Clase 6

jueves, 30 de agosto de 2018 17:42

Algoritmo EM
(Expectation - Maximization)

Este algoritmo obtiene soluciones que convergen al estimador máximo verosímil y se basa en la idea de reemplazar un problema de maximizar la verosimilitud cuando esto es difícil por una sucesión de maximizaciones fáciles y convergen a la solución del problema original.

El algoritmo EM considera dos problemas de verosimilitud diferentes, el problema de datos incompletos (y que es el problema que estamos interesados en resolver) y el problema de datos completos y que es el problema que resolveremos.

Ejemplo: Observamos $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_n$ los cuales son mutuamente independientes $X_i \sim \text{Poisson}(Z_i)$ y $Y_i \sim \text{Poisson}(\beta Z_i)$. Este modelo puede ser por ejemplo la incidencia de una enfermedad Y_i donde la tasa es una función de un efecto promedio β y un factor adicional Z_i [por ej el nivel de salud en el área i]

Z_i no se observa pero tenemos información de este a partir de X_i

La función de probabilidad conjunta de X e Y es

$$f((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \mid \beta, z_1, \ldots, z_n) =$$

$$\prod^{n} e^{-\beta z_i} (\ldots)^{y_i} \; e^{-z_i} (\ldots)^{x_i} \quad \{\{ \text{Poisson}(\lambda)$$

$$= \prod_{i=1}^{n} \frac{e^{-\beta z_i} (\beta z_i)^{y_i}}{y_i!} \frac{e^{-z_i} (z_i)^{x_i}}{x_i!}$$

Poisson (λ)

$$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Si resolvemos el problema maximizando la expresión obtenida se obtiene (ver más abajo)

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \quad y \quad \hat{z}_j = \frac{x_j + y_j}{\hat{\beta}+1} \quad j=1,2,\dots,n$$

Se ha planteado el problema en los datos completos $(x_1,y_1),\dots,(x_n,y_n)$.

Los datos perdidos hacen la estimación más difícil, por ejemplo supongamos que x_1 es missing. Podemos descartar y_1 y trabajar con una muestra de tamaño $n-1$ o sea ignoramos la información de y_1.

Pero sabemos que si pudiéramos usar esa información podemos mejorar nuestros estimadores.

La función de probabilidad cuando x_1 es missing es

$$\sum_{x_1=0}^{\infty} f((x_1,y_1),(x_2,y_2)\dots(x_n,y_n) \mid \beta, z_1, z_2, \dots, z_n)$$

este es el problema basado en los datos incompletos

Esta es la verosimilitud que queremos maximizar.

En general $Y=(Y_1,\dots,Y_n)$ son los datos incompletos y

$X^d = (X_1, \ldots, X_n)$ son los datos aumentados y entonces (Y, X) son los datos completos.
Consideremos las densidades $g(\cdot|\theta)$ de Y y $f(\cdot|\theta)$ de (Y,X). Se cumple que

$$g(y/\theta) = \int f(y, x|\theta)\, dx$$

(en el caso discreto la suma reemplaza a la integral)

O sea $L(\theta/y) = g(y/\theta)$ será la verosimilitud para los datos incompletos y $L(\theta/y,x) = f(y,x|\theta)$ la verosimilitud para los datos completos.

En nuestro caso la verosimilitud para datos incompletos es:

$$L(\beta, z_1, z_2, \ldots, z_n \mid y_1, (x_2, y_2) \ldots (x_n, y_n))$$

$$= \prod_{i=1}^{n} \frac{e^{-\beta z_i} (\beta z_i)^{y_i}}{y_i!} \prod_{i=2}^{n} \frac{e^{-z_i} (z_i)^{x_i}}{x_i!}$$

