یادگیری ماشین

نیم‌سال اول ۱۴۰۱-۱۴۰۰

مدرس: دکتر عباس حسینی

دانشکده مهندسی کامپیوتر

تمرین پنجم

محمدجواد هزاره — شماره دانشجویی: ۹۸۱۰۱۰۷۴

سوال ۱

۱. در الگوریتم k-means خوشه‌ها به صورت دایروی خواهند بود در حالی که در الگوریتم EM حالت کلی‌ای برای خوشه‌ها فرض کرده‌ایم که با ماتریس Σ برای هر خوشه مشخص می‌شود. در این الگوریتم قیدی روی دایروی بودن خوشه‌ها نیست و خوشه‌ها به شکل بیضی فرض شده‌اند. علاوه بر آن، در الگوریتم EM داده‌ها به صورت soft به خوشه‌ها تخصیص داده می‌شوند، به این معنا که برای هر داده، به میزان اعتقادی که برای تعلق آن به خوشه kام داریم، آن را در تعیین پارامترهای مدل kام دخیل می‌کنیم. این دو الگوریتم می‌توانند عملکرد یکسانی داشته باشند به شرط آن‌که در اجرای EM ، ماتریس کوواریانس مدل‌ها، Σ_kها رابه صورت $\alpha \boldsymbol{I}$ در نظر بگیریم که $\boldsymbol{I}$ ماتریس واحد است. با این کار خوشه‌ها در الگوریتم EM ، مشابه الگوریتم k-means دایروی و با شعاع برابر خواهند شد.

۲. برای توزیع احتمال x_a به شرط x_b خواهیم داشت:

$$\begin{aligned}
p(x_a|x_b) &= \frac{p(x_a, x_b)}{p(x_b)} \\
&= \frac{\sum_k \pi_k \, p(x_a, x_b|k)}{p(x_b)} \\
&= \sum_k \frac{\pi_k}{p(x_b)} \, p(x_b|k) \, p(x_a|x_b, k) \\
&= \sum_k \pi'_k \, p(x_a|x_b, k), \quad \pi'_k = \frac{\pi_k \, p(x_b|k)}{p(x_b)}
\end{aligned}$$

که $p(x_b|k) = \sum_{x_a} p(x_a, x_b|k)$ و $p(x_b) = \sum_{x_a} p(x_a, x_b)$، بنابراین برای ضرایب داریم:

$$\boxed{\pi'_k = \frac{\pi_k \sum_{x_a} p(x_a, x_b|k)}{\sum_{x_a} p(x_a, x_b)}}$$

۳. مطابق الگوریتم، کران پایین درست‌نمایی به صورت زیر خواهد بود که k اندیس مربوط به خوشه‌ها و i اندیس

مربوط به داده‌هاست:

$$F(\theta, Q) = \sum_i \left[\sum_k Q_i^{(t)}(k) \log \left(\frac{p(x^{(i)}, k; \theta)}{Q_i^{(t)}(z)} \right) \right]$$

که $Q_i^{(t)}(z)$ پارامتر مربوط به داده‌ی iام است که در مرحله‌ی t و E بدست آمده است. در واقع همان اعتقادی است که به عضویت داده‌ی iام به خوشه‌ی z داریم. احتمال داده‌ی iام و خوشه‌ی k به صورت زیر خواهد بود:

$$\begin{aligned} p(x^{(i)}, k; \theta) &= p(k)\, p(x^{(i)}|k) \\ &= \pi_k\, |2\pi\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x^{(i)} - \mu_k)^T \Sigma^{-1} (x^{(i)} - \mu_k) \right) \end{aligned}$$

بنابراین:

$$\begin{aligned} \theta^{(t+1)} &= \arg\max_\theta\ F(\theta, Q^{(t)}) \\ &= \arg\max_\theta \sum_i \left[\sum_k Q_i^{(t)}(k) \log p(x^{(i)}, k; \theta) \right] \\ &= \arg\max_\theta \sum_i \left[\sum_k Q_i^{(t)}(k) \left(\log \pi_k - \frac{d}{2} \log 2\pi + \frac{1}{2} \log |\Sigma^{-1}| - \frac{1}{2}(x^{(i)} - \mu_k)^T \Sigma^{-1} (x^{(i)} - \mu_k) \right) \right] \end{aligned}$$

