Speech Recognition (DSAI 456)

Lecture 4

Mohamed Ghalwash
mghalwash@zewailcity.edu.eg

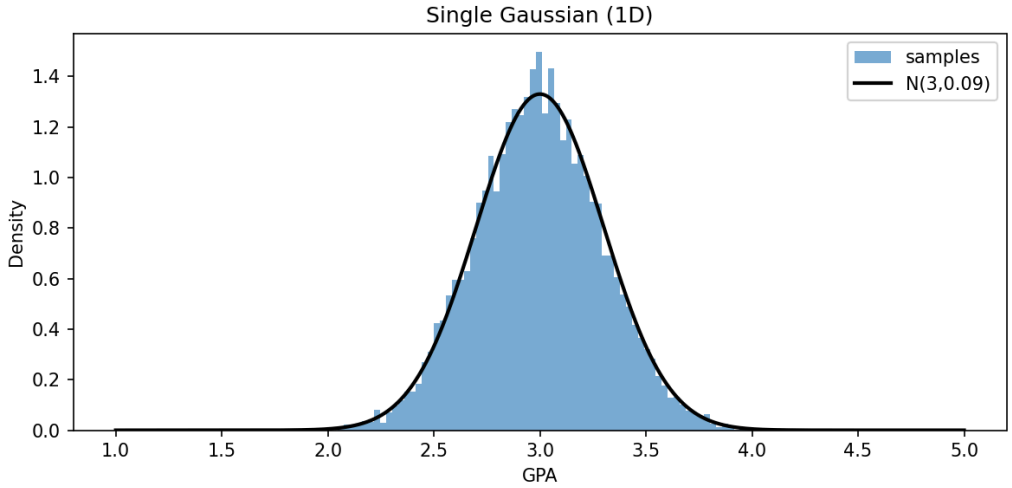
Lecture 3 Recap

- Mel Spectrum
- Mel Filter Bank
- MFCC

Agenda

- Motivation and intuition
- Mathematical formulation
- EM algorithm (E-step / M-step)
- Using GMMs in speech recognition
- Practical tips

