STA 712 Homework 7

Due: Friday, December 2, 12:00pm (noon) on Canvas.

Instructions: Submit your work as a single PDF. For this assignment, you may include written work by scanning it and incorporating it into the PDF. Include all R code needed to reproduce your results in your submission.

The EM algorithm

In this problem, we will use the EM algorithm to estimate the parameters in a mixture of two univariate Gaussian distributions.

Let $\theta \in \mathbb{R}^d$ be an unknown parameter we want to estimate. Let $Y = Y_1, ..., Y_n$ be a set of observed data, and $Z = Z_1, ..., Z_n$ a set of unobserved latent data. To estimate θ, we want to maximize the likelihood

$$L(\theta; Y) = f_Y(Y|\theta) = \int f_{Y|Z=z}(Y|\theta) f_Z(z) dz$$

However, maximizing this likelihood is challenging when Z is unobserved. Our solution is to alternate between the E and M steps of the EM algorithm:

E step: Let $\theta^{(k)}$ be the current estimate of θ. Calculate

$$Q(\theta|\theta^{(k)}) = \mathbb{E}_{Z|Y,\theta^{(k)}}[\log L(\theta; Z, Y)]$$

M step: $\theta^{(k+1)} = \text{argmax}_\theta \; Q(\theta|\theta^{(k)})$

---

1. Let $Z_i \sim Bernoulli(\alpha)$, and $Y_i|(Z_i = j) \sim N(\mu_j, \sigma_j^2)$. Then our parameter vector of interest is $\theta = (\alpha, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$, and the conditional density of $Y_i|Z_i = j$ is

$$f_{Y_i|Z_i=j}(y|\theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(y - \mu_j)^2\right\}.$$

We observe data $Y_1, ..., Y_n$, and our goal is to estimate θ. We will use the EM algorithm to estimate these parameters.

(a) Show that the *complete-data* likelihood (i.e., if we were able to observe Z_i) is

$$L(\theta; Z, Y) = \prod_{i=1}^n \alpha^{Z_i}(1-\alpha)^{1-Z_i} \frac{1}{\sqrt{2\pi\sigma_{Z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{Z_i}^2}(Y_i - \mu_{Z_i})^2\right\}$$

(b) Using (a), show that

$$Q(\theta|\theta^{(k)}) = \sum_{i=1}^n \sum_{j=0}^1 [\log \alpha_j - \frac{1}{2}\log(2\pi\sigma_j^2) - \frac{1}{2\sigma_j^2}(Y_i - \mu_j)^2] P(Z_i = j|Y_i, \theta^{(k)}),$$

where $\alpha_1 = \alpha$ and $\alpha_0 = 1 - \alpha$.

(c) Differentiate $Q(\theta|\theta^{(k)})$ with respect to μ_j to show that

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^{n} Y_i P(Z_i = j|Y_i, \theta^{(k)})}{\sum_{i=1}^{n} P(Z_i = j|Y_i, \theta^{(k)})}$$

(d) Calculate similar update rules for σ_j^2 and α_j.

(e) Now let's try it out! Generate $Y_1, ..., Y_{1000}$ from a mixture of two univariate Gaussians, with $\alpha = 0.3$, $\mu_0 = 0$, $\mu_1 = 4$, and $\sigma_0^2 = \sigma_1^2 = 1$. Beginning with $\alpha^{(0)} = 0.5$, $\mu_0^{(0)} = 0$, $\mu_1^{(1)} = 1$, and $\sigma_0^{2(0)} = \sigma_1^{2(0)} = 0.5$, run 100 iterations of the EM algorithm. What are your estimated parameters at the end?

Fisher information for ZIP models

Recall that for a ZIP model,

$$P(Y_i = y|\gamma, \beta) = \begin{cases} e^{-\lambda_i}(1-\alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1-\alpha_i) & y > 0 \end{cases}$$

with

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma^T X_i$$
$$\log(\lambda_i) = \beta^T X_i$$

2. Suppose we observe data $(X_1, Y_1), ..., (X_n, Y_n)$ and fit a ZIP model, estimating γ and β. One option for testing hypotheses about coefficients in γ and β is to use a Wald test. This relies on the fact that the distribution of $(\widehat{\gamma}, \widehat{\beta})^T$ is approximately normal, and requires us to calculate the observed information. In this probably, we will calculate the observed information matrix for the ZIP model.

