الگوریتم EM

الگوریتم EM که توسط دمپستر، لیرد و رابین[۴] در سال ۱۹۷۷ و توسط مک لاچلان و کریشنان[۵] در سال ۱۹۹۷ ارائه شد، ابزاری قدرتمند برای برآورد حداکثر درستنمایی با داده‌های ناقص می‌باشد. در اینجا معنی کلمهٔ « ناقص» حالتی کلی دارد و در موقعیتهای متفاوت می‌تواند به معانی گوناگونی مانند: داده‌های گم شده، مؤلفه‌های مجهول، مشاهدات سانسور شده، متغیرهای نهفته و نظایر آنها اشاره داشته باشد. شرح مختصری از الگوریتم EM در زیر ارائه می‌شود.

یکی از روش های براورد پارامتر روش ماکسیمم درستنمایی است

اگر بردار مشاهدات غیر کامل باشد که فقدان قسمتی از داده ها ممکن

است از مکانیزم سانسور کردن یا از مقادیر گمشده و باشد ....

استفاده نمودEMمیتوان از الگوریتم

فرض کنید $\mathbf{Y}$ بردار داده‌های مشاهده شده و $\mathbf{X}$ بردار داده‌های نامعلوم باشند. همچنین فرض کنید θ پارامتر مجهول است که می‌خواهیم برآورد شود و $l_c(\theta; \mathbf{Y}, \mathbf{X})$ لگاریتم تابع درستنمایی بر اساس کلیهٔ داده‌ها می‌باشد، که به ازای تمام مقادیر ممکن θ در فضای پارامتر Ω ، تعریف می‌شود.

الگوریتم با یک مقدار اولیهٔ $\theta^{(0)} \in \Omega$ آغاز می‌شود و دو مرحلهٔ زیر را تا رسیدن به همگرایی، تکرار می‌کند:

مرحلهٔ E :

محاسبهٔ $l^{(j)}(\theta) = E_{\mathbf{X}|\mathbf{Y},\theta^{(j-1)}}[l_c(\theta;\mathbf{Y},\mathbf{X})]$ که امید ریاضی با توجه به داده‌های گم شدهٔ $\mathbf{X}$ به شرط داده‌های مشاهده شدهٔ $\mathbf{Y}$ گرفته می‌شود و باید توجه شود که مقدار $\theta^{(j-1)}$ در این امید ریاضی جایگذاری می‌شود.

مرحلهٔ M : یافتن $\theta^{(j)} \in \Omega$ بقسمی که $l^{(j)}(\theta)$ بیشینه شود.

تکرار دو مرحلهٔ فوق به ازای $j = ۱,۲,...$ ، منجر به همگرایی دنبالهٔ $\theta^{(۱)}, \theta^{(۲)}, ...$ در ماکزیمم موضعی لگاریتم درستنمایی کلیّهٔ داده‌ها می‌شود. لازم به ذکر است که تحت شرایط خاصی این دنباله به همگرایی نمی‌رسد، برای جزئیات بیشتر به [۴] رجوع شود.

از جمله کاربردهای رایج این الگوریتم، حل مسائلی با داده‌های گم شده، یافتن نمای توزیع پسین در چهارچوب بیز[۳]، تشخیص مدلهای آمیخته و کاربردهایی در داده‌های گروه‌بندی شده، سانسور شده و یا بریده شده، می‌باشند

مثال ۱) فرض کنید طول عمر لامپ های مربوط به یک کارخانه دارای توزیع نمایی با میانگین مجهول θ باشد. از این کارخانه تعداد M+N لامپ را به تصادف انتخاب کرده و در دو آزمایش مستقل شرکت داده می شود. آزمایشگر تعداد N لامپ را در آزمایش اوّل قرار داده و طول عمر تک‌تک آنها را به صورت $y_1,\ldots,y_N$ ثبت می‌کند، در حالیکه در آزمایش دوم زمان $t > 0$ را در نظر گرفته و همه M لامپ را داخل آزمایش می‌کند و فقط لامپ هایی را که تا

زمان t هنوز نسوخته‌اند مشخص می کند. بنابراین در آزمـایش دوم طول عمر دقیق تک تک مشاهدات مشخص نیست و تنهـا مشخـصه های $E_1, E_2, \ldots, E_M$ در دسترس هستند که در آن

$$E_i = \begin{cases} ۱ & \text{اگر لامپ تا زمان } t \text{ سوخته نباشد} \\ ۰ & \text{اگر لامپ قبل از زمان } t \text{ سوخته باشد} \end{cases}$$

$$, \quad i = ۱,\ldots, \quad M$$

حال سؤال اینجاست که با این داده ها $MLE(\theta)$ چقدر می‌شود؟

فرض کنید که $X_1, \ldots, X_M$ طول عمرهای (مشاهده نشده) لامپ های شرکت کننده در آزمایش دوم باشند و $Z = \sum_{i=1}^{M} E_i$ تعداد لامپ هایی باشد که در آزمایش دوم تا زمان t هنوز نسوخته‌اند. بنابراین ترکیب داده‌های مشاهده شده در دو آزمایش به صورت زیر است: $\mathbf{Y} = (Y_1, \ldots, Y_N, E_1, \ldots, E_M)$
و داده‌های مشاهده نشده عبارتند از: $\mathbf{X} = (X_1, \ldots, X_M)$
همچنین تابع درستنمایی بر اساس کلیهٔ داده‌ها به صورت

$$L(\theta; \mathbf{Y}, \mathbf{X}) = \frac{1}{\theta^{M+N}} \exp\left(\frac{-(\sum_{i=1}^{N} Y_i + \sum_{i=1}^{M} X_i)}{\theta} \right)$$

