

Application of Skew-normal in Classification of Satellite Image

Mohammad Reza Zadkarami and Mahdi Rowhani
Shahid Chamran University

Abstract: The aim of this paper is to investigate the flexibility of the skew-normal distribution to classify the pixels of a remotely sensed satellite image. In the most of remote sensing packages, for example ENVI and ERDAS, it is assumed that populations are distributed as a multivariate normal. Then linear discriminant function (LDF) or quadratic discriminant function (QDF) is used to classify the pixels, when the covariance matrix of populations are assumed equal or unequal, respectively. However, the data was obtained from the satellite or airplane images suffer from non-normality. In this case, skew-normal discriminant function (SDF) is one of techniques to obtain more accurate image. In this study, we compare the SDF with LDF and QDF using simulation for different scenarios. The results show that ignoring the skewness of the data increases the misclassification probability and consequently we get wrong image. An application is provided to identify the effect of wrong assumptions on the image accuracy.

Key words: Classification, LDF, pixels, QDF, SDF, skew-normal.

1. Introduction

Discriminant (or classification) analysis is one of major multivariate method which is widely used in remote sensing, biostatistics, econometrics, and many other areas (Abbey and Eckstein, 2002; Salovaara *et al.* 2005; Dixon and Brerton, 2008; De la Cruz, 2008).

In many applications, the populations distribution are assumed to have multivariate normal distribution and either LDF or QDF is used, when the covariance matrix of populations are assumed equal or unequal, respectively. However the normality assumption of the data in many situations does not agree with the reality. Several authors investigated the performance of the LDF and QDF under non-normality assumptions. They identify that the LDF and QDF were greatly affected by non-normality of the populations (Lachenbruch and Sneeringer, 1973; Nakanishi and Sato, 1985; Koutras, 1987).

The most common approach adopted to solve this disagreement is transformation of the variable. However, there are many problems with the choice of the

transformation especially for multivariate data (Azzalini and Capitanio, 1999). One technique towards improving the model is to apply a more flexible multivariate distribution in order to create a logical extension to the original multivariate normal distribution. Azzalini and Capitanio (1999) firstly introduced the skew-normal discriminant function (SDF) for only two populations with equal skew parameters and variance covariance matrices, and unequal means.

Discriminant analysis is widely used in the classification of satellite image analysis. In classification of satellite or airplane images in the remote sensing, the images are taken from different channels (bands). The process is known as remote sensing since the recording of objects is done at a distance, forming the image by gathering, focusing and recording reflected light from the sun, or reflected radio waves emitted by spacecraft. A channel is a slice of wavelengths from the electromagnetic spectrum, measured by the instrument onboard the satellite. This data is usually skew in some of the components. Some authors investigated the effect of asymmetric distributions on the pixels classification (Lachenbruch 1973; Ince, 1987; Kershaw, 1987; Ripley, 1996). However, the assumption of normality played an important role in the classification of the pixels in the most of remote sensing packages. These packages, for example ENVI and ERDAS, are using only LDF or QDF methods to classify the pixels.

Azzalini and Dalla Valle (1996) introduced the multivariate skew-normal distribution which is the logical extension of the class of the multivariate normal. The SDF for two populations, when the populations have equal and known skew parameters and covariance matrices but unequal means, was introduced by Azzalini and Capitanio (1999). They notified that the LDF and SDF should be compared numerically. In this paper, we investigated the SDF for two populations when the populations have unequal skew parameter and covariance matrices. In addition, the misclassification rate for multivariate normal and multivariate skew-normal is compared in different cases. Finally, we used the real data to compare the misclassification rate of discriminant functions. Section 2 includes the description of multivariate skew-normal and SDF. In section 3 numerical results are presented. Application is presented in section 4, and in section 5 concluding remarks are offered.

