Vol.2.Issue.2.2014

http://www.bomsr.com

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

ISSN:2348-0580

RESEARCH ARTICLE

PERFORMANCE OF BAYES AND DISTANCE DISCRIMINANT FUNCTIONS IN CLASSIFICATION PROBLEM – A SIMULATION STUDY

NUNNA SRINIVASA RAO[1], A.V. DATTATREYA RAO[2]

[1]Department of Statistics, Andhra Loyola College (Autonomous),Vijayawada-520 008

[2]Professor of Statistics, Acharya Nagarjuna University, Guntur, India

NUNNA SRINIVASA RAO

*Author for Correspondence

Article Info:

Article received :10/05/2014

Revised on:28/05/2014

Accepted on:30/05/2014

ABSTRACT

The criterion of error of misclassification of Bayes and Distance linear classifiers have not been fully investigated under the influence of intra class correlations in classifying two Multivariate Normal populations [c.f p 527 Johnson and Wrichen (2001)]. Here an attempt is made to study the relative performance of Bayes and Distance classifiers in classifying a new observation X_o into one of the two Multivariate Normal populations $N_p(\mu_1, \Sigma_1)$ and $N_p(\mu_2, \Sigma_2)$in the above said context.

1. INTRODUCTION TO LINEAR CLASSIFIERS

As it is proposed to study the relative performance of Bayes and Distance classifiers in case of two Multivariate Normal populations, the decision rules are presented here for ready reference:

a) When the parameters are specified

i) Bayes Rule

When $\Sigma_1 = \Sigma_2 = \Sigma$ the Bayes Rule which is originally quadratic reduces to linear and is given by

$$(\mu_2 - \mu_1)^T \Sigma^{-1} X + \frac{1}{2}\left(\mu_1^T \Sigma^{-1} \mu_1 - \mu_2^T \Sigma^{-1} \mu_2\right) \lessgtr \ln \frac{P(\pi_1)}{P(\pi_2)} \rightarrow X \in \begin{cases} \pi_1 \\ \pi_2 \end{cases} \tag{1.1}$$

II) DISTANCE CLASSIFIER

The decision rule is

$$\|X - \mu_1\|^2 - \|X - \mu_2\|^2 \lessgtr 2 \ln \frac{P(\pi_1)}{P(\pi_2)} \rightarrow X \in \begin{cases} \pi_1 \\ \pi_2 \end{cases} \tag{1.2}$$

This decision rule has the geometrical interpretation of comparing the distances from X to μ_1 and μ_2 according to a Threshold. When $P(\pi_1)$= $P(\pi_2)$ =0.5, the decision boundary is the perpendicular bisector of the line joining μ_1 and μ_2 .

b) When the parameters are unknown

When the parameters are unknown, then they are estimated based upon two samples of sizes n_1 and n_2 respectively from π_1 and π_2 , before the classification is done. The estimates are given by

$$\hat{\mu}_i = \overline{X}_i \text{ and } \hat{\Sigma} = S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

$$\text{where } S_i^2 = \frac{1}{n_i - 1}\sum_{i=1}^{n_i}\left(X_i - \overline{X}\right)^2$$

Now, the methods described above can be used by replacing μ_1, μ_2, Σ respectively by $\overline{X}_1, \overline{X}_2$ and S_{pooled} .

2. Computation of Total Probability of Misclassification (TPM)

Computing probability of error of misclassification is somewhat difficult as it involves evaluation of multiple integrals [c.f Fukunaga (1990)]. However TPM can be estimated by means of the confusion matrix using simulation and the process is as under:

Total Probability of Misclassification (TPM) is defined as,

TPM = P(misclassifying a π_1 observation or misclassifying a π_2 observation)

Therefore the confusion matrix is of the form

		Predicted	
		π_1	π_2
Actual	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$
	π_2	$n_{2M} = n_1 - n_{2C}$	n_{2C}

n_{1C} = Number of π_1 items correctly classified as π_1 items

n_{2C} = Number of π_2 items correctly classified as π_2 items

n_{1M} = Number of π_1 items misclassified as π_2 items

n_{2M} = Number of π_2 items misclassified as π_1 items

3. METHODOLOGY

Let $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ be the two normal populations to be discriminated for the specified parameters using above said methods. Without loss of generality μ_1 is taken as Null vector, μ_2 is taken as $k\left[1\right]^T$ (where $\left[1\right]^T$ is a vector with all components as unity), *k* varying from 0.5(0.5) 3.