که d بعد داده‌هاست. با توجه به رابطه‌ی بالا، جواب این معادله برای π_kها و μ_kها تفاوتی با روش معمول EM ندارد. اما برای پیدا کردن Σ داریم:

$$\begin{aligned} \frac{\partial f(\pi_k, \mu_k, \Sigma)}{\partial \Sigma} &= \frac{\partial f(\pi_k, \mu_k, \Sigma)}{\partial \Sigma^{-1}} \times \frac{\partial \Sigma^{-1}}{\partial \Sigma} \\ &= \frac{\partial \Sigma^{-1}}{\partial \Sigma} \sum_i \left[\sum_k Q_i^{(t)}(k) \left(\frac{1}{2}\Sigma^T - \frac{1}{2}(x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T \right) \right] \end{aligned}$$

با صفر قرار دادن رابطه‌ی بالا به ازای $\hat{\Sigma}$ و $\hat{\mu_k}$ ها و توجه به این نکته که ماتریس کوواریانس ماتریسی متقارن است، خواهیم داشت: (که $\hat{\mu_k}$ تخمین μ_kها در همین مرحله و N تعداد کل داده‌ها است.)

$$\frac{\partial \Sigma^{-1}}{\partial \Sigma} \sum_i \left[\sum_k Q_i^{(t)}(k) \left(\frac{1}{2}\hat{\Sigma}^T - \frac{1}{2}(x^{(i)} - \hat{\mu_k})(x^{(i)} - \hat{\mu_k})^T \right) \right] = 0$$

$$\Longrightarrow \sum_i \left[\hat{\Sigma}^T - \sum_k Q_i^{(t)}(k)(x^{(i)} - \hat{\mu_k})(x^{(i)} - \hat{\mu_k})^T \right] = 0$$

$$\Longrightarrow \boxed{\hat{\Sigma} = \frac{1}{N} \sum_i \left[\sum_k Q_i^{(t)}(k)\ (x^{(i)} - \hat{\mu_k})(x^{(i)} - \hat{\mu_k})^T \right]}$$

سوال ۲

۱. داده‌های مسئله را $\mathcal{D} = \{x^{(i)}\}_{i=1}^{N}$ در نظر می‌گیریم که $x^{(i)} \in \mathbb{N}$. بنابراین کران پایین درست‌نمایی برابر خواهد بود با:

$$F(\theta, Q) = \sum_i \left[\sum_k Q_i(k) \log \left(\frac{p(x^{(i)}, k; \theta)}{Q_i(k)} \right) \right]$$

مطابق الگوریتم EM، در مرحله‌ی E با استفاده از پارامترها تا این لحظه، تابع F را با ثابت نگه داشتن θ، نسبت به Q بیشینه می‌کنیم. بنابراین در مرحله‌ی E داریم:

$$Q_i^{(t)} = \arg\max_Q F(\theta^{(t)}, Q)$$

که جواب این بهینه‌سازی برابر با احتمال شرطی z به شرط دیدن داده‌ی iام است. یا به عبارتی:

$$Q_i^{(t)}(z) = p(z|x^{(i)}; \theta^{(t)})$$

که برای این مسئله داریم:

$$\begin{aligned} Q_i^{(t)}(k) &= \frac{p(k, x^{(i)}; \theta^{(t)})}{p(x^{(i)}; \theta^{(t)})} \\ &= \boxed{\frac{\pi_k e^{-\theta_k} \frac{\theta_k^{x^{(i)}}}{x^{(i)}!}}{\sum_j \pi_j e^{-\theta_j} \frac{\theta_j^{x^{(i)}}}{x^{(i)}!}}} \end{aligned}$$

که پارامترهای θ_k از $\theta^{(t)}$ آمده اند که برای جلوگیری از شلوغ شدن رابطه، $^{(t)}$ برای آن‌ها نیامده است.

سپس در مرحله‌ی M با استفاده از رابطه‌ی بدست آمده برای Q، تابع F را با ثابت نگه داشتن Q نسبت به θ بیشینه می‌کنیم. بنابراین:

$$\begin{aligned} \theta^{(t+1)} &= \arg\max_\theta F(\theta, Q^{(t)}) \\ &= \arg\max_\theta \sum_i \left[\sum_k Q_i^{(t)}(k) \log p(x^{(i)}, k; \theta) \right] \end{aligned}$$

برای احتمال مشترک نیز داریم:

$$p(x^{(i)}, k; \theta) = \pi_k e^{-\theta_k} \frac{\theta_k^{x^{(i)}}}{x^{(i)}!}$$

