ロジスティック回帰・直線回帰

樋口さぶろお

龍谷大学大学院理工学研究科数理情報学専攻

理論物理学特論　L03(2016-10-05 Wed)

最終更新: Time-stamp: "2016-10-02 Sun 17:47 JST hig"

今日の目標

1. ロジスティック回帰を, 一般化線形モデルとして説明できる
2. 線形回帰を, 一般化線形モデルとして説明できる

http://hig3.net

L02-Q1

Quiz 解答:正規分布の母数の最尤推定

$\log L(\mu, \sigma^2) = \sum_i -\frac{1}{2}(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - \mu)^2$ に対して,
$\frac{\partial}{\partial\mu}\log L = \frac{\partial}{\partial(\sigma^2)}\log L = 0$ を考える.

1. $$\mu = \frac{1}{2}(x_1 + x_2), \quad \sigma^2 = \left(\frac{x_1 - x_2}{2}\right)^2.$$
2. $$\mu = \frac{1}{N}(x_1 + \cdots + x_N), \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \frac{1}{N}(x_1 + \cdots + x_N))^2.$$
   ($1/(N-1)$ ではない).

L02-Q2

Quiz 解答:ポアソン回帰

$$
\begin{aligned}
\log L =& \log\left(\frac{e^{(\beta_1+\beta_2+\beta_3)\cdot 1}}{1!}e^{-e^{\beta_1+\beta_2+\beta_3}}\right) + \log\left(\frac{e^{(\beta_1+2\beta_2)\cdot 3}}{3!}e^{-e^{\beta_1+2\beta_2}}\right) \\
&+ \log\left(\frac{e^{(\beta_1+2\beta_2)\cdot 5}}{5!}e^{-e^{\beta_1+2\beta_2}}\right) + \log\left(\frac{e^{(\beta_1+3\beta_2+\beta_3)\cdot 8}}{8!}e^{-e^{\beta_1+3\beta_2+\beta_3}}\right) \\
=&(\beta_1+\beta_2+\beta_3)\cdot 1 - e^{\beta_1+\beta_2+\beta_3} + (\beta_1+2\beta_2)\cdot 3 - e^{\beta_1+2\beta_2} \\
&+ (\beta_1+2\beta_2)\cdot 5 - e^{\beta_1+2\beta_2} + (\beta_1+3\beta_2+\beta_3)\cdot 8 - e^{\beta_1+3\beta_2+\beta_3} + \text{定数}
\end{aligned}
$$

ここまで来たよ

1. 略解:最尤推定・一般化線形モデル・ポアソン回帰
2. ロジスティック回帰・直線回帰
   - ロジスティック回帰

L03-Q1

Quiz(ロジスティック関数)

微分可能であり，単調増加であり，$\lim_{x\to-\infty} q(x) = 0$, $\lim_{x\to+\infty} q(x) = 1$ であり，簡単な式で書けるような関数を見つけよう．

L03-Q2

Quiz(ロジット関数)

ロジスティック関数の逆関数を求めよう．

L03-Q3

Quiz(ロジスティック回帰)

ロジスティック回帰の一定モデル $\mathrm{logit}(q) = \beta_1$ について，データ $\{y_i, N_i\}$ から β_1 を最尤推定しよう．
いまは β_1 しかないので，$\frac{\partial \log L}{\partial \beta_1} = 0$ は $\frac{\partial \log L}{\partial q} = 0$ と同値で，こっちのほうが計算が楽．

L03-Q4

ロジスティック回帰

ロジスティック回帰で，説明変数が因子変数 1 個である場合
$\mathrm{logit}(q) = \beta_1 + \beta_2 d$, $d = 0, 1$ について，データ $\{y_i, N_i, d_i\}$ から β_1, β_2 を最尤推定しよう．
ここで，$y^{(a)} = \sum_{d_i=a} y_i$, $N^{(a)} = \sum_{d_i=a} N_i$ などとおくと考えやすいかも．

L03-Q5

直線回帰

直線回帰で，尤度最大という条件から，‘直線からのずれの 2 乗が最小’ を導こう．データ $\{y_i, N_i\}$ から q を最尤推定しよう．