2. Multivariate skew-normal distribution

A k -dimensional random variable Z is said to have the multivariate skew-normal distribution if it is continuous with density function

$$2\phi_k(z; \Omega_z)\Phi(\alpha^T z) \quad (z \in R^k) \quad (2.1)$$

where $\phi_k(z; \Omega_z)$ is the k -dimensional normal density with zero mean and covariance matrix Ω_z , $\Phi(\cdot)$ is the standard normal distribution function, and α

is a k -dimensional vector. For simplicity, Ω_z is assumed to be full rank. The notation $Z \sim SN_k(\Omega, \alpha)$ is used to denote this distribution, which will be reduced to $N(0, \Omega_z)$ density where $\alpha = 0$. The mean vector and the covariance matrix are $\mu_z = (2/\pi)^{1/2}\delta$, $Var(Z) = \Omega_z - \mu_z\mu_z^T$ respectively, where $\delta = (1 + \alpha^T\Omega_z\alpha)^{-1/2}\Omega_z\alpha$.

Now, we introduce location and scale parameters, which have been omitted in the expression (2.1) of the density Z . Then we have $Y = \zeta + \omega Z$ where $\zeta = (\zeta_1, \dots, \zeta_k)$, $\omega = (\omega_1, \dots, \omega_k)$ are location and scale parameters, respectively; the components of ω are assumed to be positive. The density of Y is

$$2\phi_k(y - \zeta; \Omega)\Phi(\alpha^T\omega^{-1}(y - \zeta))$$

where $\Omega = \omega\Omega_z\omega^T$ is a covariance matrix. This distribution is denoted by $SN(\zeta, \Omega, \alpha)$.

2.1 Skew-normal discrimination function

Let $Y_i \sim SN_k(\zeta_i, \Omega, \alpha)$, $i = 1, 2$ denote the random variables associated to the two populations. The likelihood-based (Bayes) discrimination rule allocates a new unit with observed vector y to population one if

$$(\zeta_1 - \zeta_2)^T\Omega^{-1}\left(y - \frac{1}{2}(\zeta_1 + \zeta_2)\right) + \xi_0(w_1) - \xi_0(w_2) + \ln(\pi_1/\pi_2) > 0$$

where $w_i = \alpha^T\omega^{-1}(y - \zeta_i)$, π_i is the prior probability of i -th population ($i = 1, 2$) and $\xi_0(x) = \ln\{2\Phi(x)\}$. Nonlinearity of the left-hand side of the above inequality prevents explicit solution. However, the likelihood-based discriminant function is a linear function of when either of the following conditions holds:

$$\begin{aligned} (\zeta_1 - \zeta_2)^T\omega^{-1}\alpha &= 0 \\ \omega^{-1}\alpha &= c\Omega^{-1}(\zeta_1 - \zeta_2) \end{aligned}$$

where c is non-zero scalar constant, Azzalini and Capitanio (1999).

It can be shown that discrimination function for two multivariate skew-normal populations with location parameters ζ_1 and ζ_2 , covariance matrices Ω_1 and Ω_2 , and skew vectors α_1 and α_2 is:

$$\begin{aligned} &\frac{1}{2}\ln\left(\frac{|\Omega_1|}{|\Omega_2|}\right) - \frac{1}{2}\{y^T(\Omega_1^{-1} - \Omega_2^{-1})y - 2y^T(\Omega_1^{-1}\zeta_1 - \Omega_2^{-1}\zeta_2) \\ &\quad + \zeta_1\Omega_1^{-1}\zeta_1 - \zeta_2\Omega_2^{-1}\zeta_2\} + \xi_0(w_1) - \xi_0(w_2) \end{aligned}$$

where $w_i = \alpha_i^T\omega_i^{-1}(y - \zeta_i)$, $i = 1, 2$. Let $Y_i \sim SN_k(\zeta_i, \Omega_i, \alpha_i)$, $i = 1, 2$ denote the random variables associated with the two populations. The likelihood-based

discrimination rule allocates a new unit with observed vector to population one if

$$\frac{2\phi_k(y - \zeta_1, \Omega_1)\Phi(\alpha_1^T \omega_1^{-1}(y - \zeta_1))}{2\phi_k(y - \zeta_2, \Omega_2)\Phi(\alpha_2^T \omega_2^{-1}(y - \zeta_2))} > \frac{\pi_2}{\pi_1}$$

