宿題 1

線形モデル

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{x} \tag{1}$$

を用いた最小二乗分類を考える。いま，訓練標本の平均 $\boldsymbol{\mu}$ について次式が成立する。

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = \mathbf{0} \tag{2}$$

また，最小二乗誤差は，

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{1}{2}\sum_{i=1}^{n}\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}) - y_i\right)^2 && (3)\\ &= \frac{1}{2}||\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}||^2 && (4) \end{aligned}$$

と表されるので，これの $\boldsymbol{\theta}$ による偏微分が $\mathbf{0}$ になる条件から，

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\hat{\boldsymbol{\theta}} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{5}$$

が成立する。ここで，

$$\begin{aligned} \boldsymbol{X} &= \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}^{\mathrm{T}} && (6)\\ \boldsymbol{y} &= \begin{bmatrix}y_1 & \cdots & y_n\end{bmatrix}^{\mathrm{T}} && (7) \end{aligned}$$

である。

いま，正例を下付きの $+$ で，負例を下付きの $-$ で表すこととすると，次が成立する。

$$\begin{aligned} n &= n_+ + n_- && (8)\\ \boldsymbol{\mu}_+ &= \frac{1}{n_+}\sum_{\boldsymbol{x}_i \in D_+} x_i && (9)\\ \boldsymbol{\mu}_- &= \frac{1}{n_-}\sum_{\boldsymbol{x}_i \in D_-} x_i && (10)\\ & && (11) \end{aligned}$$

全標本の和を考えると，

$$\begin{aligned} \sum_{i=1}^{n} x_i &= \sum_{\boldsymbol{x}_i \in D_+} x_i + \sum_{\boldsymbol{x}_i \in D_-} x_i && (12)\\ n\boldsymbol{\mu} &= n_+\boldsymbol{\mu}_+ + n_-\boldsymbol{\mu}_- = \mathbf{0} && (13)\\ \boldsymbol{\mu}_- &= -\frac{n_+}{n_-}\boldsymbol{\mu}_+ && (14) \end{aligned}$$

が得られる。

また，二値分類問題を考えているので，y は正例のとき 1，負例のとき -1 となる。このことから，

$$\begin{aligned} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} &= \sum_{\boldsymbol{x}_i \in D_+} x_i - \sum_{\boldsymbol{x}_i \in D_-} x_i && (15)\\ &= n_+\boldsymbol{\mu}_+ - n_-\boldsymbol{\mu}_- && (16)\\ &= 2n_+\boldsymbol{\mu}_+ && (17) \end{aligned}$$

を得る。

また，平均 $\boldsymbol{\mu} = \mathbf{0}$ より，標本の共分散行列の推定値は次で表される。

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} \tag{18}$$

これを用いると，次が成立する。

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}}\end{bmatrix} \tag{19}$$

$$= \sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} \tag{20}$$

$$= n\hat{\boldsymbol{\Sigma}} \tag{21}$$

以上，式 (5)，(17)，(21) より，

$$n\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\theta}} = 2n_+\boldsymbol{\mu}_+ \tag{22}$$

$$\hat{\boldsymbol{\theta}} = \frac{2n_+}{n}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}_+ \tag{23}$$

を得る。一方，フィッシャー判別分析による境界の垂線方向は，

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-) = \hat{\boldsymbol{\Sigma}}\left(1 + \frac{n_+}{n_-}\right)\boldsymbol{\mu}_+ \tag{24}$$

$$= \frac{n}{n_-}\hat{\boldsymbol{\Sigma}}\boldsymbol{\mu}_+ \tag{25}$$

となる。式 (23)，(25) から，最小二乗分類により得られる識別境界の垂線 $\hat{\boldsymbol{\theta}}$ は，フィッシャー判別分析による境界の垂線 $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)$ と同じ方向になる。