

宿題 1

ガウスカーネルモデルに対して，最小二乗確率的分類を実装する。

$$q(y|\mathbf{x}; \boldsymbol{\theta}^{(y)}) = \sum_{j:y_j=y} \theta_j^{(y)} K(\mathbf{x}, \mathbf{x}_j) \quad (1)$$

$$K(\mathbf{x}, \mathbf{c}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^2}{2h^2}\right) \quad (2)$$

二乗誤差は次のようになる。

$$J_y(\boldsymbol{\theta}_y) = \frac{1}{2} \int (q(y|\mathbf{x}; \boldsymbol{\theta}_y) - p(y|\mathbf{x}))^2 p(\mathbf{x}) d\mathbf{x} \quad (3)$$

$$= \frac{1}{2} \int (q(y|\mathbf{x}; \boldsymbol{\theta}_y))^2 p(\mathbf{x}) d\mathbf{x} - \int q(y|\mathbf{x}; \boldsymbol{\theta}_y) p(y|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \int p(y|\mathbf{x})^2 p(\mathbf{x}) d\mathbf{x} \quad (4)$$

これを標本近似することを考える。第一項は，

$$\frac{1}{2} \int (q(y|\mathbf{x}; \boldsymbol{\theta}_y))^2 p(\mathbf{x}) d\mathbf{x} \rightarrow \frac{1}{2} \cdot \frac{1}{n} \sum_{i=1}^n q(y|\mathbf{x}_i; \boldsymbol{\theta}_y)^2 \quad (5)$$

第二項は，

$$\int q(y|\mathbf{x}; \boldsymbol{\theta}_y) p(y|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \int q(y|\mathbf{x}; \boldsymbol{\theta}_y) p(y) p(\mathbf{x}|y) d\mathbf{x} \quad (6)$$

$$= p(y) \int q(y|\mathbf{x}; \boldsymbol{\theta}_y) p(\mathbf{x}|y) d\mathbf{x} \quad (7)$$

$$\rightarrow \frac{n_y}{n} \cdot \frac{1}{n_y} \sum_{i:y_i=y} q(y|\mathbf{x}_i; \boldsymbol{\theta}_y) \quad (8)$$

$$= \frac{1}{n} \sum_{i:y_i=y} q(y|\mathbf{x}_i; \boldsymbol{\theta}_y) \quad (9)$$

第三項は定数なので無視できる。以上に L2 正則化項を加えると，二乗誤差は，

$$\hat{f}_y(\boldsymbol{\theta}_y) = \frac{1}{2n} \sum_{i=1}^n q(y|\mathbf{x}_i; \boldsymbol{\theta}_y)^2 - \frac{1}{n} \sum_{i:y_i=y} q(y|\mathbf{x}_i; \boldsymbol{\theta}_y) + \frac{\lambda}{2n} \|\boldsymbol{\theta}_y\|^2 \quad (10)$$

$$= \frac{1}{n} \left(\frac{1}{2} \sum_{i=1}^n q(y|\mathbf{x}_i; \boldsymbol{\theta}_y)^2 - \sum_{i:y_i=y} q(y|\mathbf{x}_i; \boldsymbol{\theta}_y) + \frac{\lambda}{2} \|\boldsymbol{\theta}_y\|^2 \right) \quad (11)$$

と表せる。

いま，ガウスカーネルモデルを考えると，カーネル行列

$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_n, \mathbf{x}_1) & \cdots & K(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \quad (12)$$

を用いれば，二乗誤差は次のように表せる。

$$\hat{f}_y(\boldsymbol{\theta}_y) = \frac{1}{n} \left\{ \frac{1}{2} \boldsymbol{\theta}_y^T \mathbf{K}^2 \boldsymbol{\theta}_y - \boldsymbol{\theta}_y^T \mathbf{K} \boldsymbol{\pi}_y + \frac{\lambda}{2} \|\boldsymbol{\theta}_y\|^2 \right\} \quad (13)$$

ここで、 \mathbf{K} は対称行列であることを利用した。また、

$$\pi_{y,i} = \begin{cases} 1 & (y_i = y) \\ 0 & (y_i \neq y) \end{cases} \quad (14)$$

である。 $\boldsymbol{\theta}_y$ による偏微分は、

$$\frac{\partial \hat{f}_y}{\partial \boldsymbol{\theta}_y} = \frac{1}{n} \{ \mathbf{K}^2 \boldsymbol{\theta}_y - \mathbf{K} \boldsymbol{\pi}_y + \lambda \boldsymbol{\theta}_y \} \quad (15)$$

$$= \frac{1}{n} \{ (\mathbf{K} + \lambda \mathbf{I}) \boldsymbol{\theta}_y - \mathbf{K} \boldsymbol{\pi}_y \} \quad (16)$$

これを $\mathbf{0}$ にするような $\boldsymbol{\theta}_y$ は、

$$\hat{\boldsymbol{\theta}}_y = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{K} \boldsymbol{\pi}_y \quad (17)$$

となる。プログラムでは、この式を用いてパラメータを求める。

バンド幅 $h = 2.0$, L2 正則化パラメータ $\lambda = 0.0001$ に対する結果を、以下の図 1 に示す。なお、プログラムは??ページの Listing ??に示した。

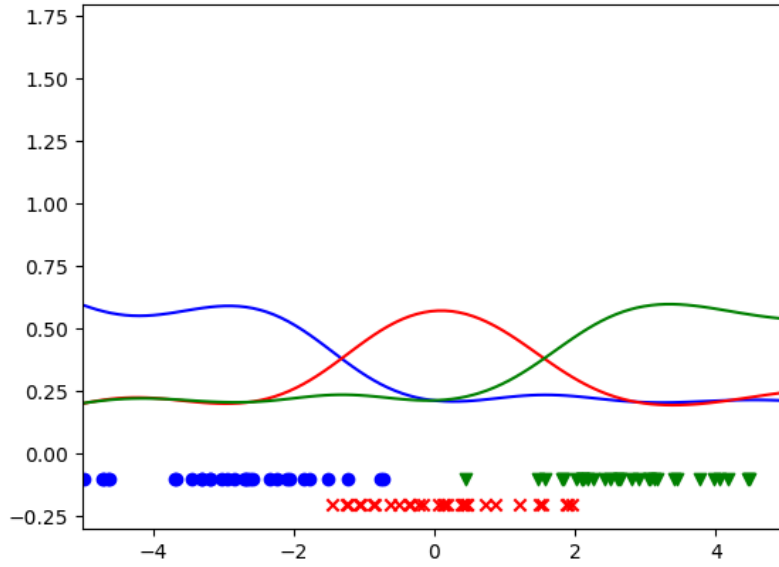


図 1: 結果