1. データセット $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ $(\boldsymbol{x}_i \in R^d,\ y_i \in \{+1, -1\})$ に対してロジスティック判別を行う．$\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \ldots, \phi_p(\boldsymbol{x}))^T$ を $\boldsymbol{x} \in R^d$ を適当な関数で p 次元に変換した関数ベクトル，$\boldsymbol{\theta} \in R^p$ をパラメーターとして

$$p(y|\boldsymbol{x}) = \frac{1}{1 + \exp(-y\boldsymbol{\theta}^T \boldsymbol{\phi}(\boldsymbol{x}))}$$

$$\left(p(+1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\boldsymbol{\theta}^T \boldsymbol{\phi}(\boldsymbol{x}))} = \frac{\exp(\boldsymbol{\theta}^T \boldsymbol{\phi}(\boldsymbol{x}))}{1 + \exp(\boldsymbol{\theta}^T \boldsymbol{\phi}(\boldsymbol{x}))},\quad p(-1|\boldsymbol{x}) = \frac{1}{1 + \exp(\boldsymbol{\theta}^T \boldsymbol{\phi}(\boldsymbol{x}))}\right)$$

と定義し，対数尤度関数

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(y_i|\boldsymbol{x}_i, \theta) \qquad (1)$$

θのパラメータを最適化したい！

を最大とするようなパラメーター $\boldsymbol{\theta}$ を見つけて判別を行うとする．

(a) 判別境界 $(p(+1|\boldsymbol{x}) = p(-1|\boldsymbol{x})$ となる $\boldsymbol{x}$ の領域$)$ を表す式を求めよ．

$$p(+1|x) = \frac{1}{1+\exp(-\theta^T\phi(x))}$$

$$= \frac{1}{1+\frac{1}{\exp(\theta^T\phi(x))}} \times \frac{\exp(\theta^T\phi(x))}{\exp(\theta^T\phi(x))}$$

$$= \frac{\exp(\theta^T\phi(x))}{\exp(\theta^T\phi(x))+1}$$

$$= \frac{\exp(\theta^T\phi(x))}{1+\exp(\theta^T\phi(x))}$$

$$p(+1|x) = p(-1|x)$$

$$\frac{\exp(\theta^T\phi(x))}{1+\exp(\theta^T\phi(x))} = \frac{1}{1+\exp(\theta^T\phi(x))}$$

$$\exp(\theta^T\phi(x)) = 1$$

$$\exp(\theta^T\phi(x)) = \exp(0)$$

$$\theta^T\phi(x) = 0$$

$\therefore$ 判別境界は $\theta^T\phi(x) = 0$ を満たす x の領域である

(b) $\boldsymbol{x} \in R^2$ とし, $\boldsymbol{\phi}(\boldsymbol{x}) = (x_1, x_2)^T$ とした時, (a) で求めた判別境界はどのような領域か?

(c) $\boldsymbol{x} \in R^2$ とし, $\boldsymbol{\phi}(\boldsymbol{x}) = (1, x_1^2, x_2^2)^T$, $\theta_2 > 0, \theta_3 > 0$ としたとき, (a) で求めた判別境界はどのような領域か?

(a)より $\theta^T \phi(x) = 0$

(b) $x \in R^2$ より $p=2$ のため $\theta \in R^2$ $\theta = (\theta_1\ \theta_2)^T$ とすると.

$$\theta^T \phi(x) = \left((\theta_1, \theta_2)^T\right)^T (x_1, x_2)^T$$

$$= \underset{1\times 2}{(\theta_1, \theta_2)} \underset{2\times 1}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = \theta_1 x_1 + \theta_2 x_2 = 0$$

∴判別境界は直線となる.

