۱- الف)

$$w^* := \arg\min\{\|w\|_2 : \forall i\in[m],\ y_i\langle w, x_i\rangle \geq 1\}$$

$$\langle w^*, w^{(t+1)}\rangle - \langle w^*, w^{(t)}\rangle = \langle w^*, w^{(t)} + y_i x_i\rangle - \langle w^*, w^{(t)}\rangle = \langle w^*, y_i x_i\rangle \geq 1$$

$$\Rightarrow \langle w^*, w^{(T+1)}\rangle \geq T$$

$$\|w^{(t+1)}\|_2^2 = \|w^{(t)} + y_i x_i\|_2^2 = \|w^{(t)}\|_2^2 + \|y_i x_i\|_2^2 + 2y_i\langle w^{(t)}, x_i\rangle \leq \|w^{(t)}\|_2^2 + \|y_i x_i\|_2^2 =$$

$$\|w^{(t)}\|_2^2 + y_i^2\|x_i\|_2^2 = \|w^{(t)}\|_2^2 + \|x_i\|_2^2 \leq \|w^{(t)}\|_2^2 + R^2$$

$$\Rightarrow \|w^{(T+1)}\|_2^2 \leq TR^2 \Rightarrow \|w^{(T+1)}\|_2 \leq \sqrt{T}R$$

$$\cos(\angle(w^*, w^{(T+1)})) = \frac{\langle w^*, w^{(T+1)}\rangle}{\|w^*\|_2\|w^{(T+1)}\|_2} \geq \frac{T}{B\sqrt{T}R} = \frac{\sqrt{T}}{BR}$$

$$\cos(\angle(w^*, w^{(T+1)})) \leq 1 \Rightarrow \frac{\sqrt{T}}{BR} \leq 1$$

$$B > 0,\ R > 0 \Rightarrow \sqrt{T} \leq BR \Rightarrow T \leq (BR)^2$$

با توجه به اینکه نامساوی بالا پس از اجرای هر تعداد باری از حلقه بروزرسانی برقرار است، پس برای آخرین مرتبه نیز برقرار است، پس تعداد کل مراحل بروزرسانی حداکثر $(RB)^2$ است.

ب) با اضافه کردن ابرپارامتر η، نامساوی‌های بدست آمده در بالا به صورت زیر در می‌آید:

$$\langle w^*, w^{(t+1)}\rangle - \langle w^*, w^{(t)}\rangle = \langle w^*, \eta y_i x_i\rangle = \eta\langle w^*, y_i x_i\rangle \geq \eta$$

$$\Rightarrow \langle w^*, w^{(T+1)}\rangle \geq \eta T$$

$$\|w^{(t+1)}\|_2^2 \leq \|w^{(t)}\|_2^2 + \eta^2\|x_i\|_2^2 \leq \|w^{(t)}\|_2^2 + \eta^2 R^2 \Rightarrow \|w^{(T+1)}\|_2^2 \leq T\eta^2 R^2$$

$$\Rightarrow \|w^{(T+1)}\|_2 \leq \sqrt{T}\eta R$$

$$\Rightarrow \cos(\angle(w^*, w^{(T+1)})) = \frac{\eta T}{B\sqrt{T}\eta R} = \frac{T}{B\sqrt{T}R} = \frac{\sqrt{T}}{BR}$$

پس تغییری در استدلال نهایی ایجاد نمی‌شود و همان نتیجه پیشین بدست می‌آید.

۲- الف)

$$\epsilon'_{t+1} := P_{i\sim D_{t+1}}[h_t(x_i) \neq y_i] = \sum_{i=1}^{m} D_{t+1}(i)\, I(y_i \neq h_t(x_i))$$

$$D_{t+1}(i) = \frac{1}{Z_t} D_t(i)\exp(-\alpha_t y_i h_t(i)) \Rightarrow \epsilon'_{t+1} = \frac{1}{Z_t}\sum_{i=1}^{m} D_t(i)\exp(-\alpha_t y_i h_t(i))\, I(y_i \neq h_t(x_i))$$

$$= \frac{1}{Z_t}\sum_{i=1}^{m} D_t(i)\exp(\alpha_t)\, I(y_i \neq h_t(x_i)) = \frac{1}{Z_t}\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{\frac{1}{2}}\sum_{i=1}^{m} D_t(i)\, I(y_i \neq h_t(x_i)) =$$

$$\frac{1}{2(\epsilon_t(1-\epsilon_t))^{\frac{1}{2}}}\left(\frac{1-\epsilon_t}{\epsilon_t}\right)^{\frac{1}{2}}\epsilon_t=\frac{1}{2}$$

