14.16

クラスがC個あるロジスティック回帰モデルをK個混合することを考える。

クラスC_cが出る事後確率は(4.104), (4.105)より

$$p(C_c|\phi,\theta)=y_c(\phi,\theta)=\frac{\exp(a_c)}{\sum_j \exp(a_j)},\quad a_c=w_c^T\phi$$

である。

t_nは1 of C 符号化した変数として、多クラスロジスティック回帰モデルのt_nの条件付分布は

$$p(t_n|\phi_n,\theta)=\prod_{c=1}^{C}p(C_c|\phi_n,\theta)^{t_{nc}}=\prod_{c=1}^{C}y_{nc}^{t_{nc}},\quad y_{nc}=\frac{\exp(a_{nc})}{\sum_j \exp(a_{nj})},\ a_{nc}=w_c^T\phi_n$$

となる。

K個の多クラスロジスティック回帰モデルの混合モデルのt_nの条件付分布は

$$p(t_n|\phi_n,\theta)=\sum_{k=1}^{K}\pi_k\prod_{c=1}^{C}y_{nkc}^{t_{nc}},\quad y_{nkc}=\frac{\exp(a_{nkc})}{\sum_j \exp(a_{nkj})},\ a_{nkc}=w_{kc}^T\phi_n \quad \cdots (14.45)'$$

となる。これより

$T=\begin{pmatrix}t_1^T\\ \vdots\\ t_N^T\end{pmatrix}$, $\Phi=\begin{pmatrix}\phi_1^T\\ \vdots\\ \phi_N^T\end{pmatrix}$とすると、尤度関数は

(t_n,ϕ_n)はiidなので

$$p(T|\Phi,\theta)=\prod_{n=1}^{N}p(t_n|\phi_n,\theta)=\prod_{n=1}^{N}\left(\sum_{k=1}^{K}\pi_k\prod_{c=1}^{C}y_{nkc}^{t_{nc}}\right)\cdots(14.46)'$$

となる。この尤度を最大化するのにEMアルゴリズムを使う。

9.2節に倣ってモデルkを決める潜在変数z_nを導入する。(9.10)(9.11)と同様に

$$p(z_n|\theta)=\prod_{k=1}^{K}\pi_k^{z_{nk}}$$

$$p(t_n|z_n,\phi_n,\theta)=\prod_{k=1}^{K}p(t_n|\phi_n,\theta)^{z_{nk}}=\prod_{k=1}^{K}\left(\prod_{c=1}^{C}y_{nkc}^{t_{nc}}\right)^{z_{nk}}$$

これよりt_n, z_nの同時分布は

$$p(t_n,z_n|\phi_n,\theta)=\prod_{k=1}^{K}\left(\pi_k\prod_{c=1}^{C}y_{nkc}^{t_{nc}}\right)^{z_{nk}}\cdots ①$$

となる。

$\mathbb{Z} = \begin{pmatrix} \mathbf{z}_1^T \\ \vdots \\ \mathbf{z}_N^T \end{pmatrix}$ とすると、完全データの尤度は

$(\mathbf{t}_n, \mathbf{z}_n, \boldsymbol{\phi}_n)$ はiid　①

$$p(\mathbb{T}, \mathbb{Z} | \boldsymbol{\Phi}, \boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{t}_n, \mathbf{z}_n | \boldsymbol{\phi}_n, \boldsymbol{\theta}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}} \right)^{z_{nk}} \quad \cdots (14.47)'$$

負担率は

①

$$\gamma_{nk} = \underset{p(\mathbb{Z}|\mathbb{T},\boldsymbol{\Phi},\boldsymbol{\theta}^{old})}{E}[z_{nk}] = p(\mathbf{z}_n = k | \mathbf{t}_n, \boldsymbol{\phi}_n, \boldsymbol{\theta}^{old}) = \frac{p(\mathbf{z}_n = k, \mathbf{t}_n | \boldsymbol{\phi}_n, \boldsymbol{\theta}^{old})}{\sum_{\mathbf{z}_n} p(\mathbf{z}_n, \mathbf{t}_n | \boldsymbol{\phi}_n, \boldsymbol{\theta}^{old})} = \frac{\pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}}}{\sum_{j=1}^{K} \pi_j \prod_{c=1}^{C} y_{njc}^{t_{nc}}} \quad \cdots (14.48)'$$

となる。Q関数は

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = \underset{p(\mathbb{Z}|\mathbb{T},\boldsymbol{\Phi},\boldsymbol{\theta})}{E}[\ln p(\mathbb{T}, \mathbb{Z} | \boldsymbol{\Phi}, \boldsymbol{\theta})] = \sum_{\mathbb{Z}} p(\mathbb{Z} | \mathbb{T}, \boldsymbol{\Phi}, \boldsymbol{\theta}) \ln p(\mathbb{T}, \mathbb{Z} | \boldsymbol{\Phi}, \boldsymbol{\theta})$$

