

クロス表のSVDおよび内積による表現と座標変換

グラフ図示による例示

改編版

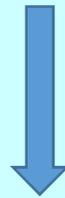
岡本安晴

(日本女子大学)

クロス表
カテゴリ変数2個の場合

		列カテゴリ		
		C_1	...	C_N
行 カ テ ゴ リ	R_1	x_{11}	...	x_{1N}
	\vdots	\vdots		\vdots
	R_M	x_{M1}	...	x_{MN}

$X = (x_{ij})$ 配列データ



情報の圧縮 (機械学習)

SVD (Singular Value Decomposition: 特異値分解)

特異値分解

$$X = U\Lambda V' = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_r) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_r \end{pmatrix} \begin{pmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_r \end{pmatrix} = \sum s_i \mathbf{u}_i \mathbf{v}'_i$$

$$\text{列ベクトル } \mathbf{x}_{\cdot j} = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{Mj} \end{pmatrix} = \sum_h \mathbf{u}_h s_h v_{jh} = \sum_h s_h v_{jh} \mathbf{u}_h, \quad \mathbf{v}_h = \begin{pmatrix} v_{1h} \\ \vdots \\ v_{Nh} \end{pmatrix}$$

$\mathbf{x}_{\cdot j}$ の要素 x_{ij} は、 \mathbf{u}_h の要素 u_{ih} の重み $s_h v_{jh}$ による加重和

すなわち、行カテゴリ変数 $(R_1 \ \cdots \ R_M)'$ の値は $\mathbf{u}_1, \dots, \mathbf{u}_r$ の加重和で表される。

この意味で、 $\mathbf{u}_1, \dots, \mathbf{u}_r$ は行カテゴリ変数の情報を含んでいる。

列カテゴリ変数 $(C_1 \ \cdots \ C_N)$ についても同様に考えて、

$\mathbf{v}_1, \dots, \mathbf{v}_r$ は列カテゴリ変数 $(C_1 \ \cdots \ C_N)$ の情報を含んでいる。

Cf. Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering*. Cambridge Univ. Press.

岡本安晴「官能評価に使える統計解析ツール Python編
日本官能評価学会誌、2022、26、2、102-106.

http://y-okamoto-psy1949.la.coccan.jp/booksetc/CataDataSVD/26-2_clm3.pdf

岡本（2022）図 3 CATA データ（島村ら，2017, Table 2 より）

岡本（2022）図 4 CATA データ（図 3）の分析結果

島村 綾、小泉昌子、峯木眞知子、市原 茂（2017） 飯の官能評価の時系列変化。
日本家政学会誌、68、9、478-485.

クロス表



SVD

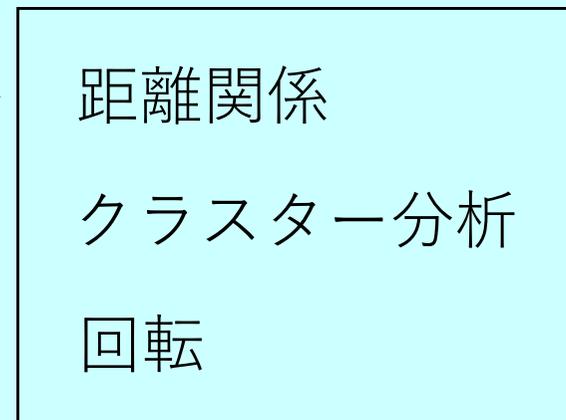


岡本安晴「官能評価に使える統計解析ツール Python編」
日本官能評価学会誌、2022、26、2、102-106.



岡太彬訓（2022、私信）

SVD → 内積モデル → 布置 →



$$X = \sum s_i \mathbf{u}_i \mathbf{v}_i' \quad s_1 \geq \dots \geq s_r > 0$$

$$X = (\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_r \end{pmatrix} \begin{pmatrix} \mathbf{v}_1' \\ \vdots \\ \mathbf{v}_r' \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_M \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}'$$

$$\begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_M \end{pmatrix} = (s_1^{\alpha_1} \mathbf{u}_1 \quad \dots \quad s_r^{\alpha_r} \mathbf{u}_r), \quad \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix} = (s_1^{\beta_1} \mathbf{v}_1 \quad \dots \quad s_r^{\beta_r} \mathbf{v}_r), \quad \alpha_t + \beta_t = 1$$

$$X = (x_{ij}) = (\mathbf{a}_i \mathbf{b}_j')$$

内積モデル

$$SS_{t,u} = \|s_t^{\alpha_t} \mathbf{u}_t\|^2 = s_t^{2\alpha_t}, \quad SS_{t,v} = \|s_t^{\beta_t} \mathbf{v}_t\|^2 = s_t^{2\beta_t}$$

$$SS_{t,u}/M = SS_{t,v}/N$$

$$\alpha_t = \frac{1}{4} \frac{\log(s_t^2 M/N)}{\log s_t}, \quad \beta_t = \frac{1}{4} \frac{\log(s_t^2 N/M)}{\log s_t}$$