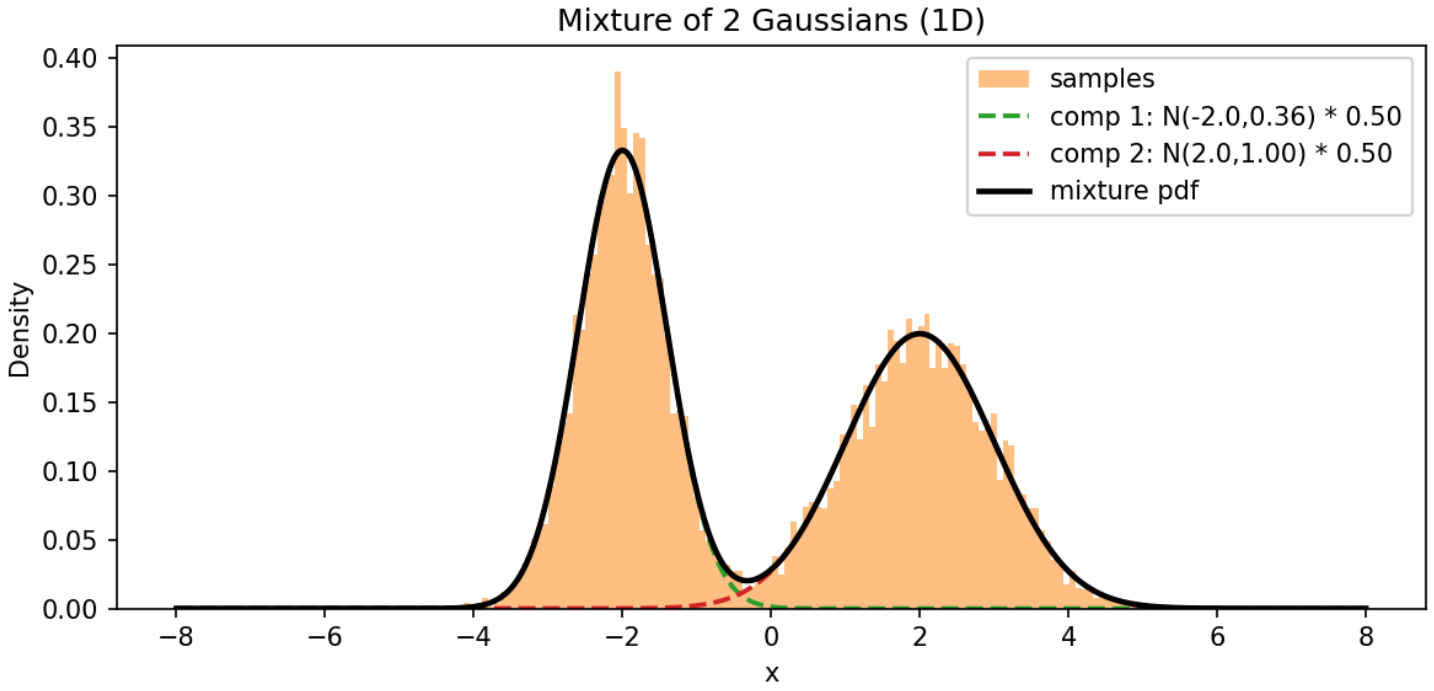
Motivation: GPA Distribution



$$GPA \sim \mathcal{N}(\mu, \sigma)$$

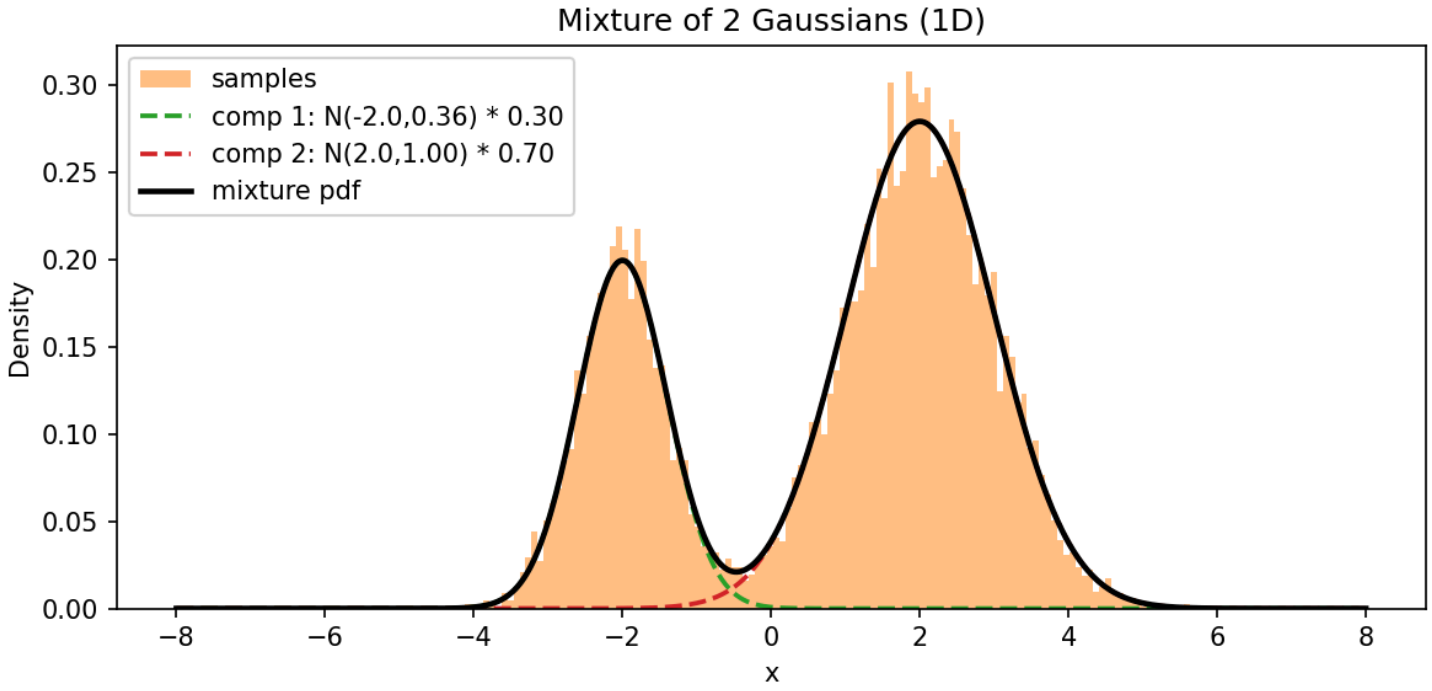
$$p(GPA) = \mathcal{N}(GPA \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(GPA-\mu)^2}{2\sigma^2}\right)$$

