統計解析II 88 ~ 100

33B21009 羽田充宏

(88)

(a) 第一主成分ベクトルは

$$\max \quad \|Xv\|^2$$

$$\text{s.t.} \quad \|v\|^2 = 1$$

の解であるのでラグランジュ緩和問題で考えると,

$$\max \quad \|Xv\|^2 - \mu(\|v\|^2 - 1)$$

vで微分して

$$2X^T X v - 2\mu v = 0$$

ここで $\lambda = \frac{\mu}{N}$ とすれば

$$\frac{1}{N} X^T X v = \lambda v$$

より第一主成分ベクトルは標本共分散行列の固有値となる.

(b) 半正定値行列に対して

$$\mathbb{R} \ni v_j^T X^T X v_i = \lambda_i v_j^T v_i$$

等しい

$$v_i^T X^T X v_j = \lambda_j v_i^T v_j$$

$$(\lambda_i - \lambda_j)\underbrace{\langle v_i, v_j \rangle}_{\text{内積}} = 0$$

従って $\lambda_i \neq \lambda_j$ ならば v_i と v_j は直交する.

また $\lambda_i = \lambda_j$ であってもグラムシュミットの直交化を行うことで固有ベクトルを直交するようにとることができる.

(c) 任意の実対称行列は直交行列によって対角化できるので
$X^T X$ は直交行列 P を用いて

$$P^T X^T X P = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_p \end{pmatrix}$$

と表せる。
ただし $\mu_1 \geq \cdots \geq \mu_p$ とする。

ここで i 番目に大きい固有値に対する固有ベクトル v_i として i 番目主成分を求めるためには、

$$\max_{\substack{\|v\|=1 \\ v_1 \perp \cdots \perp v_{i-1} \perp v}} v^T I v$$

$$= \max \; v^T (v_1 \cdots v_p) \begin{pmatrix} \mu_1 & & & & \\ & \ddots & & & \\ & & \mu_i & & \\ & & & \ddots & \\ & & & & \mu_p \end{pmatrix} (v_1^T \cdots v_p^T) v$$

の問題をとけばよく これは、
最適解 v_i, 最適値 μ_i であるので
「第1〜第$i-1$主成分ベクトルが直交」を無視して単純に対角化すればよい。

(19) $v_1 \perp \cdots \perp v_m$ であるので
$V_m = [v_1 \cdots v_m]$ として

$$V_m^T V_m = \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix} [v_1 \cdots v_m] = I \quad \text{よって}$$

$$(I - V_m V_m^T)^2 = (I - V_m V_m^T)(I - V_m V_m^T)$$

$$= I - V_m V_m^T \cancel{- V_m V_m^T} + V_m \cancel{V_m^T V_m} V_m^T$$

$$= I - V_m V_m^T$$

$$\min \sum_{i=1}^{N} x_i (I - V_m V_m^T) x_i^T$$

$\Longleftrightarrow$

$$\min_{V_m} \ \mathrm{trace}\left(X (I - V_m V_m^T) X^T \right)$$

$$\max_{V_m} \ \mathrm{trace}\left(X V_m V_m^T X^T \right)$$

$$= \mathrm{trace}\left(V_m^T X^T X V_m \right)$$

$$= \sum_{j=1}^{m} v_j^T (X^T X) v_j$$

$$= \sum_{j=1}^{m} \| X v_j \|^2$$

従って再構成誤差を最小にする V_m を求めることと主成分ベクトルを求めることは同じ操作であるといえる.

(10)

$$\max_{\|v\|_2 = 1} \{ v^T X^T X v - \lambda \|v\|_1 \} \qquad (7.30)$$

$$\max_{\|u\|_2 = \|v\|_2 = 1} \{ u^T X v - \lambda \|v\|_1 \} \qquad (7.31)$$

(a)

(7.31) のラグランジュ関数を考えて

$$L := - u^T X v + \lambda \|v\|_1 + \frac{\mu}{2} (u^T u - 1) + \frac{\delta}{2} (v^T v - 1)$$

$$\frac{\partial L}{\partial u} = - X v + \mu u = 0$$

$$\mu u = X v \qquad \|u\| = 1 \text{ より} \qquad u = \frac{X v}{\|X v\|_2}$$

改めて

$$L = -\|Xu\|_2 + \lambda\|v\|_1 + \frac{\delta}{2}(v^T v - 1)$$ が得られる.