از طرفی موقع انتخاب h_{t+1} داریم: $\epsilon_{t+1}<\frac{1}{2}<\epsilon'_t \Longrightarrow \epsilon_{t+1}\neq\epsilon'_t$

پس ممکن نیست h_t و h_{t+1} برابر باشند، پس هیچگاه تابع انتخابی در دو مرحله‌ی متوالی یکسان نیست. $\square$

ب) D_{t+1} یک توزیع احتمال است، پس مجموع اجزای آن برابر ۱ است. داریم:

$$1=\sum_{i=1}^{m}D_{t+1}(i)=\sum_{i=1}^{m}D_{t+1}(i)I(y_i\neq h_t(x_i))+\sum_{i=1}^{m}D_{t+1}(i)I(y_i=h_t(x_i))=\frac{1}{2}+\sum_{i=1}^{m}D_{t+1}(i)I(y_i=h_t(x_i))\Longrightarrow\sum_{i=1}^{m}D_{t+1}(i)I(y_i=h_t(x_i))=\frac{1}{2}$$

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

$$=-\sum_{i=1}^{m}D_{t+1}(i)I(y_i\neq h_t(x_i))+\sum_{i=1}^{m}D_{t+1}(i)I(y_i=h_t(x_i))=-\frac{1}{2}+\frac{1}{2}=0\quad\square$$

۳_ الف)

$$R(h)=P[Y\neq h(X)]=\int_X P[Y\neq h(x),X=x]dx=\int_X P[Y\neq h(x),X=x]dx$$

$$=\int_X P[Y=1,h(x)=0,X=x]+P[Y=0,h(x)=1,X=x]dx=\int_X P[h(x)=0|Y=1,X=x]P(Y=1|X=x)P_X(x)+P[h(x)=1|Y=0,X=x]P(Y=0|X=x)P_X(x)dx$$

$$=\int_X P[h(x)=0]P(Y=1|X=x)P_X(x)+P[h(x)=1]P(Y=0|X=x)P_X(x)dx=\int_X(P[h(x)=0]P(Y=1|X=x)+P[h(x)=1]P(Y=0|X=x))P_X(x)dx$$

$$=\int_X(P[h(x)=0]P(Y=1|X=x)+P[h(x)=1](1-P(Y=1|X=x)))P_X(x)dx=\int_X(P[h(x)=0]m(x)+P[h(x)=1](1-m(x)))P_X(x)dx$$

$$P[h(x)=i]=1\Longleftrightarrow h(x)=i$$

$$\Rightarrow(\operatorname{argmin}_h R(h))(x)=\begin{cases}1 & m(x)>1-m(x)\\ 0 & \text{o.w.}\end{cases}=\begin{cases}1 & m(x)>\frac{1}{2}\\ 0 & \text{o.w.}\end{cases}=h^*(x)\quad\square$$

ب)

$$R(h)=\int_X(P[h(x)=0]m(x)+P[h(x)=1](1-m(x)))P_X(x)dx=\int_X(P[h(x)=0](2m(x)-1)+1-m(x))P_X(x)dx$$

$$\Longrightarrow R(\hat{h})-R^*=\int_X(P[\hat{h}(x)=0]-P[h^*(x)=0])(2m(x)-1)P_X(x)dx=\int_X I(\hat{h}(x)\neq h^*(x))(P[\hat{h}(x)=0]-P[h^*(x)=0])(2m(x)-1)P_X(x)dx$$

$$m(x) > \frac{1}{2} \iff h^*(x) = 1 \iff P[h^*(x) = 0] = 0 \implies$$

$$I(\hat{h}(x) \neq h^*(x))(P[\hat{h}(x)=0] - P[h^*(x)=0])(2m(x)-1) = \begin{cases} I(\hat{h}(x) \neq h^*(x))(2m(x)-1) & m(x) > \frac{1}{2} \\ I(\hat{h}(x) \neq h^*(x))(1-2m(x)) & \text{o.w.} \end{cases}$$

$$= I(\hat{h}(x) \neq h^*(x))|2m(x)-1| = 2I(\hat{h}(x) \neq h^*(x))|m(x) - \tfrac{1}{2}|$$