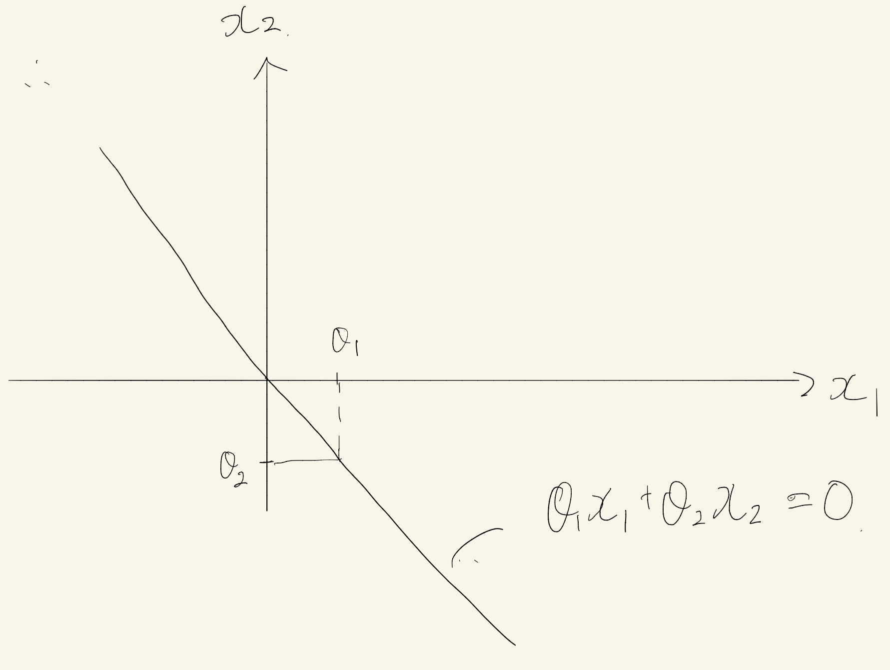
横軸を x_1, 縦軸を x_2 とすると.

$$x_2 = -\frac{\theta_1}{\theta_2} x_1$$

傾き $-\frac{\theta_1}{\theta_2}$

切片 0

∴



(c.)

$x \in \mathbb{R}^2$ より $\theta \in \mathbb{R}^2$ $\theta = (\theta_1, \theta_2, \theta_3)^T$ とすると。 ただし $\theta_2 > 0$, $\theta_3 > 0$.

$$\theta^T\phi(x) = ((\theta_1, \theta_2, \theta_3)^T)^T (1, x_1, x_2)^T$$

$$= (\theta_1, \theta_2, \theta_3)\begin{pmatrix}1\\ x_1^2\\ x_2^2\end{pmatrix}$$

$$= \theta_1 + \theta_2 x_1^2 + \theta_3 x_2^2 = 0$$

$$\theta_2 x_1^2 + \theta_3 x_2^2 = -\theta_1$$

$\theta_1 \neq 0$ のとき ※直交座標方程式

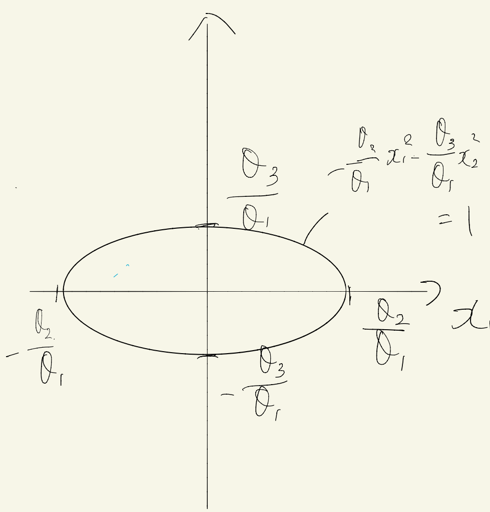
$$-\frac{\theta_2}{\theta_1}x_1^2 - \frac{\theta_3}{\theta_1}x_2^2 = 1$$

