

scikit-learn ゼミ

1.1.10. Logistic regression

Kunii Shinya

概要

- ロジスティック回帰

- 線型回帰モデルを拡張した一般化線型モデル

- 主に2値分類

- 例：入試の結果 $Y = \{\text{合格、不合格}\}$

- (x_1, x_2) = {内申点、校内偏差値}

y	x1	x2
0	3.6	60.1
1	4.1	52
0	3.7	62.5

通常重回帰分析考えると

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$\hat{y} = 0.5, 0.8, -0.4$ の予測値の場合に、どのように解釈すればよいのか不明

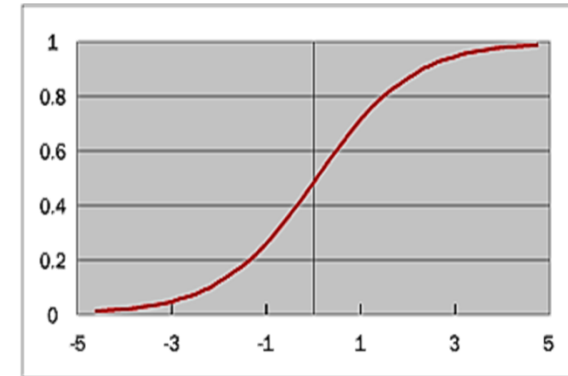
ロジスティック関数

- 生起確率 $p(y = 1)$ として考える

$$-\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

ロジット関数

$$\begin{aligned} -p(Y = 1) &= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} \\ &= \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)]} \end{aligned}$$



ロジスティック関数

関数

- LogisticRegression
 - penalty : string, 'l1' or 'l2'
 - dual : boolean
 - (双対問題)
 - C : float, optional (default=1.0)
 - Inverse of regularization strength
 - fit_intercept : bool, default: True
 - intercept_scaling : float, default: 1
 - class_weight : {dict, 'auto'}, optional
 - tol: float, optional :