数理統計 補助資料
～ロジスティック回帰モデルの推定～

2024年度2学期:　月曜1限, 水曜3限

担当教員:　石垣　司

再掲: ロジスティック回帰モデル

説明変数が複数あるロジスティック回帰モデル

- **目的変数: 変数** y
- **説明変数: 変数** $\boldsymbol{x} = [x_1, x_2, \cdots, x_P]^T$
- **データ:** $D = \{y_i, \boldsymbol{x}_i\}\ (i = 1, \cdots, N)$
  $\boldsymbol{x}_i = [1 \quad x_{i1} \quad \cdots \quad x_{iP}]^T$ **(行列** X **の** i **番目の行)**
- **回帰係数: 係数** $b_0, b_1, \cdots, b_P$ **(パラメータ)**

$$\boldsymbol{b} = [b_0 \quad b_1 \quad \cdots \quad b_P]^T$$

$$\Pr(y = 1|\boldsymbol{x}) = \frac{1}{1+e^{-(b_0+b_1x_1+,\cdots,+b_Px_P)}} = \frac{1}{1+e^{-\boldsymbol{x}^T\boldsymbol{b}}}$$

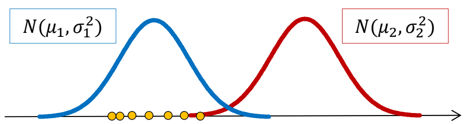
再掲: 尤度に基づくパラメータ推定

尤度: 与えられたデータがある確率分布から発生していると考えたときの尤もらしさの度合い

- **データ** y_i **を発生する確率分布** $p(y_i;\boldsymbol{\theta})$ **(パラメータ** $\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]^T$ **)を考える**
- **データ** y_i **の尤度:** $l(\boldsymbol{\theta}|y_i) = p(y_i;\boldsymbol{\theta})$
- **独立同一分布から発生したデータ** $\boldsymbol{y} = [y_1, \dots, y_N]^T$**の尤度:**

$$l(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{N} p(y_i;\boldsymbol{\theta})$$

例: 次のデータは青と赤のどちらの正規分布から発生していると考えるのが妥当か?



$$\Rightarrow \quad l_1(\mu_1,\sigma_1^2|\boldsymbol{y}) > l_2(\mu_2,\sigma_2^2|\boldsymbol{y})$$

ロジスティック回帰モデルの尤度関数 #1

ロジスティック回帰モデルの1つのデータの尤度

- $y_i = 1$ **のときの尤度:**

$$l(\boldsymbol{b}|\boldsymbol{x}_i) = \Pr(y_i = 1|\boldsymbol{x}_i) = \frac{1}{1 + e^{-\boldsymbol{x}_i^T\boldsymbol{b}}}$$

- $y_i = 0$ **のときの尤度:**

$$l(\boldsymbol{b}|\boldsymbol{x}_i) = \Pr(y_i = 0|\boldsymbol{x}_i) = 1 - \frac{1}{1 + e^{-\boldsymbol{x}_i^T\boldsymbol{b}}}$$

ロジスティック回帰モデルの尤度関数

$$l(\boldsymbol{b}|D) = \prod_{i=1}^{N} \left(\frac{1}{1 + e^{-\boldsymbol{x}_i^T\boldsymbol{b}}}\right)^{y_i} \left(1 - \frac{1}{1 + e^{-\boldsymbol{x}_i^T\boldsymbol{b}}}\right)^{1-y_i}$$

ロジスティック回帰モデルの尤度関数 #2

尤度関数（単純化のために説明変数が1つで切片が 0 の例）

$$l(b|D) = \prod_{i=1}^{N} \left(\frac{1}{1+e^{-bx_i}} \right)^{y_i} \left(1 - \frac{1}{1+e^{-bx_i}} \right)^{1-y_i}$$

#メモ　単純化のため説明変数が1つで切片が0の例で説明するが、説明変数が複数の場合でも同様の議論で最尤推定が可能

対数尤度関数

$$L(b|D) = -\sum_{i=1}^{N} y_i \log\left(1+e^{-bx_i}\right) + \sum_{i=1}^{N} (1-y_i) \log\left(1 - \frac{1}{1+e^{-bx_i}}\right)$$

