EM Algorithm

Fitting ZIP models

Last time: suppose we observe Z_i

$$\ell(\gamma, \beta) = \underbrace{\sum_{i=1}^{n} \left(Z_i \log \alpha_i + (1-Z_i) \log(1-\alpha_i) \right)}_{\ell(\gamma)}$$

$$+ \underbrace{\sum_{i=1}^{n} (1-Z_i) \left[-\lambda_i + Y_i \log \lambda_i \right]}_{\ell(\beta)}$$

$$- \sum_{i=1}^{n} (1-Z_i) \log (Y_i)!$$

$\Rightarrow$ if we know Z_i, we can separately maximize $\ell(\gamma)$ and $\ell(\beta)$

But, we don't know Z_i!

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Suppose we know γ and β instead.

$$P(Z_i = 1 \mid Y_i, \gamma, \beta) = 0 \qquad \text{if } Y_i > 0$$

$$P(Z_i = 1 \mid Y_i = 0, \gamma, \beta) = \frac{P(Y_i = 0 \mid Z_i = 1, \gamma, \beta)\, P(Z_i = 1 \mid \gamma, \beta)}{\underbrace{P(Y_i = 0 \mid \gamma, \beta)}}$$

$$P(Y_i = 0 \mid Z_i = 1, \gamma, \beta)\, P(Z_i = 1 \mid \gamma, \beta) + P(Y_i = 0 \mid Z_i = 0, \gamma, \beta)\, P(Z_i = 0 \mid \gamma, \beta)$$

$$= \frac{\alpha_i}{\alpha_i + e^{-\lambda_i}(1 - \alpha_i)}$$

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

So:

1. If we know Y_i, γ, β, we can calculate
$$P(Z_i = 1 \mid Y_i, \gamma, \beta) = \mathbb{E}[Z_i \mid Y_i, \gamma, \beta]$$

2. If we know Z_i, Y_i we can calculate $\hat{\gamma}, \hat{\beta}$ (maximizing $\ell(\gamma)$ and $\ell(\beta)$)

EM algorithm:

E-step (expectation): Given current estimates $\gamma^{(k)}$ and $\beta^{(k)}$,
$$Z_i^{(k)} = \mathbb{E}[Z_i \mid Y_i, \gamma^{(k)}, \beta^{(k)}]$$

M-step (maximization): Given $Z_i^{(k)}$,
$$\gamma^{(k+1)} = \operatorname{argmax} \ell(\gamma; Y, Z^{(k)})$$
$$\beta^{(k+1)} = \operatorname{argmax} \ell(\beta; Y, Z^{(k)})$$

iterate E & M steps until convergence

$Y = Y_1, \ldots, Y_n$

$Z^{(k)} = Z_1^{(k)}, \ldots, Z_n^{(k)}$

E-step: $\quad z_i^{(k)} = \mathbb{E}[z_i \mid Y_i, \gamma^{(k)}, \beta^{(k)}]$

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

So: Given $\gamma^{(k)}, \beta^{(k)}$

- $\hat{\alpha}_i = \dfrac{\exp\{\gamma^{(k)T} X_i\}}{1 + \exp\{\gamma^{(k)T} X_i\}}$
- $\hat{\lambda}_i = \exp\{\beta^{(k)T} X_i\}$

$$\Rightarrow \quad \underbrace{\hat{P}(z_i = 1 \mid Y_i, \gamma^{(k)}, \beta^{(k)})}_{z_i^{(k)}} = \begin{cases} 0 & Y_i > 0 \\ \dfrac{\hat{\alpha}_i}{\hat{\alpha}_i + e^{-\hat{\lambda}_i}(1-\hat{\alpha}_i)} & Y_i = 0 \end{cases}$$

M-step for β

$$\beta^{(k+1)} = \arg\max \ell(\beta; Y, Z^{(k)})$$

$$= \arg\max_{\beta} \sum_{i=1}^{n} (1 - Z_i^{(k)}) [-\lambda_i + Y_i \log \lambda_i]$$

$$= \arg\max_{\beta} \sum_{i=1}^{n} (1 - Z_i^{(k)}) [Y_i \beta^T X_i - e^{\beta^T X_i}]$$

This is weighted Poisson regression, with weights $w_i = 1 - Z_i^{(k)}$

$$U(\beta) = \sum_{i=1}^{n} w_i (Y_i X_i - e^{\beta^T X_i} X_i) = \sum_{i=1}^{n} w_i X_i (Y_i - \lambda_i)$$

$$\mathcal{I}(\beta) = \sum_{i=1}^{n} w_i \lambda_i X_i X_i^T$$

M-step for γ:

$$\gamma^{(k+1)} = \underset{\gamma}{\operatorname{argmax}}\ \ell(\gamma; Y, Z^{(k)})$$

$$= \sum_{i=1}^{n} \left(Z_i^{(k)} \log(\alpha_i) + (1 - Z_i^{(k)}) \log(1-\alpha_i) \right)$$

$$= \sum_{i=1}^{n} \left(Z_i^{(k)} \log\left(\frac{\alpha_i}{1-\alpha_i}\right) - \log\left(1 + \frac{\alpha_i}{1-\alpha_i}\right) \right)$$

$$= \sum_{i=1}^{n} Z_i^{(k)} \gamma^T X_i - \sum_{i=1}^{n} \log(1 + e^{\gamma^T X_i})$$

we want to get something like

$$\sum_i Y_i^* w_i \gamma^T X_i - \sum_i w_i \log(1 + e^{\gamma^T X_i})$$

(weighted logistic regression with response Y_i^* and weights w_i)

$$\underbrace{\sum_{i=1}^{n} Z_i^{(k)} \gamma^T x_i}_{} - \underbrace{\sum_{i=1}^{n} \log(1+e^{\gamma^T x_i})}_{}$$

$$= \sum_{i=1}^{n} Z_i^{(k)} \gamma^T x_i + \sum_{i=1}^{n} 0 \cdot (1-Z_i^{(k)}) \gamma^T x_i$$

$$= \sum_{i=1}^{n} Z_i^{(k)} \log(1+e^{\gamma^T x_i}) + \sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1+e^{\gamma^T x_i})$$

$Z_i^{(k)} = 0$ whenever $Y_i > 0$

$$Z_i^{(k)} = \mathbb{1}\{Y_i = 0\} Z_i^{(k)}$$

$$> \sum_{i=1}^{n} \mathbb{1}\{Y_i = 0\} Z_i^{(k)} \gamma^T x_i + \sum_{i=1}^{n} 0 \cdot (1-Z_i^{(k)}) \gamma^T x_i$$

$$- \sum_{i=1}^{n} Z_i^{(k)} \log(1+e^{\gamma^T x_i}) - \sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1+e^{\gamma^T x_i})$$

$$Y^* = (\mathbb{1}\{Y_1 = 0\}, \mathbb{1}\{Y_2 = 0\}, \dots, \mathbb{1}\{Y_n = 0\}, 0, 0, \dots, 0)^T \in \mathbb{R}^{2n}$$

$$X^* = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \\ x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \in \mathbb{R}^{2n \times (k+1)}$$

weights $w = (Z_1^{(k)}, \dots, Z_n^{(k)}, 1-Z_1^{(k)}, \dots, 1-Z_n^{(k)})^T \in \mathbb{R}^{2n}$

$$\boxed{\sum_i Y_i^* w_i \gamma^T x_i^* - \sum_i w_i \log(1+e^{\gamma^T x_i^*})}$$

EM algorithm in general