can belong to **multiple clusters with different probabilities**.

Why is GMM Used?

- **Clusters have different shapes and sizes:** GMM can detect clusters that are **elliptical, circular, or irregular**, while **K-Means** works best with **spherical clusters**.
- **Soft clustering is needed:** It assigns each data point a **probability of belonging** to each cluster, which is useful in areas like **recommendation systems** or **customer segmentation**.
- **Flexible clustering:** GMM performs well when **clusters overlap**, allowing data points to **partially belong to multiple clusters**.

How Does GMM Work?

GMM assumes that data comes from multiple Gaussian distributions. Using the **Expectation-Maximization (EM)** algorithm, it alternates between estimating the **probabilities of cluster membership** (E-step) and **updating the parameters** of each distribution (M-step), until the model stabilizes.



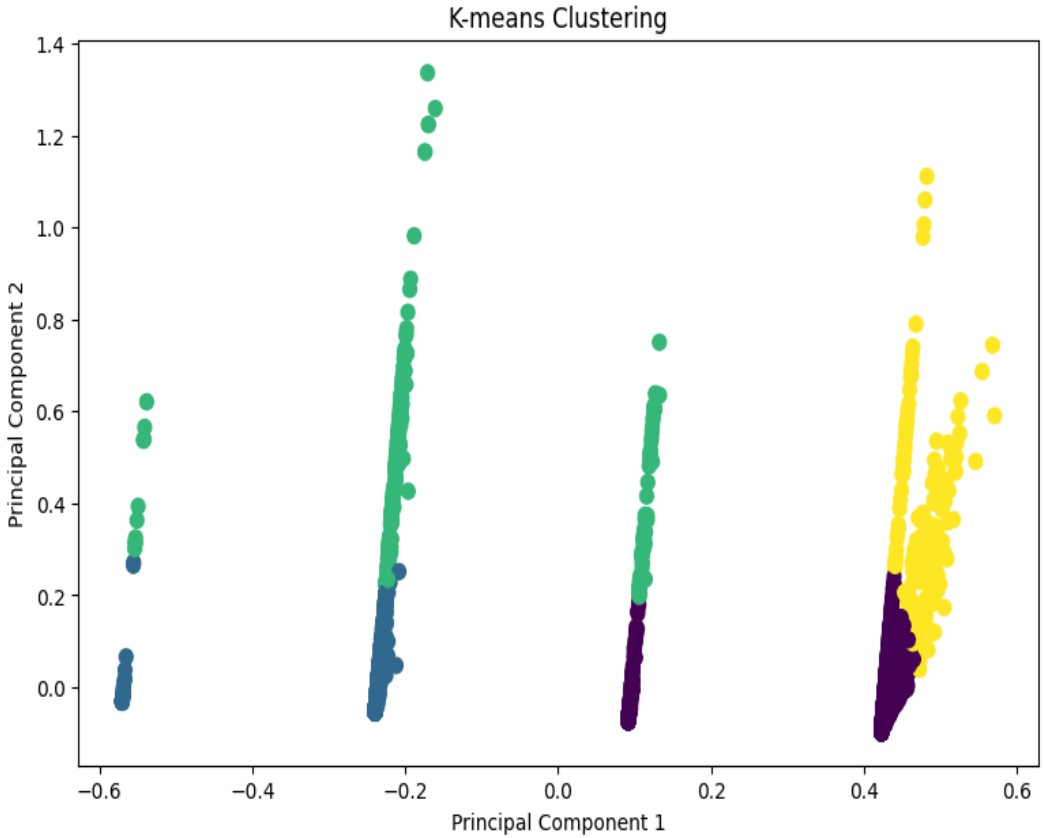
K-Means VS GMM

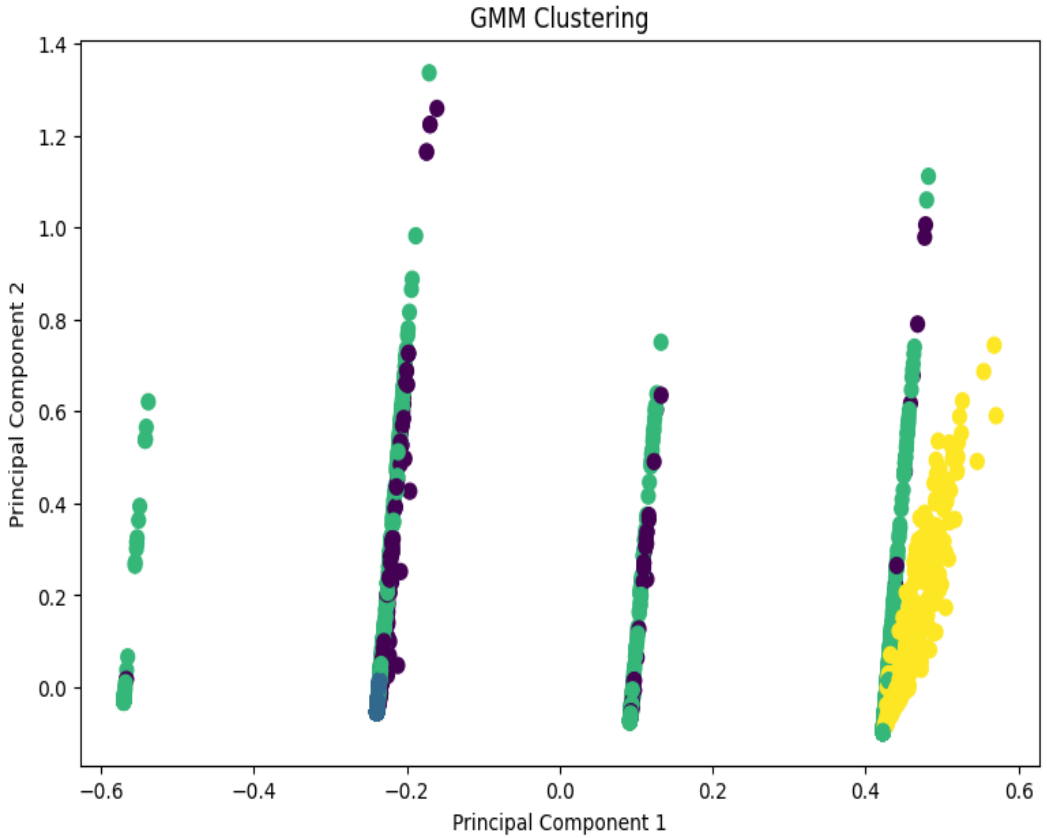
Comparison with K-Means

Feature	GMM	K-Means
Clustering Type	Soft (probability-based)	Hard (each point in one cluster only)
Cluster Shape	Any shape (elliptical, irregular)	Only spherical
Output	Probabilities for each cluster	Single cluster assignment
Speed	Slower (more complex calculations)	Faster
Use Case	Overlapping or irregular clusters	Well-separated spherical clusters

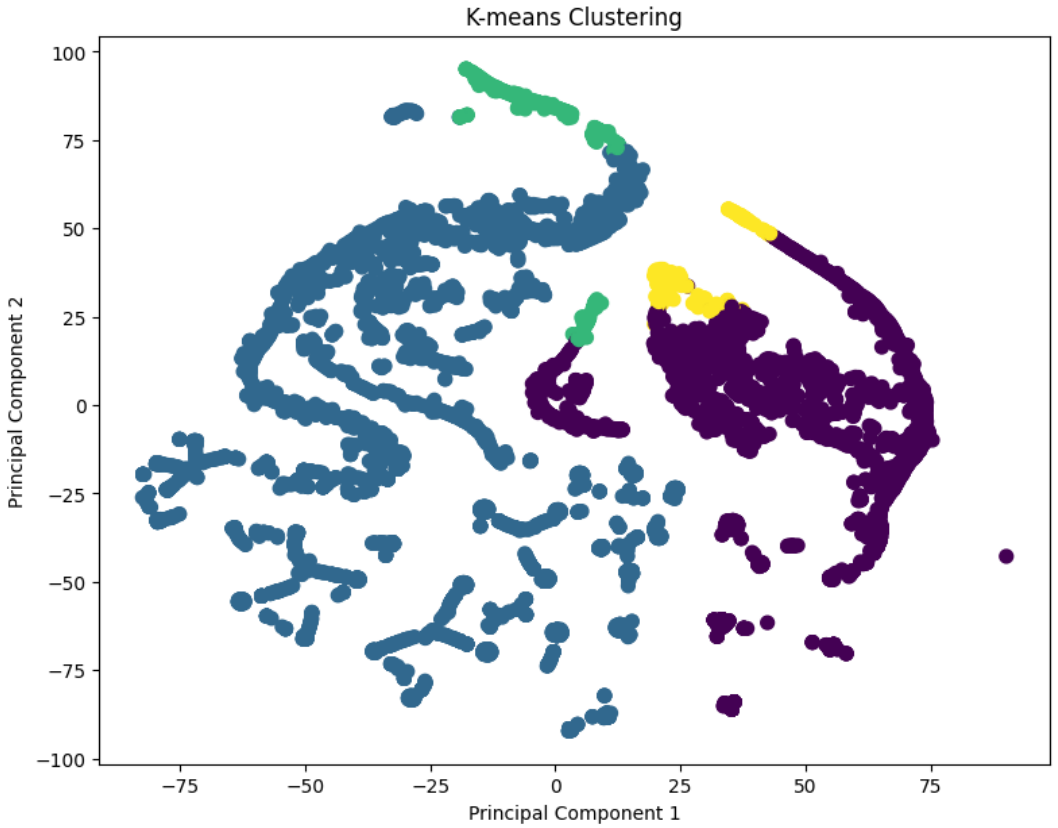
In summary, GMM is a powerful and flexible clustering algorithm that works well for complex data distributions but requires more computational resources and careful parameter selection.

