

ロジスティック判別関数と線形判別関数、 2次判別関数の性能評価

2003MM058 松下 剛士

指導教員: 田中 豊

1 はじめに

ロジスティック判別関数を求めるには、分布の密度関数を用いる方法と最尤法を用いる方法がある。また、線形判別関数、2次判別関数を求めるには、分布の密度関数を用いる方法と群の平方和の比を用いる方法がある。密度関数を用いた方法で分布の適性を調べる。

正規分布の確率密度関数を用いて2組の群から求められる線形判別関数、2次判別関数を示すと

$$\frac{1}{2}(D_{(2)}^2 - D_{(1)}^2) - \log \frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}$$

と表すことができる。ここで $D_{(k)}^2$ 、 $k = 1, 2$ はそれぞれの群の密度関数の指数部分であり、 π_k は第 k 群の出現確率であり、 $C(i|j)$ は第 j 群のものを第 i 群と間違えるときのコストである。 $\pi_2 C(1|2) = \pi_1 C(2|1)$ のときに誤判別が最小になる。

正規分布の確率密度関数を用いて2組の群から求められるロジスティック判別関数を示すと

$$\log \frac{N_1}{N_2} + \frac{1}{2}(D_{(2)}^2 - D_{(1)}^2)$$

となる。以上から、上の2つの式が定数項を除き一致することがわかる。また、ここで $\pi_2 C(1|2) = \pi_1 C(2|1)$ の時を考えると、ロジスティック判別関数の方が定数項分精度が落ちることが見て取れる。このことから、2組の群が正規分布の時ロジスティック判別関数よりも線形判別関数、2次判別関数のほうが適していることがわかる。

以上のことが実際にどうか、人工データとテストデータを作成して、人工データで作った判別式にテストデータを当てはめ、誤判別がどうなるか調べる。

2 使用データについて

2.1 人工データについて

Irisデータを線形判別分析をおこなったとき、2群と3群が一部重なることが分かる。

このことから、Rで平均0、分散1の200×4の正規分布ランダムデータに、Irisデータの2群と3群の平均と両群をプールした分散共分散行列を用いて、

A $N(2$ 群の平均, 両群をプールした分散共分散行列)

B $N(3$ 群の平均, 両群をプールした分散共分散行列)

という2つの等分散共分散正規分布データと、

C $N(2$ 群の平均, 2群の分散共分散行列)

D $N(3$ 群の平均, 3群の分散共分散行列)

という2つの不等分散共分散正規分布データを作り、その後Irisデータの2群と3群の平均の差を用いて、AとB、C

とDの平均の差を縮めたデータを作った。今回用いるデータはBとDの平均をIrisデータの2群と3群の平均の差の0.6倍と0.7倍縮めたデータを用いている。

なお、データ名27となっているものは0.6倍、データ名28となっているものは0.7倍平均が少なくなっている。また、正規分布以外のデータは、AからDのデータを用いてそれぞれの第4要素を対数正規分布化して解析に用いる。二項分布はバイナリ変数の第5要素をつくり、解析に用いる。

2.2 テストデータについて

人工データと同様の方法でデータサイズ1500×4のデータを作る。

3 等分散共分散正規分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
mean27	230	0.076667	mean28	422	0.140667
2次判別分析			2次判別分析		
mean27	230	0.076667	mean28	419	0.139667
線形判別分析			線形判別分析		
mean27	230	0.076667	mean28	419	0.139667

図 1: 正規分布データの判別結果1

図 2: 正規分布データの判別結果2

図1、2は、正規分布データを分析結果を示したものである。

図1において、ロジスティック判別分析の誤判別率は0.07667である。2次判別分析の誤判別率は0.07667である。線形判別分析の誤判別率は0.07667である。

図2において、ロジスティック判別分析の誤判別率は0.14067である。2次判別分析の誤判別率は0.13967である。線形判別分析の誤判別率は0.13967である。

図1において、全ての誤判別率が等しくなり、図2において、線形判別分析と2次判別分析が共に等しくなっている。

4 等分散共分散二項分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
nikou27	230	0.076667	nikou28	420	0.14
2次判別分析			2次判別分析		
nikou27	233	0.077667	nikou28	417	0.139
線形判別分析			線形判別分析		
nikou27	233	0.077667	nikou28	417	0.139

