

Department of Statistics

STAT 4560/7560 Applied Multivariate Data Analysis (Spring 2021)

HW assignment 4

Due Tuesday, Apr. 6, 11:59 pm

Use R (or other software) if necessary.

- Recall the data set “T5_5_FBEETLES.DAT” (can be found in Canvas) from HW 3.

Table 5.5. Four Measurements on Two Species of Flea Beetles

<i>Haltica oleracea</i>					<i>Haltica carduorum</i>				
Experiment Number	y_1	y_2	y_3	y_4	Experiment Number	y_1	y_2	y_3	y_4
1	189	245	137	163	1	181	305	184	209
2	192	260	132	217	2	158	237	133	188
3	217	276	141	192	3	184	300	166	231
4	221	299	142	213	4	171	273	162	213
5	171	239	128	158	5	181	297	163	224
6	192	262	147	173	6	181	308	160	223
7	213	278	136	201	7	177	301	166	221
8	192	255	128	185	8	198	308	141	197
9	170	244	128	192	9	180	286	146	214
10	201	276	146	186	10	177	299	171	192
11	195	242	128	192	11	176	317	166	213
12	205	263	147	192	12	192	312	166	209
13	180	252	121	167	13	176	285	141	200
14	192	283	138	183	14	169	287	162	214
15	200	294	138	188	15	164	265	147	192
16	192	277	150	177	16	181	308	157	204
17	200	287	136	173	17	192	276	154	209
18	181	255	146	183	18	181	278	149	235
19	192	287	141	198	19	175	271	140	192
					20	197	303	170	205

- Find the Fisher’s classification function $z = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}_p^{-1} \mathbf{y}$ and the cutoff point $\frac{1}{2}(\bar{z}_1 + \bar{z}_2)$.
 - Find the classification table (confusion matrix) using the Fisher’s linear classification rule found in part (a), applied to the training data.
 - (Graduate students only) Suppose, in nature, the proportions of two species of flea beetles are given as *Haltica oleracea* : *Haltica carduorum* = 0.9 : 0.1, and hence the prior is given as $p_1 = 0.9$, and $p_2 = 0.1$. If we want to reflect this prior to our classification rule, we can apply the normal-based classification rule. Assume the misclassification costs are the same (i.e., $c_{12} = c_{21}$). Update the cutoff point accordingly, and describe how does the imbalanced prior affect the classification rule.
- The data set “T6_2_ROOT.DAT” (can be found in Canvas) includes the measurements of apple trees of six different rootstocks.
 - Find the eigenvalues and eigenvectors of $\mathbf{E}^{-1} \mathbf{H}$.

Table 6.2 Rootstock Data

Rootstock	y_1	y_2	y_3	y_4
1	1.11	2.569	3.58	.760
1	1.19	2.928	3.75	.821
1	1.09	2.865	3.93	.928
1	1.25	3.844	3.94	1.009
1	1.11	3.027	3.60	.766
1	1.08	2.336	3.51	.726
1	1.11	3.211	3.98	1.209
1	1.16	3.037	3.62	.750
2	1.05	2.074	4.09	1.036
2	1.17	2.885	4.06	1.094
2	1.11	3.378	4.87	1.635
2	1.25	3.906	4.98	1.517
2	1.17	2.782	4.38	1.197
2	1.15	3.018	4.65	1.244
2	1.17	3.383	4.69	1.495
2	1.19	3.447	4.40	1.026
3	1.07	2.505	3.76	.912
3	.99	2.315	4.44	1.398
3	1.06	2.667	4.38	1.197
3	1.02	2.390	4.67	1.613
3	1.15	3.021	4.48	1.476
3	1.20	3.085	4.78	1.571
3	1.20	3.308	4.57	1.506
3	1.17	3.231	4.56	1.458
4	1.22	2.838	3.89	.944
4	1.03	2.351	4.05	1.241
4	1.14	3.001	4.05	1.023
4	1.01	2.439	3.92	1.067
4	.99	2.199	3.27	.693
4	1.11	3.318	3.95	1.085
4	1.20	3.601	4.27	1.242
4	1.08	3.291	3.85	1.017
5	.91	1.532	4.04	1.084
5	1.15	2.552	4.16	1.151
5	1.14	3.083	4.79	1.381
5	1.05	2.330	4.42	1.242
5	.99	2.079	3.47	.673
5	1.22	3.366	4.41	1.137
5	1.05	2.416	4.64	1.455
5	1.13	3.100	4.57	1.325
6	1.11	2.813	3.76	.800
6	.75	.840	3.14	.606
6	1.05	2.199	3.75	.790
6	1.02	2.132	3.99	.853
6	1.05	1.949	3.34	.610
6	1.07	2.251	3.21	.562
6	1.13	3.064	3.63	.707

- (b) Find the relative importance of each discriminant function. Determine the number of discriminant functions required to account for 90% of relative importance.
 - (c) (Graduate students only) Plot the first two discriminant function for each observation and for the mean vectors. Describe what you observed.
 - (d) Assuming equal misclassification costs and priors, conduct the linear classification and find the classification table (confusion matrix).
3. Use the data set “T3_8_GLUCOSE.DAT” (can be found in Canvas).
- (a) Find canonical correlations between (y_1, y_2, y_3) and (x_1, x_2, x_3) .
 - (b) Find the standardized coefficients for the canonical variates, and interpret the results.

Table 3.8. Blood Glucose Measurements on Three Occasions

Fasting			One Hour after Sugar Intake		
y_1	y_2	y_3	x_1	x_2	x_3
60	69	62	97	69	98
56	53	84	103	78	107
80	69	76	66	99	130
55	80	90	80	85	114

(continued)

Table 3.8. (Continued)

Fasting			One Hour after Sugar Intake		
y_1	y_2	y_3	x_1	x_2	x_3
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
103	77	77	160	117	121
77	68	74	144	71	153
66	77	68	77	82	89
70	70	72	114	93	122
75	65	71	77	70	109
91	74	93	118	115	150
66	75	73	170	147	121
75	82	76	153	132	115
74	71	66	143	105	100
76	70	64	114	113	129
74	90	86	73	106	116
74	77	80	116	81	77
67	71	69	63	87	70
78	75	80	105	132	80
64	66	71	83	94	133
71	80	76	81	87	86
63	75	73	120	89	59
90	103	74	107	109	101
60	76	61	99	111	98
48	77	75	113	124	97
66	93	97	136	112	122
74	70	76	109	88	105
60	74	71	72	90	71
63	75	66	130	101	90
66	80	86	130	117	144
77	67	74	83	92	107
70	67	100	150	142	146
73	76	81	119	120	119
78	90	77	122	155	149
73	68	80	102	90	122
72	83	68	104	69	96
65	60	70	119	94	89
52	70	76	92	94	100

Note: Measurements are in mg/100 ml.