از طرفی زمانی که داشته باشیم $\hat{h}(x) \neq h^*(x)$، $\underline{m(x)}$ و $\underline{\hat{m}(x)}$ در دو طرف $\frac{1}{2}$ قرار دارند، پس می‌توانیم بنویسیم:

$$|m(x) - \hat{m}(x)| = |m(x) - \tfrac{1}{2}| + |\hat{m}(x) - \tfrac{1}{2}| \geq |m(x) - \tfrac{1}{2}|$$

$$\implies 2I(\hat{h}(x) \neq h^*(x))|m(x) - \tfrac{1}{2}| \leq 2I(\hat{h}(x) \neq h^*(x))|m(x) - \hat{m}(x)| \leq |m(x) - \hat{m}(x)|$$

$$\implies R(\hat{h}) - R^* = \int_X I(\hat{h}(x) \neq h^*(x))(P[\hat{h}(x)=0] - P[h^*(x)=0])(2m(x)-1)P_X(x)\,dx \leq$$

$$\int_X 2|m(x) - \hat{m}(x)|P_X(x)\,dx = 2\int_X |m(x) - \hat{m}(x)|P_X(x)\,dx \quad \square$$

۴_ الف) $X_4 :=$ محدودیت‌ها، $X_3 :=$ محل رستوران، $X_2 :=$ رنج قیمت، $X_1 :=$ نوع غذا، $Y :=$ رضایتمندی

$$H(Y|X_1) = -\left(\tfrac{3}{11}\left(\tfrac{2}{3}\log\tfrac{2}{3} + \tfrac{1}{3}\log\tfrac{1}{3}\right) + \tfrac{3}{11}\left(\tfrac{2}{3}\log\tfrac{2}{3} + \tfrac{1}{3}\log\tfrac{1}{3}\right) + \tfrac{5}{11}\left(\tfrac{3}{5}\log\tfrac{3}{5} + \tfrac{2}{5}\log\tfrac{2}{5}\right)\right) = 0{,}9416$$

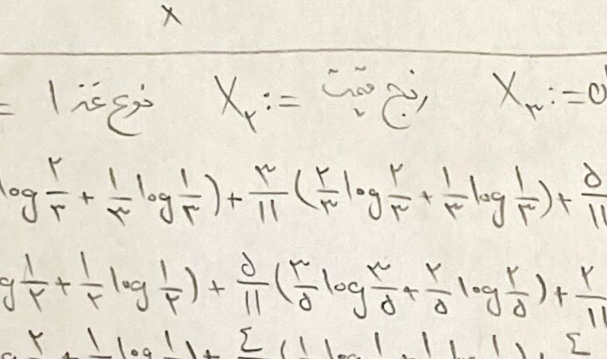
$$H(Y|X_2) = -\left(\tfrac{4}{11}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right) + \tfrac{5}{11}\left(\tfrac{3}{5}\log\tfrac{3}{5} + \tfrac{2}{5}\log\tfrac{2}{5}\right) + \tfrac{2}{11}(1\log 1)\right) = 0{,}8045$$

$$H(Y|X_3) = -\left(\tfrac{3}{11}\left(\tfrac{2}{3}\log\tfrac{2}{3} + \tfrac{1}{3}\log\tfrac{1}{3}\right) + \tfrac{4}{11}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right) + \tfrac{4}{11}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right)\right) = 0{,}9776$$

$$H(Y|X_4) = -\left(\tfrac{4}{11}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right) + \tfrac{3}{11}\left(\tfrac{2}{3}\log\tfrac{2}{3} + \tfrac{1}{3}\log\tfrac{1}{3}\right) + \tfrac{4}{11}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right)\right) = 0{,}9776$$

$$\implies \operatorname{argmin}_X H(Y|X) = X_2$$

پس با تقسیم راس اولیه برحسب قیمت به درخت زیر می‌رسیم:



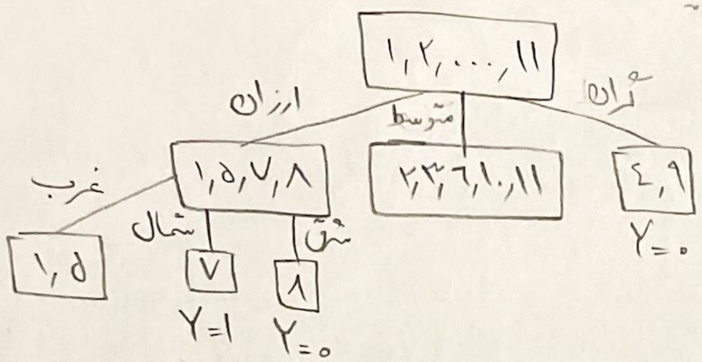
برای فرزند سمت چپ داریم:

$$H(Y|X_1) = -\left(\tfrac{1}{4}(1\log 1) + \tfrac{1}{2}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right) + \tfrac{1}{4}(1\log 1)\right) = 0{,}5$$

$$H(Y|X_3) = -\left(\tfrac{1}{2}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right) + \tfrac{1}{4}(1\log 1) + \tfrac{1}{4}(1\log 1)\right) = 0{,}5$$

$$H(Y|X_4) = -\left(\frac{1}{2}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)+\frac{1}{4}(1\log 1)+\frac{1}{4}(1\log 1)\right) = 0{,}5$$