クロス表における距離と布置における距離

行カテゴリ間

$$X = (\mathbf{a}_i \mathbf{b}'_j) = \begin{pmatrix} \vdots \\ \mathbf{x}_i. \\ \vdots \end{pmatrix}, \quad \mathbf{x}_i. = (\mathbf{a}_i \mathbf{B}'), \quad B = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}$$

$$\|\mathbf{x}_i. - \mathbf{x}_h.\|^2 = \|(\mathbf{a}_i - \mathbf{a}_h) \mathbf{B}'\|^2 = (\mathbf{a}_i - \mathbf{a}_h) \mathbf{B}' \mathbf{B} (\mathbf{a}_i - \mathbf{a}_h)' = \|\mathbf{a}_i - \mathbf{a}_h\|_{\mathbf{B}' \mathbf{B}}^2$$

$$\mathbf{B}' \mathbf{B} = \begin{pmatrix} s_1^{\beta_1} \mathbf{v}_1 & \cdots & s_r^{\beta_r} \mathbf{v}_r \end{pmatrix}' \begin{pmatrix} s_1^{\beta_1} \mathbf{v}_1 & \cdots & s_r^{\beta_r} \mathbf{v}_r \end{pmatrix} = \text{diag}(s_1^{2\beta_1} \quad \cdots \quad s_r^{2\beta_r})$$

列カテゴリ間

$$X = (\mathbf{a}_i \mathbf{b}'_j) = (\cdots \quad \mathbf{x}_{.j} \quad \cdots), \quad \mathbf{x}_{.j} = (\mathbf{A} \mathbf{b}'_j), \quad A = \begin{pmatrix} \vdots \\ \mathbf{a}_i \\ \vdots \end{pmatrix}$$

$$\|\mathbf{x}_{.j} - \mathbf{x}_{.h}\|^2 = \|A(\mathbf{b}'_j - \mathbf{b}'_h)\|^2 = \left(A(\mathbf{b}'_j - \mathbf{b}'_h) \right)' \left(A(\mathbf{b}'_j - \mathbf{b}'_h) \right) = \|\mathbf{b}'_j - \mathbf{b}'_h\|_{A' A}^2$$

$$A' A = \begin{pmatrix} s_1^{\alpha_1} \mathbf{u}_1 & \cdots & s_r^{\alpha_r} \mathbf{u}_r \end{pmatrix}' \begin{pmatrix} s_1^{\alpha_1} \mathbf{u}_1 & \cdots & s_r^{\alpha_r} \mathbf{u}_r \end{pmatrix} = \text{diag}(s_1^{2\alpha_1} \quad \cdots \quad s_r^{2\alpha_r})$$

行カテゴリと列カテゴリ間の距離と内積

$$\|\mathbf{a}_i - \mathbf{b}_j\|^2 = \|\mathbf{a}_i\|^2 + \|\mathbf{b}_j\|^2 - 2\mathbf{a}_i\mathbf{b}_j'$$

$$= \|\mathbf{a}_i\|^2 + \|\mathbf{b}_j\|^2 - 2x_{ij}$$

$$x_{ij} = \frac{\|\mathbf{a}_i\|^2 + \|\mathbf{b}_j\|^2 - \|\mathbf{a}_i - \mathbf{b}_j\|^2}{2}$$

$$x_{ij} = \mathbf{a}_i\mathbf{b}_j' = \|\mathbf{a}_i\| \|\mathbf{b}_j\| \cos \theta_{ij}$$

