

# 宿題 1

ガウスカーネルモデル

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{j=1}^{n+n'} \theta_j K(\mathbf{x}, \mathbf{x}_j) = \mathbf{K} \boldsymbol{\theta} \quad (1)$$

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

$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_{n+n'}, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_{n+n'}, \mathbf{x}_1) & \cdots & K(\mathbf{x}_{n+n'}, \mathbf{x}_{n+n'}) \end{bmatrix} \quad (3)$$

に対してラプラス正則化最小二乗分類を実装する。

$$\sum_{i, i'=1}^m W_{i, i'} (a_i - a_{i'})^2 = 2 \sum_{i, i'=1}^m L_{i, i'} a_i a_{i'} \quad (4)$$

$$\mathbf{D} = \text{diag}\left(\sum_{i=1}^m W_{1, i}, \cdots, \sum_{i=1}^m W_{m, i}\right) \quad (5)$$

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (6)$$

近傍グラフの重みにはガウスカーネルを用い、

$$W_{i, i'} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_{i'}\|^2}{2h^2}\right) \quad (7)$$

とする。

目的関数は、

$$J(\boldsymbol{\theta}) = \sum_{i=1}^n (f_{\boldsymbol{\theta}}(\mathbf{x}_i) - y_i)^2 + \lambda \|\boldsymbol{\theta}\|^2 + \nu \sum_{i, i'=1}^{n+n'} W_{i, i'} (f_{\boldsymbol{\theta}}(\mathbf{x}_i) - f_{\boldsymbol{\theta}}(\mathbf{x}_{i'}))^2 \quad (8)$$

$$= \|\tilde{\mathbf{K}} \boldsymbol{\theta} - \mathbf{y}\|^2 + \lambda \|\boldsymbol{\theta}\|^2 + 2\nu \boldsymbol{\theta}^T \mathbf{K}^T \mathbf{L} \mathbf{K} \boldsymbol{\theta} \quad (9)$$

となる。これを  $\boldsymbol{\theta}$  で偏微分すると、

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = 2\tilde{\mathbf{K}}^T \tilde{\mathbf{K}} \boldsymbol{\theta} - 2\tilde{\mathbf{K}} \mathbf{y} + 2\lambda \boldsymbol{\theta} + 4\nu \mathbf{K}^T \mathbf{L} \mathbf{K} \boldsymbol{\theta} \quad (10)$$

$$= 2\left(\tilde{\mathbf{K}}^T \tilde{\mathbf{K}} + \lambda \mathbf{I} + 2\nu \mathbf{K}^T \mathbf{L} \mathbf{K}\right) \boldsymbol{\theta} - 2\tilde{\mathbf{K}} \mathbf{y} \quad (11)$$

$$= \mathbf{0} \quad (12)$$

これより、

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \left(\tilde{\mathbf{K}}^T \tilde{\mathbf{K}} + \lambda \mathbf{I} + 2\nu \mathbf{K}^T \mathbf{L} \mathbf{K}\right)^{-1} \tilde{\mathbf{K}} \mathbf{y} \quad (13)$$

である。

$h = 1.0$ ,  $\lambda = 1.0$ ,  $\nu = 1.0$  としてこれを実装したものが??ページの Listing ??である。

結果は図 1 に示した通りである。教師ありのデータは各クラス 1 つずつしかないが、ラベルなしのデータについても上手く分けられている。

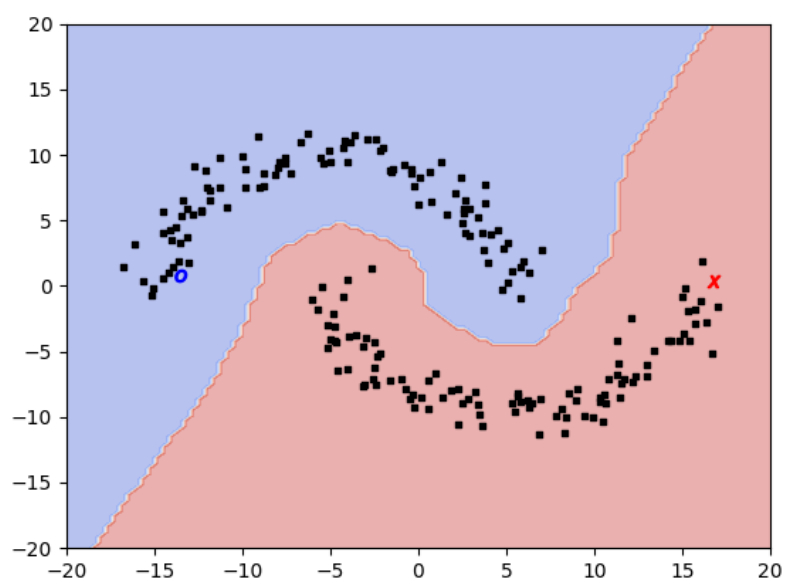


图 1: 結果