Exploiting the relationship $V^{1/2}\rho V^{1/2} = \Sigma$, (where *V* and ρ are respectively the diagonal matrix of variances and intra class correlations), here it is considered $\Sigma = \rho$ and the dimensionality (p) ranging from 3(1)10 and the correlations in the correlation matrix ranging from 0.2 to 0.8 spreading with equidistant along the rows are taken in ascending fashion and in another case it is taken in the descending order. The priori probabilities $p(\pi_1) = p_1$ and $p(\pi_2) = p_2 = 1 - p_1$. Here we have taken $p_1 = 0.1(0.1)0.9$.

4. GENERATION OF MULTIVARIATE NORMAL DATA

The vectors of multivariate normal data with the specified parameters can be obtained starting from univariate standard normal data using Box and Muller (1958) technique and from this, we can generate multivariate normal data for any specified set of parameters μ, Σ [c.f Fukunaga (1990)].

When the parameters are not specified then the simulation study involves two phases - *Training* (estimation) and *Validation* (classification). In training phase, based on samples of size n_1 and n_2 drawn respectively from the specified populations, estimation of parameters is done, then the classifiers are constructed, while in validation phase another set of pseudo random vectors are drawn from Multivariate Normal and is used to study the performance of classifiers under consideration.

5. SALIENT OBSERVATIONS ON THE SIMULATION RESULTS

The tolerance limit for TPM is taken as 10% and the following observations were made:

PARAMETERS ARE SPECIFIED

I. Decreasing pattern of correlations in $\sum$ matrix.

1. Distance classifier performs better than Bayes classifier under orthogonal transformation.

II. Increasing pattern of correlations in $\sum$ matrix

2. Distance classifier under orthogonal transformation is the only classifier satisfying the tolerance limit condition.

PARAMETERS ARE NOT SPECIFIED

I. *Decreasing pattern of correlations in* $\sum$ *matrix.*

3. Bayes classifier performs better than Distance classifier under given vector.
4. Under Orthogonal transformation, Distance classifier performs better than Bayes classifier even when the priori probabilities $p(\pi_1)$ are increasing.
5. Under *Jackknifing* Bayes classifier performs better with the gradual increase in $p(\pi_1)$.

II. *Increasing pattern of correlations in* $\sum$ *matrix.*

6. Under *Jackknifing* Bayes classifier is performing better when compared to the Distance classifier.
7. In general *given vector* under any classifier is performing better when compared to Orthogonal transformation and Jackknifing, However *Jackknifing* under Bayes classifier is equally good when compared to Bayes classifier *under given* vector with the increase in priori probability $p(\pi_1)$.

Performance of *jackknifing* very much depends on size of the samples and priori probabilities. It is observed that *Jackknifing* gives better results with the increase in sample sizes and priori probabilities $p(\pi_1)$.