بنابراین خواهیم داشت:

$$l_c(\theta;\mathbf{Y},\mathbf{X}) = \ln L(\theta)$$

$$= -(M+N)\log\theta - \frac{1}{\theta}\left(\sum_{i=1}^{N} Y_i + \sum_{i=1}^{M} X_i\right) \quad (1)$$

حـــال $E(X_i|\mathbf{Y})$ را در دو حالـــت $\mathrm{E_i} = 0$ و $\mathrm{E_i} = 1$ محاسبه می‌کنیم. با توجه به این موضوع که پیشامد $\mathrm{E_i} = 1$ معادل $X_i > t$ می‌باشد، می‌توان نوشت:

$$f(x_i \mid x_i > t) = \frac{f(x_i)}{1 - F(t)} = \frac{1}{\theta} e^{\frac{t}{\theta} - \frac{x_i}{\theta}}$$

$$E(X_i \mid E_i = 1) = E(X_i \mid X_i > t)$$

$$= \int_t^\infty x_i f(x_i \mid x_i > t) dx_i = t + \theta$$

بطور مشابه در حالت $E_i = 0$ خواهیم داشت:

$$E(X_i \mid E_i = 0) = \int_t^t x_i f(x_i \mid x_i \le t) dx_i = \theta - \frac{te^{t/\theta}}{1 - e^{t/\theta}}$$

در نتیجه می‌توان نوشت:

$$E(X_i|\mathbf{Y}) = E(X_i|E_i) = \begin{cases} t+\theta & \mathrm{E_i} = ۱ \\ \theta - \dfrac{te^{t/\theta}}{۱-e^{t/\theta}} & \mathrm{E_i} = ۰ \end{cases} \qquad (۲)$$

حال می‌توان گفت که j امین مرحله از الگوریتم شامل جایگذاری $E(X_i|E_i)$ از رابطهٔ (۲)، بجای X_i در رابطهٔ (۱) می‌باشد. لذا خواهیم داشت:

$$l^{(j)}(\theta) = -(N+M)\log\theta - \frac{۱}{\theta}[N\bar{Y} + Z(t+\theta^{(j-۱)}) + (M-Z)(\theta^{(j-۱)} - tp^{(j-۱)})] \qquad (۳)$$

که در آن $p^{(j)} = \frac{e^{-t/\theta^{(j)}}}{1-e^{-t/\theta^{(j)}}}$ و منظور از $\theta^{(j)}$ ، θی بدست آمده در jامین تکرار می‌باشد. در jامین تکرار مرحلهٔ M ، $\theta^{(j)}$ را چنان برآورد می‌کنیم که رابطهٔ (۳) را بیشینه می‌سازد، به عبارت دیگر با فرض ثابت بودن $\theta^{(j+1)}$، برآورد بیشترین درستنمایی $\theta^{(j)}$ را بدست می‌آوریم:

$$\frac{\partial l^{(j)}(\theta)}{\partial \theta} = ۰ \qquad \Rightarrow \qquad (۴)$$

$$\theta^{(j)} = f(\theta^{(j-۱)})$$

$$\equiv \frac{N\bar{Y} + Z(t + \theta^{(j-۱)}) + (M - Z)(\theta^{(j-۱)} - tp^{(j-۱)})}{N + M}$$

حال می‌توان با انتخاب یک $\theta^{(۰)} > ۰$ معادلهٔ (۴) را تا رسیدن به همگرایی تکرار کرد. خاصیّت خودسازگاری الگوریتم وقتی آشکار می‌شود که $Z = M$ باشد (یعنی کلیهٔ لامپ هایی که در

آزمایش دوّم شرکت داده‌ایم تا زمان t سالم باشند)، در این حالت به جواب معروف زیر می‌رسیم:

$$\hat{\theta} = \frac{N\bar{Y} + Mt}{N}$$

مراجع

[1] Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, (with discussion), *Journal of the Royal Statistical Society,* Ser. B. 39, 1-38.
[2] McLachlan, G., and Krishnan, T., 1997, *The EM-algorithm and Extensions*, New York: Wiley.
[3] Tanner, M. A., 1996, *Tools for Statistical Inference,* (3rd ed.), New York: Springer-Verlag.
[4] Wu, C. F. J., 1983, On the Convergence Properties of the EM Algorithm, *The Annals of Statistics*, 11, 95-103.