بنابراین:

$$\theta^{(t+1)} = \arg\max_\theta \sum_i \left[\sum_k Q_i^{(t)}(k) \left(\log \pi_k - \theta_k + x^{(i)} \log \theta_k \right) \right]$$

همچنین می‌دانیم که $\sum_k \pi_k = 1$. بنابراین از روش ضرایب لاگرانژ استفاده می‌کنیم. برای لاگرانژ داریم:

$$L(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda) = \sum_i \left[\sum_k Q_i^{(t)}(k) \left(\log \pi_k - \theta_k + x^{(i)} \log \theta_k \right) \right] - \lambda (\sum_k \pi_k - 1)$$

برای مشتق‌ها خواهیم داشت:

$$\begin{cases} \dfrac{\partial L}{\partial \pi_j} = \sum_i \left[Q_i^{(t)}(j) \dfrac{1}{\pi_j} \right] - \lambda \\ \dfrac{\partial L}{\partial \theta_j} = \sum_i \left[Q_i^{(t)}(j) (-1 + \dfrac{x^{(i)}}{\theta_j}) \right] \\ \dfrac{\partial L}{\partial \lambda} = \sum_k \pi_k - 1 \end{cases}$$

که با صفر قرار دادن روابط بالا به ازای $\theta_j^{(t+1)}$ و $\pi_j^{(t+1)}$، به نتایج زیر می‌رسیم:

$$\boxed{\begin{cases} \pi_k^{(t+1)} = \dfrac{\sum_i Q_i^{(t)}(k)}{\sum_j \sum_i Q_i^{(t)}(j)} = \dfrac{\sum_i Q_i^{(t)}(k)}{N} \\ \theta_k^{(t+1)} = \dfrac{\sum_i x^{(i)} Q_i^{(t)}(k)}{\sum_i Q_i^{(t)}(k)} \end{cases}}$$

۲. در این قسمت می‌خواهیم بهینه‌سازی‌ای که در قسمت قبل و در مرحله‌ی E استفاده کردیم را اثبات کنیم. در واقع می‌خواهیم نشان دهیم جواب بهینه‌سازی $\arg\max_Q F(\theta^{(t)}, Q)$ به صورت $Q_i^{(t)}(z) = p(z|x^{(i)}; \theta^{(t)})$ است. برای این منظور داریم:

$$Q^{(t)} = \arg\max_Q \sum_i \left[\sum_k Q_i(k) \left(\log p(x^{(i)}, k; \theta^{(t)}) - \log Q_i(k) \right) \right]$$

همچنین برای هر i قیدی به صورت $\sum_k Q_i(k) = 1$ داریم. بنابراین برای بهینه‌سازی از روش ضرایب لاگرانژ استفاده می‌کنیم. لاگرانژ به صورت زیر خواهد بود:

$$L(\boldsymbol{Q}, \boldsymbol{\lambda}) = \sum_i \left[\sum_k Q_i(k) \left(\log p(x^{(i)}, k; \theta^{(t)}) - \log Q_i(k) \right) \right] - \sum_i \left[\lambda_i \left(\sum_k Q_i(k) - 1 \right) \right]$$

برای حل این بهینه‌سازی، باید دقت کنیم که Q_iها تابع هستند. اما برای حل می‌توان آن‌ها را به صورت سطرهای یک ماتریس در نظر گرفت چرا که دامنه‌ی k گسسته و محدود است. اگر N تعداد داده‌ها و K تعداد خوشه‌ها باشد، ماتریس $A_{N\times K}$ را به این صورت که $A_{ij} = Q_i(j)$ در نظر می‌گیریم. هدف پیدا کردن درایه‌های این

ماتریس خواهد بود. بنابراین:

$$L(A,\boldsymbol{\lambda}) = \sum_i \left[\sum_k A_{ik}\left(\log p(x^{(i)},k;\theta^{(t)}) - \log A_{ik}\right)\right] - \sum_i \left[\lambda_i \left(\sum_k A_{ik} - 1\right)\right]$$

برای مشتق‌ها خواهیم داشت:

$$\begin{cases} \dfrac{\partial L}{\partial A_{ij}} = \log p(x^{(i)},j;\theta^{(t)}) - \log A_{ij} - 1 - \lambda_i \\ \dfrac{\partial L}{\partial \lambda_i} = \displaystyle\sum_j A_{ij} - 1 \end{cases}$$

