宿題 1

二乗ヒンジ損失に対する適応正則化分類の $\boldsymbol{\mu}$ の解が次式で与えられることを示す。

$$\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}} + \frac{y\max(0,\ 1-\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\phi(\boldsymbol{x})y)}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x}) \tag{1}$$

損失は，

$$\begin{aligned} J(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) &= \left(\max(0,\ 1-\boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})y)\right)^2 + \phi(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\Sigma}\phi(\boldsymbol{x}) \\ &\quad + \gamma\left\{\log\frac{\det(\tilde{\boldsymbol{\Sigma}})}{\det\{\boldsymbol{\Sigma}\}} + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\right) + (\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}}) - d\right\} \end{aligned} \tag{2}$$

と表される。f の x による劣微分を $\partial_x f$ などと表すと，この損失の $\boldsymbol{\mu}$ による偏微分は，

$$\frac{\partial J}{\partial \boldsymbol{\mu}} = 2\cdot\max(0,\ 1-\boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})y)\cdot\partial_{\boldsymbol{\mu}}\max(0,\ 1-\boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})y) + \gamma\left(2\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} - 2\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}\right) \tag{3}$$

となる。$z = \boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})y$ とおくと，

$$\frac{\partial z}{\partial \boldsymbol{\mu}} = y\phi(\boldsymbol{x}) \tag{4}$$

であり，また，

$$\max(0,\ 1-z) = \begin{cases} 1-z & (z<1) \\ 0 & (z\geq 1) \end{cases} \tag{5}$$

$$\partial_{\boldsymbol{\mu}}\max(0,\ 1-z) = \begin{cases} -y\phi(\boldsymbol{x}) & (z<1) \\ \mathbf{0} & (z>1) \\ [-y\phi(\boldsymbol{x}),\ \mathbf{0}] & (z=1) \end{cases} \tag{6}$$

$$\tag{7}$$

である。

$z>1$ のとき，式 (3) – (6) から，

$$\frac{\partial J}{\partial \boldsymbol{\mu}} = 2\gamma\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} - \tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}\right) = \mathbf{0} \tag{8}$$

よって，

$$\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}} \tag{9}$$

$z<1$ のとき，式 (3) – (6) から，

$$\frac{\partial J}{\partial \boldsymbol{\mu}} = 2(1-\boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})y)(-y\phi(\boldsymbol{x})) + 2\gamma\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} - \tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}\right) = \mathbf{0} \tag{10}$$

$\boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{x})$ はスカラーなので転置しても値は変わらないことから，

$$\left(-y\phi(\boldsymbol{x})\right)(1-y\phi(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\mu})+\gamma\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}-\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}\right)=\mathbf{0} \tag{11}$$

$$y^2\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\mu}-y\phi(\boldsymbol{x})+\gamma\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}-\gamma\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}=\mathbf{0} \tag{12}$$

$$\left(\gamma\tilde{\boldsymbol{\Sigma}}^{-1}+y^2\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\right)\boldsymbol{\mu}=\gamma\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}+y\phi(\boldsymbol{x}) \tag{13}$$

$$\left(\tilde{\boldsymbol{\Sigma}}^{-1}+\frac{y^2}{\gamma}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\right)\boldsymbol{\mu}=\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}+\frac{y}{\gamma}\phi(\boldsymbol{x}) \tag{14}$$

いま，$y\in\{-1,\ 1\}$ を考えるから，$y^2=1$ となる。また，ShermanMorrison-Woodbury 公式から，

$$\left(\tilde{\boldsymbol{\Sigma}}^{-1}+\frac{1}{\gamma}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\right)^{-1}=\left(\tilde{\boldsymbol{\Sigma}}-\frac{\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\right) \tag{15}$$

が成立する。よって，

$$\begin{aligned}
\hat{\boldsymbol{\mu}}&=\left(\tilde{\boldsymbol{\Sigma}}^{-1}+\frac{1}{\gamma}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\right)^{-1}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}+\frac{y}{\gamma}\phi(\boldsymbol{x})\right) &(16)\\
&=\left(\tilde{\boldsymbol{\Sigma}}-\frac{\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\right)\left(\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}+\frac{y}{\gamma}\phi(\boldsymbol{x})\right) &(17)\\
&=\tilde{\boldsymbol{\mu}}-\frac{\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\tilde{\boldsymbol{\mu}}+\frac{y}{\gamma}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})\left(1-\frac{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\right) &(18)\\
&=\tilde{\boldsymbol{\mu}}-\frac{\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^{\mathrm{T}}}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\tilde{\boldsymbol{\mu}}+\frac{y\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma} &(19)\\
&=\tilde{\boldsymbol{\mu}}+\frac{\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\left(y-\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\mu}}\right) &(20)
\end{aligned}$$

$y\in\{-1,\ 1\}$ から，

$$\begin{aligned}
\hat{\boldsymbol{\mu}}&=\tilde{\boldsymbol{\mu}}+\frac{y\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\left(1-\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\mu}}y\right) &(21)\\
&=\tilde{\boldsymbol{\mu}}+\frac{y\left(1-\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\phi(\boldsymbol{x})y\right)}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x}) &(22)
\end{aligned}$$

$z=1$ のときの劣勾配を $\mathbf{0}$ としてしまえば，式 (9)，(22) をまとめて，

$$\hat{\boldsymbol{\mu}}=\tilde{\boldsymbol{\mu}}+\frac{y\max(0,\ 1-\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\phi(\boldsymbol{x})y)}{\phi(\boldsymbol{x})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x})+\gamma}\tilde{\boldsymbol{\Sigma}}\phi(\boldsymbol{x}) \tag{23}$$