機械学習とデータマイニングの基礎　原先生課題

28G23027 川原尚己

課題 1

$\boldsymbol{x}^{(n)} \in \mathbb{R}^d, y^{(n)} \in \{-1,1\}, \boldsymbol{\theta} \in \mathbb{R}^d$とする．

$$\phi(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N} \log\left(1 + exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top \boldsymbol{x}^{(n)}\right)\right)\right) \tag{1}$$

とする．任意のデータ$\left\{\boldsymbol{x}^{(n)} \in \mathbb{R}^d, y^{(n)} \in \{-1,1\}\right\}_{n=1}^{N}$及び任意の$\boldsymbol{\theta} \in \mathbb{R}^d$について$\|\nabla\phi(\boldsymbol{\theta})\| \leq R$となる$R$の最小値を求めよ．ただし$\left\|\boldsymbol{x}^{(n)}\right\| \leq K$とする．

課題 2

混合世紀分布の EM アルゴリズムについて，Mステップの最適なπ_k, μ_k, Σ_kが書きで表されることを示せ．

$$\hat{\pi}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)}}{\sum_{n=1}^{N}\sum_{k=1}^{K} \alpha_k^{(n)}} \tag{2}$$

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)} \boldsymbol{x}^{(n)}}{\sum_{n=1}^{N} \alpha_k^{(n)}} \tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)} \left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}\right)\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}\right)^\top}{\sum_{n=1}^{N} \alpha_k^{(n)}} \tag{4}$$

回答

課題 1

$\boldsymbol{x}^{(n)} = \left(x_1^{(n)}, \dots, x_d^{(n)}\right)^\top, \boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$とする．このとき，$\phi(\boldsymbol{\theta})$の$\theta_i$による偏微分は，

$$\frac{\partial\phi(\boldsymbol{\theta})}{\partial\theta_i} = -\frac{1}{N}\sum_{n=1}^{N} y^{(n)} x_i^{(n)} \frac{\exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top \boldsymbol{x}^{(\boldsymbol{n})}\right)\right)}{1 + \exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top \boldsymbol{x}^{(\boldsymbol{n})}\right)\right)} \tag{5}$$

となる．(5)式より，$\|\nabla\phi(\boldsymbol{\theta})\|$は以下のように表される．

$$\begin{aligned}
\|\nabla\phi(\boldsymbol{\theta})\| &= \sqrt{\sum_{i=1}^{d}\left(\frac{\partial\phi(\boldsymbol{\theta})}{\partial\theta_i}\right)^2} \\
&= \frac{1}{N}\sqrt{\sum_{i=1}^{d}\left(\sum_{n=1}^{N} y^{(n)} x_i^{(n)} \frac{\exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top \boldsymbol{x}^{(\boldsymbol{n})}\right)\right)}{1 + \exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top \boldsymbol{x}^{(\boldsymbol{n})}\right)\right)}\right)^2}
\end{aligned} \tag{6}$$

コーシーシュワルツの不等式及び，$\left(y^{(n)}\right)^2 = 1$より，(6)式は以下のように評価できる．

$$\|\nabla\phi(\boldsymbol{\theta})\| \leq \frac{1}{N}\sqrt{\sum_{i=1}^{d}\left(\sum_{n=1}^{N}\left(y^{(n)}x_i^{(n)}\right)^2\right)\left(\sum_{n=1}^{N}\left(\frac{\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}{1+\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}\right)^2\right)}$$

$$= \frac{1}{N}\sqrt{\sum_{i=1}^{d}\left(\sum_{n=1}^{N}\left(x_i^{(n)}\right)^2\right)\left(\sum_{n=1}^{N}\left(\frac{\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}{1+\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}\right)^2\right)} \tag{7}$$

また，任意のデータ $\left\{\boldsymbol{x}^{(n)} \in \mathbb{R}^d, y^{(n)} \in \{-1,1\}\right\}_{n=1}^{N}$ 及び任意の $\boldsymbol{\theta} \in \mathbb{R}^d$に対し $\exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})}\right)\right) > 0$であることから，

$$\left(\frac{\exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})}\right)\right)}{1+\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}\right)^2 < 1, \sup\left(\left(\frac{\exp\left(-y^{(n)}\left(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})}\right)\right)}{1+\exp\left(-y^{(n)}(\boldsymbol{\theta}^\top\boldsymbol{x}^{(\boldsymbol{n})})\right)}\right)^2\right) = 1$$

が成り立つ．これを(7)式に適用することで，

$$\|\nabla\phi(\boldsymbol{\theta})\| < \frac{1}{N}\sqrt{\sum_{i=1}^{d}\left(\sum_{n=1}^{N}\left(x_i^{(n)}\right)^2\right)N}$$

$$= \frac{1}{N}\sqrt{N\sum_{i=1}^{d}\sum_{n=1}^{N}\left(x_i^{(n)}\right)^2} \tag{8}$$

を得る．さらに，$\left\|\boldsymbol{x}^{(n)}\right\| \leq K$を用いることで(8)式は，

$$\|\nabla\phi(\boldsymbol{\theta})\| = \frac{1}{N}\sqrt{N\sum_{n=1}^{N}\sum_{i=1}^{d}\left(x_i^{(n)}\right)^2}$$

$$\leq \frac{1}{N}\sqrt{N\sum_{n=1}^{N}K^2}$$

$$= K$$

と変形できる．すなわち，$\|\nabla\phi(\boldsymbol{\theta})\|$の上限は$K$であるから，求める値は$K$に等しい．

課題２

関数$f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$を以下のように定義する．

$$f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{n=1}^{N} \sum_{k=1}^{K} \alpha_k^{(n)} \left(\log \pi_k + \log p\left(\boldsymbol{x}^{(n)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right). \tag{9}$$