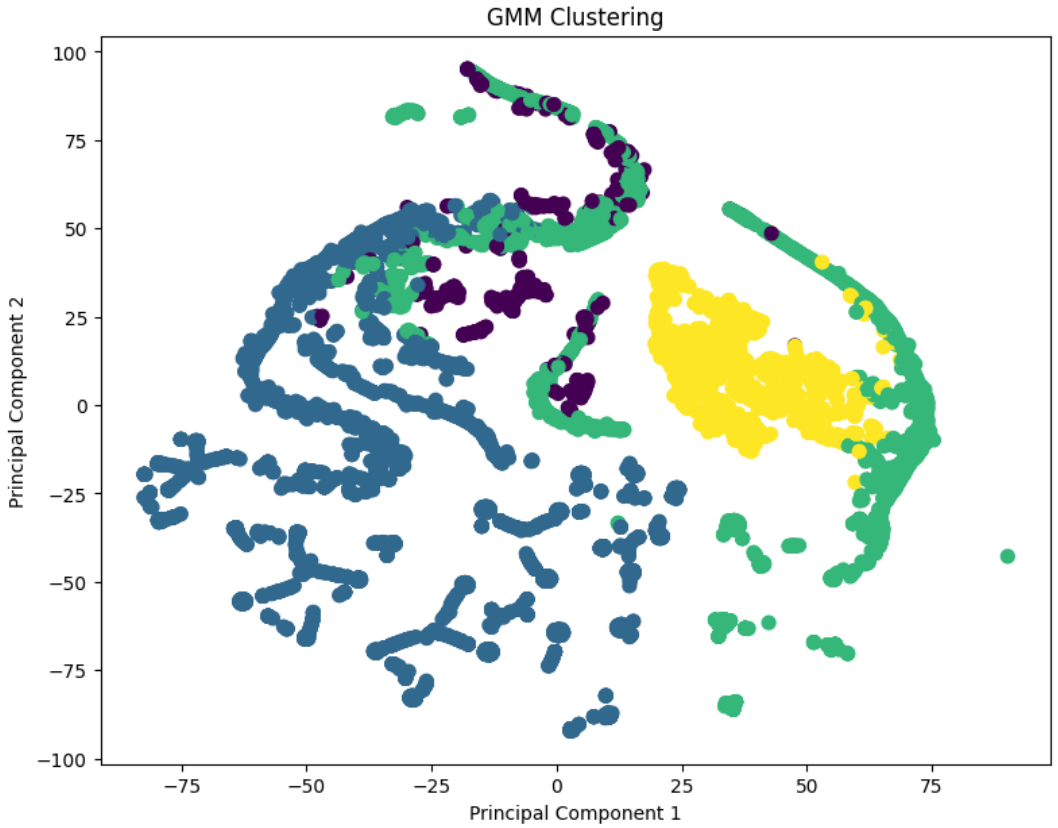
Example: Clustering Analysis using K-Means and GMM with PCA





The same example but with T-SNE





General Conclusion:

Both clustering approaches, **K-means** and **GMM (Gaussian Mixture Model)**, successfully identified patterns in the data but differ in how they form groups.

- With **K-means**, the clusters are **well separated but rigid**, making it suitable for data with **simple and spherical shapes**.
- In contrast, **GMM** provides **greater flexibility** by modeling the **probability of belonging** to each cluster, allowing it to better adapt to **complex or overlapping structures**.
- The **PCA visualizations** show the overall separation, while the **t-SNE plots** reveal that GMM captures the **non-linear structure** of the data more effectively.