$$\theta_{ij} = \cos^{-1} \frac{x_{ij}}{\|\mathbf{a}_i\| \|\mathbf{b}_j\|}$$

表 1. クロス表-1

	C1	C2	C3	C4	C5
R1	10	5	0	0	0
R2	5	10	5	0	0
R3	0	5	10	3	0
R4	0	0	3	10	5
R5	0	0	0	5	10

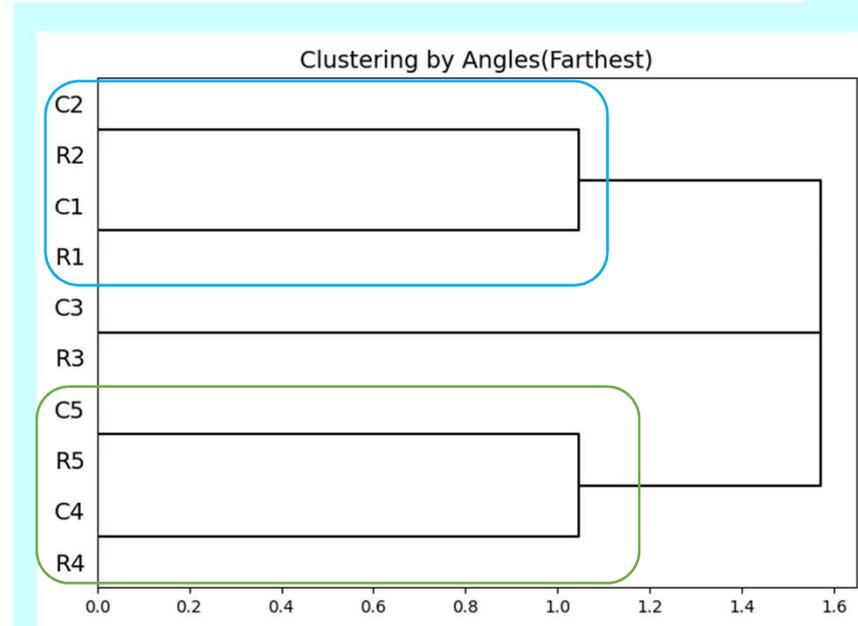
