علیرضا دهقان پور فراشاه ۹۸۱۰۱۵۵۵

سوال یک :

الف) فرض کنید w^* برداری باشد که $\|w^*\|_2 = B$ و به ازای همه i داشته باشیم $y_i w_*^T x_i \geq 1$.

$$w^{(1)} = 0$$

برای داده‌های misclassified

$$w^{K+1} = w^K + y_i x_i \xRightarrow{w^{*T}x} w^{*T} w^{K+1} = w^{*T} w^K + y_i w^{*T} x_i$$

$$\Rightarrow w^{*T} w^{K+1} = w^{*T} w^K + y_i w^{*T} x_i \underset{\text{طبق فرض}}{\geq} w^{*T} w^K + 1 \geq w^{*T} w^{K-1} + 2$$

به وسیله استقرا و با توجه به این که $w^{(1)}$ صفر است داریم:

$$w^{*T} w^{K+1} \geq K$$

همچنین داریم

$$w^{*T} w^{K+1} \leq \|w^*\|_2 \cdot \|w^{K+1}\| \leq B \|w^{K+1}\|_2$$

بنابراین

$$B \|w^{K+1}\|_2 \geq w^{*T} w^{K+1} \geq K \Rightarrow \|w^{K+1}\| \geq \frac{K}{B}$$

از طرف دیگر داریم :

$$\|w^{K-1}\|^2 = \|w^K + y_i x_i\|^2 = (w^K + y_i x_i)^T (w^K + y_i x_i) =$$

$$= \|w^K\|^2 + \|y_i x_i\|^2 + 2 \underbrace{y_i w^{K^T} x_i} < \|w^K\|^2 + \|y_i x_i\|^2 =$$

چون misclassified است منفی است

$$= \|w^K\|^2 + \|x_i\|^2 \leq \|w^K\|^2 + R^2 \leq \|w^{K-1}\|^2 + 2R^2$$

زیرا y_i ۱ یا ۱- است

بنابراین :

$$\frac{K^2}{B^2} \leq \|w^{K+1}\|^2 < K R^2 \Rightarrow$$

$$\Rightarrow \frac{K^2}{B^2} < K R^2 \Rightarrow \boxed{K < R^2 B^2}$$

اثبات بخش ب در صفحه بعد

ب) اگر در هر مرحله $w^{t+1} = w^t + \eta y_i x_i$ آن گاه خواهیم داشت:

$$w^{K+1} = w^K + \eta y_i x_i \Rightarrow w^{*T} w^{K+1} = w^{*T} w^K + \eta y_i w^{*T} x_i \geq w^{*T} w^K + \eta$$

$$\Rightarrow w^{*T} w^{K+1} \geq K\eta$$

$$\left.\begin{array}{l} w^{*T} w^{K+1} \leq \|w^*\| \, \|w^{K+1}\| \leq B\|w^{K+1}\| \end{array}\right\} \Rightarrow \|w^{K+1}\| \geq \frac{K\eta}{B}$$

همچنین

$$\|w^{K+1}\|^2 = \|w^K + \eta y_i x_i\|^2 = (w^K + \eta y_i x_i)^T (w^K + \eta y_i x_i) =$$

$$= \|w^K\|^2 + \|\eta y_i x_i\|^2 + 2\eta y_i w^{K^T} x_i < \|w^K\|^2 + \eta^2 \|x_i\|^2 \leq$$

$$\leq \|w^K\|^2 + \eta^2 R^2 \Rightarrow \|w^{K+1}\|^2 < K\eta^2 R^2$$

$$\frac{K^2 \cancel{\eta^2}}{B^2} \leq \|w^{K+1}\|^2 < K \cancel{\eta^2} R^2 \Rightarrow$$

$$\Rightarrow \boxed{K < B^2 R^2}$$

سؤال دو :

الف) با برهان خلف اثبات می کنیم. ابتدا فرض کنید $h_t = h_{t+1}$. توجه کنید چون هر مرحله وزن ها را نرمال می کنیم جمع کل وزن ها یک خواهد بود بنابراین داریم :

$$\varepsilon_{t+1} = \frac{\sum_{i \in \text{misclassified}} D_{t+1}(i)}{1 = \sum_{i=1}^{m} D_{t+1}(i)} = \sum_{i \in \text{misclassified}} D_{t+1}(i)$$

