ロジスティック回帰
Logistic Regression
LR

明治大学 理工学部 応用化学科
データ化学工学研究室　金子 弘昌

ロジスティック回帰 (LR) とは？

- ✓(名前に “回帰” とあるが) クラス分類手法
- ✓0 or 1 の 2 クラス分類において、予測値が 1 になる確率として出力される
- ✓目的変数のロジット変換で、説明変数との間の関係が線形になる
- ✓目的変数がロジスティックシグモイド関数で表される

ロジスティック回帰モデル

✓クラス t : 0 もしくは 1

✓目的変数 y

✓説明変数 x (m 個) : $x_1, x_2, ..., x_m$

$$y = \frac{1}{1 + \exp(-z)}$$ ・・・ロジスティックシグモイド関数

$$z = b_1 x_1 + b_1 x_2 + \cdots + b_1 x_m + c$$ ・・・線形関数

$b_1, b_2, ..., b_m, c$: ロジスティック回帰モデルのパラメータ

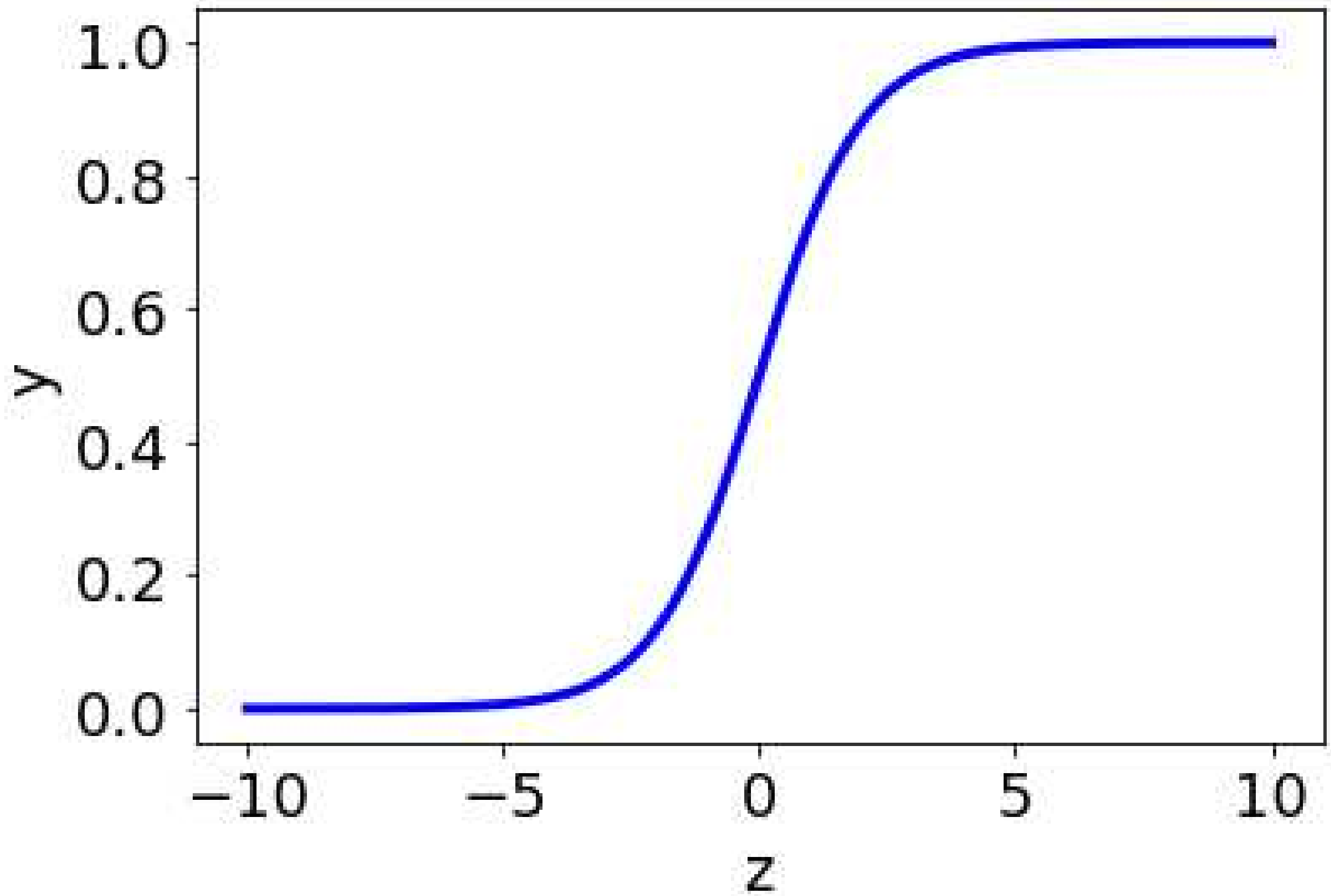
⇨ y が 0.5 以上・・・クラス 1 と予測
y が 0.5 未満・・・クラス 0 と予測

$$z = \ln\left(\frac{y}{1 - y}\right)$$ ・・・ロジット関数 [1]

[1] https://datachemeng.com/post-3543/

ロジスティックシグモイド関数

$$y = \frac{1}{1 + \exp(-z)}$$



なぜロジスティックシグモイド関数？

$p(\mathrm{C}_1 \mid \mathrm{x})$ ・・・ x が与えられたときのクラス 1 の事後確率

ベイズの定理

$$p(\mathrm{C}_1 \mid \mathrm{x}) = \frac{p(\mathrm{x} \mid \mathrm{C}_1)\, p(\mathrm{C}_1)}{p(\mathrm{x} \mid \mathrm{C}_1)\, p(\mathrm{C}_1) + p(\mathrm{x} \mid \mathrm{C}_2)\, p(\mathrm{C}_2)}$$

$$= \frac{1}{1 + \dfrac{p(\mathrm{x} \mid \mathrm{C}_2)\, p(\mathrm{C}_2)}{p(\mathrm{x} \mid \mathrm{C}_1)\, p(\mathrm{C}_1)}}$$

$$= \frac{1}{1 + \exp(-\mathrm{z})}$$

$p(\mathrm{C}_1)$ ・・・ クラス 1 の事前確率

ただし、 $\mathrm{z} = \ln\left(\dfrac{p(\mathrm{x} \mid \mathrm{C}_1)\, p(\mathrm{C}_1)}{p(\mathrm{x} \mid \mathrm{C}_2)\, p(\mathrm{C}_2)} \right)$

パラメータの求め方

✓ロジスティック回帰モデルのパラメータ $\mathbf{b} = [b_1, b_2, ..., b_m, c]$

→ 最尤法を用いて決定

尤度関数 $$p(\mathbf{t} \mid \mathbf{b}) = \prod_{i=1}^{n} y_i^{t_i} (1 - y_i)^{1-t_i}$$

n : サンプル数
t_i : i 番目のサンプルにおいて、クラス 0 なら 0、クラス 1 なら 1
y_i : i 番目のサンプルにおける y の値
$\mathbf{t}$: $[t_1, t_2, .., t_n]$