

# Gaussian Mixture Model Tutorial

September 17, 2025

## 1 Introduction

Gaussian Mixture Models (GMMs) describe data as a weighted sum of multiple Gaussian components. Each component captures a latent cluster with its own mean, covariance, and mixing weight. GMMs excel at modeling elliptical clusters and overlapping distributions, offering richer shapes than K-means while retaining probabilistic assignments and density estimates.

## 2 Theory and Formulas

### 2.1 Mixture Likelihood

For observation  $\mathbf{x}$  in  $\mathbb{R}^d$ , the mixture density is

$$p(\mathbf{x} | \Theta) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (1)$$

where  $\Theta = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$  and mixing weights satisfy  $\sum_k \pi_k = 1$  with  $\pi_k \geq 0$ .

### 2.2 Expectation-Maximization

Maximum likelihood estimation is commonly performed via the EM algorithm:

1. **E-step**: compute responsibilities  $\gamma_{ik} = p(z_i = k | \mathbf{x}_i, \Theta^{(t)})$  using Bayes' rule.
2. **M-step**: update parameters with responsibility-weighted sufficient statistics:

$$\pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \gamma_{ik}, \quad (2)$$

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\sum_{i=1}^n \gamma_{ik} \mathbf{x}_i}{\sum_{i=1}^n \gamma_{ik}}, \quad (3)$$

$$\boldsymbol{\Sigma}_k^{(t+1)} = \frac{\sum_{i=1}^n \gamma_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k^{(t+1)}) (\mathbf{x}_i - \boldsymbol{\mu}_k^{(t+1)})^\top}{\sum_{i=1}^n \gamma_{ik}}. \quad (4)$$

3. Iterate until the log-likelihood converges.

## 2.3 Model Selection

Criteria such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) penalize model complexity when choosing the number of components  $K$  and covariance structure (full, tied, diagonal, or spherical). Posterior responsibilities provide soft assignments that can be thresholded for anomaly detection.

## 3 Applications and Tips

- **Density estimation:** model multi-modal patterns in speech, finance, or sensor signals.
- **Soft clustering:** responsibilities capture uncertainty and overlapping clusters, useful for customer segmentation or topic discovery.
- **Anomaly detection:** low likelihood under the mixture indicates unusual behavior relative to learned components.
- **Best practices:** scale features, initialize with K-means or multiple random starts, monitor covariance conditioning, and guard against singularities by adding regularization.

## 4 Python Practice

The script `gen_clustering_gmm_figures.py` fits GMMs to synthetic data, visualizes contour lines of the learned density, and plots BIC for different component counts.

Listing 1: Excerpt from `gen_clustering_gmm_figures.py`

```
1 from sklearn.mixture import GaussianMixture
2
3 model = GaussianMixture(n_components=3, covariance_type="full",
4                           init_params="kmeans", random_state=42)
5 model.fit(points)
6 responsibilities = model.predict_proba(points)
7
8 bic_scores = []
9 for k in range(1, 7):
10     gm = GaussianMixture(n_components=k, covariance_type="full",
11                           init_params="kmeans", random_state=42)
12     gm.fit(points)
13     bic_scores.append(gm.bic(points))
```