ただし，$\pi_{\mathrm{k}} \geq 0, \sum_{k=1}^{K} \pi_k = 1$であり，$p\left(\boldsymbol{x}^{(n)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$は以下のように表される．

$$p\left(\boldsymbol{x}^{(n)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right) \tag{10}$$

ラグランジュの未定定数法により，関数$f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$を最大化する$(\hat{\pi}_k, \widehat{\boldsymbol{\mu}}_{\mathrm{k}}, \widehat{\boldsymbol{\Sigma}}_k)$を求める．式(12)で表される制約の下において，以下の関数$L(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \lambda)$を考える．

$$L(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \lambda) = f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) - \lambda \left(-1 + \sum_{k=1}^{K} \pi_k \right) \tag{11}$$

$$\sum_{k=1}^{K} \pi_k = 1 \tag{12}$$

すると，以下の連立方程式の解が$(\hat{\pi}_k, \widehat{\boldsymbol{\mu}}_{\mathrm{k}}, \widehat{\boldsymbol{\Sigma}}_k)$となる．

$$\frac{\partial L}{\partial \pi_k} = 0 \tag{13}$$

$$\frac{\partial L}{\partial \boldsymbol{\mu}_k} = \mathbf{0} \tag{14}$$

$$\frac{\partial L}{\partial \boldsymbol{\Sigma}} = \boldsymbol{O} \tag{15}$$

まず，(13)式を解く．

$$\begin{aligned} \frac{\partial L}{\partial \pi_k} &= \frac{\partial}{\partial \pi_k} f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) - \frac{\partial}{\partial \pi_k} \left(\lambda \left(-1 + \sum_{k=1}^{K} \pi_k \right) \right) \\ &= \frac{1}{\pi_k} \sum_{n=1}^{N} \alpha_k^{(n)} - \lambda = 0 \end{aligned}$$

$$\therefore \pi_k = \frac{1}{\lambda} \sum_{n=1}^{N} \alpha_k^{(n)} \tag{16}$$

(12),(16)式よりλを求めると，以下を得る．

$$\lambda = \sum_{k=1}^{K} \sum_{n=1}^{N} \alpha_k^{(n)} \tag{17}$$

(16),(17)式より(2)式を得る．

$$\hat{\pi}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)}}{\sum_{n=1}^{N} \sum_{k=1}^{K} \alpha_k^{(n)}} \tag{2}$$

次に，(14)式を解く．(9),(10),(11)式及び$\boldsymbol{\Sigma_k}$は対称行列であることより，(14)式は以下のよ

うに表される．ただし，c=$\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\boldsymbol{\Sigma}}}$とする．

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\mu}_k} &= \frac{\partial}{\partial \boldsymbol{\mu}_k} f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \frac{\partial}{\partial \boldsymbol{\mu}_k} \log p\left(\boldsymbol{x}^{(n)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \frac{\partial}{\partial \boldsymbol{\mu}_k} \left(\log c - \frac{1}{2}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)^{\top} \boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)\right) \\
&= \frac{1}{2}\sum_{n=1}^{N} \alpha_k^{(n)} \left(\boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right) + \left(\boldsymbol{\Sigma}_k^{-1}\right)^{\top}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)\right) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right) \\
&= \boldsymbol{\Sigma}_k^{-1} \sum_{n=1}^{N} \alpha_k^{(n)} \boldsymbol{x}^{(n)} - \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \sum_{n=1}^{N} \alpha_k^{(n)} = \boldsymbol{0}
\end{aligned}
$$

これより,(3)式が得られる．

$$
\widehat{\boldsymbol{\mu}}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)} \boldsymbol{x}^{(n)}}{\sum_{n=1}^{N} \alpha_k^{(n)}} \tag{3}
$$

最後に(15)式を解く．(9),(10),(11)式及び$\boldsymbol{\Sigma_k}$は対称行列であることより，(15)式は以下のように表される．

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\Sigma}_k} &= \frac{\partial}{\partial \boldsymbol{\Sigma}_k} f(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \log p\left(\boldsymbol{x}^{(n)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}_k) - \frac{1}{2}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)^{\top} \boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)\right) \\
&= \sum_{n=1}^{N} \alpha_k^{(n)} \left(-\frac{1}{2}(\boldsymbol{\Sigma}_k^{-1})^{\top} + \frac{1}{2}(\boldsymbol{\Sigma}_k^{-1})^{\top}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)^{\top}(\boldsymbol{\Sigma}_k^{-1})^{\top}\right) \\
&= -\frac{\boldsymbol{\Sigma}_k^{-1}}{2}\sum_{n=1}^{N} \alpha_k^{(n)} + \frac{\boldsymbol{\Sigma}_k^{-1}}{2}\left(\sum_{n=1}^{N} \alpha_k^{(n)}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_k\right)^{\top}\right)\boldsymbol{\Sigma}_k^{-1} = \boldsymbol{O}
\end{aligned}
$$

これより，(4)式が得られる．

$$
\widehat{\boldsymbol{\Sigma}}_k = \frac{\sum_{n=1}^{N} \alpha_k^{(n)}\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}\right)\left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}\right)^{\top}}{\sum_{n=1}^{N} \alpha_k^{(n)}} \tag{4}
$$