چون $h_t = h_{t+1}$ هر دو دسته بند یکسان عمل می کنند و داده های یکسانی را نادرست پیش بینی می کنند

$$\varepsilon_{t+1} = \sum_{i \in mis} D_{t+1}(i) = \sum_{i \in mis} \frac{D_t(i) \times \exp(-\alpha y_i h_t(x_i))}{Z_t} =$$

$$\overset{y_i h_t(x_i) = -1}{=} \sum_{i \in mis} \frac{D_t(i)\, e^{\alpha}}{Z_t} = \sum_{i \in mis} \frac{D_t(i)\, e^{\frac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t}}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} =$$

$$= \sum_{i \in mis} \frac{D_t(i)\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} = \frac{\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} \underbrace{\sum_{i \in mis} D_t(i)}_{\varepsilon_t} =$$

$$= \frac{\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t} \times \varepsilon_t}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} = \frac{\sqrt{\varepsilon_t(1-\varepsilon_t)}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}} = \frac{1}{2} \Rightarrow \varepsilon_{t+1} = \frac{1}{2}$$

اگر $\varepsilon_{t+1} = \frac{1}{2}$ بشود α صفر می شود و دیگر وزن ها تغییر نخواهند کرد و همان جا متوقف می شویم.

بنابراین تا زمانی که ε کمتر از ۰/۵ باشد دو دسته بند یکسان متوالی انتخاب نخواهند شد.

ب) ابتدا با فرض جلوی رویم و به یک رابطه درست می‌رسیم و مراحل را برمی‌گردیم.

تمامی مراحل بازگشت پذیر است:

$$\sum_{i=1}^{m} y_i h_t(x_i) D_{t+1}(i) = 0 =$$

$$= \sum_{i=1}^{m} y_i h_t(x_i) \frac{e^{-\alpha y_i h_t(x_i)} D_t(i)}{Z_t} = \sum_{i=1}^{m} y_i h_t(x_i) \frac{e^{-\frac{1}{2}\log\frac{1-\varepsilon}{\varepsilon} y_i h_t(x_i)} D_t(i)}{Z_t}$$

$$= \sum_{i=1} \frac{y_i h_t(x_i) D_t(i) \sqrt{\frac{\varepsilon}{1-\varepsilon}}^{\,y_i h_t(x_i)}}{Z_t} = 0 \iff$$

Z_t ثابت

$$\iff \sum_{\substack{i \in \text{misclassified} \\ \{y_i h_t(x_i) = -1\}}} -D_t(i) \sqrt{\frac{1-\varepsilon}{\varepsilon}} + \sum_{i \in \{y_i h_t(x_i) = 1\}} D_t(i) \sqrt{\frac{\varepsilon}{1-\varepsilon}} = 0$$

$$\iff \left(\sqrt{\frac{\varepsilon}{1-\varepsilon}}\right) \sum_{i \in \{y_i h_t(x_i) = 1\}} D_t(i) = \left(\sqrt{\frac{1-\varepsilon}{\varepsilon}}\right) \sum_{i \in \{y_i h_t(x_i) = -1\}} D_t(i) \iff$$

$$\iff \frac{\varepsilon}{1-\varepsilon} = \frac{\sum_{i \in \{y_i h_t(x_i) = -1\}} D_t(i)}{\sum_{i \in \{y_i h_t(x_i) = 1\}} D_t(i)}$$

توجه کنید که $\varepsilon = \sum_{\text{misclassified}} D_t(i)$ و

و چون جمع وزن ها یک است

نسبت وزن داده‌های نادرست برچسب خورده

به وزن داده‌های درست برچسب خورده

$\frac{\varepsilon}{1-\varepsilon}$ است که به همان رسیدیم بنابراین تمامی مراحل را بازمی‌گردیم.