Resolviendo se obtiene $\hat{\beta} = \dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} \hat{z}_i}$

$$y_1 = \hat{z}_1 \hat{\beta}$$

$$x_j + y_j = \hat{z}_j (\hat{\beta} + 1) \qquad j = 2, 3, \ldots, n$$

Veamos este problema resuelto por el algoritmo E.M

El algoritmo EM nos permite maximizar $L(\theta|y)$ trabajando solo con $L(\theta|y,x)$ y la función de densidad condicional de X dado Y, θ $k(x|\theta,y)$

$$L(\theta|y,x) = f(y,x|\theta)$$

$$k(x|\theta,y) = \frac{f(y,x|\theta)}{g(y|\theta)}$$

$$\log L(\theta|y) = \log g(y/\theta) = \log \frac{f(y,x|\theta)}{k(x|\theta,y)} =$$

$$= \log f(y,x|\theta) - \log k(x|\theta,y) =$$

$$= \log L(\theta|y,x) - \log k(x|\theta,y)$$

Como x no es observado reemplazamos el lado derecho de la igualdad por su esperanza bajo $k(x|\theta',y)$ creando una nueva identidad

$$\log L(\theta|y) = E(\log L(\theta|y,X)|\theta',y) - E(\log k(X|\theta,y)|\theta',y)$$

A partir de un valor inicial $\theta^{(0)}$ creamos una sucesión $\theta^{(r)}$ como:

$$(\ast\ast\ast)\quad \theta^{(r+1)} = \underset{\theta}{\operatorname{argmax}}\, E\left(\log L(\theta|y,x)\,\middle|\,\theta^{(r)}, y\right)$$

En el algoritmo el "paso E" calcula el valor esperado del logaritmo de la verosimilitud y el "paso M" encuentra el máximo.

En nuestro ejemplo $(x,y) = ((x_1,y_1),(x_2,y_2)\ldots(x_n,y_n))$ notamos a los datos completos y $(x_{(-1)}, y)$ notamos a los datos incompletos.

obs $(x_{(-1)}, y) = (y_1,(x_2,y_2),\ldots,(x_n,y_n))$

$$E\left(\log L(\beta, z_1, z_2, \ldots, z_n|(x,y)\,|\,z^{(r)}, (x_{(-1)}, y)\right) =$$

$$= \sum_{x_1=0}^{+\infty} \log\left(\prod_{i=1}^{n} \frac{e^{-\beta z_i}(\beta z_i)^{y_i}}{y_i!}\cdot\frac{e^{-z_i}(z_i)^{x_i}}{x_i!}\right)\frac{e^{-z_1^{(r)}}(z_1^{(r)})^{x_1}}{x_1!}$$

$$= \sum_{i=1}^{n}\left[-\beta z_i + y_i(\log\beta + \log z_i) - \log y_i!\right] +$$

$$+ \sum_{i=2}^{n}\left[-z_i + x_i\log z_i - \log x_i!\right] +$$

$$+ \sum_{x_1=0}^{+\infty}\left[-z_1 + x_1\log z_1 - \log x_1!\right]\frac{e^{-z_1^{(r)}}(z_1^{(r)})^{x_1}}{x_1!} =$$

$$\left(\sum^{n} \left[-\beta z_i + y_i(\log\beta + \log z_i)\right] + \sum^{n}\left[-z_i + x_i\log z_i\right]\right.$$

$$= \left(\sum_{i=1}^{n} [-\beta z_i + y_i (\log \beta + \log z_i)] + \sum_{i=2}^{n} [-z_i + x_i \log z_i] + \sum_{x_1=0}^{\infty} [-z_1 + x_1 \log z_1] \frac{e^{-z_1^{(r)}} (z_1^{(r)})^{x_1}}{x_1!} \right) -$$

$$\underbrace{- \left(\sum_{i=1}^{n} \log y_i! + \sum_{i=1}^{n} \log x_i! + \sum_{x_1=0}^{\infty} \log x_1! \frac{e^{-z_1^{(r)}} (z_1^{(r)})^{x_1}}{x_1!} \right)}_{(*)}$$

(*) no depende de β y z_i luego la podemos desechar en la optimización.