Or

$$\ln\left(\frac{\phi_k(y - \zeta_1; \Omega_1)}{\phi_k(y - \zeta_2; \Omega_2)}\right) + \xi_0(w_1) - \xi_0(w_2) - \ln\left(\frac{\pi_2}{\pi_1}\right) > 0.$$

Now

$$\begin{aligned} & \ln\left(\frac{\phi_k(y - \zeta_1; \Omega_1)}{\phi_k(y - \zeta_2; \Omega_2)}\right) \\ &= \frac{1}{2}\ln\left(\frac{|\Omega_1|}{|\Omega_2|}\right) - \frac{1}{2}(y - \zeta_1)^T \Omega_1^{-1}(y - \zeta_1) + \frac{1}{2}(y - \zeta_2)^T \Omega_2^{-1}(y - \zeta_2) \\ &= \frac{1}{2}\ln\left(\frac{|\Omega_1|}{|\Omega_2|}\right) - \frac{1}{2}\{y^T \Omega_1^{-1}y - y^T \Omega_1^{-1}\zeta_1 - \zeta_1^T \Omega_1^{-1}y + \zeta_1^T \Omega_1^{-1}\zeta_1 - y^T \Omega_2^{-1}y \\ & \quad + y^T \Omega_2^{-1}\zeta_2 + \zeta_2^T \Omega_2^{-1}y - \zeta_2^T \Omega_2^{-1}\zeta_2\} \\ &= \frac{1}{2}\ln\left(\frac{|\Omega_1|}{|\Omega_2|}\right) - \frac{1}{2}\{y^T(\Omega_1^{-1} - \Omega_2^{-1})y - 2y^T(\Omega_1^{-1}\zeta_1 - \Omega_2^{-1}\zeta_2) \\ & \quad + \zeta_1^T \Omega_1^{-1}\zeta_1 - \zeta_2^T \Omega_2^{-1}\zeta_2\} \end{aligned}$$

Because $y^T \Omega_i^{-1} \zeta_i$ is scalar, then $y^T \Omega_i^{-1} \zeta_i = \zeta_i^T \Omega_i^{-1} y, i = 1, 2$. Therefore, the observed vector y is allocated to population one if

$$\begin{aligned} & \frac{1}{2}\ln\left(\frac{|\Omega_1|}{|\Omega_2|}\right) - \frac{1}{2}\{y^T(\Omega_1^{-1} - \Omega_2^{-1})y - 2y^T(\Omega_1^{-1}\zeta_1 - \Omega_2^{-1}\zeta_2) \\ & \quad + \zeta_1^T \Omega_1^{-1}\zeta_1 - \zeta_2^T \Omega_2^{-1}\zeta_2\} + \xi_0(w_1) - \xi_0(w_2) - \ln\left(\frac{\pi_2}{\pi_1}\right) > 0. \end{aligned}$$

3. Simulation Study

A simulation study is conducted in order to compare the LDF or QDF with corresponding types of the SDF. We considered 8 different situations to compare the misclassification probability of classification methods in the various scenarios. The simulation procedure is done using R package.

Comparing LDF with the corresponding type of SDF, the various cases are considered different for the relative positions of the skew parameters and covariance matrices. However, the maximum likelihood estimators were replaced with unknown parameters in the discriminant functions. Similarly, in comparing QDF with the corresponding type of SDF, we considered unequal covariance matrices and means. However, the skew parameter of two populations could be equal. It

should be notified that we replaced the unknown parameters of the discriminant functions by the maximum likelihood estimators (MLE).

For the evaluation of the discriminant functions in each case, we simulated randomly two training samples of size $n_1 = n_2 = 1000$ from the k -dimensional multivariate skew-normal ($k = 2, 3, 4, 5$) and built the corresponding discriminant functions with these training samples. Now, two another random samples of size 500 with the same parameters of the first step are generated as test samples; the test samples individuals are classified and the probability of misclassification is recorded. The procedure is repeated 100,000 times. Finally, the mean of misclassification probabilities for the LDF or QDF and the corresponding type of SDF are calculated.