با صفر قرار دادن روابط بالا به ازای A^*_{ij} و λ^*_j خواهیم داشت:

$$\begin{cases} (1) \quad \log p(x^{(i)},j;\theta^{(t)}) - \log A^*_{ij} - 1 - \lambda^*_i = 0 \\ (2) \quad \displaystyle\sum_j A^*_{ij} - 1 = 0 \end{cases}$$

با حل رابطه‌ی (1) برای A^*_{ij} بر حسب λ^*_i به نتیجه‌ی زیر می‌رسیم:

$$A^*_{ij} = p(x^{(i)},j;\theta^{(t)})\, e^{-(1+\lambda^*_i)}$$

با قرار دادن این رابطه در معادله‌ی (2) ادامه می‌دهیم:

$$e^{-(1+\lambda^*_i)} \sum_j p(x^{(i)},j;\theta^{(t)}) = 1 \implies 1 + \lambda^*_i = \log p(x^{(i)};\theta^{(t)})$$

حال این نتیجه را در معادله‌ی (1) قرار می‌دهیم:

$$\begin{aligned} &\log p(x^{(i)},j;\theta^{(t)}) - \log A^*_{ij} - \log p(x^{(i)};\theta^{(t)}) = 0 \\ &\implies \log\left(\frac{p(x^{(i)},j;\theta^{(t)})}{A^*_{ij}\, p(x^{(i)};\theta^{(t)})}\right) = 0 \\ &\implies A^*_{ij} = \frac{p(x^{(i)},j;\theta^{(t)})}{p(x^{(i)};\theta^{(t)})} \\ &\implies \boxed{A^*_{ij} = p(j|x^{(i)};\theta^{(t)})} \end{aligned}$$

بنابراین با توجه به تعریف A، داریم $Q^{(t)}_i(k) = p(k|x^{(i)};\theta^{(t)})$ و خواسته‌ی مسئله اثبات می‌شود.

سوال ۳

۱. احتمال این‌که w_t برد در روز t داشته باشیم به شرط آن‌که بدانیم در این روز در کل m_t بازی انجام شده و بازیکن kام بازی می‌کرده است، با فرض این‌که نتیجه بازی‌ها از یکدیگر مستقل بوده و بازیکن kام با احتمال p_k بازی خود را می‌برد، برابر خواهد بود با:

$$w_t|m_t, k \sim Bin(m_t, p_k) \implies p(w_t|m_t, k) = \binom{m_t}{w_t} p_k^{w_t}(1-p_k)^{m_t-w_t}$$

برای احتمال این‌که w_t برد داشته باشیم به شرط آن‌که بدانیم در این روز تعداد m_t بازی صورت گرفته، برابر خواهد بود با:

$$\begin{aligned} p(w_t|m_t) &= \sum_k p(w_t, k|m_t) \\ &= \sum_k p(k|m_t)\, p(w_t|k, m_t) \\ &= \sum_k \pi_k \binom{m_t}{w_t} p_k^{w_t}(1-p_k)^{m_t-w_t} \end{aligned}$$

۲. می‌دانیم در مرحله‌ی E، مقدار Q به صورت $Q_t^{(i)}(z_t) = p(z_t|x_t; \theta^{(i-1)})$ آپدیت می‌شود. بنابراین در این مسئله داریم:

$$\begin{aligned} Q_t^{(i)}[k] &= p(k|w_t, m_t; \theta^{(i-1)}) \\ &= \frac{p(k, w_t, m_t)}{p(w_t, m_t)} \\ &= \frac{p(k)\, p(m_t|k)\, p(w_t|m_t, k)}{p(m_t)\, p(w_t|m_t)} \\ &= \boxed{\frac{\pi_k^{(i-1)} \binom{m_t}{w_t} \left(p_k^{(i-1)}\right)^{w_t} \left(1-p_k^{(i-1)}\right)^{m_t-w_t}}{\sum_j \pi_j^{(i-1)} \binom{m_t}{w_t} \left(p_j^{(i-1)}\right)^{w_t} \left(1-p_j^{(i-1)}\right)^{m_t-w_t}}} \end{aligned}$$

که با توجه به مستقل بودن m_t از k، $p(m_t|k)$ از صورت با $p(m_t)$ از مخرج ساده شده است.

