Lecture 33: Gaussian mixture models and the EM algorithm

Ciaran Evans

Previously: Gaussian mixture model

- Observe data $X_1, ..., X_n$
- Assume each observation i comes from one of k groups. Let $Z_i \in \{1, ..., k\}$ denote the group assignment
  - The group Z is an unobserved (**latent**) variable

Model:

- $P(Z_i = j) = \lambda_j$
- $X_i | (Z_i = j) \sim N(\mu_j, \sigma_j^2)$

Posterior probabilities and parameter estimation

- Posterior probabilities:

$$P(Z_i = j|X_i) = \frac{\lambda_j f(X_i|Z_i = j)}{\lambda_1 f(X_i|Z_i = 1) + \cdots + \lambda_k f(X_i|Z_i = k)}$$

- Parameter updates:
  - $\widehat{\lambda}_j = \frac{1}{n}\sum_{i=1}^{n} P(Z_i = j|X_i)$
  - $\widehat{\mu}_j = \dfrac{\sum_{i=1}^{n} X_i P(Z_i = j|X_i)}{\sum_{i=1}^{n} P(Z_i = j|X_i)}$
  - $\widehat{\sigma}_j = \sqrt{\dfrac{\sum_{i=1}^{n} (X_i - \widehat{\mu}_j)^2 P(Z_i = j|X_i)}{\sum_{i=1}^{n} P(Z_i = j|X_i)}}$

Today: where do these estimates come from??

Parameter estimation

The quantity we **want** to optimize is called the **log-likelihood**, and it is given by

$$\ell(\lambda, \mu, \sigma) = \sum_{i=1}^{n} \log \left(\sum_{j=1}^{k} \lambda_j f(X_i | Z_i = j) \right)$$

Question: We have discussed optimization extensively in this course. How do we usually try to optimize a function?

Parameter estimation

Let $\theta = (\lambda, \mu, \sigma)$ be the collection of all parameters we are trying to estimate for the Gaussian mixture model. Let $\theta^{(t)}$ be our current estimates of these parameters, at iteration t, and let

$$\gamma_{ij}^{(t)} = P^{(t)}(Z_i = j | X_i, \theta^{(t)})$$

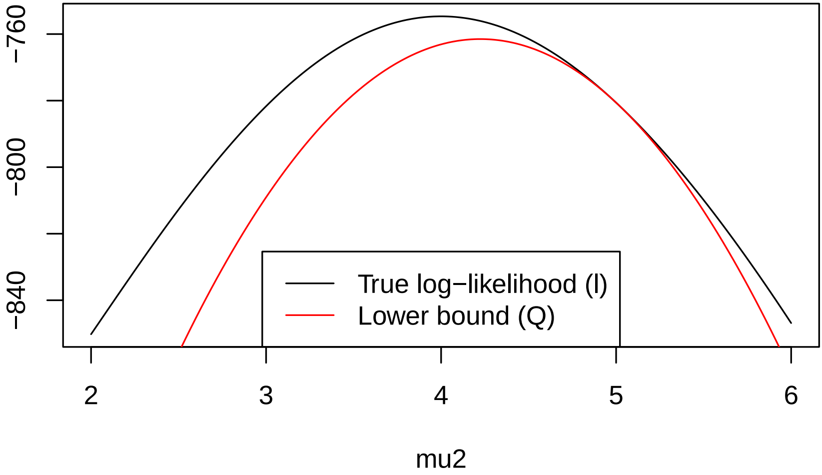
be the posterior probabilities calculated with $\theta^{(t)}$. Then define

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij}^{(t)} \log(\lambda_j f(X_i | Z_i = j)) - \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij}^{(t)} \log(\gamma_{ij}^{(t)})$$

- $\ell(\lambda, \mu, \sigma) \geq Q(\theta|\theta^{(t)})$
- Maximizing $Q(\theta|\theta^{(t)})$ helps us maximize $\ell(\lambda, \mu, \sigma)$

Example

- $\mu^{(t)} = (1, 5)$
- Want to find μ_1, μ_2 to maximize Q. Look at different possibilities for μ_2:



Doing the calculus

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{k}\gamma_{ij}^{(t)}\log\left(\lambda_j\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left\{-\frac{1}{2\sigma_j^2}(X_i-\mu_j)^2\right\}\right) - \sum_{i=1}^{n}\sum_{j=1}^{k}\gamma_{ij}^{(t)}\log(\gamma_{ij}^{(t)})$$

Summary: EM algorithm

Want to maximize

$$\ell(\theta) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{k} \lambda_j f(X_i | Z_i = j)\right)$$

Define

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{k} \gamma_{ij}^{(t)} \log(\lambda_j f(X_i|Z_i = j)) - \sum_{i=1}^{n}\sum_{j=1}^{k} \gamma_{ij}^{(t)} \log(\gamma_{ij}^{(t)})$$

EM algorithm:

1. Begin with $\theta^{(0)}$
2. Calculate $Q(\theta|\theta^{(0)})$
3. $\theta^{(1)}$ maximizes $Q(\theta|\theta^{(0)})$
4. Iterate between steps 2 and 3 until $\ell(\theta)$ stops changing