図 3: 二項分布データの判別結果1

図 4: 二項分布データの判別結果2

図3、4は、二項分布データを分析結果を示したものである。

図3において、ロジスティック判別分析の誤判別率は0.07667である。2次判別分析の誤判別率は0.07667である。線形判別分析の誤判別率は0.07667である。

図4において、ロジスティック判別分析の誤判別率は0.14である。2次判別分析の誤判別率は0.139である。線形判別分析の誤判別率は0.139である。

全ての誤判別率がほぼ等しくなっていることがわかる。

5 等分散共分散対数正規分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
exp27	234	0.078	exp28	426	0.142
2次判別分析			2次判別分析		
exp27	236	0.078667	exp28	437	0.145667
線形判別分析			線形判別分析		
exp27	234	0.078	exp28	426	0.142

図 5: 対数正規分布データの判別結果1 図 6: 対数正規分布データの判別結果2

図5、6は、対数正規分布データを分析結果を示したものである。

図5において、ロジスティック判別分析の誤判別率は0.078である。2次判別分析の誤判別率は0.07867である。線形判別分析の誤判別率は0.078である。

図6において、ロジスティック判別分析の誤判別率は0.142である。2次判別分析の誤判別率は0.14567である。線形判別分析の誤判別率は0.142である。

ロジスティック判別分析と線形判別分析の誤判別率が等しく、共に2次判別分析の誤判別率よりも低い値を出していることがわかる。

6 不等分散正規分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
var27	191	0.063667	var28	368	0.122667
2次判別分析			2次判別分析		
var27	173	0.057667	var28	338	0.112667
線形判別分析			線形判別分析		
var27	191	0.063667	var28	376	0.125333

図 7: 正規分布データの判別結果1 図 8: 正規分布データの判別結果2

図7、8は、正規分布データを分析結果を示したものである。

図7において、ロジスティック判別分析の誤判別率は0.06367である。2次判別分析の誤判別率は0.05767である。線形判別分析の誤判別率は0.06367である。

図8において、ロジスティック判別分析の誤判別率は0.12267である。2次判別分析の誤判別率は0.11267である。線形判別分析の誤判別率は0.12533である。誤判別率から2次判別分析が優位であると思われる。

7 不等分散共分散二項分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
Nikou27	193	0.064333	Nikou28	372	0.124
2次判別分析			2次判別分析		
Nikou27	174	0.058	Nikou28	334	0.111333
線形判別分析			線形判別分析		
Nikou27	195	0.065	Nikou28	375	0.125

図 9: 不等分散共分散データの判別結果1 図 10: 不等分散共分散データの判別結果2

図9、10は、二項分布データを分析結果を示したものである。

図9において、ロジスティック判別分析の誤判別率は0.06433である。2次判別分析の誤判別率は0.058である。線形判別分析の誤判別率は0.065である。

図10において、ロジスティック判別分析の誤判別率は0.124である。2次判別分析の誤判別率は0.11133である。線形判別分析の誤判別率は0.125である。

誤判別率から2次判別分析が優位であると思われる。

8 不等分散共分散対数正規分布データでの判別結果

ロジスティック判別分析			ロジスティック判別分析		
データ名	誤判別数	誤判別率	データ名	誤判別数	誤判別率
exp27	188	0.062667	exp28	364	0.121333
2次判別分析			2次判別分析		
exp27	184	0.061333	exp28	341	0.113667
線形判別分析			線形判別分析		
exp27	213	0.071	exp28	381	0.127

図 11: 対数正規分布データの判別結果1 図 12: 対数正規分布データの判別結果2

図11、12は、対数正規分布データを分析結果を示したものである。

図11において、ロジスティック判別分析の誤判別率は0.06267である。2次判別分析の誤判別率は0.06133である。線形判別分析の誤判別率は0.071である。

図12において、ロジスティック判別分析の誤判別率は0.12133である。2次判別分析の誤判別率は0.11367である。線形判別分析の誤判別率は0.127である。

誤判別率から2次判別分析が優位であると思われる。

9 おわりに

1. 等分散共分散行列が等しいときの正規分布のときに、ロジスティック判別分析よりも線形判別分析のほうが誤判別率が小さかった。
2. 等分散共分散行列が等しいときの対数正規分布のときに、2次判別分析よりも他の2つの分析の誤判別率が小さかった。
3. 等分散共分散行列が等しくないときに、2次判別分析が他の2つの分析の誤判別率が小さかった。

ということがいえた。