۳. اگر کران پایین درست‌نمایی را با $F(\theta, Q)$ نشان دهیم، آن‌گاه خواهیم داشت:

$$F(\theta, Q) = \sum_t \left[\sum_k Q_t[k] \log\left(\frac{p(x_t, k; \theta)}{Q_t[k]}\right)\right]$$

حال برای پیدا کردن بیشینه‌ی این تابع به روش الگوریتم EM، باید Q را ثابت و برابر مقدار به دست آمده در مرحله‌ی

E در نظر گرفت و سپس F را نسبت به θ بیشینه کنیم. بنابراین:

$$\begin{aligned}
\theta^{(i)} &= \arg\max_{\theta} F(\theta, Q^{(i)}) \\
&= \arg\max_{\theta} \sum_t \left[\sum_k Q_t^{(i)}[k] \left(\log p(x_t, k; \theta) - \log Q_t^{(i)}[k] \right) \right] \\
&= \arg\max_{\theta} \sum_t \left[\sum_k Q_t^{(i)}[k] \log p(x_t, k; \theta) \right]
\end{aligned}$$

برای احتمال داده‌ی tام نیز داریم:

$$\begin{aligned}
p(w_t, m_t, k; \theta) &= p(k)\, p(m_t|k)\, p(w_t|m_t, k) \\
&= \pi_k \binom{m_t}{w_t} p_k^{w_t} (1 - p_k)^{m_t - w_t}\, p(m_t|k)
\end{aligned}$$

با جایگذاری این رابطه در رابطه‌ی مربوط به تتا و در نظر گرفتن این موضوع که $p(m_t|k)$ مستقل(از لحاظ تابع، نه از لحاظ احتمالاتی) از θ است، خواهیم داشت:

$$\theta^{(i)} = \arg\max_{\pi_k, p_k} \sum_t \left[\sum_k Q_t^{(i)}[k] \left(\log \pi_k + w_t \log p_k + (m_t - w_t) \log(1 - p_k) \right) \right]$$

برای پیداکردن $p_j^{(i)}$ کافیست از رابطه‌ی بالا مشتق گرفته و برابر صفر قرار دهیم. بنابراین:

$$\frac{\partial f(\pi_k, p_k)}{\partial p_j} = \sum_t \left[Q_t^{(i)}[j] \left(\frac{w_t}{p_j} - \frac{(m_t - w_t)}{1 - p_j} \right) \right]$$

و با صفر قرار دادن رابطه‌ی بالا به ازای $p_j^{(i)}$ به نتیجه‌ی زیر می‌رسیم:

$$\boxed{p_j^{(i)} = \frac{\sum_{t=1}^{n} Q_t^{(i)}[j]\, w_t}{\sum_{t=1}^{n} Q_t^{(i)}[j]\, m_t}}$$

سوال ۴

۱. لگاریتم احتمال پسین به صورت زیر به احتمال پیشین پارامتر و درست‌نمایی بستگی خواهد داشت:

$$\log p(\theta|X) = \log p(\theta) + \log p(X|\theta) + Const$$

بنابراین θای که احتمال پیشین را بیشینه کند، عبارت سمت راست را نیز بیشینه خواهد کرد. هدف حل مسئله‌ی زیر است:

$$\theta^* = \arg\max_\theta \left(\log p(\theta) + \log p(X|\theta)\right)$$

برای حل این مسئله به روش EM، نخست کران پایینی روی تابع هدف پیدا می‌کنیم. برای این کار تابع زیر را ارائه می‌کنیم:

$$\bar{F}(\theta, Q) = F(\theta, Q) + \log p(\theta)$$

که F همان تابعی است که در روش MLE استفاده می‌کردیم و برابر با:

$$F(\theta, Q) = \sum_i \left[\sum_k Q_i(k) \log\left(\frac{p(x^{(i)}, k; \theta)}{Q_i(k)}\right)\right]$$

اثبات کران پایین بودن تابع ارائه شده نیز سر راست است. برای F که با توجه به محاسبات قبلی و نامساوی Jesnen می‌دانیم $F(\theta, Q) \geq \log p(X|\theta)$، و با اضافه کردن $\log p(\theta)$ به دو طرف نامعادله نیز به نتیجه‌ی خواسته شده یعنی $\bar{F}(\theta, Q) \geq \log p(X|\theta) + \log p(\theta) = \log p(\theta|X)$ می‌رسیم.

بنابراین مراحل الگوریتم EM به صورت زیر خواهد بود:

$$\begin{aligned}
\text{E}: &\quad Q^{(t)} = \arg\max_Q \bar{F}(\theta^{(t)}, Q) = \arg\max_Q F(\theta^{(t)}, Q) \\
\text{M}: &\quad \theta^{(t+1)} = \arg\max_\theta \bar{F}(\theta, Q^{(t)})
\end{aligned}$$

که در تساوی دوم مرحله‌ی E، از استقلال جبری $\log p(\theta)$ از Q استفاده شده است.

