

第9回 母数の推定

パラメタの分割 その他

今後の話題

- 交互最小2乗解 Alternating Least Squares
 - ◆ 最尤解にも応用可能

- 応用例
 - ◆ 因子分析、多次元尺度法、クラスター分析、ニューラルネット等
 - ◆ <http://mayekawa.in.coocan.jp/titech/docs/fa12.pdf>
 - ◆ <http://mayekawa.in.coocan.jp/titech/docs/mds2.pdf>
 - ◆ Mayekawa, Shin-ichi. (1995) An efficient algorithm for feed-forward neural network regression analysis. {¥s} Behaviormetrika}. Vol.22, No. 1, 67-90

- 周辺最尤解

Gauss-Newton 法

最適化されるべき関数が最小2乗基準の場合 $r(\boldsymbol{\xi}) = \text{RSS}(\boldsymbol{\xi})$

いま、求めるべきパラメタの仮の値を $\boldsymbol{\xi}_{(l)}$ とする。モデルの各要素 $g(x_i|\boldsymbol{\xi})$ を $\boldsymbol{\xi}_{(l)}$ のまわりで1次の項まで Taylor 展開したものを $N \times 1$ のベクトルとしてまとめれば、

$$g(\boldsymbol{\xi}) \approx g(\boldsymbol{\xi}_{(l)}) + \mathbf{F}(\boldsymbol{\xi}_{(l)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(l)}) \quad (102)$$

となるが、ここに、

$$\mathbf{F}(\boldsymbol{\xi}) = \frac{\partial \mathbf{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \left[\frac{\partial g(x_i|\boldsymbol{\xi})}{\partial \xi_k} \right], \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, Q \quad (103)$$

は、ヤコビアン行列と呼ばれる $N \times Q$ の1次微分行列、また、 $\mathbf{F}(\boldsymbol{\xi}_{(l)})$ はそれを $\boldsymbol{\xi}_{(l)}$ で評価した行列

$$\mathbf{F}(\boldsymbol{\xi}_{(l)}) = \left. \frac{\partial \mathbf{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{(l)}} \quad (104)$$

である。

3

この1次近似を RSS に代入すれば、

$$\text{RSS}(\boldsymbol{\xi}) = \sum_{i=1}^N (y_i - g(x_i|\boldsymbol{\xi}))^2 = \|\mathbf{y} - \mathbf{g}(\boldsymbol{\xi})\|^2 \quad (105)$$

$$\approx \|\mathbf{y} - (\mathbf{g}(\boldsymbol{\xi}_{(l)}) + \mathbf{F}(\boldsymbol{\xi}_{(l)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(l)}))\|^2 \quad (106)$$

$$= \|\mathbf{y} - \mathbf{g}(\boldsymbol{\xi}_{(l)}) - \mathbf{F}(\boldsymbol{\xi}_{(l)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(l)})\|^2 \quad (107)$$

となるから、

$$\frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \approx -2\mathbf{F}(\boldsymbol{\xi}_{(l)})'(\mathbf{y} - \mathbf{g}(\boldsymbol{\xi}_{(l)})) + 2\mathbf{F}(\boldsymbol{\xi}_{(l)})'\mathbf{F}(\boldsymbol{\xi}_{(l)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(l)}) = \mathbf{0}$$

の解は

$$\boldsymbol{\xi} = \boldsymbol{\xi}_{(l)} + (\mathbf{F}(\boldsymbol{\xi}_{(l)})'\mathbf{F}(\boldsymbol{\xi}_{(l)}))^{-1}\mathbf{F}(\boldsymbol{\xi}_{(l)})'(\mathbf{y} - \mathbf{g}(\boldsymbol{\xi}_{(l)})) \quad (108)$$

となる。

4

また、先に示したグレイディエントベクトルが

$$\mathbf{g} = \frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -2\mathbf{F}(\boldsymbol{\xi})'(\mathbf{y} - \mathbf{g}(\boldsymbol{\xi})) \quad (109)$$

となるため、Gauss-Newton 法ではパラメタの更新式において

$$\mathbf{d}_{(\ell)} = -\frac{1}{2}(\mathbf{F}'(\boldsymbol{\xi}_{(\ell)})'\mathbf{F}(\boldsymbol{\xi}_{(\ell)}))^{-1}\mathbf{g}(\boldsymbol{\xi}_{(\ell)}) \quad (110)$$