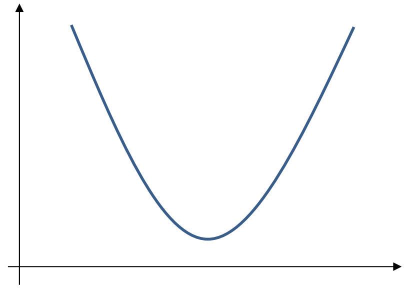
- この関数を最大化したい。しかし，最適解はあるのか？
- 最適解があるとしたら，どのように求めるのか？

凸関数 #1

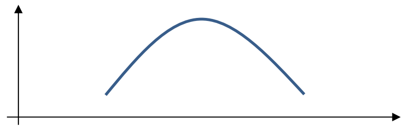
定義

- 関数 f が区間内の任意の2点 x_1, x_2 と $0 < t < 1$ を満たす任意の t に対して，常に次の関係を満たすとき，関数 f を凸関数とよぶ

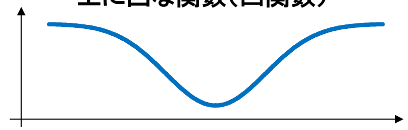
$$tf(x_1) + (1-t)f(x_2) \geq f(tx_1 + (1-t)x_2)$$



下に凸な関数（凸関数）



上に凸な関数（凹関数）

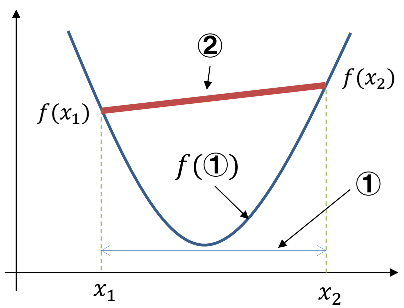


一見、凸関数に見えるが非凸関数の例

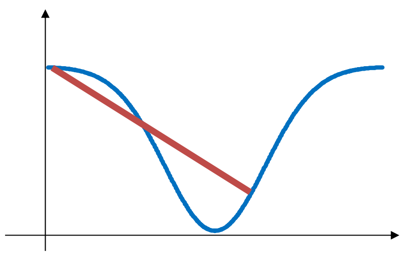
凸関数 #2

凸関数の定義の意味（ $x_1 < x_2$ となる2点 x_1, x_2 を考える）

- ① $tx_1 + (1-t)x_2$ は x_1 と x_2 の内分点
- ② $tf(x_1) + (1-t)f(x_2)$ は $f(x_1)$ と $f(x_2)$ の内分点
- 凸関数とは常に②が f(①)よりも上にある関数



下に凸な関数（凸関数）



一見，凸関数に見えるが
非凸関数の例

凸関数の性質

命題1

- 関数 f が $f''(x) > 0 \Leftrightarrow$ 関数 f は下に凸な関数
- 関数 f が $f''(x) < 0 \Leftrightarrow$ 関数 f は上に凸な関数

命題2

- 関数 f が凸関数のとき $f'(x^*) = 0$ ならば，関数 f は $x = x^*$ で最小となる

ロジスティック回帰モデルの尤度関数 #3

ロジスティック回帰モデルの尤度関数の特徴

1. **対数尤度関数** $L(b|D)$ **は凸関数(上に凸)であるため最適解は存在する** check!

2. **対数尤度関数** $L(b|D)$ **を最大にする** b **は解析的に求めることができない** check!

   ⇒ **コンピュータを利用した数値的最適化**

#メモ　ロジスティック回帰モデルの回帰係数を最小2乗法＋数値計算で求めても最尤法＋数値計算で求めても推定値の良さは同じ程度。ただし、効率性は最尤法の方がよい　Breslawa and Irvine (1982)

演習問題

次の確率密度関数をもつ指数分布

$$f(y;\lambda) = \begin{cases} \lambda \exp(-\lambda y) & (y \geq 0) \\ 0 & (y < 0) \end{cases}$$

と $f(y;\lambda)$ **から独立に発生したデータ** $D = \{y_1, y_2, \cdots, y_N\}$ **を考える。このとき，その対数尤度関数** $L(\lambda|D)$ **は凸関数となるかどうかを確かめよ**