سؤال سه:

الف)

$$R(h) = P[Y \neq h(x)]$$

ابتدا برای $X = x$ می‌نویسیم و بعد امید ریاضی می‌گیریم:

$$P[Y \neq h(x)] = P[Y=0, h(x)=1 \mid x] + P[Y=1, h(x)=0 \mid x] =$$

$h(x)$ با توجه به x تعیین می‌شود

$$= (1-P[Y=1 \mid x])(1\{h(x)=1\}) + (P[Y=1 \mid x])(1 - 1\{h(x)=1\})$$

$$= P[Y=1 \mid x] + 1\{h(x)=1\}(1-2P[Y=1 \mid x])$$

برای هر h دلخواه و x

$$R(h(x)) - R(h^*(x)) = P[Y=1 \mid x] + 1\{h(x)=1\}(1-2P[Y=1 \mid x]) + - P[Y=1 \mid x] - 1\{h^*(x)=1\}(1-2P[Y=1 \mid x]) =$$

$$= \underbrace{(1-2P[Y=1 \mid x])}_{A}(1\{h(x)=1\} - 1\{h^*(x)=1\})$$

① اگر $P[Y=1 \mid x] > \frac{1}{2}$ باشد $h^*(x)=1$ است پس A منفی خواهد شد و B می‌شود $(1\{h(x)=1\} - 1)$

پس $0 \leq B$ خواهد شد و $0 \leq AB$ خواهد شد.

② اگر $P[Y=1 \mid x] \leq \frac{1}{2}$ باشد آن گاه $h^*(x)=0$ است و A نامنفی است و B می‌شود $1\{h(x)=1\}$ که نامنفی است پس $0 \leq AB$ است.

بنابراین $R(h) - R(h^*)$ همواره مثبت خواهد بود و اگر به ازای هر x امید بگیریم خواهیم داشت:

$$R(h) - R(h^*) = \int_x (1-2P[Y=1 \mid x]) P(x) (1\{h(x)=1\} - 1\{h^*(x)=1\}) \, dx \geqslant 0$$

پس به ازای هر h داریم $\underline{R(h) \geqslant R(h^*)}$ یعنی $\underline{h^*}$ جواب بهینه است.

اثبات بخش ب در صفحه بعد.

ب) از بخش قبل ثابت کردیم:

$$R(h) - R(h^*) = \int_x \underbrace{(1-2P[Y=1|x])}_{A}\underbrace{(1\{h(x)=1\} - 1\{h^*(x)=1\})}_{B} P(x)\, dx =$$

A و B هم علامت اند

$$= \int_x |1-2P[Y=1|x]|\ |1\{h(x)=1\} - 1\{h^*(x)=1\}|\ P(x)\, dx =$$

$$= \int_x \underbrace{|1-2P[Y=1|x]|}_{|A|}\ 1\{h(x) \neq h^*(x)\}\ P(x)\, dx$$

اگر $h(x) \neq h^*(x)$ باشد یعنی $m(x) = P[Y=1|x] > \frac{1}{2}$ و $m^*(x) \leq \frac{1}{2}$ یا

$P[Y=1|x] \leq \frac{1}{2}$ و $m^*(x) > \frac{1}{2}$ بنابراین اگر $h(x) \neq h^*(x)$:

$$m(x) > \frac{1}{2} \text{ و } m^*(x) \leq \frac{1}{2} \Rightarrow m(x) - m^*(x) \geqslant m(x) - \frac{1}{2}$$

$$m(x) \leq \frac{1}{2} \text{ و } m^*(x) > \frac{1}{2} \Rightarrow m^*(x) - m(x) \geqslant \frac{1}{2} - m(x)$$

$$\Rightarrow |m(x) - m^*(x)| \geqslant \left|\frac{1}{2} - m(x)\right| = \left|\frac{1}{2} - P[Y=1|x]\right| \Rightarrow$$

$$\Rightarrow 2|m(x) - m^*(x)| \underset{*}{\geqslant} |1-2P(Y=1|x]| = |A|$$

$$\Rightarrow R(h) - R(h^*) = \int_x \underbrace{|1-2P[Y=1|x]|}_{|A|}\ 1\{h(x) \neq h^*(x)\}\ P(x)\, dx \leq$$

با توجه به *

$$\leq 2\int_x |m(x) - m^*(x)|\ 1\{h(x) \neq h(x)\}\ P(x)\, dx \leq$$

$$\overset{1\{h(x) \neq h^*(x)\} \leq 1}{\leq} 2\int_x |m(x) - m^*(x)|\ P(x)\, dx$$

سوال چهار:

$H(Y|X)$ را برای هر ویژگی محاسبه می کنیم.