۲. داده‌ی ورودی مسئله را به صورت (n_b, n_d) در نظر می‌گیریم. متغیر نهان مسئله نیز همان n_a خواهد بود. با مشخص شدن این سه متغیر مقدار n_c از روی معادله‌ی $n_a + n_b + n_c + n_d = n$ مشخص خواهد شد. یک سمپل در این مسئله برابر با یک دوتایی (n_b, n_d) خواهد بود. بنابراین در این نمونه از مسئله‌ی داده شده فقط یک سمپل داریم. (بنابراین اندیس‌های i را می‌توان از تمام معادلات حذف کرد.) برای احتمال دیدن n_a ، n_b ، n_c و n_d خواهیم

داشت:

$$p(n_a, n_b, n_c, n_d; \theta) = \frac{n!}{n_a! n_b! n_c! n_d!} \left(\frac{1}{3}\right)^{n_a} \left(\frac{1}{3}(1-\theta)\right)^{n_b} \left(\frac{2}{3}\theta\right)^{n_c} \left(\frac{1}{3}(1-\theta)\right)^{n_d}$$

با توجه به قسمت اول سوال، برای مرحله‌ی E داریم:

$$Q^{(t)} = \arg\max_{Q} \sum_k Q(k) \log \left(\frac{p(n_b, n_d, n_a = k; \theta)}{Q(k)}\right)$$

که حل این معادله مشابه قبل بوده و جواب آن برابر $p(n_a = k | n_b, n_d; \theta^{(t)})$ خواهد بود. بنابراین:

$$\begin{aligned} Q^{(t)} &= p(n_a = k | n_b, n_d; \theta^{(t)}) \\ &= \frac{p(n_a = k, n_b, n_d)}{\sum_j p(n_a = j, n_b, n_d)} \\ &= \frac{\frac{n!}{k! n_b! n_c! n_d!} 3^{-n} (1-\theta)^{n_b + n_d} (2\theta)^{n_c}}{\sum_j \frac{n!}{j! n_b! n_c! n_d!} 3^{-n} (1-\theta)^{n_b + n_d} (2\theta)^{n_c}}, \quad n_c = n - (n_a + n_b + na) \end{aligned}$$

برای مرحله‌ی M خواهیم داشت:

$$\begin{aligned} \theta^{(t+1)} &= \arg\max_{\theta} \ \log p(\theta) + \sum_k Q^{(t)}(k) \log \left(\frac{p(n_b, n_d, n_a = k; \theta)}{Q^{(t)}(k)}\right) \\ &= \arg\max_{\theta} \ \log p(\theta) + \sum_k Q^{(t)}(k) \log p(n_b, n_d, n_a = k; \theta) \\ &= \arg\max_{\theta} \ (v_1 - 1) \log \theta + (v_2 - 1) \log(1-\theta) + \sum_k Q^{(t)}(k) \left[(n_b + n_d) \log(1-\theta) + n_c \log \theta\right] \end{aligned}$$

برای حل این بهینه‌سازی نیز کافیست نسبت به θ مشتق گرفته و برابر صفر قرار دهیم. بنابراین:

$$\begin{aligned} \frac{df}{d\theta} &= \frac{v_1 - 1}{\theta} - \frac{v_2 - 1}{1-\theta} + \sum_k Q^{(t)}(k) \left(\frac{n_c}{\theta} - \frac{n_b + n_d}{1-\theta}\right) \\ &= \frac{v_1 - 1}{\theta} - \frac{v_2 - 1}{1-\theta} - \frac{n_b + n_d}{1-\theta} + \sum_k Q^{(t)}(k) \frac{n - n_b - n_d - k}{\theta} \\ &= \frac{v_1 - 1 + n - n_b - n_d - \sum_k k Q^{(t)}(k)}{\theta} - \frac{v_2 - 1 + n_b + n_b}{1-\theta} \end{aligned}$$

با صفر قرار دادن رابطه‌ی بالا برای $\theta^{(t+1)}$ خواهیم داشت:

$$\boxed{\theta^{(t+1)} = \frac{(v_1 - 1) + (n - n_b - n_d) - \sum_k kQ^{(t)}(k)}{(v_1 + v_2 - 2) + n - \sum_k kQ^{(t)}(k)}}$$