A8 Principal Component Analysis

Problems

1. Prove that the first principal component represents the line on which the sum of squared projections of the points is a maximum.
2. Show that the principal components are not scale invariant.
3. Show that if y_p is uncorrelated (in the sample) with the other variables, then its variance s_p^2 is an eigenvalue of $\mathbf{S}$ with the corresponding eigenvector $(0,\ 0,\ \ldots,\ 1)'$.
4. If $\mathbf{z}_i = \mathbf{A}\mathbf{y}_i$, where $\mathbf{A}$ is an orthogonal matrix, show that the principal components represent a rotation of axes.
5.

    1. Show that $\left(\mathbf{A} - b\mathbf{I}\right)\mathbf{x} = \left(\lambda - b\right)\mathbf{x}$
    2. Show that the eigenvalues of $c\mathbf{A}$ (c a scalar) are $c\lambda_1, \ldots, c\lambda_n$, with the eigenvectors same as those of $\mathbf{A}$.
    3. Let $\mathbf{R}_k = k\mathbf{R} - (k-1)\mathbf{I}$
    4. Using the results (1) and (2), show that $\lambda = 1 + \dfrac{\lambda_k - 1}{k}$ where λ and λ_k are eigenvalues of $\mathbf{R}$ and $\mathbf{R}_k$ respectively.

Problems based on datasets

1. Carry out a principal component analysis of the probe word data given below.

51 36 50 35 42;

27 20 26 17 27;

37 22 41 37 30;

42 36 32 34 27;

27 18 33 14 29;

43 32 43 35 40;

41 22 36 25 38;

38 21 31 20 16;

36 23 27 25 28;

26 31 31 32 36;

29 20 25 26 25;

Use both $\mathbb{S}$ and $\mathbb{R}$. Show the percent of variance explained for each of the five components. Based on the average eigenvalue, decide how many components to retain.

2. Carry out principal component analysis on the glucose data given below. Use both $\mathbb{S}$ and $\mathbb{R}$. Show the percent of variance explained for each of the six components. Based on the average eigenvalue, decide how many components to retain.

60	69	62	97	69	98
56	53	84	103	78	107
80	69	76	66	99	130
55	80	90	80	85	114
62	75	68	116	130	91
74	64	70	109	101	103
64	71	66	77	102	130
73	70	64	115	110	109
68	67	75	76	85	119
69	82	74	72	133	127
60	67	61	130	134	121
70	74	78	150	158	100
66	74	78	150	131	142
83	70	74	99	98	105
68	66	90	119	85	109
78	63	75	164	98	138
77	68	74	144	71	153
66	77	68	77	82	89
70	70	72	114	93	122
75	65	71	77	70	109

3. Carry out principal component analysis on the biochemical data given below. Use both $\mathbb{S}$ and $\mathbb{R}$. Show the percent of variance explained for each of the 11 components. Based on the average eigenvalue, decide how many components to retain.

5.7	4.67	17.6	1.50	.104	1.50	1.88	5.15	8.40	7.5	.11
5.5	4.67	13.4	1.65	.245	1.32	2.24	5.75	4.50	7.1	.11
4.6	2.70	20.3	.90	.097	.89	1.28	4.35	1.20	2.3	.11
5.7	3.49	22.3	1.75	.174	1.50	2.24	7.55	2.75	4.0	.12
5.6	3.49	20.5	1.40	.210	1.19	2.00	8.50	3.30	2.0	.12
4.0	3.49	18.5	1.20	.275	1.03	1.84	10.25	2.00	2.0	.12
5.3	4.84	12.1	1.90	.170	1.87	2.40	5.95	2.60	14.8	.14
5.4	4.84	12.0	1.64	.164	1.68	3.00	4.30	2.72	14.5	.14
5.4	4.84	10.1	2.30	.275	2.08	2.68	5.45	2.40	.9	.20
5.6	4.48	14.7	2.35	.210	2.55	3.00	3.75	7.00	2.0	.21
5.6	4.48	14.8	2.35	.050	1.32	2.84	5.10	4.00	.4	.12
5.6	4.48	14.4	2.50	.143	2.38	2.84	4.05	8.00	3.8	.18
5.2	3.48	18.1	1.50	.153	1.20	2.60	9.00	2.35	14.5	.13
5.2	3.48	19.7	1.65	.203	1.73	1.88	5.30	2.52	12.5	.20
5.6	3.48	16.9	1.40	.074	1.15	1.72	9.85	2.45	8.0	.07
5.8	2.63	23.7	1.65	.155	1.58	1.60	3.60	3.75	4.9	.10
4.0	2.63	19.2	.90	.155	.96	1.20	4.05	3.30	.2	.10
5.3	2.63	18.0	1.60	.129	1.68	2.00	4.40	3.00	3.6	.18
5.4	4.46	14.8	2.45	.245	2.15	3.12	7.15	1.81	12.0	.13
5.6	4.46	15.6	1.65	.422	1.42	2.56	7.25	1.92	5.2	.15
5.3	2.80	14.2	1.65	.063	1.62	2.04	5.30	3.90	10.2	.12
5.4	2.80	14.1	1.25	.042	1.62	1.84	3.10	4.10	8.5	.30
5.5	2.80	17.5	1.05	.030	1.56	1.48	2.40	2.10	9.6	.20
5.4	2.57	14.1	2.70	.194	2.77	2.56	4.25	2.60	6.9	.17
5.4	2.57	19.1	1.60	.139	1.59	1.88	5.80	2.30	4.7	.16
5.2	2.57	22.5	.85	.046	1.65	1.20	1.55	1.50	3.5	.21
5.5	1.26	17.0	.70	.094	.97	1.24	4.55	2.90	1.9	.12
5.9	1.26	12.5	.80	.039	.80	.64	2.65	.72	.7	.13
5.6	2.52	21.5	1.80	.142	1.77	2.60	4.50	2.48	8.3	.17
5.6	2.52	22.2	1.05	.080	1.17	1.48	4.85	2.20	9.3	.14
5.3	2.52	13.0	2.20	.215	1.85	3.84	8.75	2.40	13.0	.11
5.6	3.24	13.0	3.55	.166	3.18	3.48	5.20	3.50	18.3	.22
5.5	3.24	10.9	3.30	.111	2.79	3.04	4.75	2.52	10.5	.21

5.6	3.24	12.0	3.65	.180	2.40	3.00	5.85	3.00	14.5	.21
5.4	1.56	22.8	.55	.069	1.00	1.14	2.85	2.90	3.3	.15
5.3	1.56	14.5	2.05	.222	1.49	2.40	4.55	3.90	6.3	.11
5.2	1.56	18.4	1.05	.267	1.17	1.36	4.60	2.00	4.9	.11
5.8	4.12	12.5	5.90	.093	3.80	3.84	2.90	3.00	22.5	.24
5.7	4.12	8.7	4.25	.147	3.62	5.32	3.00	3.55	19.5	.20
5.7	4.12	9.4	3.85	.217	3.36	5.52	3.40	5.20	1.3	.31
5.4	2.14	15.0	2.45	.418	2.38	2.40	5.40	1.81	20.0	.17
5.4	2.14	12.9	1.70	.323	1.74	2.48	4.45	1.88	1.0	.15
4.9	2.03	12.1	1.80	.205	2.00	2.24	4.30	3.70	5.0	.19
5.0	2.03	13.2	3.65	.348	1.95	2.12	5.00	1.80	3.0	.15
4.9	2.03	11.5	2.25	.320	2.25	3.12	3.40	2.50	5.1	.18