ANNEXURE

100*TPM OF BAYES AND DISTANCE CLASSIFIERS IN CLASSIFYING TWO POPULATIONS

$\pi_1 : N_p(0,\Sigma)$ and $\pi_2 : N_p(\mu_2,\Sigma)$

TABLE NO: 1.1: DECREASING ORDER OF CORRELATIONS IN Σ MATRIXWHEN THE PARAMETERS ARE SPECIFIED

$\pi_1 : N_{10}(0,\Sigma)$ $\pi_2 : N_{10}(\mu_2,\Sigma)$

$p(\pi_1)$	μ_2	**(3.0)1ᵀ**		**(2.5)1ᵀ**		**(2.0)1ᵀ**		**(1.5)1ᵀ**		**(1)1ᵀ**		**(0.5)1ᵀ**	
	Classifier	❶	❷	❶	❷	❶	❷	❶	❷	❶	❷	❶	❷
0.1	**Bayes**	2.8	0.0	6.2	0.2	10.8	4.1	21.5	17.0	36.3	41.0	50.0	50.0
	Distance	5.8	0.0	10.1	0.0	13.7	0.3	19.8	1.5	27.9	8.4	40.3	36.1
0.2	**Bayes**	3.0	0.0	5.1	3.0	7.2	1.2	16.8	7.8	29.1	27.3	46.8	49.5
	Distance	8.4	0.0	9.2	0.0	11.5	0.0	18.6	0.9	28.7	6.7	38.8	29.3
0.3	**Bayes**	2.9	0.1	5.9	0.2	9.7	1.0	17.5	5.1	26.9	19.3	39.3	44.1
	Distance	6.8	0.0	11.5	0.0	15.2	0.1	20.5	0.9	30.5	8.0	36.7	24.3
0.4	**Bayes**	3.1	0.0	5.7	0.1	9.6	0.7	15.2	3.6	26.5	12.5	37.9	32.1
	Distance	7.9	0.0	11.2	0.0	13.5	0.1	20.8	0.9	29.7	5.4	37.6	22.5
0.5	**Bayes**	4.1	0.0	6.2	0.0	10.1	0.6	16.4	2.4	25.3	9.5	36.9	26.4
	Distance	7.9	0.0	11.9	0.0	13.4	0.2	22.2	1.3	30.1	5.3	39.7	21.7
0.6	**Bayes**	4.2	0.0	7.1	0.1	10.2	0.4	17.9	2.9	26.7	11.1	41.3	32.1
	Distance	7.7	0.0	11.2	0.0	15.5	0.0	21.4	0.9	30.6	5.3	40.0	23.3
0.7	**Bayes**	3.6	0.0	6.1	0.1	12.3	0.5	20.5	5.2	27.6	17.4	43.9	43.5
	Distance	7.4	0.0	9.6	0.0	14.6	0.0	21.6	1.3	29.8	5.8	41.5	25.3
0.8	*Bayes*	4.7	0.0	9.6	0.3	14.2	1.1	24.4	6.6	33.8	26.1	47.3	49.5
	Distance	7.1	0.0	11.8	0.0	15.1	0.1	25.2	1.1	29.5	6.3	40.9	28.8
0.9	**Bayes**	6.4	0.0	12.2	0.4	19.9	3.9	28.5	17.1	40.9	40.9	49.9	49.9
	Distance	8.8	0.0	12.4	0.0	16.6	0.0	22.5	1.3	34.5	10.0	41.9	38.0

$p(\pi_1)$: Priori Probability of π_1 ❶ For a Given Vector ❷ Using Orthogonal Transformation

