

1 One Dimensional Mixture of Two Gaussians

Suppose we have a mixtures of two Gaussians in \mathbb{R} that can be described by a pair of random variables (X, Z) where X takes values in \mathbb{R} and Z takes value in the set $\{1, 2\}$. The joint-distribution of the pair (X, Z) is given to us as follows:

$$Z \sim \text{Bernoulli}(\beta),$$

$$(X|Z = k) \sim \mathcal{N}(\mu_k, \sigma_k), \quad k \in \{1, 2\},$$

We use θ to denote the set of all parameters $\beta, \mu_1, \sigma_1, \mu_2, \sigma_2$.

- Write down the expression for the joint likelihood $p_{\theta}(X = x_i, Z_i = 1)$ and $p_{\theta}(X = x_i, Z_i = 2)$. What is the marginal likelihood $p_{\theta}(X = x_i)$?
- What is the log-likelihood $\ell_{\theta}(\mathbf{x})$? Why is this hard to optimize?
- (Optional) You'd like to optimize the log likelihood: $\ell_{\theta}(x)$. However, we just saw this can be hard to solve for an MLE closed form solution. Show that a lower bound for the log likelihood is $\ell_{\theta}(x_i) \geq \mathbb{E}_q \left[\log \left(\frac{p_{\theta}(X=x_i, Z_i=k)}{q_{\theta}(Z_i=k|X=x_i)} \right) \right]$.
- (Optional) The EM algorithm first initially starts with two randomly placed Gaussians (μ_1, σ_1) and (μ_2, σ_2) , which are both particular realizations of θ .
 - E-step: $q_{i,k}^{t+1} = p_{\theta}(Z_i = k|X = x_i)$. For each data point, determine which Gaussian generated it, being either (μ_1, σ_1) or (μ_2, σ_2) .
 - M-step: $\theta^{t+1} = \text{argmax}_{\theta} \sum_{i=1}^n \mathbb{E}_q \left[\log(p_{\theta}(X = x_i, Z_i = k)) \right]$. After labeling all datapoints in the E-step, adjust (μ_1, σ_1) and (μ_2, σ_2) .

Why does alternating between the E-step and M-step result in maximizing the lower bound?

- E-step: What are expressions for probabilistically imputing the classes for all the datapoints, i.e. $q_{i,1}^{t+1}$ and $q_{i,2}^{t+1}$?
- What is the expression for μ_1^{t+1} for the M-step?
- Compare and contrast k-means, soft k-means, and mixture of Gaussians fit with EM.