Table 1: Comparison of the LDF and the corresponding type of SDF (e_1 and e_2 are the mean of misclassification probability for the LDF and the SDF, respectively).

Dim.	Known Parameters				Unknown Parameters			
	$\Omega_1 = \Omega_2$ $\alpha_1 = \alpha_2$		$\Omega_1 = \Omega_2$ $\alpha_1 \neq \alpha_2$		$\Omega_1 = \Omega_2$ $\alpha_1 = \alpha_2$		$\Omega_1 = \Omega_2$ $\alpha_1 \neq \alpha_2$	
	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2
k								
2	0.1323	0.1067	0.1262	0.0980	0.1176	0.0831	0.1220	0.0897
3	0.1400	0.1115	0.1548	0.1273	0.1376	0.1013	0.1543	0.1158
4	0.1714	0.1403	0.2182	0.1482	0.1457	0.1404	0.2394	0.1537
5	0.2288	0.1929	0.2531	0.1918	0.1841	0.1715	0.2489	0.2298

Tables 1 and 2 contain summary values of the numerical work, in particular misclassification probabilities. The main conclusions from these tables are as follows:

1. The mean misclassification probability of the LDF or QDF is higher than that for the corresponding type of SDF.
2. The mean of misclassification probabilities increases when the dimension of multivariate skew-normal becomes higher in all situations.
3. The mean of misclassification probabilities decreased when the parameters of dicriminant functions are unknown and we replaced them by the MLEs.

In general, QDF or LDF performs more misclassification probability when compared with corresponding type of SDF.

Table 2: Comparison of the QDF and the corresponding type of SDF (e_1 and e_2 are the mean of misclassification probability for the QDF and the SDF, respectively).

Dim.	Known Parameters				Unknown Parameters			
	$\Omega_1 = \Omega_2$		$\Omega_1 = \Omega_2$		$\Omega_1 = \Omega_2$		$\Omega_1 = \Omega_2$	
	$\alpha_1 = \alpha_2$		$\alpha_1 \neq \alpha_2$		$\alpha_1 = \alpha_2$		$\alpha_1 \neq \alpha_2$	
k	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2
2	0.1272	0.1107	0.1114	0.1067	0.1063	0.0982	0.1462	0.1090
3	0.1636	0.1381	0.1403	0.1115	0.1262	0.1085	0.1591	0.1341
4	0.1920	0.1413	0.1584	0.1403	0.1812	0.1506	0.1862	0.1608
5	0.2281	0.1824	0.2181	0.1863	0.2328	0.1744	0.2420	0.2003

4. Application

For numerical illustration, we have applied the discrimination functions to classify the pixels of a three-channel satellite image of one area of the Shadegan wetland. The Shadegan wetland is in the south-west of Iran at the head of the Persian Gulf. It is the largest wetland of Iran covering about 400,000 hectares. The wetland plays a significant hydrological and ecological role in the natural functioning of the northern Gulf. The image was obtained by Landsat ETM+ satellite. The data is used to compare the pixels classification by the two discrimination functions: LDF and the corresponding type of SDF.

In image classification, the input image should be a digital data for digitally processing the remotely sensed data. Digital data can be obtained from instruments that calibrated onto the satellite or airplane by recording the reflected or emitted radiation from individual patches of ground, known as pixels. Digital data is composed of these pixels which are recorded digitally by numeric values. These values are popularly known as digital numbers (DN) or brightness values and these values do not represent the true radiometric values and because of the radiometric distortions. The digital data can be expressed as $L = y(i, j)$ where $y(i, j) = (y_1(i, j), y_2(i, j), \dots, y_d(i, j))$ is a vector representing the features of the pixel with a location (i, j) . Here, $y_1(i, j), y_2(i, j), \dots, y_d(i, j)$ represents the features describing the object which may be spectral reflectance or emittance values from optical or infrared imagery, radar backscatter values, secondary measurements derived from the image, or geographical features such as terrain elevation, slope and aspect. This set of gray-scale values for a single pixel is known as a pattern. Thus, a pattern is a set of measurements on the chosen features for the individual that is to be classified. We identified four classes of homogeneous regions by a topographic map or other knowledge about the region as described

in Table 3.