نوع غذا:

$$\frac{3}{11}\times\Big(\overbrace{\frac{1}{3}\log\frac{1}{3}+\frac{2}{3}\log\frac{2}{3}}^{-0/2504}\Big)+\frac{3}{11}\times\Big(\overbrace{\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}}^{-0/8504}\Big)+\frac{5}{11}\Big(\overbrace{\frac{2}{5}\log\frac{2}{5}+\frac{3}{5}\log\frac{3}{5}}^{-0/4413}\Big)\simeq -0/94$$

رنج قیمت:

$$\frac{4}{11}\Big(\overbrace{\frac{1}{2}\log+\frac{1}{2}\log\frac{1}{2}}^{-1}\Big)+\frac{2}{11}\Big(\overbrace{1\log 1}^{0}\Big)+\frac{5}{11}\Big(\overbrace{\frac{3}{5}\log\frac{3}{5}+\frac{2}{5}\log\frac{2}{5}}^{-0/4413}\Big)=-0/80$$

محل رستوران:

$$\frac{4}{11}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{4}{11}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{3}{11}\Big(\frac{1}{3}\log\frac{1}{3}+\frac{2}{3}\log\frac{2}{3}\Big)\simeq -0/98$$

محدودیت:

$$\frac{4}{11}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{4}{11}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{3}{11}\Big(\frac{1}{3}\log\frac{1}{3}+\frac{2}{3}\log\frac{2}{3}\Big)\simeq -0/98$$

پس اولین ویژگی رنج قیمت است.

گران	متوسط	ارزان
2 0	2 3	2 2

نود گران را دیگر باز نمی کنیم ولی برای ارزان و متوسط

برای متوسط:

نوع غذا:

$$\frac{1}{5}\log 1+\frac{2}{5}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{2}{5}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)=-0/8$$

محل:

$$\frac{3}{5}\times\Big(\frac{1}{3}\log\frac{1}{3}+\frac{2}{3}\log\frac{2}{3}\Big)+\frac{2}{5}(\log 1)=0/55$$

محدودیت ها:

$$\frac{1}{5}(\log 1)+\frac{2}{5}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{2}{5}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)=-0/8$$

پس برای متوسط از محل استفاده می کنیم:

متوسط

↙ شرق ↘ شمال

شرق: 0 2 — شمال: 2 1

حال شمال را دسته بندی می کنیم

نوع غذا:

$$\frac{1}{3}\log 1+\frac{1}{3}\log 1+\frac{1}{3}\log 1=0$$

محدودیت:

$$\frac{2}{3}\Big(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\Big)+\frac{1}{3}(\log 1)=-0/66$$

نوع غذا را انتخاب می کنیم پس داریم:

شمال

↙ ایرانی ↓ فست فود ↘ آسیایی

ایرانی: 1 0 — فست فود: 0 1 — آسیایی: 1 0

پس سمت متوسط کامل شد.

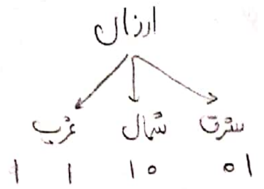
ادامه صفحه بعد

برای دسته ارزان نیز همین کار را انجام می دهیم.

نوع غذا : $\frac{1}{4}\log 1 + \frac{1}{4}\log 1 + \frac{1}{2}\times\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = -0.5$

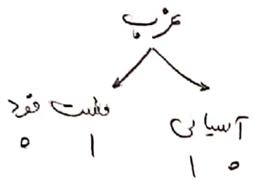
محل : $\frac{1}{4}\log 1 + \frac{1}{4}\log 1 + \frac{1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = -0.5$

محدودیت : $\frac{1}{4}\log 1 + \frac{1}{4}\log 1 + \frac{1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = -0.5$

