宿題 1

線形モデル

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{\theta} \tag{1}$$

に対して，重み付き最小二乗法を考える。ここで，

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \cdots & \theta_b \end{bmatrix}^{\mathrm{T}} \tag{2}$$

$$\boldsymbol{\phi}(\boldsymbol{x}) = \begin{bmatrix} \phi_1(\boldsymbol{x}) & \cdots & \phi_b(\boldsymbol{x}) \end{bmatrix}^{\mathrm{T}} \tag{3}$$

である。このとき，損失関数は以下のように表される。

$$\begin{aligned}
J(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{i=1}^{n} \tilde{w}_i (f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)^2 && (4)\\
&= \frac{1}{2} \sum_{i=1}^{n} \tilde{w}_i \left(\boldsymbol{\phi}(\boldsymbol{x}_i)^{\mathrm{T}} \boldsymbol{\theta} - y_i\right)^2 && (5)\\
&= \frac{1}{2} \sum_{i=1}^{n} \left(\boldsymbol{\phi}(\boldsymbol{x}_i)^{\mathrm{T}} \boldsymbol{\theta} - y_i\right) \tilde{w}_i \left(\boldsymbol{\phi}(\boldsymbol{x}_i)^{\mathrm{T}} \boldsymbol{\theta} - y_i\right) && (6)\\
&= \frac{1}{2} (\boldsymbol{\Phi\theta} - \boldsymbol{y})^{\mathrm{T}} \tilde{\boldsymbol{W}} (\boldsymbol{\Phi\theta} - \boldsymbol{y}) && (7)\\
&= \frac{1}{2} (\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} - \boldsymbol{y}^{\mathrm{T}}) \tilde{\boldsymbol{W}} (\boldsymbol{\Phi\theta} - \boldsymbol{y}) && (8)\\
&= \frac{1}{2} \left\{ \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi\theta} - \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi\theta} + \boldsymbol{y}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} \right\} && (9)
\end{aligned}$$

ここで，

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_1(\boldsymbol{x}_1) & \cdots & \phi_b(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{x}_n) & \cdots & \phi_b(\boldsymbol{x}_n) \end{bmatrix} \tag{10}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^{\mathrm{T}} \tag{11}$$

$$\tilde{\boldsymbol{W}} = \mathrm{diag}\{\tilde{w}_1, \cdots, \tilde{w}_n\} \tag{12}$$

である。また，$\boldsymbol{y}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi\theta}$ はスカラーであるから転置しても値は変わらず，$\tilde{\boldsymbol{W}}$ は対称行列であるので転置しても $\tilde{\boldsymbol{W}}$ のままであることに注意すると，結局，損失関数は次のように表される。

$$J(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi\theta} - \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} + \frac{1}{2} \boldsymbol{y}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} \tag{13}$$

これを $\boldsymbol{\theta}$ で偏微分すると，

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi\theta} - \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} \tag{14}$$

となる。これを $\boldsymbol{0}$ とするような $\boldsymbol{\theta}$ が J に最小を与えるので，$\boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi}$ の逆行列が存在することを仮定すると，

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \tilde{\boldsymbol{W}} \boldsymbol{y} \tag{15}$$

となる。