Table 3: Characters of the samples

Class	C_1	C_2	C_3	C_4
Color of class	Blue	Brown	Green	White
Sample size	120	50	80	60

The digital number (DN) matrices for three channels: red, green and blue are obtained using ERDAS software. The data are classified by LDF, using ERDAS, and corresponding type of SDF, using R.

The highest classification producer's and user's accuracy were obtained when SDF was used (Tables 4 and 5), where producer's accuracy is the percentage of the sampling units predicted to belong to the correct class, and user's accuracy is the percentage of the sampling units predicted to belong to a particular class that actually belongs to that class. In fact, the overall classification accuracy for the four-class classification for LDF and SDF were 78.06 and 89.03, respectively.

The results of LDF show low classification accuracy for class C_2 : producer's accuracy was 68% and user's accuracy 69.39% (Table 4). However, class C_1 showed the highest producer's and user's accuracy values (81.67% and 85.96%, respectively).

Table 4: The number of individuals classified by LDF (Fisher discriminant function). The diagonal shows the number of correctly classified sample units for each class.

Allocated class	Actual class					User's accuracy (%)
	C_1	C_2	C_3	C_4	Total	
C_1	98	6	5	5	114	85.96
C_2	8	34	5	2	49	69.39
C_3	6	8	64	7	85	75.29
C_4	8	2	6	46	62	74.19
Total	120	50	80	60		
Producer's accuracy (%)	81.67	68	80	76.67		

When using SDF, class C_2 again showed the lowest producer's and user's accuracy (both 80%). On the other hand, class C_1 showed the highest producer's accuracy (93.33%), but class C_3 showed the highest user's accuracy (91.06%). Therefore, SDF achieved high accuracy for all classes, considering this fact that there are no generally accepted limits on how accurate a classification should be in order to qualify as reliable, but usually an overall accuracy exceeding 85% is considered reasonable, often with the additional criterion that the accuracy should not be lower than 70% for any class (Salovaara *et al.*, 2005).

Table 5: The number of individuals classified by SDF. The diagonal shows the number of correctly classified sample units for each class.

Allocated class	Actual class				Total	User's accuracy (%)
	C_1	C_2	C_3	C_4		
C_1	112	5	3	3	123	91.06
C_2	2	40	4	4	50	80
C_3	1	5	71	0	77	92.21
C_4	5	0	2	53	60	88.33
Total	120	50	80	60		
Producer's accuracy (%)	93.33	80	88.75	88.33		

Then the assumption of the skew-normal distribution increases the accuracy of the pixels classification and consequently it succeeded in improving the overall accuracy of the image. The misclassification probability (producer) for classes: C_1, C_2, C_3 and C_4 for the SDF are 0.067, 0.20, 0.1125 and 0.1167 respectively. However those probabilities for the LDF are 0.1833, 0.32, 0.20 and 0.2333 respectively.

We can see the effect of classification methods on the producing of the Shadegan wetland image from Figure 1. Consider the skewness of the data when using SDF to classify the data results in getting the more accurate image.

5. Conclusion

The skew-normal approach has been proposed in this paper in order to perform a pixels classification of the satellite image (see Figures 1). The approach has the advantage of being appropriate for classifying satellite image pixels when the data suffers from the non-normality. We demonstrate our approach with real data set and it is showed that the skew-normal models have better performance than that of the normal models. The results indicated that the skew-normal models proposed here are quite flexible and accurate.

Acknowledgments

The authors would like to thank referee and the editor for providing thoughtful comments and suggestions which led to substantial improvement of the presentation of the material in this paper. The authors acknowledge Dr K Rangzan and Mr. R Rozbahani's help (GIS centre, Shahid Chamran University) for providing the image of the Shadegan wetland region.