Motivation: Two Distributions



$$x \sim \mathcal{N}_1(\mu_1, \sigma_1) + \mathcal{N}_2(\mu_2, \sigma_2)$$

Motivation: Two Distributions

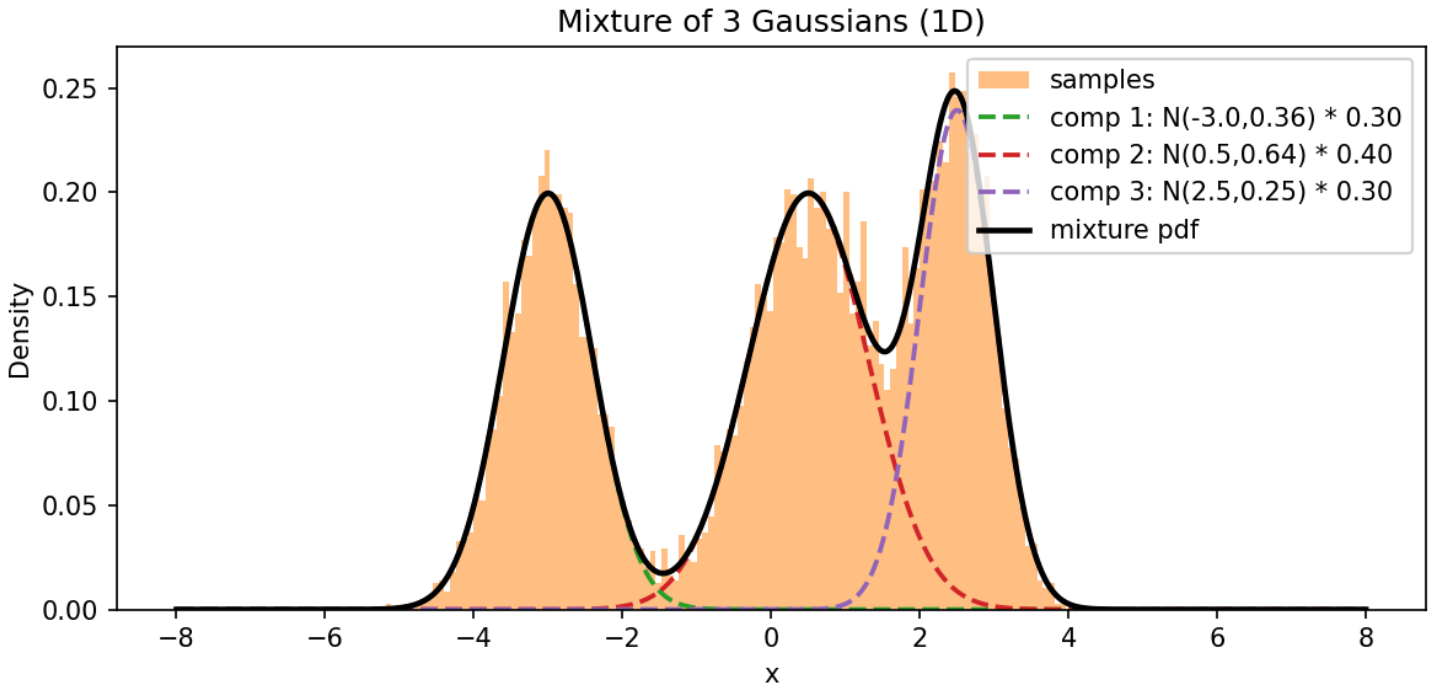


$$x \sim \mathcal{N}_1(\mu_1, \sigma_1) + \mathcal{N}_2(\mu_2, \sigma_2)$$

Should be weighted sum

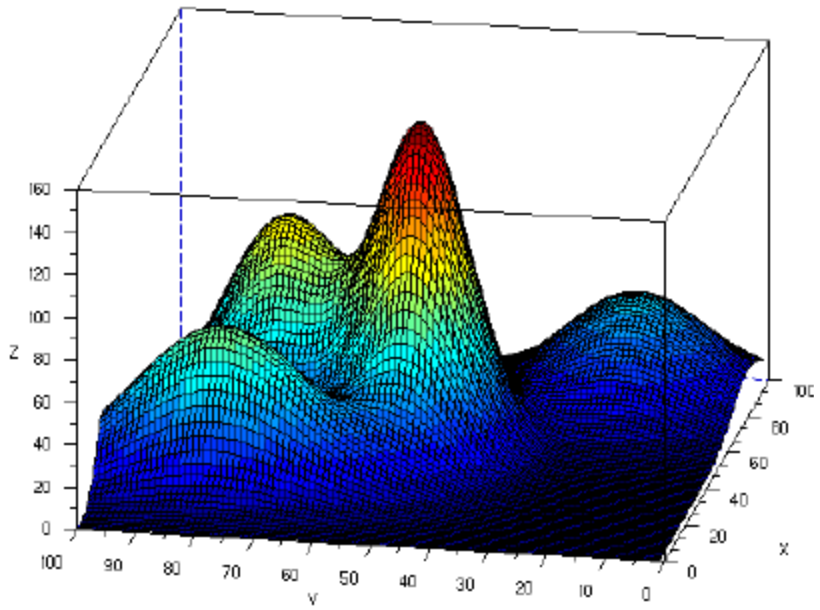
$$x \sim 0.3 * \mathcal{N}_1(\mu_1, \sigma_1) + 0.7 * \mathcal{N}_2(\mu_2, \sigma_2)$$