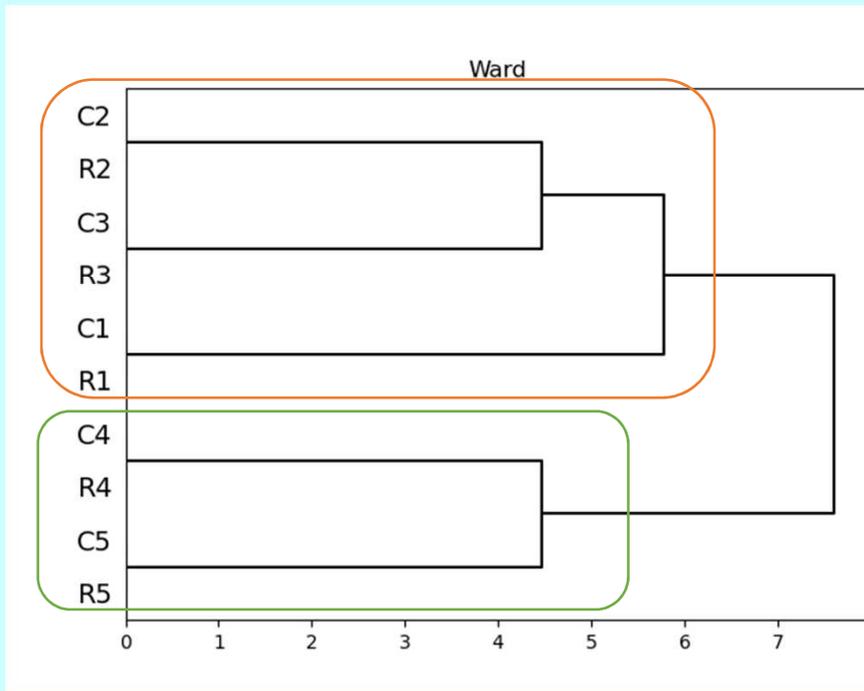
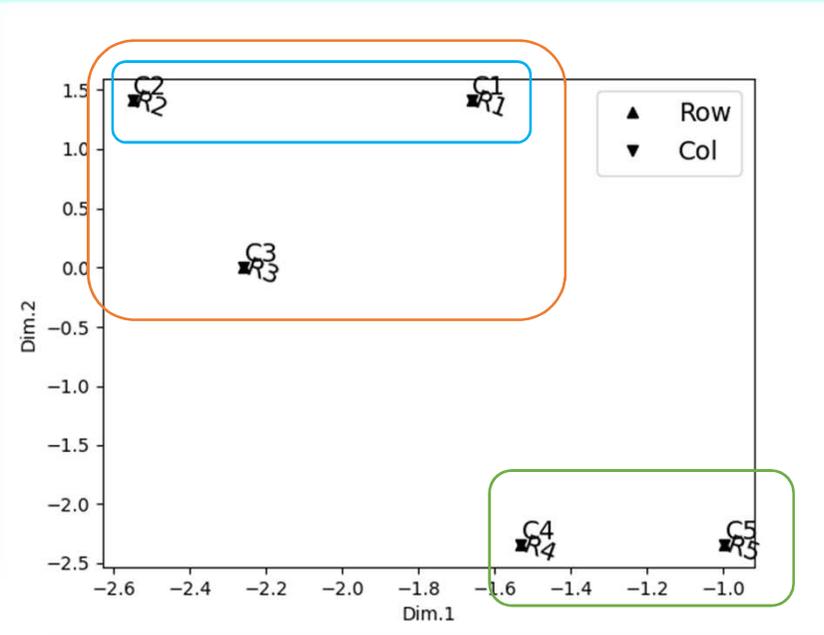
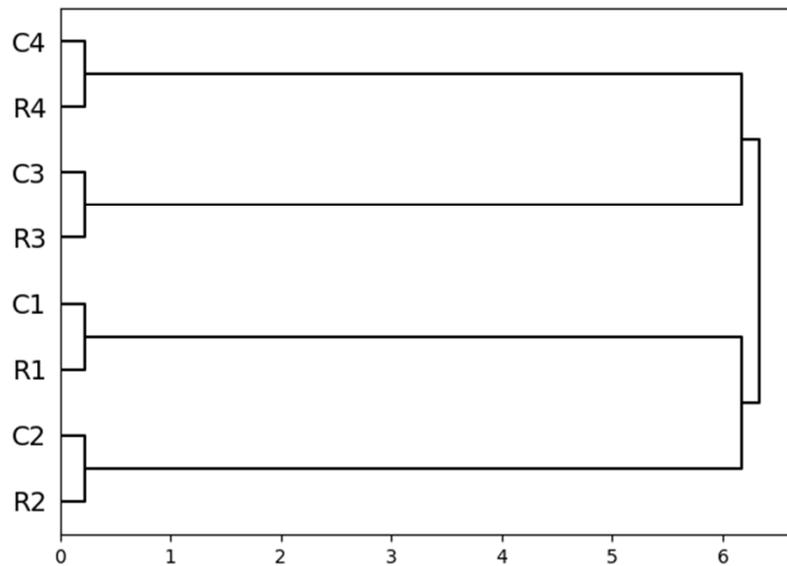
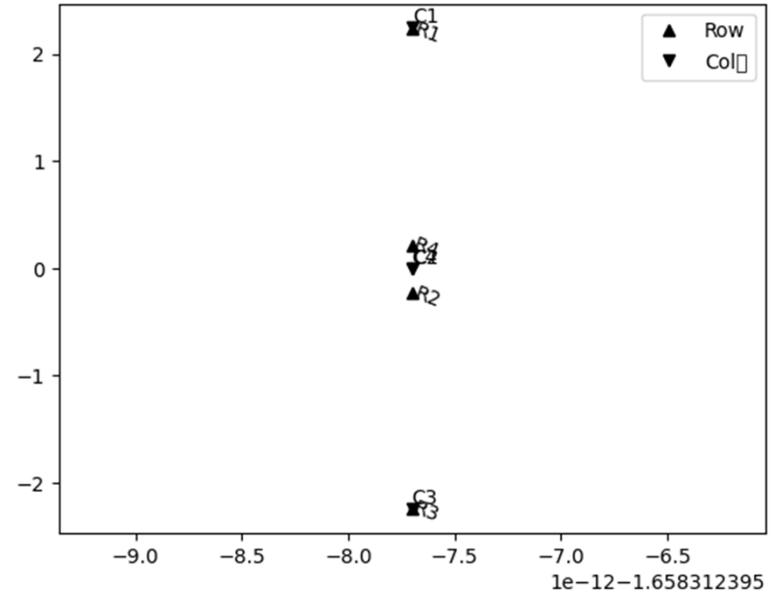
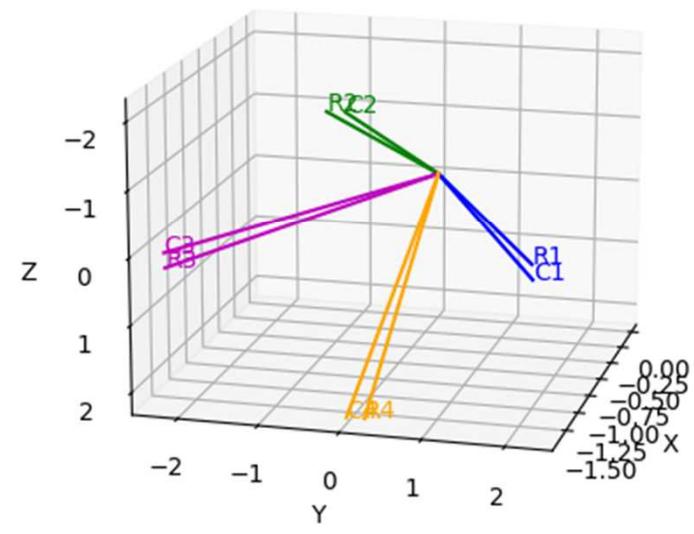
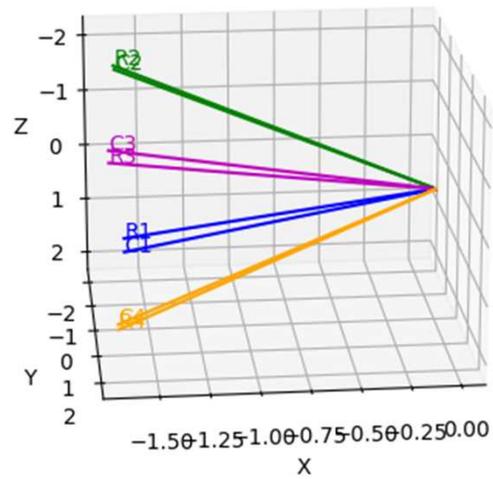