Además $$\sum_{x_1=0}^{+\infty} [-z_1 + x_1 \log z_1] \frac{e^{-z_1^{(r)}} (z_1^{(r)})^{x_1}}{x_1!} =$$

$$= -z_1 \underbrace{\sum_{x_1=0}^{+\infty} \frac{e^{-z_1^{(r)}} (z_1^{(r)})^{x_1}}{x_1!}}_{=1} + \log z_1 \underbrace{\sum_{x_1=0}^{\infty} x_1 \frac{e^{-z_1^{(r)}} (z_1^{(r)})^{x_1}}{x_1!}}_{z_1^{(r)} \text{ pues}} =$$

es la esperanza de una Poisson.

$$X \sim P(\lambda) \Rightarrow E(X) = \lambda$$

$$= -z_1 + z_1^{(r)} \log z_1$$

entonces el valor esperado de la verosimilitud de los datos completos es

$$\sum_{i=1}^{n}\left[-\beta z_i + y_i(\log\beta + \log z_i)\right] + \sum_{i=2}^{n}\left[-z_i + x_i \log z_i\right] +$$

$$+\left(-z_1 + z_1^{(r)} \log z_1\right).$$

o sea este valor esperado es igual a la verosimilitud de los datos completos con la excepción de que x_1 se reemplaza por $z_1^{(r)}$ luego

$$\hat{\beta}^{(r+1)} = \frac{\sum_{i=1}^{n} y_i}{z_1^{(r)} + \sum_{i=2}^{n} x_i}$$

$$\hat{z}_1^{(r+1)} = \frac{\hat{z}_1^{(r)} + y_1}{\hat{\beta}^{(r+1)} + 1}$$

$$\hat{z}_j^{(r+1)} = \frac{x_j + y_j}{\hat{\beta}^{(r+1)} + 1} \qquad j = 2, 3, \ldots$$

El paso E resulta pues sustituir $\hat{z}_1^{(r)}$ por x_1 y el paso M es el cálculo de los estimadores en la iteración r. $(\hat{\beta}^{(r)}, \hat{z}_1^{(r)}, \hat{z}_2^{(r)}, \ldots, \hat{z}_n^{(r)})$

La prueba de que estos estimadores convergen a los estimadores de máxima verosimilitud no la daremos solo enunciamos el siguiente teorema

Teorema 45: (Sucesión EM monótona)

La sucesión $\{\hat{\theta}_{(r)}\}$ definida por (***) cumple

$$L(\hat{\theta}^{(r+1)} | y) \geq L(\hat{\theta}^{(r)} | y)$$

con igualdad si y solo si dos iteraciones sucesivas obtienen el mismo valor del máximo del valor esperado de la verosimilitud de los datos completos o sea

$$E\left(\log L\left(\hat{\theta}^{(r+1)} | y, X\right) \Big| \hat{\theta}^{(r)}, y\right) = E\left(\log L\left(\hat{\theta}^{(r)} | y, X\right) \Big| \hat{\theta}^{(r)}, y\right)$$

Solución del primer problema

$$L(\beta, \tau_1, \dots, \tau_n | (x_1, y_1), (x_2, y_2), \dots (x_n, y_n)) =$$

$$= \prod_{i=1}^{n} \frac{e^{-\beta \tau_i} (\beta \tau_i)^{y_i}}{y_i!} \frac{e^{-\tau_i} (\tau_i)^{x_i}}{x_i!}$$

$$\log L = \sum_{i=1}^{n} \left[(-\beta \tau_i) + y_i \log \beta \tau_i - \log y_i! + (-\tau_i) + x_i \log \tau_i - \log x_i! \right]$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \left[(-\tau_i) + y_i \frac{\cancel{\tau_i}}{\beta \cancel{\tau_i}} \right] = 0 \qquad (1)$$

$$\frac{\partial \log L}{\partial \tau_i} = -\beta + y_i \frac{\beta}{\beta \tau_i} - 1 + \frac{x_i}{\tau_i} = 0 \quad \forall i \text{ (n ecuaciones)}$$

de (1) $\frac{1}{\beta}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i \Rightarrow \beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} z_i}$ (2)

$\forall i$ tenemos $\frac{y_i + x_i}{z_i} = 1+\beta \Rightarrow z_i = \frac{y_i + x_i}{1+\beta}$

o sea $\sum_{i=1}^{n} z_i = \frac{\sum_{i=1}^{n}(y_i + x_i)}{1+\beta}$ y

$$\beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i + x_i}\cdot(1+\beta) \Rightarrow \text{de donde}$$

$$\beta\left(1 - \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i + x_i}\right) = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i + x_i}$$

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}} \quad \text{y} \quad \boxed{\hat{z}_i = \frac{y_i + x_i}{1+\hat{\beta}}} \ \forall i$$