Motivation: Three Distributions



$$x \sim 0.3 * \mathcal{N}_1(\mu_1, \sigma_1) + 0.4 * \mathcal{N}_2(\mu_2, \sigma_2) + 0.3 * \mathcal{N}_2(\mu_2, \sigma_2)$$

Motivation: Mixture of Multivariate Distributions



Univariate vs Multivariate

Each component is modeled as

- Univariate

$$p(x) = \mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Multivariate

$$p(x) = \mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

Gaussian Mixture Model

A weighted sum of Gaussian components used to model complex continuous distributions

What is a GMM?

- Density

$$p(x) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

$$\text{such that } \sum_{k=1}^{K} \pi_k = 1, \quad \pi_k \geq 0$$

- Parameters:
  - weight π_k
  - mean μ_k
  - covariance Σ_k

Parameter Estimation

How to learn parameters of the GMM?

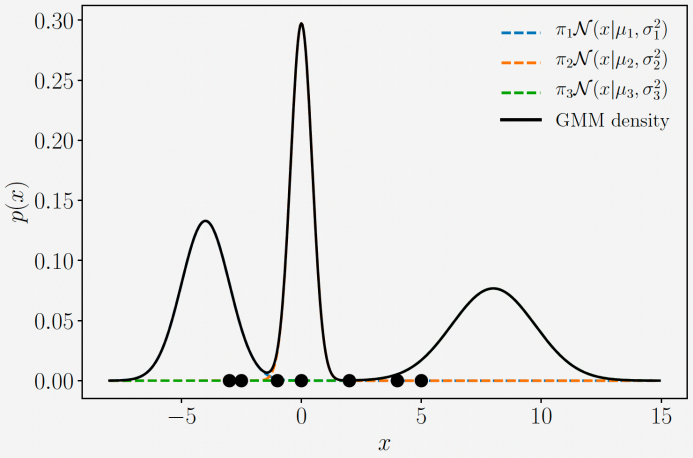
- **Problem:** Parameters are multiplied

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

- **Solution:** Iterative algorithm

Responsibilities (soft assignments)

The responsibility r_{nk} represents the probability that x_n has been generated by the k^{th} component



$$\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \\ 0.057 & 0.943 & 0.0 \\ 0.001 & 0.999 & 0.0 \\ 0.0 & 0.066 & 0.934 \\ 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \in \mathbb{R}^{N \times K}$$

$$r_{nk} = \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

$$N_k = \sum_{k=1}^{K} r_{nk}$$

Optimization: Maximum Likelihood

- Full data likelihood

$$p(X \mid \theta) = \prod_{n=1}^{N} p(x_n|\theta)$$

$$\Rightarrow \log p(X \mid \theta) = \log \prod_{n=1}^{N} p(x_n|\theta)$$

$$= \sum_{n=1}^{N} \log p(x_n|\theta)$$

$$\Rightarrow \mathcal{L} = \sum_{n=1}^{N} \log p(x_n|\theta) + \lambda\Big(\sum_{k=1}^{K} \pi_k - 1\Big)$$

Derivation: closed-form for π_k (E-step)

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \pi_j} &= \sum_{n=1}^{N} \frac{\partial \log p(x_n|\theta)}{\partial \pi_j} + \lambda \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial p(x_n|\theta)}{\partial \pi_j} + \lambda \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\partial \pi_j} + \lambda \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \mathcal{N}(x_n \mid \mu_j, \Sigma_j) + \lambda \\
&= \sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} + \lambda
\end{aligned}
$$

Derivation: closed-form for π_k (E-step)