با توجه به اینکه تفاوتی بین ویژگی‌ها وجود ندارد، به دلخواه باید یکی را انتخاب کنیم. ما در اینجا محل رستوران را انتخاب می‌کنیم. درخت به صورت زیر در می‌آید:

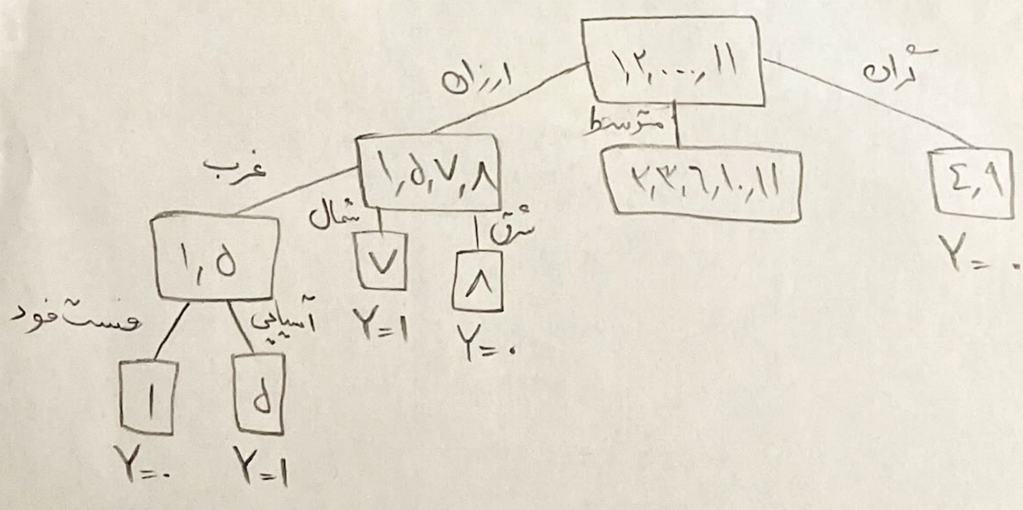


برای رأس ۱,۵ داریم:

$$H(Y|X_1) = -\left(\frac{1}{2}(1\log 1)+\frac{1}{2}(1\log 1)\right) = 0$$

$$H(Y|X_4) = -\left(1\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)\right) = 1$$

پس برحسب نوع غذا این رأس را تقسیم می‌کنیم. داریم:



برای رأس ۲,۳,۶,۱۰,۱۱ داریم:

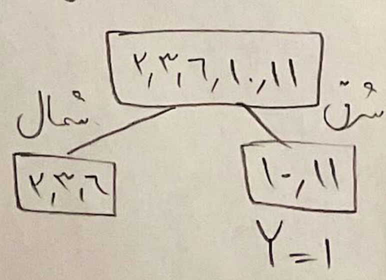
$$H(Y|X_1) = -\left(\frac{1}{5}(1\log 1)+\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)+\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)\right) = 0{,}8$$

$$H(Y|X_3) = -\left(\frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right)+\frac{2}{5}(1\log 1)\right) = 0{,}55$$

$$H(Y|X_4) = -\left(\frac{1}{5}(1\log 1)+\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)+\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)\right) = 0{,}8$$

$\operatorname{argmin}_x H(Y|X) = X_3 \longrightarrow$ محل رستوران

این رأس به صورت زیر در می‌آید:



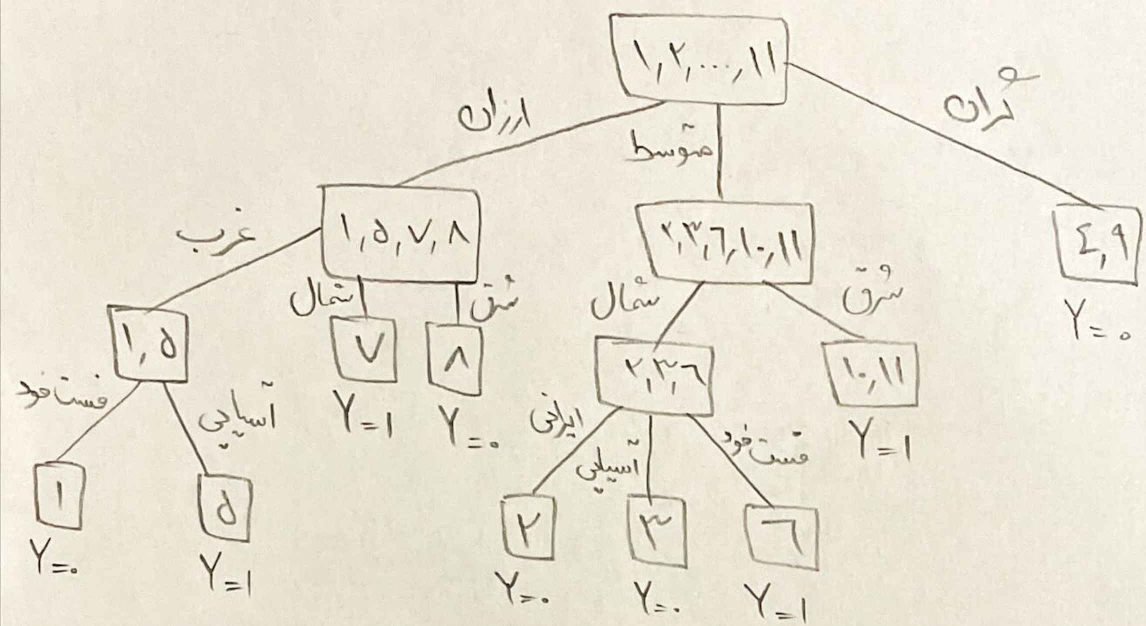
برای راس ۲،۳،۶ داریم:

$$H(Y|X_1) = -\left(\frac{1}{3}(1\log 1) + \frac{1}{3}(1\log 1) + \frac{1}{3}(1\log 1)\right) = 0$$

$$H(Y|X_4) = -\left(\frac{2}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) + \frac{1}{3}(1\log 1)\right) = \frac{2}{3}$$

$\arg\min_X H(Y|X) = X_1 \rightarrow$ نوع غذا

نهایتاً درخت به صورت زیر در می آید:



ب)

۱۲: ارزان $\rightarrow$ شمال $\rightarrow \hat{Y}=1$

۱۳: متوسط $\rightarrow$ شرق $\rightarrow \hat{Y}=1$

۱۴: ارزان $\rightarrow$ غرب $\rightarrow$? $\rightarrow$ با توجه به اینکه به راس پایانی ای نمی رسیم، در پایین ترین راسی که اکثریت وجود دارد، این رستوران را طبق آن اکثریت پیشبینی می کنیم. $\rightarrow Y=0$ (در راس ریشه)

۱۵: ارزان $\rightarrow$ شرق $\rightarrow \hat{Y}=0$

۱۶: ارزان $\rightarrow$ شمال $\rightarrow \hat{Y}=1$

پ)

$$\left.\begin{aligned}\hat{Y} &= [1,1,0,0,1]\\ Y &= [0,1,0,1,0]\end{aligned}\right\} \left.\begin{aligned}TP&=1\\ FP&=2\\ FN&=1\end{aligned}\right\} \begin{aligned}\text{Precision} &= \frac{TP}{TP+FP} = \frac{1}{1+2} = \frac{1}{3}\\ \text{Recall} &= \frac{TP}{TP+FN} = \frac{1}{1+1} = \frac{1}{2}\end{aligned}$$

$$F_1 = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} = 2\times\frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{3}+\frac{1}{2}} = \frac{\frac{1}{3}}{\frac{5}{6}} = \frac{6}{5}\times\frac{1}{3} = \frac{2}{5} = \boxed{0.4}$$