TABLE 1.2: INCREASING ORDER OF CORRELATIONS IN Σ MATRIX WHEN THE PARAMETERS ARE SPECIFIED

$\pi_1 : N_{10}(0,\Sigma)$ $\pi_2 : N_{10}(\mu_2,\Sigma)$

$p(\pi_1)$	μ_2	(3.0)1ᵀ		(2.5)1ᵀ		(2.0)1ᵀ		(1.5)1ᵀ		(1.0)1ᵀ		(0.5)1ᵀ	
	Classifier	❶	❷	❶	❷	❶	❷	❶	❷	❶	❷	❶	❷
0.1	Bayes	16.6	20.1	23.4	25.0	28.7	32.4	39.8	40.4	46.6	48.1	50.0	50.0
	Distance	6.7	0.0	11.1	0.0	13.9	0.4	23.6	1.2	30.3	8.4	41.4	37.7
0.2	Bayes	18.9	23.4	21.1	28.1	27.0	32.5	36.5	35.4	41.3	44.3	50.0	50.0
	Distance	6.7	0.0	9.9	0.0	16.8	0.4	24.6	0.8	30.5	8.1	41.0	27.4
0.3	Bayes	15.1	25.0	19.4	29.6	25.5	29.5	31.4	34.5	41.4	38.1	47.3	49.3
	Distance	8.0	0.0	9.6	0.0	15.2	0.2	22.7	0.6	32.7	7.4	39.6	24.6
0.4	Bayes	17.9	26.2	20.3	29.5	26.1	30.3	32.2	34.4	37.2	39.3	46.8	46.3
	Distance	6.9	0.0	10.5	0.0	17.2	0.2	23.8	1.5	31.6	7.6	42.1	23.3
0.5	Bayes	16.4	33.1	21.6	34.0	25.2	33.4	31.0	36.8	36.3	43.0	44.1	46.1
	Distance	6.8	0.0	10.5	0.0	14.2	0.1	19.9	1.0	29.6	8.3	40.4	23.6
0.6	Bayes	19.7	32.5	22.6	33.8	26.5	35.3	31.5	38.8	37.1	40.2	45.8	47.2
	Distance	7.2	0.0	10.9	0.0	16.7	0.2	21.9	1.9	30.6	5.5	42.5	25.0
0.7	Bayes	16.9	35.8	25.3	35.8	30.9	39.3	34.0	41.7	41.4	44.3	47.3	49.4
	Distance	6.6	0.0	11.7	0.0	17.5	0.4	24.6	1.4	30.5	8.2	39.4	25.1
0.8	Bayes	19.5	39.6	27.7	40.1	29.3	43.3	37.2	45.0	45.1	46.8	49.7	50.0
	Distance	6.9	0.0	14.7	0.1	15.0	0.6	25.8	2.0	31.5	9.0	42.2	29.2
0.9	Bayes	23.8	42.7	27.2	44.6	34.8	46.7	41.6	48.1	47.8	49.7	50.1	50.0
	Distance	8.3	0.0	10.6	0.0	15.4	0.1	25.4	2.4	30.9	14.4	40.4	40.9

$p(\pi_1)$: Priori Probability of π_1 ❶ For a Given Vector ❷ Using Orthogonal Transformation

TABLE 1.3: DECEASING ORDER OF CORRELATIONS IN Σ MATRIX
WHEN THE PARAMETERS ARE NOT SPECIFIED

$\pi_1 : N_{10}(0,\Sigma)$; $\pi_2 : N_{10}(\mu_2,\Sigma)$ Size of First Sample: 10 ; Size of Second Sample: 20

