

<svdデータ解析>

多変量データに対する特異値分解

右の表は種々の色サンプルの分光反射率のデータです。
第1行目は400nmから700nmまでの40nmごとの波長の値、
第1列目は色サンプルの番号です。
それらサンプルの分光反射率の値を1000倍した値が
データとして記入されています。
このデータに対して特異値分解を行ってみましょう。

Data	400	440	480	520	560	600	640	680
1	307	470	456	422	517	862	891	893
2	201	226	214	205	234	508	532	530
3	72	73	69	58	56	279	317	312
4	53	53	44	35	37	296	489	492
5	236	309	332	445	568	817	845	847
6	83	90	102	178	273	507	525	523
7	37	38	48	140	243	590	643	649
8	22	22	24	47	98	233	231	222
9	128	147	175	489	671	687	697	697
10	35	45	86	407	734	739	736	757
11	42	45	58	235	338	367	349	348
12	20	19	22	87	134	149	140	133
13	109	153	266	650	680	539	480	527
14	45	55	86	320	367	248	190	217
15	27	28	34	137	146	99	74	81
16	245	388	580	758	632	393	298	341
17	109	162	305	445	323	144	93	114
18	44	60	112	223	125	39	25	29
19	15	17	28	66	29	11	9	10
20	303	548	755	772	607	365	271	305
21	160	267	451	452	276	122	80	96
22	63	109	246	228	92	35	24	27
23	19	27	70	61	20	10	9	9
24	243	442	604	477	222	135	136	153
25	114	202	322	219	69	37	37	41
26	30	52	121	55	14	9	9	9
27	323	569	550	376	244	206	247	282
28	199	301	286	165	90	70	86	102
29	63	121	130	49	18	13	15	16
30	321	518	424	269	230	376	385	646
31	208	261	194	103	83	167	171	337
32	353	603	558	484	460	826	874	876
33	265	336	285	213	190	524	572	697
34	145	150	112	70	58	247	279	371

Step 1 データ行列の準備

表の名前はDataとしています。

データ格納用の行列を準備します。

$$m = \text{table_row}(\text{Data}) - 1$$

$$n = \text{table_column}(\text{Data}) - 1$$

$$A = 0_{m,n} \quad m\text{行}n\text{列の零行列を代入定義}$$

表データを行列に格納します。

$$\begin{cases} \text{for } i = 1 \text{ to } m \text{ step } 1 \\ \text{for } j = 1 \text{ to } n \text{ step } 1 \\ A_{i,j} = \text{Data}_{j+1,i+1} / 1000 \end{cases}$$

備考

表の名前のDataについては、第1添字は列、第2添字は行を参照して、
Dataの第1行目と第1列目は項目名になっているので、それぞれ添字変数に1を加えて、
それらをスキップしています。

Step 2 特異値分解の計算

データ行列の特異値分解を行います。

$$\{w, U, V\} = \text{svd}(A)$$

wに特異値、Uに右行列、Vに左行列が格納されます。
これによって行列Aは次のように特異値分解されました。

$$A = UWV^T$$

ここでWは w_1, w_2, \dots, w_n を要素とする対角行列です。

成分で書くと

$$A_{p,q} = \sum_{r=1}^n w_r U_{p,r} V_{q,r}$$

この式の展開において、 w_r の小さい項を無視することによって、データ近似を検討します。
総和が100になるよう基準化した相対特異値 f を計算します。

特異値の相対値の計算
$$tr = \sum_{r=1}^n w_r \quad f = \frac{w}{tr} \times 100$$

累積相対特異値gの計算
$$g = \left\{ \sum_{r=1}^k f_r \mid k \in \mathbb{N}_{1..n} \right\}$$

r	w	f	g
1	5.3822	58.41	58.41
2	1.8225	19.78	78.19
3	1.1106	12.05	90.24
4	0.2875	3.12	93.36
5	0.2665	2.89	96.25
6	0.1572	1.71	97.96
7	0.1125	1.22	99.18
8	0.0757	0.82	100.00

右の表は、特異値の値(e)と相対特異値(f)と累積相対特異値寄与率(g)を示しています。右表を参照し、3までとって近似します。

$$A_{p,q} = \sum_{r=1}^3 S_{p,r} V_{q,r} \quad \text{ここで} \quad S_{p,r} = e_r U_{p,r}$$

これらの形式から、Vは主成分分析の主成分ベクトル、Sは主成分得点に対応しています。

右行列Vの縦ベクトル

Vの最初の3つの縦ベクトルを表示します。

$$V_{*,1} = \begin{pmatrix} -0.168225 \\ -0.250212 \\ -0.279051 \\ -0.329634 \\ -0.342307 \\ -0.429556 \\ -0.443182 \\ -0.473776 \end{pmatrix} \quad V_{*,2} = \begin{pmatrix} 0.177016 \\ 0.364690 \\ 0.537935 \\ 0.429746 \\ 0.121717 \\ -0.268414 \\ -0.384779 \\ -0.355943 \end{pmatrix} \quad V_{*,3} = \begin{pmatrix} -0.305183 \\ -0.458014 \\ -0.243431 \\ 0.413687 \\ 0.621933 \\ 0.108389 \\ -0.114790 \\ -0.234445 \end{pmatrix}$$

結果を主成分との結果を比較してみましょう。
このため、適当に-1を掛けて主成分ベクトルと符号を合わせます。

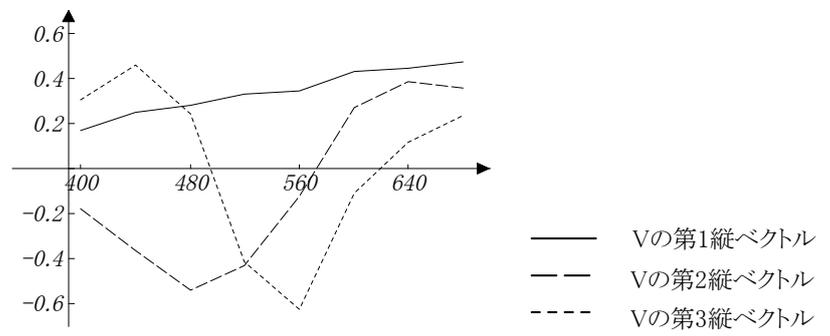
$$V_{j,1} = -V_{j,1} \quad V_{j,2} = -V_{j,2} \quad V_{j,3} = -V_{j,3}$$

このデータではベクトルの要素番号は波長に対応していました。
そこで各番号に対応する波長を設定します。

$$\lambda = \sum_{k=1}^n (400 + 40(k-1))$$

下の表は各波長に対するVの3つの縦ベクトルの値を示しています。グラフはそれらの図示です。

λ	$V_{*,1}$	$V_{*,2}$	$V_{*,3}$
400	0.168	-0.177	0.305
440	0.250	-0.365	0.458
480	0.279	-0.538	0.243
520	0.330	-0.430	-0.414
560	0.342	-0.122	-0.622
600	0.430	0.268	-0.108
640	0.443	0.385	0.115
680	0.474	0.356	0.234



右のグラフは同じデータの主成分分析に対する固有ベクトルです。
主成分ベクトルと特異値分解ベクトルが類似な形状であることが興味深く思えます。

