宿題 1

線形モデル $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$ を用いた L2 正則化回帰に，一つ抜き交差確認法を適用したときの二乗誤差 L が次式のように表されることを示す。

$$L = \frac{1}{n} ||\tilde{\boldsymbol{H}}^{-1} \boldsymbol{H} \boldsymbol{y}||^2 \tag{1}$$

$$\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}^{\mathrm{T}} \tag{2}$$

$$\tilde{\boldsymbol{H}} : \boldsymbol{H} \text{ と同じ対角成分を持ち，非対角成分は } 0 \tag{3}$$

訓練誤差 L_{train} は，

$$L_{\mathrm{train}} = ||\boldsymbol{\Phi\theta} - \boldsymbol{y}||^2 + \frac{\lambda}{2}||\boldsymbol{\theta}||^2 \tag{4}$$

これを最小にする $\boldsymbol{\theta}$ は，上式の $\boldsymbol{\theta}$ による偏微分が 0 である条件から得られて，

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} \tag{5}$$

となる。これより，標本 $(\boldsymbol{x}_i, y_i)$ を除いた標本群を用いて学習したときに得られるパラメータ $\boldsymbol{\theta}$ は，

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_i &= (\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{y}_i && (6)\\
&= (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda \boldsymbol{I} - \boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}})^{-1}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (7)\\
&= (\boldsymbol{U} - \boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}})^{-1}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (8)
\end{aligned}$$

と表される。ここで，

$$\boldsymbol{U} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda \boldsymbol{I} \tag{9}$$

とおいた。また，ShermanMorrison-Woodbury 公式を用いると，

$$(\boldsymbol{U} - \boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}})^{-1} = \boldsymbol{U}^{-1} + \frac{\boldsymbol{U}^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}}{1 - \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_i} \tag{10}$$

となることから，結局，

$$\hat{\boldsymbol{\theta}}_i = \left(\boldsymbol{U}^{-1} + \frac{\boldsymbol{U}^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}}{1 - \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_i}\right)(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) \tag{11}$$

これより，$\alpha_i = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_i$ とおくと，y_i の予測値 $\hat{y}_i$ は，

$$\begin{aligned}
\hat{y}_i &= \boldsymbol{\phi}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}}_i && (12)\\
&= \boldsymbol{\phi}_i^{\mathrm{T}}\left(\boldsymbol{U}^{-1} + \frac{\boldsymbol{U}^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}}{1 - \alpha_i}\right)(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (13)\\
&= \frac{1}{1-\alpha_i}\left(\boldsymbol{\phi}_i^{\mathrm{T}}(1-\alpha_i)\boldsymbol{U}^{-1} + \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\right)(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (14)\\
&= \frac{1}{1-\alpha_i}\left(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1} - \alpha_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1} + \alpha_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\right)(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (15)\\
&= \frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}}{1-\alpha_i}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\phi}_i y_i) && (16)\\
&= \frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \alpha_i y_i}{1-\alpha_i} && (17)
\end{aligned}$$

よって，テスト誤差 L は，

$$
\begin{aligned}
L &= \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 && (18)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - \alpha_i y_i}{1-\alpha_i} - y_i\right) && (19)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - y_i}{1-\alpha_i}\right) && (20)
\end{aligned}
$$

となる。

　また，$\boldsymbol{H}$ と $\tilde{\boldsymbol{H}}$ について，

$$
\begin{aligned}
\boldsymbol{H} &= \boldsymbol{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda\boldsymbol{I})^{-1}\boldsymbol{\Phi}^{\mathrm{T}} && (21)\\
&= \boldsymbol{I} - \boldsymbol{\Phi}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}} && (22)\\
&= \boldsymbol{I} - \begin{bmatrix} \boldsymbol{\phi}_1^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_1^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_n^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_n^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_n \end{bmatrix} && (23)
\end{aligned}
$$

より，$\boldsymbol{H}$（及び $\tilde{\boldsymbol{H}}$）の i 番目対角成分は $1-\alpha_i$，$\tilde{\boldsymbol{H}}^{-1}$ の i 番目対角成分は $1/(1-\alpha_i)$ となる。また，

$$
\begin{aligned}
\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - y_i &= \sum_{j=1}^{n}\left(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\phi}_j y_j\right) - y_i && (24)\\
&= -(\boldsymbol{H}\boldsymbol{y})_i && (25)
\end{aligned}
$$

である。以上より，テスト誤差は，

$$
\begin{aligned}
L &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{U}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} - y_i}{1-\alpha_i}\right) && (26)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{-(\boldsymbol{H}\boldsymbol{y})_i}{1-\alpha_i}\right)^2 && (27)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{H}}_i^{-1}(\boldsymbol{H}\boldsymbol{y})_i\right)^2 && (28)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\boldsymbol{y}\right)_i^2 && (29)\\
&= \frac{1}{n}||\tilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\boldsymbol{y}||^2 && (30)
\end{aligned}
$$

となる。