Gaussian Mixture Models

ALT-OPT & EM Algorithm Quiz

Objective for MLE in a GMM

For a GMM with parameters $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$ and data $\{x_n\}_{n=1}^N$, what is the **true** MLE objective (incomplete data log-likelihood)?

A. $\Sigma_{n=1}^N \Sigma_{k=1}^K \log[\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)]$

B. $\Sigma_{n=1}^N \log \Sigma_{k=1}^K \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)$

C. $\Sigma_{k=1}^K \log \Sigma_{n=1}^N \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)$

D. $\Sigma_{n=1}^N \Sigma_{k=1}^K z_{nk}[\log \pi_k + \log \mathcal{N}(x_n|\mu_k,\Sigma_k)]$

Why is MLE for GMM hard?

Why can't we get simple closed-form expressions for π_k, μ_k, Σ_k when doing MLE directly for a GMM?

A. Gaussians are not differentiable.

B. The covariance matrices are always singular.

C. The objective involves a "log of sum", which prevents factorization.

D. The mixture weights π_k must sum to 1.

Incomplete vs Complete Data Log-Likelihood

Which statement best describes the **difference** between ILL and CLL in a latent variable model like a GMM?

A. ILL uses only priors; CLL uses only likelihoods.

B. ILL is log $p(X|\Theta)$ and CLL is log $p(X,Z|\Theta)$.

C. ILL is always larger than CLL.

D. They are always identical.

What does ALT-OPT actually maximize?

In ALT-OPT for GMM, we alternately update cluster assignments (Z) and parameters (Θ). What objective is this algorithm directly maximizing?

A. $\log p(X|\Theta)$ (ILL)

B. $\mathbb{E}_{p(Z|X,\Theta)}[\log p(X,Z|\Theta)]$

C. $\log p(X,Z|\Theta)$ (CLL)

D. $\mathbb{E}_{q(Z)}[\log p(X|\Theta)]$

E-step in EM for GMM

In the EM algorithm for GMM, which quantity is computed in the **E-step**?

A. $\hat{\mu}_k = (1/N_k)\ \Sigma_n\ z_{nk}\ x_n$

B. $\gamma_{nk} = p(z_n{=}k \mid x_n, \Theta)$

C. $\pi_k = N_k/N$

D. $\Sigma_k = (1/N_k)\ \Sigma_n\ z_{nk}\ (x_n-\mu_k)(x_n-\mu_k)^T$

Formula for Responsibilities

For a GMM, the responsibility $\gamma_{nk} = p(z_n = k \mid x_n, \Theta)$ is given by:

A. $\gamma_{nk} = \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)$

B. $\gamma_{nk} = \mathcal{N}(x_n|\mu_k,\Sigma_k) / \Sigma_j \mathcal{N}(x_n|\mu_j,\Sigma_j)$

C. $\gamma_{nk} = \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k) / \Sigma_{j=1}^K \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)$

D. $\gamma_{nk} = 1/K$

M-step Updates with Soft Assignments

Given responsibilities γ_{nk}, what is the correct expression for the **mean** of component k in the EM M-step?

A. $\mu_k = (1/N) \Sigma_{n=1}^{N} x_n$

B. $\mu_k = (1/N_k) \Sigma_{n=1}^{N} \gamma_{nk} x_n$, where $N_k = \Sigma_{n=1}^{N} \gamma_{nk}$

C. $\mu_k = \Sigma_{n=1}^{N} \gamma_{nk} x_n$

D. $\mu_k = \arg\max_x \mathcal{N}(x|\mu_k,\Sigma_k)$

When does EM reduce to k-means?

Under which conditions does EM for a GMM behave most like **k-means** clustering?

A. Arbitrary covariances and unequal mixture weights.

B. Diagonal covariances with arbitrary scales.

C. All covariances Σ_k are equal spherical matrices and very small, and all π_k are equal.

D. When responsibilities are initialized uniformly and never updated.

Artificial Latent Variables in GMMs

Even if we define a mixture density without explicit latent variables, we can "artificially introduce" them. Why is this useful for GMM MLE?

A. It makes the Gaussians become independent.

B. It allows us to rewrite the mixture as a single Gaussian.

C. It converts a hard marginalization problem into an alternating estimation problem over Θ and Z.

D. It removes the need to estimate Σ_k.

What does EM guarantee to improve?

In the EM algorithm for GMM, which quantity is guaranteed **not to decrease** at each iteration (assuming exact E and M steps)?

A. The complete-data log-likelihood $\log p(X,Z|\Theta)$

B. The incomplete-data log-likelihood $\log p(X|\Theta)$

C. The entropy of the posterior $H(p(Z|X,\Theta))$

D. The sum of squared distances to cluster means