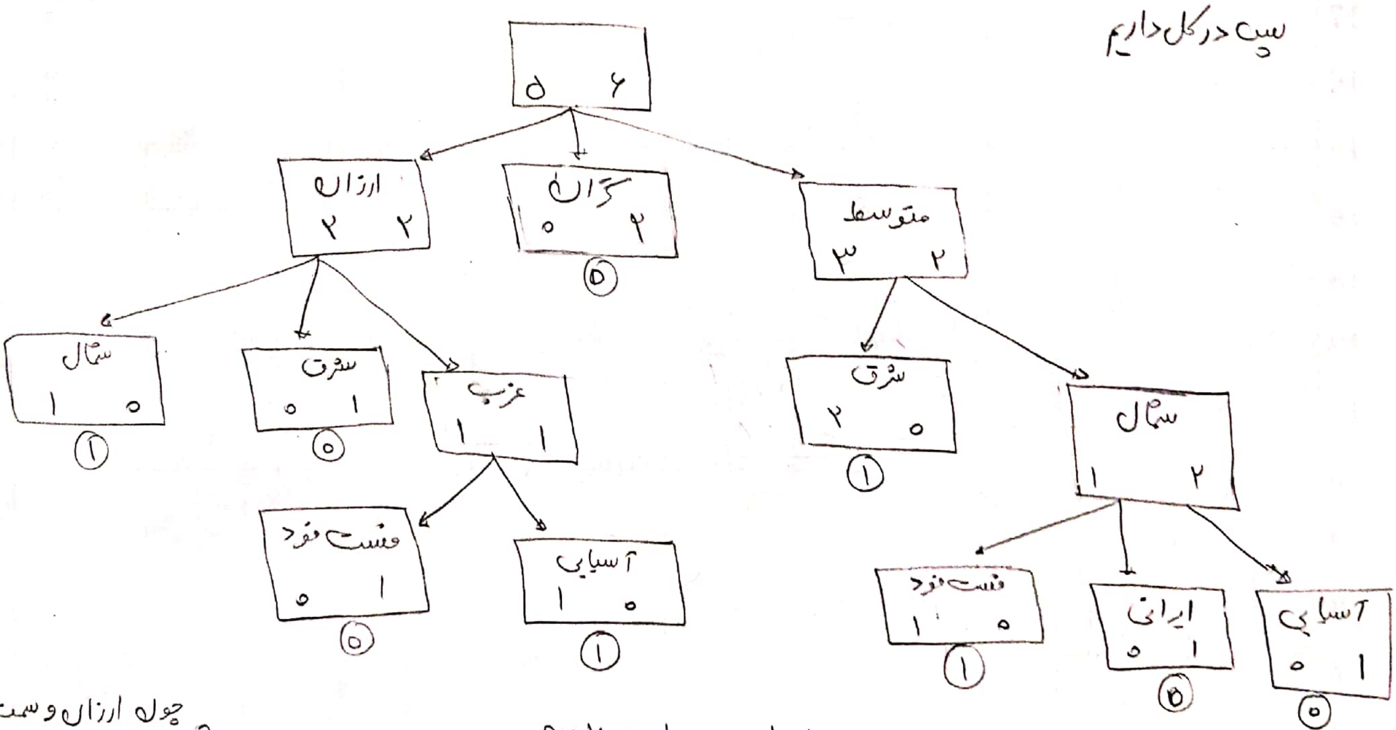


فرقی ندارد و رندوم محل را انتخاب می کنیم:

پس غرب را باید باز کنیم که چون هر دو از لحاظ محدودیت vegeterian هستند براساس نوع غذا باز می کنیم:



پس در کل داریم



ب)

	Predicted	label
رستوران ۱۲	۱	۰
″ ۱۳	۱	۱
″ ۱۴	۱ یا ۰	۰
″ ۱۵	۰	۱
″ ۱۶	۱	۰

چون ارزان و سمت غرب است به ۲ برگ سمت چپ درخت که بعد از غرب آمده است می رسیم. چون ایرانی نداریم به صورت رندوم پیش بینی می کنیم.

