

A6 Classification Analysis

Problem

For the two multivariate normal distribution populations G_1 and G_2 with means μ_1 and μ_2 and with $\Sigma_1 = \Sigma_2 = \Sigma$, show that

$$\frac{f(\mathbf{y} | G_1)}{f(\mathbf{y} | G_2)} = e^{(\mu_1 - \mu_2)' \Sigma^{-1} \mathbf{y} - (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) / 2}$$

Problem

$$\text{Show that } P(\text{classify as } G_2 | G_1) = \Phi \left[\frac{-\frac{1}{2} \Delta^2 + \ln \left(\frac{p_2}{p_1} \right)}{\Delta} \right]$$

Problem

Four measurements were made on two species of flea beetles. The variables were

y_1 = distance of transverse groove from posterior border of prothorax (μm)

y_2 = length of elytra (in 0.01 mm)

y_3 = length of second antennal joint (μm)

y_4 = length of third antennal joint (μm)

Do a classification analysis on the beetle data below as follows.

- (a) Find the classification function $z = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} \mathbf{y}$ and the cutoff point

$$\frac{1}{2} (\bar{z}_1 + \bar{z}_2).$$

- (b) Find the classification table using the linear classification function in part (a).

<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

Problem

The data in the table below are judges' scores on fish prepared by three methods. Twelve fish were cooked by each method, and several judges tasted fish samples and rated each on four variables: y_1 = aroma, y_2 = flavour, y_3 = texture, and y_4 = moisture. Each entry in the table is an average score for the judges on that fish.

Do a classification analysis on the fish data below as follows. Assume

$$p_1 = p_2 = p_3.$$

- Find the linear classification functions.
- Find the classification table using the linear classification functions in part (a) (assuming $\Sigma_1 = \Sigma_2 = \Sigma_3$).
- Find the classification table using the linear classification functions and the holdout method.

Method 1				Method 2				Method 3			
y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
5.4	6.0	6.3	6.7	5.0	5.3	5.3	6.5	4.8	5.0	6.5	7.0
5.2	6.2	6.0	5.8	4.8	4.9	4.2	5.6	5.4	5.0	6.0	6.4
6.1	5.9	6.0	7.0	3.9	4.0	4.4	5.0	4.9	5.1	5.9	6.5
4.8	5.0	4.9	5.0	4.0	5.1	4.8	5.8	5.7	5.2	6.4	6.4
5.0	5.7	5.0	6.5	5.6	5.4	5.1	6.2	4.2	4.6	5.3	6.3
5.7	6.1	6.0	6.6	6.0	5.5	5.7	6.0	6.0	5.3	5.8	6.4
6.0	6.0	5.8	6.0	5.2	4.8	5.4	6.0	5.1	5.2	6.2	6.5
4.0	5.0	4.0	5.0	5.3	5.1	5.8	6.4	4.8	4.6	5.7	5.7
5.7	5.4	4.9	5.0	5.9	6.1	5.7	6.0	5.3	5.4	6.8	6.6
5.6	5.2	5.4	5.8	6.1	6.0	6.1	6.2	4.6	4.4	5.7	5.6
5.8	6.1	5.2	6.4	6.2	5.7	5.9	6.0	4.5	4.0	5.0	5.9
5.3	5.9	5.8	6.0	5.1	4.9	5.3	4.8	4.4	4.2	5.6	5.5