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} + \lambda$$

$$= \frac{1}{\pi_j} \sum_{n=1}^{N} \frac{\pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} + \lambda$$

$$= \frac{1}{\pi_j} \sum_{n=1}^{N} r_{nj} + \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = 0 \Rightarrow \frac{1}{\pi_j} \sum_{n=1}^{N} r_{nj} = -\lambda \Rightarrow \pi_j = -\frac{1}{\lambda} \sum_{n=1}^{N} r_{nj}$$

$$\Rightarrow \boxed{\pi_j = -\frac{N_j}{\lambda}}$$

Derivation: closed-form for λ (E-step)

$$\mathcal{L} = \sum_{n=1}^{N} \log p(x_n|\theta) + \lambda\Big(\sum_{k=1}^{K} \pi_k - 1\Big)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{k=1}^{K} \pi_k - 1$$

$$\Rightarrow \sum_{k=1}^{K} \pi_k = 1 \xRightarrow{\pi_k = -\frac{N_k}{\lambda}} \sum_{k=1}^{K} -\frac{N_k}{\lambda} = 1$$

$$\Rightarrow \lambda = -\sum_{k=1}^{K} N_k = -N \Rightarrow \boxed{\pi_j = \frac{N_j}{N}}$$

Derivation: closed-form for μ_k (M-step)

$$\mathcal{L} = \sum_{n=1}^{N} \log p(x_n|\theta) + \lambda\Big(\sum_{k=1}^{K} \pi_k - 1\Big)$$

$$\begin{aligned}
\frac{\partial \log p(X \mid \theta)}{\partial \mu_j} &= \sum_{n=1}^{N} \frac{\partial \log p(x_n|\theta)}{\partial \mu_j} \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial p(x_n|\theta)}{\partial \mu_j} \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\partial \mu_j} \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\partial \mu_j}
\end{aligned}$$

Derivation: closed-form for μ_k (M-step)

$$
\begin{aligned}
\frac{\partial \log p(X \mid \theta)}{\partial \mu_j} &= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \frac{\partial \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\partial \mu_j} \\
&= \sum_{n=1}^{N} \frac{1}{p(x_n|\theta)} \pi_j (x_n - \mu_j)^T \Sigma_j^{-1} \mathcal{N}(x_n \mid \mu_j, \Sigma_j) \\
&= \sum_{n=1}^{N} \frac{1}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} \pi_j (x_n - \mu_j)^T \Sigma_j^{-1} \mathcal{N}(x_n \mid \mu_j, \Sigma_j) \\
&= \sum_{n=1}^{N} (x_n - \mu_j)^T \Sigma_j^{-1} \frac{\pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} \\
&= \sum_{n=1}^{N} (x_n - \mu_j)^T \Sigma_j^{-1} r_{nj}
\end{aligned}
$$

Derivation: closed-form for μ_k (M-step)

$$\Rightarrow \sum_{n=1}^{N} (x_n - \mu_j)^T \Sigma_j^{-1} r_{nj} = 0$$

$$\left[\sum_{n=1}^{N} (r_{nj} x_n - r_{nj} \mu_j)^T \right] \Sigma_j^{-1} = 0$$

$$\sum_{n=1}^{N} r_{nj} x_n = \sum_{n=1}^{N} r_{nj} \mu_j$$

$$\mu_j = \frac{1}{\sum_{n=1}^{N} r_{nj}} \sum_{n=1}^{N} r_{nj} x_n$$

$$\boxed{\mu_j = \frac{1}{N_j} \sum_{n=1}^{N} r_{nj} x_n}$$

Derivation: closed-form for Σ_k (M-step)

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} (x_n - \mu_k)^T (x_n - \mu_k)$$

Expectation-Maximization (EM)

- Initialize π_k, μ_k, Σ_k
- Loop until convergence
  - E-step: Evaluate responsibilities r_{nk} for every data point x_n using current parameters π_k, μ_k, Σ_k

$$r_{nk} = \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

  - M-step: Re-estimate parameters π_k, μ_k, Σ_k using the current responsibilities r_{nk} (from E-step):

$$\boxed{\pi_k = \frac{N_k}{N}}, \boxed{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} x_n}, \boxed{\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} (x_n - \mu_k)^T (x_n - \mu_k)}$$

  - Evaluate log-likelihood:

$$\mathcal{L} = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k) \right)$$

Practical Consideration

- **Choosing K:** cross-validation, AIC, BIC

$$\text{BIC} = -2 \log L + p \log N$$

- Initialize with k-means for faster convergence
- **Regularization:** floor covariances to avoid singularities (add ϵI)
- **Covariance choices:** full, tied, diagonal — trade-off accuracy vs. computation

GMMs in Speech Recognition (Overview)

- Treat each target class (e.g., a speaker identity) as a separate generative model $p(x|class)$
- Recognition chooses the class that maximizes the class-conditional likelihood given the observed acoustic features (or the cumulative likelihood across a window of frames)

```
from sklearn.mixture import GaussianMixture
gmm = GaussianMixture(n_components=64, covariance_type='diag', ...)
gmm.fit(X)
ll = gmm.score_samples(X_test)
probs = gmm.predict_proba(X_test) # responsibilities (posteriors)
```

Learn More

Course Homepage