

## 数量化Ⅳ類

n個の対象の間に親近性が定義できるとして、 $e_{ij}$  を対象 i と対象 j との間の親近性とする。

ここでは  $e_{ij} > 0$  とする。 親近性が強いものほど  $e_{ij}$  は値が大きいものとする。

各対象 i について数量  $x_i$  を与える。

数量の付与基準は、

- ・ 親近性の高いペア(i,j)には  $\|x_i - x_j\|^2$  が小さく
- ・ 親近性の低いペア(i,j)には  $\|x_i - x_j\|^2$  が大きく

なるように決める。

つまり、

$$Q \equiv - \sum_{i=1}^n \sum_{j=1}^n e_{ij} (x_i - x_j)^2$$

$$\text{但し } \bar{x} = \sum_{i=1}^n x_i = 0, \quad \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i)^2 = 1$$

として Q を最大化する問題となる。

Lagrange の未定乗数法を用い、

$$G = - \sum_{i=1}^n \sum_{j=1}^n e_{ij} (x_i - x_j)^2 - \lambda \left( \sum_{i=1}^n x_i^2 - 1 \right) \quad \text{として、}$$

$$\frac{\partial G}{\partial x_\ell} = 0 \quad (\ell = 1 \dots n) \quad \text{となる} \quad \{x_\ell\} \quad \text{を求める。}$$

すなわち  $\ell = 1 \dots n$  について

$$\begin{aligned} \frac{\partial G}{\partial x_\ell} &= - \sum_{i=1}^n \sum_{j=1}^n \frac{\partial}{\partial x_\ell} \{e_{ij} (x_i - x_j)^2\} - \lambda \left( \sum_{i=1}^n \frac{\partial}{\partial x_\ell} x_i^2 \right) \\ &= - \sum_{j=1}^n \frac{\partial}{\partial x_\ell} \{e_{\ell j} (x_\ell - x_j)^2\} - \sum_{i=1}^n \frac{\partial}{\partial x_\ell} \{e_{i\ell} (x_i - x_\ell)^2\} - \lambda (2x_\ell) \\ &= - 2 \sum_{j=1}^n e_{\ell j} (x_\ell - x_j) + 2 \sum_{i=1}^n e_{i\ell} (x_i - x_\ell) - 2\lambda x_\ell = 0 \end{aligned}$$

したがって

$$-\sum_{j=1}^n e_{\ell j} (x_{\ell} - x_j) + \sum_{i=1}^n e_{i\ell} (x_i - x_{\ell}) = \lambda x_{\ell} \quad (\ell = 1 \dots n)$$

を得る。

$\ell = 1 \dots n$  について、

$$\sum_{j=1}^n e_{\ell j} (x_j - x_{\ell}) + \sum_{i=1}^n e_{i\ell} (x_i - x_{\ell}) = \lambda x_{\ell} \quad (\text{左辺第1項符号を逆に})$$

$$\sum_{i=1}^n e_{\ell i} (x_i - x_{\ell}) + \sum_{i=1}^n e_{i\ell} (x_i - x_{\ell}) = \lambda x_{\ell} \quad (\text{左辺第1項符号添字を } j \rightarrow i)$$

$$\sum_{i=1}^n (e_{\ell i} + e_{i\ell}) (x_i - x_{\ell}) = \lambda x_{\ell} \quad \text{が成り立つ。}$$

$a_{\ell i} = e_{\ell i} + e_{i\ell}$  として、 $\ell = 1 \dots n$  について

$$\begin{aligned} \sum_{i=1}^n a_{\ell i} (x_i - x_{\ell}) &= \lambda x_{\ell} \\ \sum_{i=1}^n a_{\ell i} x_i - \sum_{i=1}^n a_{\ell i} x_{\ell} &= \lambda x_{\ell} \\ \sum_{i=1}^n a_{\ell i} x_i - x_{\ell} \sum_{i=1}^n a_{\ell i} &= \lambda x_{\ell} \end{aligned}$$

上式で  $i = \ell$  の時、左辺第1項は  $a_{\ell\ell} x_{\ell}$  であり、左辺第2項は  $-x_{\ell} a_{\ell\ell}$  であるから  $a_{\ell\ell} x_{\ell} - x_{\ell} a_{\ell\ell} = 0$  となる。

従って  $\ell = 1 \dots n$  について

$$\begin{aligned} \sum_{i \neq \ell} a_{\ell i} x_i - x_{\ell} \sum_{i \neq \ell} a_{\ell i} &= \lambda x_{\ell} \\ -\left(\sum_{i \neq \ell} a_{\ell i}\right) x_{\ell} + \sum_{i \neq \ell} a_{\ell i} x_i &= \lambda x_{\ell} \end{aligned}$$

となるので、

$$\begin{bmatrix} -\left(\sum_{k \neq 1} a_{1k}\right) & a_{12} & \cdots & a_{1n} \\ a_{21} & -\left(\sum_{k \neq 2} a_{2k}\right) & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & -\left(\sum_{k \neq n} a_{nk}\right) \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

と表現される。

すなわち

$$i \neq j \text{ の時は } h_{ij} = a_{ij} \quad , \quad i = j \text{ の時は } h_{ij} = - \sum_{k \neq i} a_{ki}$$

を成分とする行列を  $H$  とすると、 $H\mathbf{x} = \lambda\mathbf{x}$  となる固有値問題となる。  
すなわち固有方程式の最大の固有値に対応する固有ベクトルを求めれば  
良い。

この時、 $\mathbf{x}^t H \mathbf{x} = \lambda \mathbf{x}^t \mathbf{x} = \lambda$  となる。

上記で求めた 1 次元の数量により、対象間の関係が十分に説明できない  
場合には、2 次元の数量化を行う。

これを解くには、 $H\mathbf{x} = \lambda\mathbf{x}$  の 2 番目に大きい固有値の固有ベクトルを  
求めれば良い。

それでも、説明が十分でなければ、3 番目に大きい固有値の固有ベクトル  
というように計算する。