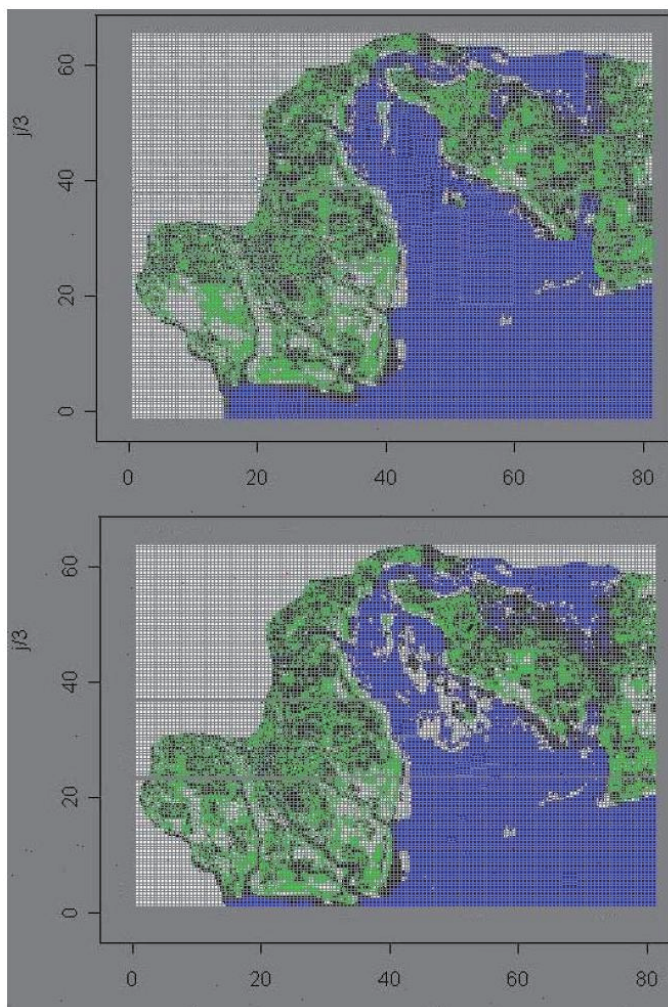


Figure 1: The image which is produced by LDF (above) and Corresponding type of SDF (below).

References

- Azzalini, A. and Capitanio, A. (1999). Statistical application of the multivariate skew-normal distribution. *Journal of Royal Statistical Society B* **65**, 367-389.
- Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.
- Abbey, C. K. and Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision* **2**, 66-78.

- De la Cruz, R. (2008). Bayesian non-linear regression models with skew-elliptical errors: Applications to the classification of longitudinal profiles. *Computational Statistics and Data Analysis* **53**, 436-449.
- Diciccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of American Statistical Association* **99**, 439-450.
- Dixon, S. J. and Brereton, R. G. (2009). Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemo-metrics Intelligent Laboratory Systems* **95**, 1-17.
- Ince, F. (1987). Maximum likelihood classification, optimal or problematic? A comparison with nearest neighbour classification. *International Journal Remote Sensing* **8**, 1829-1838.
- Kershaw, C. D. (1987). Discrimination problem for satellite images. *International Journal Remote Sensing* **8**, 1377-1383. Koutras, M. (1987). On the performance of the linear discriminant function for spherical distributions. *Journal of Multivariate Analysis* **22**, 1-12.
- Lachenbruch, P. A. (1973). Robustness of the linear and quadratic discrimination function to certain types of non-normality. *Communications in Statistics – Theory and Methods* **1**, 39-56.
- Nakanishi, H. and Sato, Y. (1985). The performance of the linear and quadratic discriminant functions for three types of non-normal distribution. *Communications in Statistics – Theory and Methods* **14**, 1181-1200.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.
- Salovaara, K.J. et al. (2005). Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment* **97**, 39-51.

Received December 10, 2008; accepted May 6, 2009.

Mohammad Reza Zadkarami
Department of Statistics
Faculty of Mathematical Sciences and Computer
Shahid Chamran University
Ahvaz, Iran
zadkarami_m@scu.ac.ir or zadkarami@yahoo.co.uk

Mahdi Rowhani
Department of Statistics
Faculty of Mathematical Sciences and Computer
Shahid Chamran University
Ahvaz, Iran
mehdirowhani@yahoo.com