表2 クロス表-2

	C1	C2	C3	C4
R1	10	1	0	0
R2	0	10	1	0
R3	0	0	10	1
R4	1	0	0	10





内積モデルの座標軸の回転

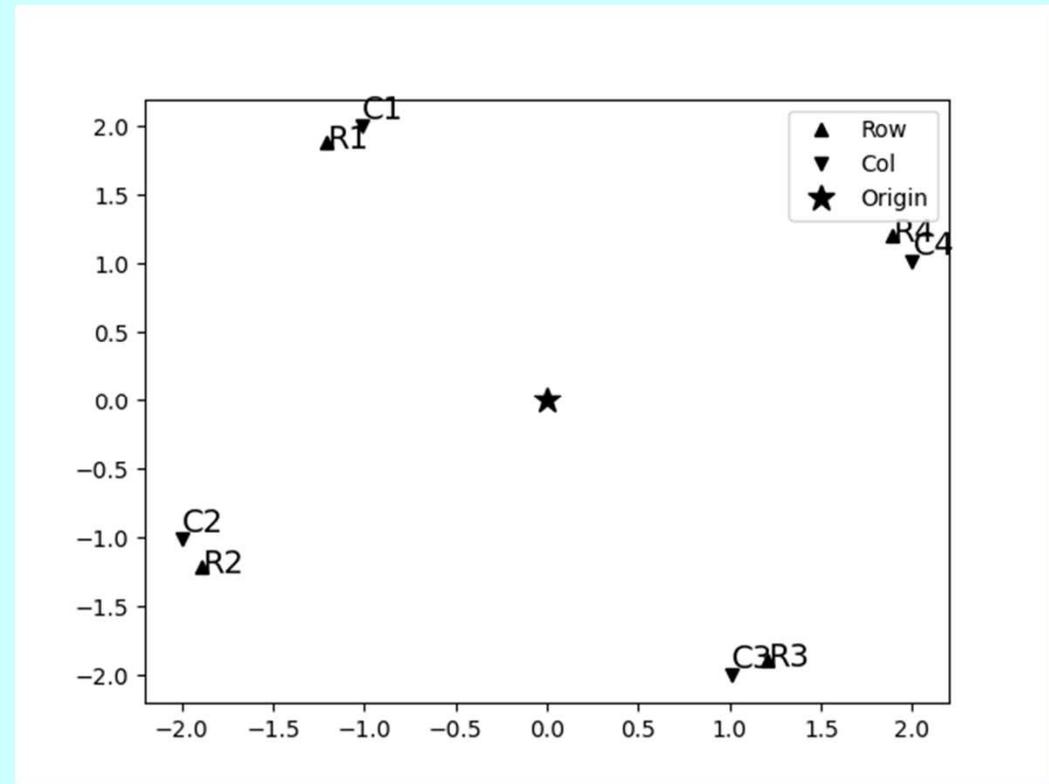
$$Y = \begin{pmatrix} A \\ B \end{pmatrix} \quad Z = Y - 1_{M+N}\bar{Y} \quad \bar{Y} = \frac{1}{M+N} 1'_{M+N}Y$$

$$\mathbf{w}_1 \quad \cdots \quad \mathbf{w}_r = \text{Arg} \max_{\mathbf{w}, \mathbf{w}'\mathbf{w}=1} (Z\mathbf{w}')'(Z\mathbf{w}'), \quad \mathbf{w}_s \perp \mathbf{w}_t$$

$$P = (\mathbf{w}'_1 \quad \cdots \quad \mathbf{w}'_r)$$

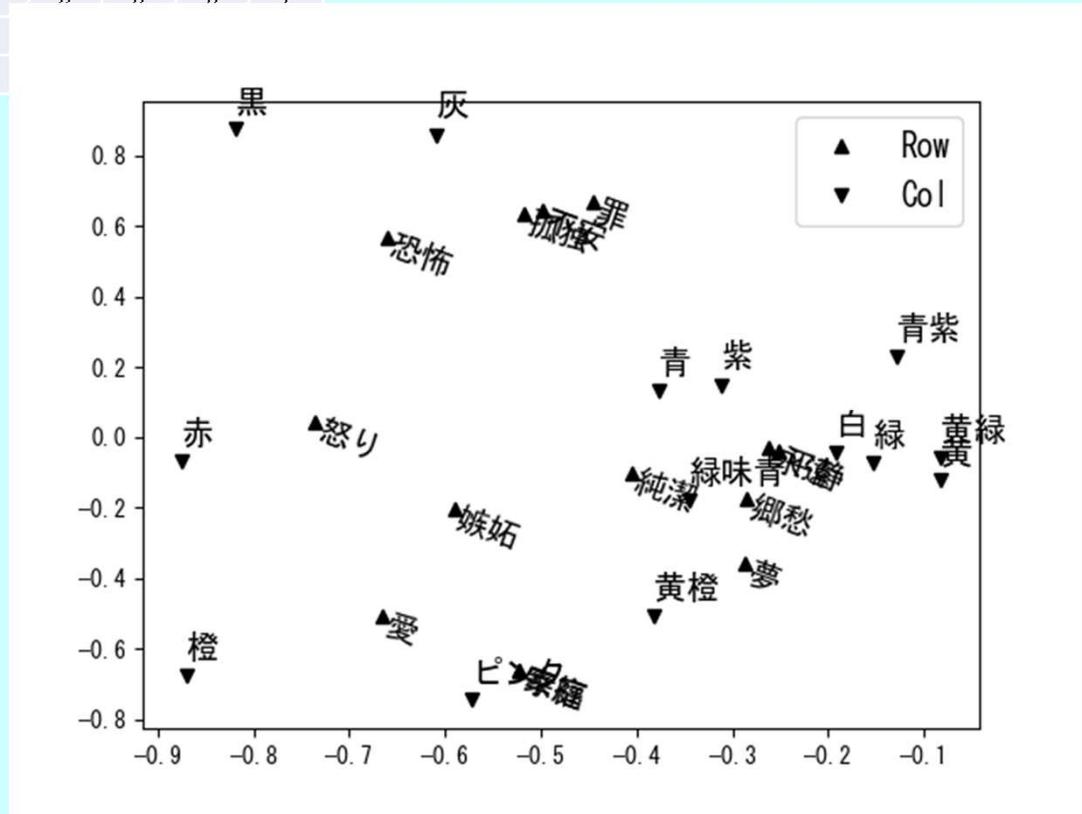
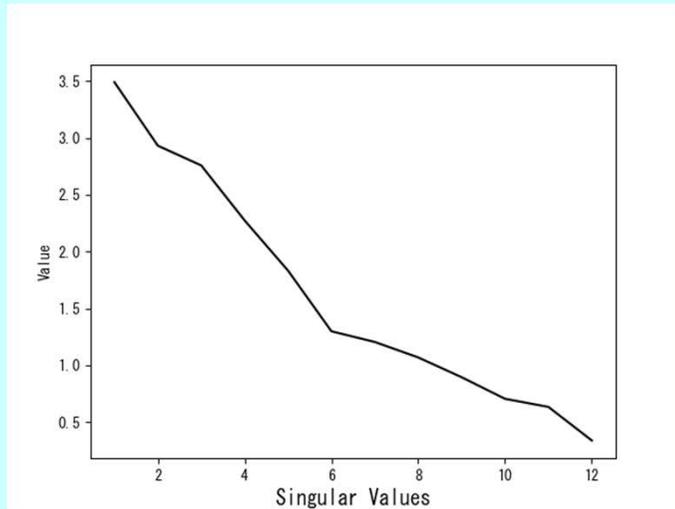
$$\mathbf{a}_i P = (\mathbf{a}_i \mathbf{w}'_1, \dots, \mathbf{a}_i \mathbf{w}'_r)$$

