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。

$$q\left(y \mid \boldsymbol{x};\, \boldsymbol{\theta}^{(y)}\right) = \sum_{j:y_j=y} \theta_j^{(y)} K(\boldsymbol{x},\, \boldsymbol{x}_j) \tag{1}$$

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

二乗誤差は次のようになる。

$$\begin{aligned} J_y(\boldsymbol{\theta}_y) &= \frac{1}{2}\int (q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y) - p(y \mid \boldsymbol{x}))^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} & (3)\\ &= \frac{1}{2}\int (q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y))^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \int q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y)p(y|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \frac{1}{2}\int p(y|\boldsymbol{x})^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} & (4) \end{aligned}$$

これを標本近似することを考える。第一項は，

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

第二項は，

$$\begin{aligned} \int q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y)p(y|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \int q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y)p(y)p(\boldsymbol{x}|y)\mathrm{d}\boldsymbol{x} & (6)\\ &= p(y)\int q(y|\boldsymbol{x};\, \boldsymbol{\theta}_y)p(\boldsymbol{x}|y)\mathrm{d}\boldsymbol{x} & (7)\\ &\to \frac{n_y}{n}\cdot\frac{1}{n_y}\sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y) & (8)\\ &= \frac{1}{n}\sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y) & (9) \end{aligned}$$

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，

$$\begin{aligned} \hat{J}_y(\boldsymbol{\theta}_y) &= \frac{1}{2n}\sum_{i=1}^{n} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y)^2 - \frac{1}{n}\sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y) + \frac{\lambda}{2n}||\boldsymbol{\theta}_y||^2 & (10)\\ &= \frac{1}{n}\left(\frac{1}{2}\sum_{i=1}^{n} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y)^2 - \sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\, \boldsymbol{\theta}_y) + \frac{\lambda}{2}||\boldsymbol{\theta}_y||^2\right) & (11) \end{aligned}$$

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

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

を用いれば，二乗誤差は次のように表せる。

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

ここで，$\boldsymbol{K}$ は対称行列であることを利用した。また，

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

である。$\boldsymbol{\theta}_y$ による偏微分は，

$$\frac{\partial \hat{J}_y}{\partial \boldsymbol{\theta}_y} = \frac{1}{n}\left\{\boldsymbol{K}^2\boldsymbol{\theta}_y - \boldsymbol{K}\boldsymbol{\pi}_y + \lambda\boldsymbol{\theta}_y\right\} \tag{15}$$

$$= \frac{1}{n}\left\{\left(\boldsymbol{K}^2 + \lambda\boldsymbol{I}\right)\boldsymbol{\theta}_y - \boldsymbol{K}\boldsymbol{\pi}_y\right\} \tag{16}$$

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

$$\hat{\boldsymbol{\theta}}_y = \left(\boldsymbol{K}^2 + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{K}\boldsymbol{\pi}_y \tag{17}$$

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。

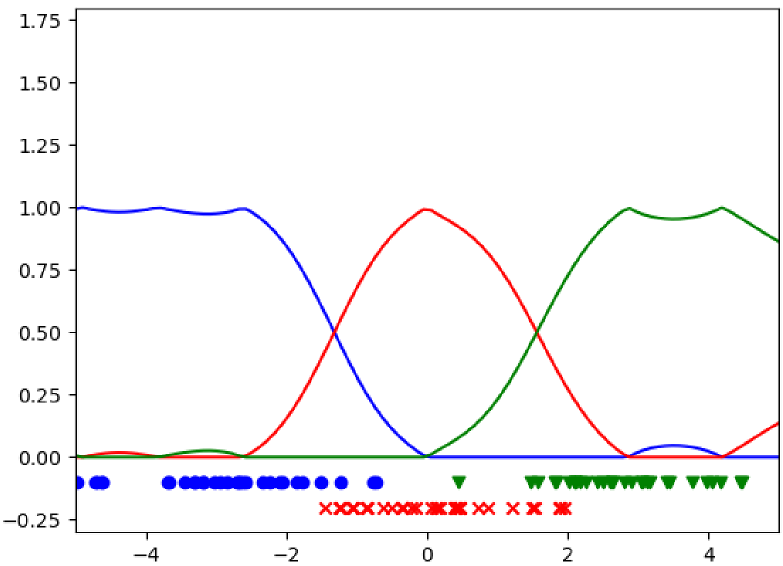

図 1: 結果