## 5 Result

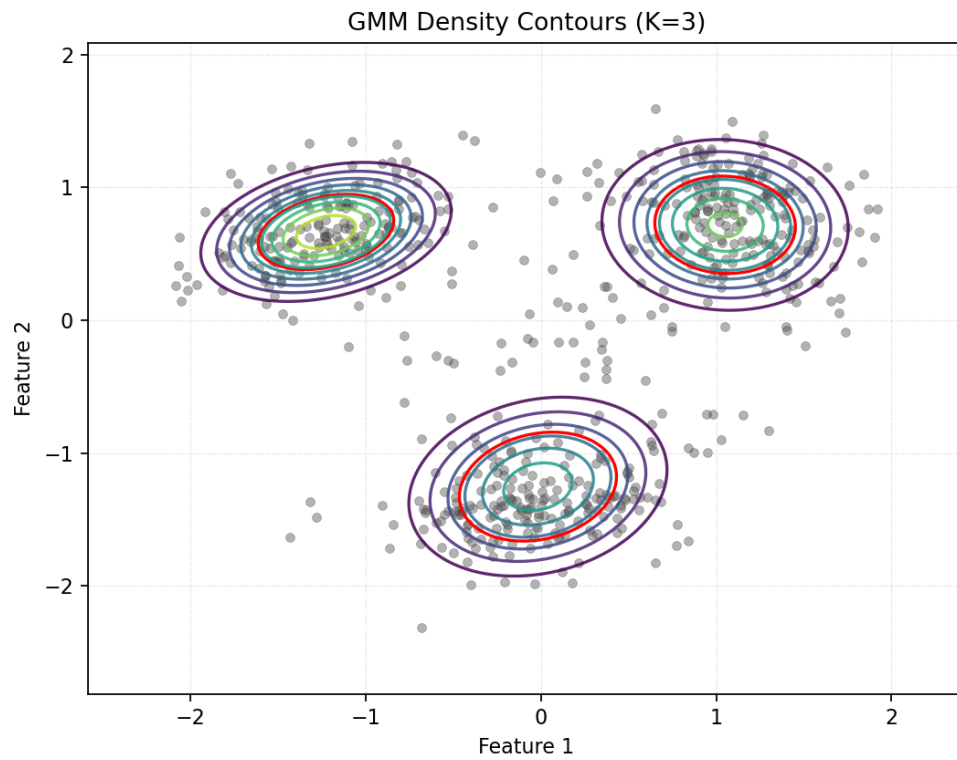


Figure 1: Density contours from a three-component GMM fitted to synthetic data

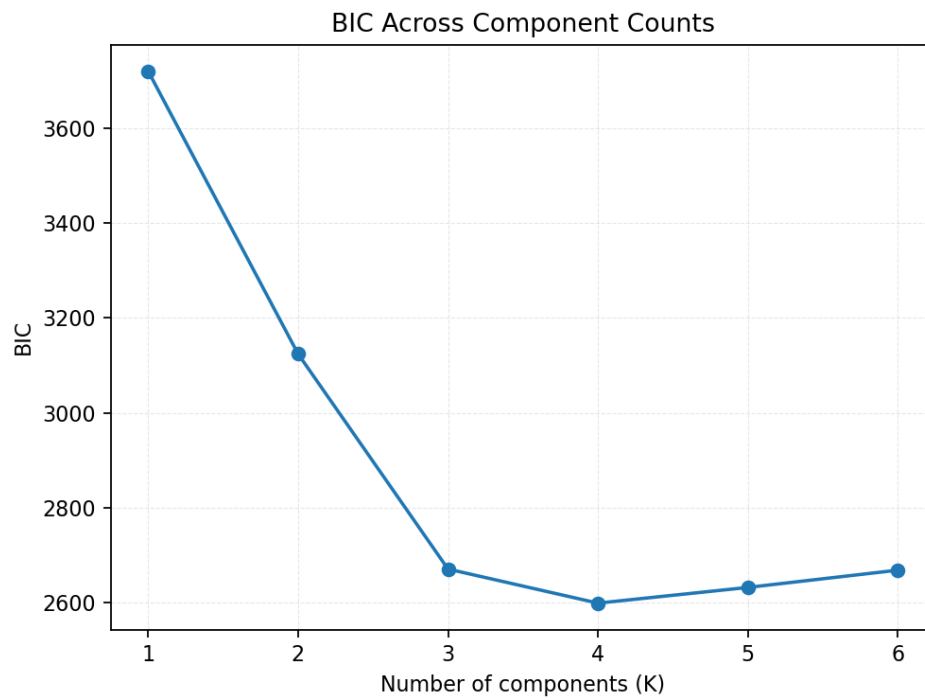


Figure 2: BIC across candidate component counts

## 6 Summary

Gaussian Mixture Models generalize K-means by modeling covariance structure and providing probabilistic assignments. EM alternates between computing responsibilities and updating parameters until convergence. Diagnostics such as BIC and likelihood surfaces help tune component counts and detect degenerate solutions. The synthetic example highlights how density visualization and model selection complement each other in practice.