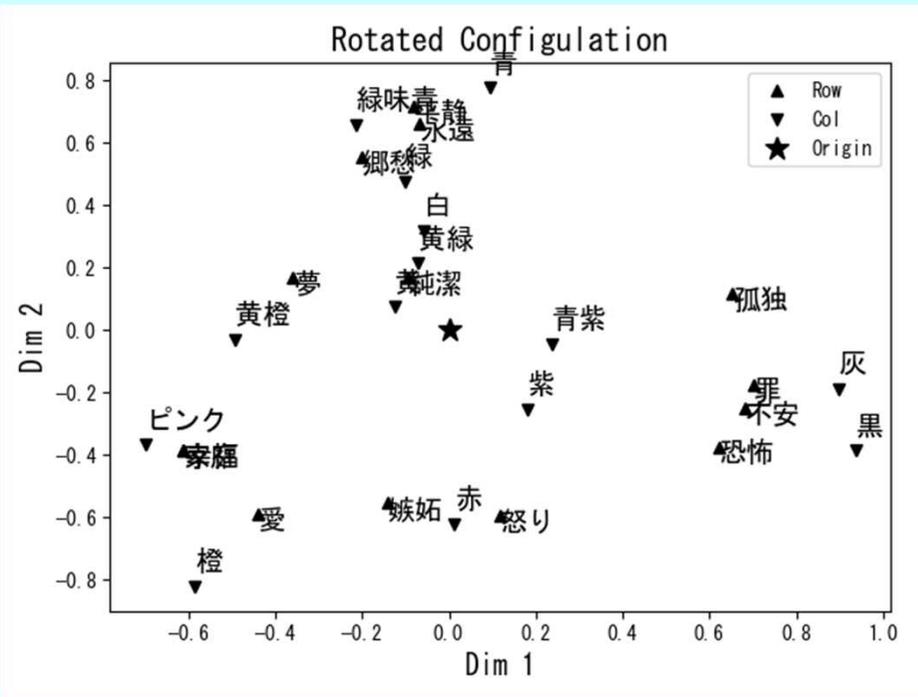
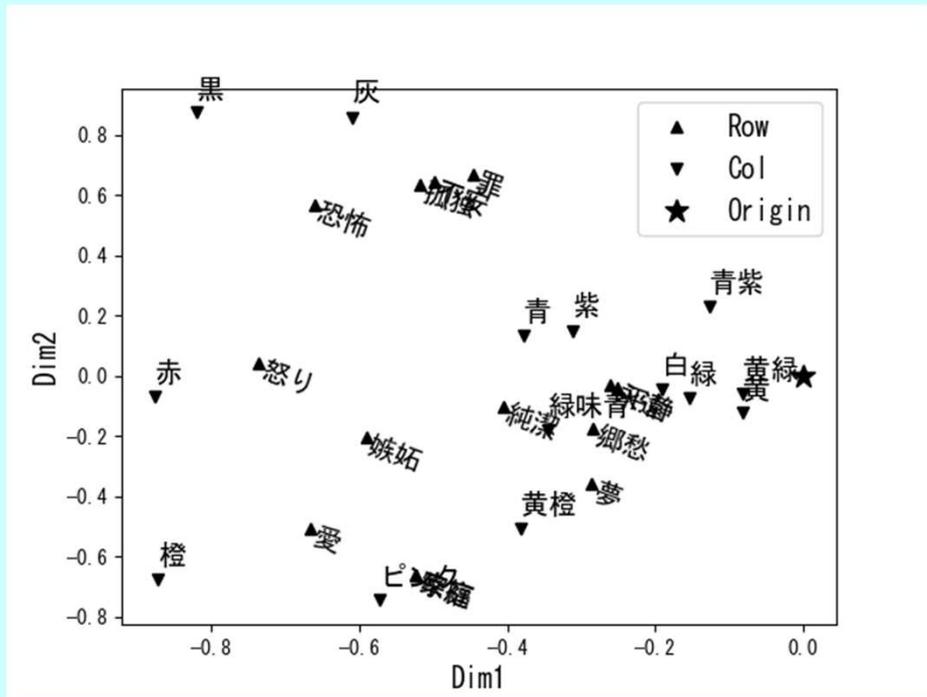
$$\mathbf{b}_j P = (\mathbf{b}_j \mathbf{w}'_1, \dots, \mathbf{b}_j \mathbf{w}'_r)$$