$\theta_1 < 0$ のとき 判別境界 隋円となる。

長軸を a^2 短軸を b^2 とすると

半長軸 $a = \frac{\theta_2}{\theta_1}$

半短軸 $b = \frac{\theta_3}{\theta_1}$



1. データセット $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ $(\boldsymbol{x}_i \in R^d,\ y_i \in \{+1, -1\})$ に対してロジスティック判別を行う．$\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \ldots, \phi_p(\boldsymbol{x}))^T$ を $\boldsymbol{x} \in R^d$ を適当な関数で $\boxed{p}$ 次元に変換した関数ベクトル，$\boldsymbol{\theta} \in R^p$ をパラメーターとして

これが使うの！

$$p(y|\boldsymbol{x}) = \frac{1}{1 + \exp(-y\boldsymbol{\theta}^T\boldsymbol{\phi}(\boldsymbol{x}))}$$

$$\left(p(+1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\boldsymbol{\theta}^T\boldsymbol{\phi}(\boldsymbol{x}))} = \frac{\exp(\boldsymbol{\theta}^T\boldsymbol{\phi}(\boldsymbol{x}))}{1 + \exp(\boldsymbol{\theta}^T\boldsymbol{\phi}(\boldsymbol{x}))},\quad p(-1|\boldsymbol{x}) = \frac{1}{1 + \exp(\boldsymbol{\theta}^T\boldsymbol{\phi}(\boldsymbol{x}))}\right)$$

と定義し，対数尤度関数

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(y_i|\boldsymbol{x}_i, \theta) \tag{1}$$

(d)

$x_i \in \mathbb{R}^d$

$y_i \in \{-1, 1\}$

$$L(\theta) = \sum_{i=1}^{n} \log p(y_i | x_i ; \theta)$$

$$= \sum_{i=1}^{n} \log \frac{1}{1 + \exp(-y_i \theta^T \phi(x_i))}$$

$$= \log \left(\frac{1}{1+\exp(-y_1 \theta^T x_1)} + \frac{1}{1+\exp(-y \theta^T x_2)} + \cdots \frac{1}{1+\exp(-y_n \theta^T x_n)} \right)$$

ここですべて正しく判別できるため誤差は0となる．

つまり，$\frac{\partial}{\partial \theta} L(\theta) = 0$ となる．

$$L(\theta) = \sum_{i=1}^{n} \log \frac{1}{1+\exp(-y_i \theta^T \phi(x_i))}$$

$g(f_1 \cdot f_2)$ log $1+\exp(-y_i\theta^T$

$$= \sum_{i=1}^{n} -\log(1+\exp(-y_i \theta^T \phi(x_i)))$$

$\mu = -y_i \theta^T \phi(x_i)$

$s = 1+\exp(\mu)$

$L = -\sum_{i=1}^{n} \log s$ とすると

$$\frac{\partial \mu}{\partial \theta} = -y_i \phi(x_i)$$

$$\frac{\partial s}{\partial \mu} = \exp(\mu)$$

$$\frac{\partial L}{\partial s} = -\sum_{i=1}^{n} \frac{1}{s}$$

$$\frac{\partial L}{\partial \theta} = \frac{\partial \mu}{\partial \theta} \cdot \frac{\partial s}{\partial \mu} \cdot \frac{\partial L}{\partial s}$$ で

計算することができる

$$\frac{\partial L}{\partial \theta} = -\sum_{i=1}^{n} \left(-y_i \phi(x_i) \cdot \exp(-y_i \theta^T \phi(x_i)) \cdot \frac{1}{1+\exp(-y_i \theta^T \phi(x_i))} \right)$$

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} \frac{y_i \phi(x_i) \exp(-y_i \theta^T \phi(x_i))}{1+\exp(-y_i \theta^T \phi(x_i))}$$

$\frac{\partial L}{\partial \theta} = 0$ より

$$\sum_{i=1}^{n} y_i \phi(x_i) \exp(-y_i \theta^T \phi(x_i)) = 0$$

$\therefore \theta$ は $\sum_{i=1}^{n} y_i \phi(x_i) \exp(-y_i \theta^T \phi(x_i)) = 0$

を満たす必要がある.