B: Bayes Classifier D: Distance Classifier

$p(\pi_1)$	μ_2	$(3.0)1^T$			$(2.5)1^T$			$(2.0)1^T$			$(1.5)1^T$			$(1.0)1^T$			$(0.5)1^T$		
		❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸
0.1	**B**	0	49	50	0	47.5	50	0.5	44	49.7	0.7	46.7	49.2	8.7	42.2	49	30.7	45	48.5
	D	0	0.5	12.5	0.2	0.5	13.7	4.2	0.7	21.5	9	5.5	29	21.5	14.7	42.2	36	37.7	48.5
0.2	**B**	0	49.2	48.2	0	46.5	47.7	0	46.7	46	0.2	44.2	45.2	7.5	50.7	44.2	24.2	51.7	45.5
	D	1	1	8.2	1.2	0.5	11.5	3.7	3.2	16.5	8.5	4.2	23.7	19	15	34.5	33	38	46.5
0.3	**B**	0	48.5	30	0	46	34	0	42.7	34.5	1.2	43.2	34.5	6	41.7	33	23.5	45.5	41
	D	0	0.2	6.2	1.2	1.2	11.2	4.7	0.7	15	10.7	6.5	22	19.5	10.7	29	34	31.5	40.5
0.4	**B**	0	48.2	14.2	0	47.7	14	0	47.5	17.5	1	44.2	18.5	7.7	44.5	24	26.7	46	33.5
	D	0	0.5	6.2	0.7	0.2	8.7	3.2	1.7	11.2	8.2	6.2	17.2	16.7	15	24.7	34.5	28.5	35.7
0.5	**B**	0	47.5	3.2	0	47.7	4	0	45.5	6.5	2.2	44.5	7.7	5	41.5	14.2	22	47.7	25.5
	D	0.2	0	6.7	0	1.5	6.5	3	3	9.7	9.5	7.2	13.2	15.7	13.7	20.7	28	33.2	32.5
0.6	**B**	0	48.2	0.2	0	47.2	0.7	0	47	0.7	1.2	45.2	3.5	5.2	47	9.5	21.5	45.2	21.7
	D	0.2	0.2	5.5	0.7	1.2	6.7	3.7	1.7	9.2	7.5	8	14.7	16.7	14.5	18	26.5	30	30.2
0.7	**B**	0	49	0	0	48	0	0	47	0	0.7	47	1.5	4.5	43.2	4	27.7	48.5	29.5
	D	0.2	0.7	2.7	0.7	2.5	6	3	3.5	8.7	5.7	5.7	12	18.5	19	18	33.5	36.5	35.7
0.8	**B**	0	48.7	0	0	47.5	0.5	0	46	1.5	1.2	44.5	3.5	6.7	47.7	10.7	21.2	46.2	31.5
	D	0.2	0.7	4.7	0.2	3	5.5	2.7	3.6	5.7	7.2	8.5	8.5	16.2	26	16.2	31	35.5	36.7
0.9	**B**	0	49.7	9.5	0	48.2	11	0	47	17	1	47	18.5	7.7	45.2	30.2	31.5	49.2	42.2
	D	0.2	1	3.5	1.2	1	4.7	2	4.7	5.5	9	8.2	7.7	17.5	23	18.5	37.5	42.5	46

$p(\pi_1)$ **: Priori Probability of** π_1 **❶ For a Given Vector ❷ Using Orthogonal Transformation ❸ Jackknife Method**

TABLE 1.4: INCEASING ORDER OF CORRELATIONS IN Σ MATRIX
WHEN THE PARAMETERS ARE NOT SPECIFIED

$\pi_1 : N_{10}(0,\Sigma)$; $\pi_2 : N_{10}(\mu_2,\Sigma)$ Size of First Sample: 10 ; Size of Second Sample: 20