ج - الف)

$$P(\phi_j, t_j|\pi) = P(\phi_j|t_j,\pi)P(t_j|\pi) = P(\phi_j|t_j)P(t_j|\pi) = \prod_{i=1}^{K}(P(\phi_j|C_i)\pi_i)^{t_{ji}}$$

$$\Rightarrow P(\phi, t|\pi) = \prod_{j=1}^{N} P(\phi_j, t_j|\pi) = \prod_{j=1}^{N}\prod_{i=1}^{K}(P(\phi_j|C_i)\pi_i)^{t_{ji}}$$

$$\Rightarrow \log P(\phi, t|\pi) = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(\log P(\phi_j|C_i) + \log \pi_i)$$

$$\sum_{i=1}^{K} \pi_i = 1 \Rightarrow \sum_{i=1}^{K}\pi_i - 1 = 0 \Rightarrow \mathcal{L}(\pi, \lambda) = \log P(\phi, t|\pi) + \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right)$$

$$\frac{\partial \mathcal{L}(\pi,\lambda)}{\partial \pi_i} = \sum_{j=1}^{N} t_{ji}\frac{1}{\pi_i} + \lambda = \frac{1}{\pi_i}\sum_{j=1}^{N} t_{ji} + \lambda = \frac{1}{\pi_i}N_i + \lambda = 0 \Rightarrow \pi_i = \frac{-N_i}{\lambda}$$

$$\frac{\partial \mathcal{L}(\pi,\lambda)}{\partial \lambda} = \sum_{i=1}^{K}\pi_i - 1 = 0 \Rightarrow \sum_{i=1}^{K}\pi_i = 1 \Rightarrow \sum_{i=1}^{K}\frac{-N_i}{\lambda} = 1 \Rightarrow \frac{-1}{\lambda}\sum_{i=1}^{K} N_i = 1 \Rightarrow \frac{-1}{\lambda}N = 1$$

$$\Rightarrow \frac{-N}{\lambda} = 1 \Rightarrow \lambda = -N \Rightarrow \pi_i = \frac{-N_i}{-N} = \frac{N_i}{N} \quad \square$$

ب)
$$\mathcal{L}(\mu, \Sigma, \pi, \lambda) = \log P(\phi, t|\mu, \Sigma, \pi) + \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right) = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(\log \mathcal{N}(\phi_j|\mu_i, \Sigma) + \log \pi_i) + \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right) = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}\left(\frac{-M}{2}\log 2\pi - \frac{1}{2}\log\det(\Sigma) - \frac{1}{2}(\phi_j - \mu_i)^T\Sigma^{-1}(\phi_j - \mu_i) + \log\pi_i\right) + \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right)$$

$$\Rightarrow \frac{\partial \mathcal{L}(\mu,\Sigma,\pi,\lambda)}{\partial \mu_i} = \sum_{j=1}^{N} t_{ji}\left(\frac{-1}{2}(-2\Sigma^{-1}(\phi_j - \mu_i))\right) = \Sigma^{-1}\sum_{j=1}^{N} t_{ji}(\phi_j - \mu_i) = 0$$

$$\Rightarrow \sum_{j=1}^{N} t_{ji}(\phi_j - \mu_i) = 0 \Rightarrow \boxed{\mu_i = \frac{1}{N_i}\sum_{j=1}^{N} t_{ji}\phi_j}$$

$$\frac{\partial \mathcal{L}(\mu,\Sigma,\pi,\lambda)}{\partial \Sigma} = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}\left(\frac{-1}{2}(\Sigma^{-1} - \Sigma^{-1}(\phi_j - \mu_i)(\phi_j - \mu_i)^T\Sigma^{-1})\right) = \frac{-1}{2}\Sigma^{-1}\sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(I - (\phi_j - \mu_i)(\phi_j - \mu_i)^T\Sigma^{-1}) = 0 \Rightarrow \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(I - (\phi_j - \mu_i)(\phi_j - \mu_i)^T\Sigma^{-1}) = 0$$

$$\Rightarrow NI = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(\phi_j - \mu_i)(\phi_j - \mu_i)^T\Sigma^{-1} \Rightarrow N\Sigma = \sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(\phi_j - \mu_i)(\phi_j - \mu_i)^T$$

$$\Rightarrow \Sigma = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K} t_{ji}(\phi_j - \mu_i)(\phi_j - \mu_i)^T = \boxed{\left(\sum_{i=1}^{K}\sum_{j=1}^{N} t_{ji}(\phi_j - \mu_i)(\phi_j - \mu_i)^T\right)\frac{1}{N}}$$

$$Q_i := \frac{1}{N_i}\sum_{j=1}^{N} t_{ji}(\phi_j - \mu_i)(\phi_j - \mu_i)^T =$$ ماتریس کوواریانس ویژگی‌های داده‌های متعلق به دسته‌ی i ام

$$\Rightarrow \boxed{\Sigma = \frac{1}{N}\sum_{i=1}^{K} N_i Q_i} \longrightarrow$$ میانگین وزن‌دار ماتریس کوواریانس دسته‌های مختلف