と置いたものになっていることがわかる。

多変量の場合

基準変数が多変量の場合は、

$$\text{RSS} = \|\mathbf{Y} - \mathbf{G}(\boldsymbol{\xi})\|^2 = \sum_{j=1}^p \|\mathbf{y}_{(j)} - \mathbf{g}_{(j)}(\boldsymbol{\xi})\|^2 \quad (111)$$

と書けるため、変数ごとのヤコビアン行列を

$$\mathbf{F}_j(\boldsymbol{\xi}) = \frac{\partial \mathbf{g}_{(j)}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \left[\frac{\partial g_{ij}(\boldsymbol{\xi})}{\partial \xi_k} \right], \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, Q$$

と定義すれば、グレイディエントならびに更新の方向は

$$\mathbf{g} = \frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -2 \sum_{j=1}^p \mathbf{F}_{(j)}(\boldsymbol{\xi})'(\mathbf{y}_{(j)} - \boldsymbol{\theta}_{(j)}(\boldsymbol{\xi})) \quad (112)$$

$$\mathbf{d}_{(\ell)} = -\frac{1}{2} \left(\sum_{j=1}^p \mathbf{F}_{(j)}(\boldsymbol{\xi}_{(\ell)})' \mathbf{F}_{(j)}(\boldsymbol{\xi}_{(\ell)}) \right)^{-1} \mathbf{g}(\boldsymbol{\xi}_{(\ell)}) \quad (113)$$

で与えられる。

もしくは多変量の \mathbf{Y} を変数ごとに縦に並べたものを \mathbf{y} 、それに対応するモデルを $\mathbf{g}(\boldsymbol{\xi})$ とすれば、これから定義されるヤコビアン行列は

$$\mathbf{F}(\boldsymbol{\xi}) = \frac{\partial \mathbf{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \left[\frac{\partial g_{ij}(\boldsymbol{\xi})}{\partial \xi_k} \right], \quad i = 1, 2, \dots, N, k = 1, 2, \dots, Q$$

となるが、これは $\mathbf{F}_j(\boldsymbol{\xi})$ を縦に並べたものとなる。したがって、

$$\mathbf{g} = \frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -2\mathbf{F}(\boldsymbol{\xi})'(\mathbf{y} - \boldsymbol{\theta}(\boldsymbol{\xi})) \quad (114)$$

$$\mathbf{d}_{(\ell)} = -\frac{1}{2}(\mathbf{F}(\boldsymbol{\xi}_{(\ell)})'\mathbf{F}(\boldsymbol{\xi}_{(\ell)}))^{-1}\mathbf{g}(\boldsymbol{\xi}_{(\ell)}) \quad (115)$$

7

パラメタの分割

主成分分析

$$\mathbf{Y} = \widehat{\mathbf{Y}} + \mathbf{E} \quad (116)$$

$$\widehat{\mathbf{Y}} = \mathbf{F}\mathbf{A}' \quad (117)$$

最小2乗基準

$$\text{RSS} = \sum_{i=1}^N \sum_{j=1}^p (y_{ij} - \widehat{y}_{ij})^2 \quad (118)$$

$$= \|\mathbf{Y} - \mathbf{F}\mathbf{A}'\|^2 = \|\mathbf{Y}' - \mathbf{A}\mathbf{F}'\|^2 \quad (119)$$

8

線型モデル： $\|Y - X\beta\|^2$ の最小2乗解は $(X'X)^{-1}X'Y$

F が所与の場合 $\|Y - FA'\|^2$

$$A' = (F'F)^{-1}F'Y \quad \text{or} \quad A = Y'F(F'F)^{-1} \quad (120)$$

A が所与の場合 $\|Y' - AF'\|^2$

$$F' = (A'A)^{-1} * A'Y' \quad \text{or} \quad F = YA(A'A)^{-1} \quad (121)$$

数値例有り

9

■ その他の例

◆ 項目反応理論における能力値パラメタ θ と項目パラメタの同時推定

- ★ θ が所与の場合、(a,b) は項目ごとに独立なロジスティック回帰
- ★ (a,b) が所与の場合は個人ごとに独立な1変量の最適化

◆ 非計量多次元尺度法

- ★ 最適変換が所与の場合、座標 X を計量的 MDS として求める。
- ★ X が所与の場合、データをの最適変換を最小2乗単調回帰で求める。

◆ ニューラルネット

- ★ 各層の重みごとに推定