これを最小化して得られる最適解は

$$\max_{\|v\|_2=1} \{v^T X^T X v - \lambda\|v\|_1\}$$ の最適解でもある言える.

(b) $\frac{\partial L}{\partial u} = 0$ より $u = \frac{Xv}{\|Xv\|}$ は (a) で示した。

$$\frac{\partial L}{\partial v_j} = (X^T u)_j - \lambda \partial_j \|v\|_1 + \delta v_j$$

$\frac{\partial L}{\partial v_j} = 0$ を考えると,

$$0 \in (X^T u)_j + \begin{cases} +\lambda + \delta v_j & (X^T u)_j \geq \lambda \\ +\lambda[-1,1] & -\lambda \leq (X^T u)_j \leq \lambda \\ -\lambda + \delta v_j & (X^T u)_j \leq -\lambda \end{cases}$$

つまり $\delta v = S_\lambda(X^T u)$

以上より

$$v = \frac{S_\lambda(X^T u)}{\|S_\lambda(X^T u)\|_2}$$ が成り立つ.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (a-\lambda)(d-\lambda) - bc$$

$$\lambda^2 - \underbrace{(a+d)}_{Tr}\lambda + \det$$

(91)

$$L(u,v) = -u^T X v + \lambda \|v\|_1 + \frac{\mu}{2}(u^T u - 1) + \frac{\delta}{2}(v^T v - 1)$$

(a)

$$\left.\begin{array}{l} \dfrac{\partial^2 L}{\partial u_i^2} = \mu \\[2em] \dfrac{\partial^2 L}{\partial v_i^2} = \delta \end{array}\right\} \quad \text{従って (7.32) は } u, v \text{ に関して双凸である.}$$

(b) $N = P = 1 \quad X > \sqrt{\mu\delta}$ のとき.

$$\nabla^2 L = \begin{bmatrix} \frac{\partial^2 L}{\partial u^2} & \frac{\partial^2 L}{\partial u \partial v} \\ \frac{\partial^2 L}{\partial u \partial v} & \frac{\partial^2 L}{\partial v^2} \end{bmatrix} = \begin{bmatrix} \mu & X \\ X & \delta \end{bmatrix}$$

$\det(\nabla^2 L) < 0$ であれば固有値の積が負であり, 負の固有値を含んでおり凸でなくなるが

$$\det(\nabla^2 L) = \mu\delta - X^2$$

$$X^2 - \mu\delta > 0 \text{ の時負となり.}$$

双凸 => 凸 は一般には成り立たない.

(92)

$\Psi: \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}$ が $f: \mathbb{R}^P \to \mathbb{R}$ に対して

$$\begin{cases} f(\beta) \le \Psi(\beta, \theta), & \theta \in \mathbb{R}^P \\ f(\beta) = \Psi(\beta, \beta) \end{cases}$$

の時 β において Ψ は f より優勢という.

(a) Ψ が f より優勢である時, $\beta^0 \in \mathbb{R}^P$ を選び

$$\beta^{t+1} = \underset{\beta \in \mathbb{R}^P}{\arg\min}\ \Psi(\beta, \beta^t) \quad \text{と更新する時.}$$

まず定義より

$$f(\beta^t) = \Psi(\beta^t, \beta^t) \ge \underbrace{\Psi(\beta^{t+1}, \beta^t)}_{\text{更新式の定義より}} \ge \underbrace{f(\beta^{t+1})}_{\text{優勢の定義より.}}$$

(b)

$$f(u) := -\|Xu\|_2 + \lambda\|u\|_1$$

$$\Phi(u, u') := -\frac{(Xu)^T(Xu')}{\|Xu'\|_2} + \lambda\|u\|$$ とする.