   (a) Show that the log likelihood of γ and β is

   $$\ell(\gamma, \beta; Y) = \sum_{i:Y_i=0} \log\left(e^{-\lambda_i}(1-\alpha_i) + \alpha_i\right) + \sum_{i:Y_i>0} (Y_i \log \lambda_i - \lambda_i) + \sum_{i:Y_i>0} \log(1-\alpha_i) - \sum_{i:Y_i>0} \log(Y_i!)$$

   (b) Rearrange (a) to show that

   $$\ell(\gamma, \beta; Y) = \sum_{i=1}^{n} \log(\exp\{-e^{\beta^T X_i}\} + \exp\{\gamma^T X_i\})\mathbb{1}\{Y_i = 0\} + \sum_{i=1}^{n} (Y_i \beta^T X_i - \exp\{\beta^T X_i\})\mathbb{1}\{Y_i > 0\}$$
   $$- \sum_{i=1}^{n} \log(1 + \exp\{\gamma^T X_i\}) - \sum_{i:Y_i>0} \log(Y_i!)$$

(c) The score function is

$$U(\gamma, \beta) = \begin{pmatrix} \dfrac{\partial \ell}{\partial \gamma} \\ \dfrac{\partial \ell}{\partial \beta} \end{pmatrix},$$

where both $\dfrac{\partial \ell}{\partial \gamma}$ and $\dfrac{\partial \ell}{\partial \beta}$ are vectors. Find $\dfrac{\partial \ell}{\partial \gamma}$ and $\dfrac{\partial \ell}{\partial \beta}$.

(d) The observed information matrix is

$$\mathcal{J}(\gamma, \beta) = -\begin{pmatrix} \dfrac{\partial^2 \ell}{\partial \gamma^2} & \dfrac{\partial^2 \ell}{\partial \gamma \partial \beta} \\ \dfrac{\partial^2 \ell}{\partial \beta \partial \gamma} & \dfrac{\partial^2 \ell}{\partial \beta^2} \end{pmatrix}$$

where each entry is itself a matrix. Calculate $\mathcal{J}(\gamma, \beta)$.

Multivariate EDMs

Recall that a multivariate EDM has probability function

$$f(y; \theta, \phi) = a(y, \phi) \exp\left\{ \frac{y^T \theta - \kappa(\theta)}{\phi} \right\},$$

where $\phi > 0$, $\theta, y \in \mathbb{R}^d$, and $\kappa : \mathbb{R}^d \to \mathbb{R}$. As in a univariate EDM,

$$\frac{\partial \kappa}{\partial \theta} = \mu \qquad \frac{\partial \mu}{\partial \theta} = V(\mu),$$

with $\mu = \mathbb{E}[Y] \in \mathbb{R}^d$ and $V(\mu) = \frac{1}{\phi} Var(Y) \in \mathbb{R}^{d \times d}$.

3. Suppose that $Y \sim Categorical(\pi_1, ..., \pi_J)$. Then $\mu = (\pi_1, ..., \pi_{J-1})^T$,
$\theta = \left(\log\left(\dfrac{\pi_1}{1 - \sum_{j=1}^{J-1} \pi_j} \right), ..., \log\left(\dfrac{\pi_{J-1}}{1 - \sum_{j=1}^{J-1} \pi_j} \right) \right)$, and $\kappa(\theta) = -\log\left(1 - \sum\limits_{j=1}^{J-1} \pi_j\right)$.

   (a) By differentiating κ, confirm that $\dfrac{\partial \kappa}{\partial \theta} = \mu$ for the categorical distribution.

   (b) For the categorical distribution, show that

$$V(\mu) = \begin{bmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_{J-1} \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_{J-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{J-1}\pi_1 & -\pi_{J-1}\pi_2 & \cdots & \pi_{J-1}(1-\pi_{J-1}) \end{bmatrix}$$