	赤	橙	黒	紫	灰	青紫	白	緑味青	ピンク	黄橙	青	緑	黄緑	黄
怒り	1	1	1	0	0	0	0	0	0	0	0	0	0	0
嫉妬	1	1	0	1	0	0	0	0	0	0	0	0	0	0
罪	0	0	1	0	1	1	0	0	0	0	0	0	0	0
永遠	0	0	0	0	0	0	1	1	0	0	1	0	0	0
幸福	0	1	0	0	0	0	0	0	1	1	0	0	0	0
孤独	0	0	1	0	1	0	0	0	0	0	1	0	0	0
平静	0	0	0	0	0	0	0	1	0	0	1	1	0	0
郷愁	0	0	0	0	0	0	0	0	0	1	1	1	1	0
家庭	0	1	0	0	0	0	0	0	1	1	0	0	0	0
愛	1	1	0	0	0	0	0	0	1	0	0	0	0	0
純潔	1	0	0	0	0	0	1	1	0	0	0	0	0	0
夢	0	0	0	0	0	0	0	1	1	0	0	0	0	1
不安	0	0	1	1	1	0	0	0	0	0	0	0	0	0
恐怖	1	0	1	0	1	0	0	0	0	0	0	0	0	0

表14.1.1 象徴語と色の適合関係
 (岡本「心理学データ分析と測定」2014)
 注：大山「色彩心理学入門」(1994)から
 適合するときを1、適合しないときを0
 で表したものの。