Xu, Xu' に対してシュワルツの不等式を考えると,

$$|\langle Xu, Xu'\rangle|^2 \leq \|Xu\|_2^2 \cdot \|Xu'\|_2^2$$

$$(Xu)^T(Xu') \leq \|Xu\|_2 \|Xu'\|_2$$

$$-\frac{(Xu)^T(Xu')}{\|Xu'\|} + \lambda\|u\|_1 \geq -\|Xu\|_2 + \lambda\|u\|_1$$

またシュワルツの不等式より等号成立は $u = u'$ であるので
左辺を Φ, 右辺を f とすれば

$$\begin{cases} f(u) \leq \Phi(u, u') \\ f(u) = \Phi(u, u) \end{cases}$$ をみたす.

$X \in \mathbb{R}^{N \times P}$

(c) u^0 を任意に定め,

$$u^{t+1} := \underset{\|u\|_2 = 1}{\arg\max}\ \Phi(u, u^t)$$

$1 \times N$

ここで $$\Phi(u, u^t) = -\frac{(Xu^t)^T}{\|Xu^t\|}(Xu) + \lambda\|u\|_1$$

より

$\frac{\partial \Phi}{\partial u} = 0$ となる u を求めればよい

$u = Xu$.

$$\frac{\partial \Phi}{\partial u} = -\frac{X^T X u^t}{\|Xu^t\|} + \lambda\partial\|u\|_1$$

$$u^{t+1} = \frac{S_\lambda(X^T X u^t)}{\|Xu^t\|}$$

よって $$(\text{右辺}) = \frac{S_\lambda(X^T u)}{\|S_\lambda(X^T u)\|_2}$$

(94)

$$\min_{\substack{u,v \in \mathbb{R}^p \\ \|u\|_2 = 1}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \| x_i - x_i v u^T \|_2^2 + \lambda_1 \|v\|_1 + \lambda_2 \|v\|_2^2 \right\}$$

(a) ラグランジュ関数を考えて

$$L := \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T - \frac{2}{N} \sum_{i=1}^{N} u^T x_i^T x_i v + \frac{1}{N} \sum_{i=1}^{N} v^T x_i^T x_i v$$

$$+ \lambda_1 \|v\|_1 + \lambda_2 \|v\|_2 + \mu (u^T u - 1)$$

$$\frac{\partial^2 L}{\partial u_k^2} = \mu \qquad \frac{\partial^2 L}{\partial v_k^2} = \frac{2}{N} x_k^2 + \lambda_2$$

よって L は u, v に関して双凸である.

(b)

$$\frac{\partial L}{\partial u_j} = -\frac{2}{N} \sum_{i=1}^{N} x_i^T x_i v + 2\mu u = 0$$

$$\begin{pmatrix} z_1 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} x_1 v \\ \vdots \\ x_N v \end{pmatrix}$$ とおけば $\|u\|_2 = 1$ より

$$u = \frac{X^T z}{\|X^T z\|_2}$$ となる.

(96) $\|v_k\|_2 \le 1$, $\|v_k\|_1 \le c$, $\|v_k\|_2 \le 1$, $u_k^T u_j = 0 \ (j=1,2,\cdots k-1)$
のもとで
$u_k^T X v_k$ の最大化をはかる.

ここで
$$P_{k-1}^{\perp} := I - \sum_{i=1}^{k-1} u_i u_i^T$$ とすれば

$$u_j P_{k-1}^{\perp} X v_k = u_j^T \left(I - \sum_{i=1}^{k-1} u_i u_i^T \right) X v_k$$
$$= \left(u_j^T - \sum_{i=1}^{k-1} u_j^T u_i u_i^T \right) X v_k$$
$$= 0$$
$$\because \ u_i^T u_j = \delta_{ij}$$

また
$L = u_k^T X v_k - \mu (u_k^T u_k - 1)$ を最大にする u_k は
$$\frac{\partial L}{\partial u_k} = X v_k - 2 \mu u_k = 0$$

$\|u_k\| \le 1$ をみたすように
$$u_k = \frac{X v_k}{\| u_k^T X v_k \|}$$ と表せる

従って制約をみたしつつ
L を最大にする u_k は

$$u_k = \frac{P_{k-1}^{\perp} X v_k}{\| P_{k-1}^{\perp} X v_k \|_2}$$ となる ■