ج)

$$\left.\begin{array}{l} recall = \frac{1}{2} \\ Precision = \frac{1}{3} \end{array}\right\} \Rightarrow F_1 = \frac{2\times Pr\times Re}{Pr+Re} = \boxed{\frac{2}{5}}$$

سوال پنج :

الف) log-likelihood : $\log P(\Phi, t) = \log \prod_{i=1}^{N} \prod_{K=1}^{K} \left(P(\Phi_i | C_K) P(C_K)\right)^{t_{iK}} =$

$$= \sum_{i=1}^{N} \sum_{K=1}^{K} t_{iK} \left(\ln P(\Phi_i | C_K) + \ln(\pi_K)\right)$$

$$\frac{\partial l}{\partial \pi_K} = \sum_{i=1}^{N} t_{iK} \times \frac{1}{\pi_K}$$

چون $\sum_{i}^{K} \pi_i = 1$ است باید لاگرانژ استفاده کنیم :

$$\frac{\partial l}{\partial \pi_K} + \lambda \frac{\partial}{\partial \pi_K}\left(\sum_i \pi_i - 1\right) = 0 \Rightarrow \sum_{i=1}^{N} t_{iK} \frac{1}{\pi_K} + \lambda = 0$$

$$\Rightarrow \pi_K = \frac{\sum_{i=1}^{N} t_{iK}}{-\lambda} \quad , \quad \sum_{K=1}^{K} \pi_K = 1 = \frac{\sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK}}{-\lambda} = 1$$

$$\Rightarrow -\lambda = \sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} = N \quad , \quad \sum_{i=1}^{N} t_{iK} = N_K$$

$$\Rightarrow \pi_K = \frac{\sum_{i=1}^{N} t_{iK}}{-\lambda} = \frac{N_K}{N}$$

ب) برای M_K و Σ گرادیان را برابر صفر می‌گذاریم :

$$l = \sum_{i=1}^{N} \sum_{K=1}^{K} t_{iK} \left(\ln \frac{1}{(2\pi)^{d/2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - M_K)^T \Sigma^{-1} (x_i - M_K)\right) + \ln \pi_K\right)$$

$$\frac{\partial l}{\partial M_K} = \sum_{i=1}^{N} t_{iK} \underbrace{\frac{-1}{2} \times (-2) \times \Sigma^{-1}}_{\text{ثابت}} (x_i - M_K) = 0 \Rightarrow$$

$$\sum_{i=1}^{N} t_{iK} (x_i - M_K) = 0 \Rightarrow \sum_{i=1}^{N} t_{iK} x_i = \sum_{i=1}^{N} t_{iK} \underbrace{M_K}_{\text{ثابت}}$$

$$\Rightarrow \sum_{i=1}^{N} t_{iK} x_i = N_K \times M_K \Rightarrow \boxed{M_K = \frac{\sum_{i=1}^{N} t_{iK} x_i}{N_K}}$$

ادامه در صفحه بعد

برای بدست آوردن Σ باید :

$$\frac{\partial l}{\partial \Sigma} = 0 \Rightarrow \frac{\partial l}{\partial \Sigma} = \sum_{i=1}^{N} \sum_{K=1}^{K} \left(-\frac{1}{2} \frac{\partial |\Sigma|}{\partial \Sigma} + \frac{\partial}{\partial \Sigma} \frac{-1}{2} (x_i - \mu_K)^T \Sigma^{-1} (x_i - \mu_K) \right) t_{iK}$$

$$= \sum_{i=1}^{N} \sum_{K=1}^{K} \left(-\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} (x_i - \mu_K)(x_i - \mu_K)^T \Sigma^{-1} \right) t_{iK} = 0$$

$$= \sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} \left(\underbrace{-\frac{1}{2} \Sigma^{-1}}_{\text{ثابت}} \left(I + (x_i - \mu_K)(x_i - \mu_K)^T \right) \Sigma^{-1} \right) \overset{\times \Sigma}{=} 0$$

$$\Rightarrow \sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} \left(\Sigma + (x_i - \mu_K)(x_i - \mu_K)^T \right) = 0$$

$$\Rightarrow \underbrace{\sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} \Sigma}_{N \Sigma} = \sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} (x_i - \mu_K)(x_i - \mu_K)^T$$

$$\Rightarrow \Sigma = \frac{1}{N} \sum_{K=1}^{K} \sum_{i=1}^{N} t_{iK} (x_i - \mu_K)(x_i - \mu_K)^T$$

اگر :

$$\begin{cases} S_K := \dfrac{1}{N_K} \displaystyle\sum_{i=1}^{N} t_{iK} (x_i - \mu_K)(x_i - \mu_K)^T \\ \Sigma = \dfrac{1}{N} \displaystyle\sum_{K=1}^{K} N_K S_K \end{cases}$$