$p(\pi_1)$	μ_2	$(3.0)1^T$			$(2.5)1^T$			$(2.0)1^T$			$(1.5)1^T$			$(1.0)1^T$			$(0.5)1^T$		
		❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸	❶	❷	❸
0.1	**B**	0.0	31.7	50.0	0.0	18.7	50.0	0.2	24.5	49.7	0.5	28.7	50.0	5.7	29.2	49.2	29.0	35.5	48.2
	D	0.7	0.2	10.7	3.0	0.0	14.5	6.2	2.0	20.2	15.5	4.0	29.2	25.0	11.7	38.5	36.0	35.7	47.0
0.2	**B**	0.0	19.7	48.5	0.0	21.7	48.0	0.0	25.0	45.7	0.5	28.7	44.0	7.0	31.0	40.5	26.5	42.2	45.0
	D	1.5	0.2	9.0	2.7	1.7	12.5	5.5	2.2	18.5	11.5	7.0	22.5	19.2	17.7	30.5	46.0	37.5	46.0
0.3	**B**	0.0	27.5	34.2	0.0	19.0	31.7	0.2	20.7	35.2	1.0	29.2	31.2	7.2	31.2	33.7	22.2	42.0	35.5
	D	1.0	0.2	7.5	1.7	0.7	13.2	6.0	3.5	16.2	10.7	5.7	20.5	23.0	11.7	25.2	37.5	25.2	40.5
0.4	**B**	0.0	28.5	16.0	0.0	26.7	15.7	0.2	25.5	16.5	1.0	25.7	23.5	6.0	38.2	21.0	20.5	37.7	30.7
	D	1.2	1.2	8.5	0.5	1.2	8.5	4.2	2.0	12.5	17.5	8.2	22.2	18.5	11.2	24.7	35.5	28.5	34.2
0.5	**B**	0.0	34.5	2.5	0.5	33.7	3.0	0.2	28.5	6.0	1.7	28.5	10.2	6.2	34.5	13.5	24.5	44.0	26.0
	D	0.7	0.5	9.5	2.7	2.2	11.2	3.7	6.2	13.0	15.7	4.5	20.0	19.7	13.0	24.0	37.2	26.5	35.0
0.6	**B**	0.0	27.0	0.0	0.0	27.2	1.5	0.0	27.0	1.7	0.2	26.2	3.5	3.7	28.2	6.2	19.5	35.2	19.0
	D	0.5	0.0	8.0	3.0	2.0	9.0	5.7	4.7	11.5	12.7	5.0	14.5	19.2	11.7	23.0	33.0	27.7	33.0
0.7	**B**	0.0	26.2	0.0	0.0	30.2	0.0	0.2	28.5	0.2	1.0	30.0	2.0	5.2	32.7	5.2	21.0	41.0	22.5
	D	0.2	0.5	7.7	2.7	1.0	7.5	4.2	2.7	12.0	9.7	7.2	11.7	21.5	15.7	21.7	38.2	29.7	35.7
0.8	**B**	0.0	35.2	0.0	0.0	31.2	0.0	0.2	32.0	1.2	0.2	33.2	3.0	6.5	29.2	15.0	20.2	41.0	29.5
	D	0.5	2.0	6.0	2.5	0.7	9.5	4.2	7.0	9.2	11.5	8.5	11.0	25.0	18.5	23.5	34.0	38.7	37.2
0.9	**B**	0.0	32.0	8.5	0.0	38.7	8.0	0.0	29.7	15.0	0.7	37.2	21.0	8.5	36.2	30.5	27.2	40.0	40.2
	D	1.2	0.5	6.2	2.5	3.0	4.5	4.0	4.0	6.5	9.7	16.2	11.7	24.5	22.2	25.2	37.2	41.7	45.7

B: Bayes Classifier; D: Distance Classifier

$p(\pi_1)$ **: Priori Probability of** π_1 **❶ For a Given Vector ❷ Using Orthogonal Transformation ❸ Jackknife Method**

Note: The typical tables of TPM values are included here to save the space, and the exhaustive tables are available with the authors.

REFERENCES

[1]. Box GEP and Muller ME(1958) A Note on the generation of Normal Variates. Ann.Math.Statist,29, pp610-611.

[2]. Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition. Academic Press, 2nd edition, New York

[3]. Johnson Richard A., and Wichern Dean. W(2001). Applied Multivariate Statistical Analysis, Prentice Hall International Inc., Englewood Cliffs, New Jersey.

[4]. A.V.Dattatreya Rao, and N.Srinivasa Rao (2003) Performance of Bayes classifier in Pattern Recognition, Gujarat Statistical Review, Vol.30, pp.59-66.

[5]. A.V.Dattatreya Rao, and N.Srinivasa Rao (2007) Performance of Fishers discriminant function in classification problem – A simulation study, The Aligarh Journal of Statistics, vol. 27, pp1-9.

[6]. N. Srinivasa Rao (2013) Performance of Distance Classifier in Pattern Recognition, Bulletin of Mathematics and Statistics Research (BOMSR) E-ISSN:2348-0580, Vol.1,Issue1 pp 50-54.