大山 正・岩脇三良・宮埜嘉夫（2005）「心理学研究法」サイエンス社
 図4-4：「幸福」に対する4地域の大学生のセマンティック・プロフィール
 より

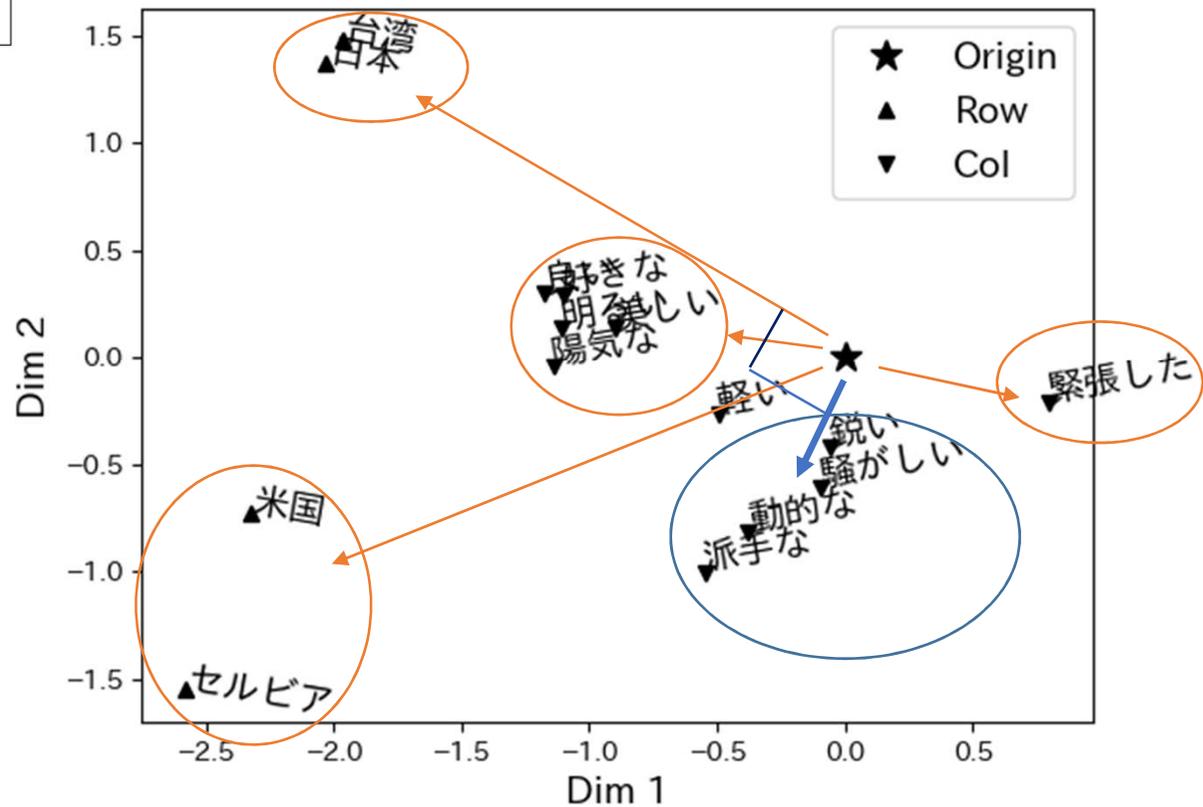
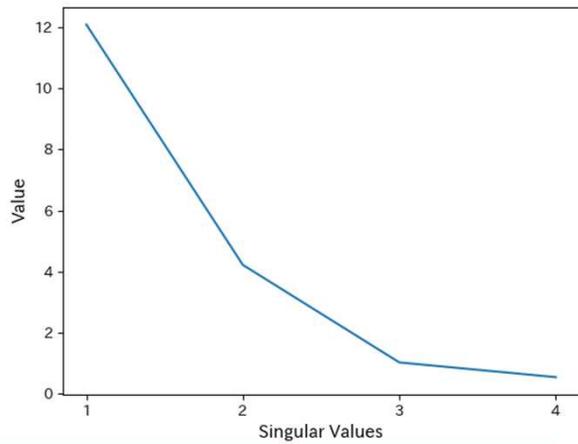
	良い —わるい	好きな —嫌いな	美しい —汚い	騒がしい —静かな	動的な —静的な	派手な —地味な	軽い —重い	明るい —暗い	陽気な —陰気な	緊張した —ゆるんだ	鋭い —鈍い
日本	2.75	2.5	1.8	-0.6	-0.7	0	0.75	2.75	2.2	-2	-0.6
セルビア	2.4	2.45	2.3	1.3	2.3	2.75	1.8	2.7	3	-1.7	0.95
台湾	2.74	2.75	2.25	-0.7	-0.1	-0.75	0.5	2.1	2.25	-1.8	-0.3
米国	2.74	2.3	1.75	0.5	1.4	2.3	1.2	2.4	2.7	-1.75	0.25

評定値例

良い：

非常に良い（3）、かなり良い（2）、やや良い（1）、
 どちらでもない（0）

やや悪い（-1）、かなり悪い（-2）、非常に悪い（-3）



SVD・内積モデル	対応分析
数値データ	度数データ
前処理なし	前処理 $S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$

対応分析とSVDによる内積モデル

	A	B	C	D	
1	Names	C1	C2	C3	
2	R1		1	2	3
3	R2		2	4	6
4	R3		3	6	9
5	R4		4	8	12

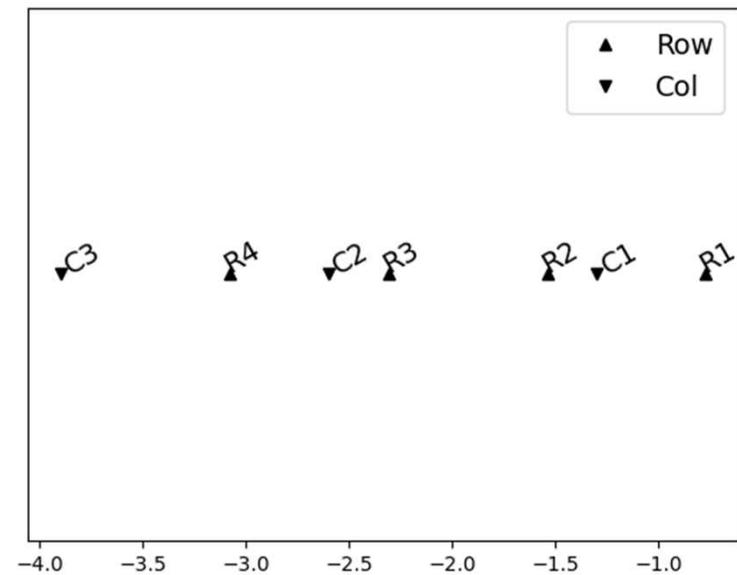
SVDによる内積モデル



対応分析 (Greenacre, 2017, e. (A.4))



$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2} = 0$$



まとめ

1. クロス表などの配列データのSVDにより、行情報および列情報を得ることができる。
2. 配列データは度数以外の数値データでもよい。
3. SVDから内積モデルを得ることができる。
4. 内積モデルは座標軸の直交回転ができる。
5. 内積モデルから、座標値あるいはベクトルのなす角に基づいたクラスタ分析ができる。