先端データ解析論　2

48-196406　電子情報学専攻

磯川　弘基

HW.1

ガウスカーネルモデルに対する l_2 -正則化を用いた最小二乗回帰の交差確認法を実装し，正則化パラメータとガウス幅を決定せよ

実装コードは別ファイル参照.
一つ抜きの場合，各パラメータを 0.1~1.0 の範囲で 0.1 ステップで探索したところ，$h = 0.3, l = 0.1$ で誤差が最小となった.. このパラメータを用いてフィッティングを行った結果を図 1 に示す.

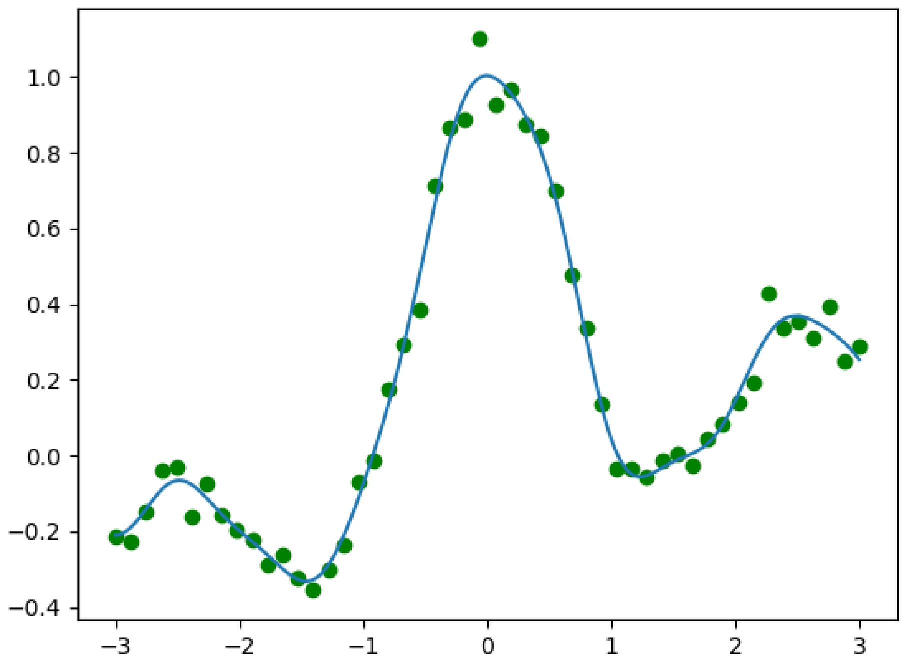

図 1: 一つ抜き交差確認法により得られた正則化パラメータ，及びガウス幅を用いて最小二乗回帰を行った結果.

HW.2

線形モデをル l_2-正則化回帰に対する一つ抜き交差確認による二乗誤差は，次式で解析的に計算できることを示せ．

$$\frac{1}{n}\|\tilde{H}^{-1}Hy\|^2 \tag{1}$$

i を除いた際の線形モデルを l_2-正則化回帰を行う．

$$\begin{aligned}\hat{\theta}_i &= \arg\min_{\theta}[\tfrac{1}{2}\textstyle\sum_{j=1(without j=i)}^{n}(f_{\theta_i}(x_j)-y_j)^2+\tfrac{\lambda}{2}\|\theta_i\|^2]\\ &= \arg\min_{\theta_i}[\tfrac{1}{2}\|\Phi_i\theta_i-\boldsymbol{y_i}\|^2+\tfrac{\lambda}{2}\|\theta_i\|^2]\end{aligned}$$

$[\cdot]$ 部分を θ_i について微分する．

$$\nabla_{\theta_i}[\cdot] = \quad \Phi_i^{\mathsf{T}}\Phi_i\theta_i-\Phi_i^{\mathsf{T}}\boldsymbol{y_i}+\lambda\theta_i$$

したがって，(x_i, y_i) を抜いて学習したパラメータ $\hat{\theta}_i$ は，以下を満たす．

$$\Phi_i^{\mathsf{T}}\Phi_i\hat{\theta}_i-\Phi_i^{\mathsf{T}}\boldsymbol{y_i}+\lambda\boldsymbol{\hat{\theta}_i}=0$$

これを，$\hat{\theta}_i$ について整理すると，以下の式が得られる．

$$\hat{\theta}_i=(\Phi_i^{\mathsf{T}}\Phi_i+\lambda I)^{-1}\Phi_i^{\mathsf{T}}\boldsymbol{y_i}$$

ここで，Φ_i と Φ の関係について，以下の関係が成り立つ．

$$\begin{aligned}\Phi_i^{\mathsf{T}}\Phi_i &= \Phi^{\mathsf{T}}\Phi-\phi_i^{\mathsf{T}}\phi_i\\ \Phi_i^{\mathsf{T}}\boldsymbol{y_i} &= \Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i\end{aligned}$$

これらを用いて，$\hat{\theta}_i$ を書き換える．

$$\begin{aligned}\hat{\theta}_i &= (\Phi^{\mathsf{T}}\Phi-\phi_i^{\mathsf{T}}\phi_i+\lambda I)^{-1}(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i)\\ &= (U-\phi_i^{\mathsf{T}}\phi_i)^{-1}(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i)\\ &= (U^{-1}+\tfrac{U^{-1}\phi_i\phi_i^{\mathsf{T}}U^{-1}}{1-\phi_i^{\mathsf{T}}U^{-1}\phi_i})(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i) \qquad (\because \text{逆行列の公式})\end{aligned}$$

ここで，$\hat{\theta}_i$ を用いて得られる，x_i に対する y_i の予測値 $y_{i(pred)}$ を考える．

$$\begin{aligned}y_{i(pred)} &= \phi_i^{\mathsf{T}}\hat{\theta}_i\\ &= \phi_i^{\mathsf{T}}(U^{-1}+\tfrac{U^{-1}\phi_i\phi_i^{\mathsf{T}}U^{-1}}{1-\phi_i^{\mathsf{T}}U^{-1}\phi_i})(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i)\\ &= \tfrac{\phi_i^{\mathsf{T}}U^{-1}-\phi_i^{\mathsf{T}}U^{-1}\phi_i\phi_i^{\mathsf{T}}U^{-1}+\phi_i^{\mathsf{T}}U^{-1}\phi_i\phi_i^{\mathsf{T}}U^{-1}}{1-\phi_i^{\mathsf{T}}U^{-1}\phi_i}(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i)\\ &= \tfrac{\phi_i^{\mathsf{T}}U^{-1}(\Phi^{\mathsf{T}}\boldsymbol{y}-y_i\phi_i)}{1-\phi_i^{\mathsf{T}}U^{-1}\phi_i}\end{aligned}$$

したがって，すべての i に関する平均二乗誤差の平均を取ると，

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\|y_{i(pred)} - y_i\right\|^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\phi_i^{\mathsf{T}}U^{-1}\Phi^{\mathsf{T}}y - y_i}{1-\phi_i^{T}U^{-1}\phi_i}\right)^2 \\
&= \frac{1}{n}\left\|\tilde{H}^{-1}Hy\right\|^2
\end{aligned}
$$

が得られる．