(14.47)'

$$= \sum_{\mathbb{Z}} p(\mathbb{Z} | \mathbb{T}, \boldsymbol{\Phi}, \boldsymbol{\theta}) \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}} \right)^{z_{nk}}$$

$$= \sum_{\mathbb{Z}} p(\mathbb{Z} | \mathbb{T}, \boldsymbol{\Phi}, \boldsymbol{\theta}) \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \ln \pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{\mathbb{Z}} p(\mathbb{Z} | \mathbb{T}, \boldsymbol{\Phi}, \boldsymbol{\theta}) z_{nk} \ln \pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}}$$

$\gamma_{nk} = \underset{p(\mathbb{Z}|\mathbb{T},\boldsymbol{\Phi},\boldsymbol{\theta})}{E}[z_{nk}] = \sum_{\mathbb{Z}} p(\mathbb{Z}|\mathbb{T},\boldsymbol{\Phi},\boldsymbol{\theta}) z_{nk}$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \ln \pi_k \prod_{c=1}^{C} y_{nkc}^{t_{nc}}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left(\ln \pi_k + \sum_{c=1}^{C} t_{nc} \ln y_{nkc} \right) \quad \cdots (14.49)'$$

となる。

(π_kについて最適化)

Q関数のπ_kを含む項は本文と同じなのでQを最大化するπ_kは(14.50)と同じ

(w_k について最適化)

Q の w_{kc} についての勾配は

$$\nabla_{kc} Q = \frac{\partial}{\partial w_{kc}} Q = \frac{\partial}{\partial w_{kc}} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left(\ln \pi_k + \sum_{c=1}^{C} t_{nc} \ln y_{nkc} \right)$$

$$= \sum_{n=1}^{N} \gamma_{nk} t_{nc} \frac{\partial}{\partial w_{kc}} \ln y_{nkc}$$

$$= \sum_{n=1}^{N} \gamma_{nk} t_{nc} (1 - y_{nkc}) \phi_n \quad \cdots (14.51)'$$

$$\frac{\partial}{\partial w_{kc}} \ln y_{nkc} = \frac{\partial}{\partial w_{kc}} \ln \frac{\exp(a_{nkc})}{\sum_j \exp(a_{nkj})}$$

$$= \frac{\partial}{\partial w_{kc}} a_{nkc} - \frac{\partial}{\partial w_{kc}} \ln \sum_j \exp(a_{nkj})$$

$$\frac{\partial a_{nkc}}{\partial w_{kc}} = \frac{\partial}{\partial w_{kc}} w_{kc}^T \phi_n = \phi_n$$

$$= \phi_n - \frac{1}{\sum_j \exp(a_{nkj})} \frac{\partial}{\partial w_{kc}} \sum_j \exp(a_{nkj})$$

$$= \phi_n - \frac{1}{\sum_j \exp(a_{nkj})} \exp(a_{nkc}) \frac{\partial}{\partial w_{kc}} a_{nkc}$$

$$= \phi_n - y_{nkc} \phi_n = (1 - y_{nkc}) \phi_n$$

ヘッセ関数は

$$\mathbb{H}_{kc} = -\nabla_{kc} \nabla_{kc} Q$$

$$= -\frac{\partial}{\partial w_{kc}} \sum_{n=1}^{N} \overbrace{\gamma_{nk} t_{nc} (1 - y_{nkc}) \phi_n}^{(14.51)'}$$

$$= -\sum_{n=1}^{N} \left\{ \frac{\partial}{\partial w_{kc}} \gamma_{nk} t_{nc} (1 - y_{nkc}) \right\} \phi_n^T$$

$f = f(x)$, u は x に依らないとき

$$\frac{\partial f u}{\partial x} = \frac{\partial f}{\partial x} u^T \quad \text{(分母レイアウト)}$$

$$= \sum_{n=1}^{N} \gamma_{nk} t_{nc} \frac{\partial y_{nkc}}{\partial w_{kc}} \phi_n^T$$

$$= \sum_{n=1}^{N} \gamma_{nk} t_{nc} y_{nkc} (1 - y_{nkc}) \phi_n \phi_n^T \quad \cdots (14.52)'$$

$$\frac{\partial}{\partial w_{kc}} y_{nkc} = \frac{\partial}{\partial w_{kc}} \frac{\exp(a_{nkc})}{\sum_j \exp(a_{nkj})}$$

$$= \frac{\exp(a_{nkc}) \frac{\partial a_{nkc}}{\partial w_{kc}} \sum_j \exp(a_{nkj}) - \exp(a_{nkc}) \exp(a_{nkc}) \frac{\partial a_{nkc}}{\partial w_{kc}}}{\{\sum_j \exp(a_{nkj})\}^2}$$

$$= \frac{\exp(a_{nkc}) \phi_n \sum_j \exp(a_{nkj}) - \exp(a_{nkc}) \exp(a_{nkc}) \phi_n}{\{\sum_j \exp(a_{nkj})\}^2}$$

$$= y_{nkc} \phi_n - y_{nkc}^2 \phi_n = y_{nkc} (1 - y_{nkc}) \phi_n$$

となる。

勾配とヘッセ行列を用いて IRLS で w_{kc} を更新する。