Using Gaussian Copulas in Supervised Probabilistic Classification

Rogelio Salinas-Gutiérrez, Arturo Hernández-Aguirre, Mariano J. J. Rivera-Meraz, and Enrique R. Villa-Diharce

Center for Research in Mathematics (CIMAT), Guanajuato, México
`{rsalinas,artha,mrivera,villadi}@cimat.mx`

Abstract. This chapter introduces copula functions and the use of the Gaussian copula function to model probabilistic dependencies in supervised classification tasks. A copula is a distribution function with the implicit capacity to model non linear dependencies via concordance measures, such as Kendall's τ. Hence, this chapter studies the performance of a simple probabilistic classifier based on the Gaussian copula function. Without additional preprocessing of the source data, a supervised pixel classifier is tested with a 50-images benchmark; the experiments show this simple classifier has an excellent performance.

Key words: Gaussian copula, supervised classification.

1 Introduction

In Pattern Recognition applications many algorithms and models have been proposed for many tasks, specially for *clustering*, *regression* and *classification*. Applications in which a *training data set* with categories and attributes is available and the goal is to assign a new object to one of a finite number of discrete categories are known as *supervised classification* problems [2, 12, 15]. In this work we present the use of the Gaussian copula function as an alternative for modeling dependence structure in a supervised probabilistic classifier.

Copula functions are suitable tools in statistics for modeling multiple dependence, not necessarily linear dependence, in several random variables. For this reason, copula functions have been widely used in economics and finance [5, 7, 9, 25, 26]. More recently copula function have been used in other fields such as climate [22], oceanography [6], hydrology [10], geodesy [1], reliability [17], evolutionary computation [20, 21] and engineering [11]. By using copula theory, a joint distribution can be built with a copula function and, possibly, several different marginal distributions. Copula theory has been used also for modeling multivariate distributions in *unsupervised learning* problems such as image segmentation [4, 8] and retrieval tasks [16, 19, 24]. In [13], the bivariate copula functions Ali-Mikhail-Haq, Clayton, Frank and Gumbel are used for unsupervised classification. These copulas are well defined for two variables but when extended to three or more variables several complications arise (for instance, undefined copula parameters), preventing their generalization and applicability. For the Gaussian copula however, there exist a simple "general formula" for any number of variables. This work introduces the use of Gaussian copula in supervised classification, and compares an independent probabilistic classifier with a copula-based probabilistic classifier.

The content of the chapter is the following: Section 2 is a short introduction to copula functions, Section 3 presents a copula based probabilistic model for classification. Section 4 presents the experimental setting to classify an image database, and Section 5 summarizes the conclusions.

2 Copula Functions

The copula concept was introduced 50 years ago by Sklar [23] to separate the effect of dependence from the effect of marginal distributions in a joint distribution. Although copula functions can model linear and nonlinear dependencies, they have been barely used in computer science applications where nonlinear dependencies are common and need to be represented.

Definition 1. *A copula C is a joint distribution function of standard uniform random variables. That is,*

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d) \ ,$$

where $U_i \sim U(0,1)$ for $i = 1, \ldots, d$.

For a more formal definition of copula functions, the reader is referred to [14, 18]. The following result, known as Sklar's theorem, states how a copula function is related to a joint distribution function.

Theorem 1 (Sklar's theorem). *Let F be a d-dimensional distribution function with marginals $F_1, F_2, \ldots, F_d$, then there exists a copula C such that for all x in $\overline{\mathbb{R}}^d$,*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \ ,$$

where $\overline{\mathbb{R}}$ denotes the extended real line $[-\infty, \infty]$. If $F_1(x_1)$, $F_2(x_2)$, $\ldots, F_d(x_d)$ are all continuous, then C is unique. Otherwise, C is uniquely determined on $Ran(F_1) \times Ran(F_2) \times \cdots \times Ran(F_d)$, where Ran stands for the range.

According to Theorem 1, any joint distribution function F with continuous marginals $F_1, F_2, \ldots, F_d$ has associated a copula function C. Moreover, the associated copula C is a function of the marginal distributions $F_1, F_2, \ldots, F_d$. An important consequence of Theorem 1 is that the d-dimensional joint density f and the marginal densities $f_1, f_2, \ldots, f_d$ are also related:

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \cdot \prod_{i=1}^{d} f_i(x_i) \ , \tag{1}$$

where c is the density of the copula C. The Equation (1) shows that the product of marginal densities and a copula density builds a d-dimensional joint density. Notice that the dependence structure is given by the copula function and the marginal densities can be of different distributions. This contrasts with the usual way to construct multivariate distributions, which suffers from the restriction that the marginals are usually of the same type. The separation between marginal distributions and a dependence structure explains the modeling flexibility given by copula functions.

2.1 Gaussian Copula Function

There are several parametric families of copula functions, such as Student's t copula and Archimedean copulas. One of these families is the Gaussian copula function.

Definition 2. *The copula associated to the joint standard Gaussian distribution is called Gaussian copula.*

According to Definition 2 and Theorem 1, if the d-dimensional distribution of a random vector $(Z_1, \ldots, Z_d)$ is a joint standard Gaussian distribution, then the associated Gaussian copula has the following expression:

$$C(\Phi(z_1), \ldots, \Phi(z_d); \Sigma) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_d} \frac{e^{-\frac{1}{2}t' \Sigma^{-1} t}}{(2\pi)^{(n/2)} |\Sigma|^{1/2}} dt_d \cdots dt_1 \ ,$$

or equivalently,

$$C(u_1, \ldots, u_d; \Sigma) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{e^{-\frac{1}{2}t' \Sigma^{-1} t}}{(2\pi)^{(n/2)} |\Sigma|^{1/2}} dt_d \cdots dt_1 \ ,$$

where Φ is the cumulative distribution function of the marginal standard Gaussian distribution and Σ is a symmetric matrix with main diagonal of ones. The elements outside the main diagonal of matrix Σ are the pairwise correlations ρ_{ij} between variables Z_i and Z_j, for $i, j = 1, \ldots, d$ and $i \neq j$. It can be noticed that a d-dimensional standard Gaussian distribution has mean vector zero and a correlation matrix Σ with $d(d-1)/2$ parameters.

The dependence parameters ρ_{ij} of a d-dimensional Gaussian copula can be estimated using the maximum likelihood method. To do so, we follow the steps of Algorithm 1.

Due to Equation (1), the d-dimensional Gaussian copula density can be calculated as:

$$\begin{aligned} c(\Phi(z_1), \ldots, \Phi(z_d); \Sigma) &= \frac{\frac{1}{(2\pi)^{(d/2)}|\Sigma|^{1/2}} e^{-\frac{1}{2}z' \Sigma^{-1} z}}{\prod_{i=1}^{d} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}z_i^2}} \\ &= \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}z'(\Sigma^{-1} - I)z} \ . \end{aligned} \tag{2}$$

Algorithm 1 Pseudocode for estimating parameters

1: for each random variable X_i, $i = 1, \ldots, d$, estimate its marginal distribution function $\hat{F}_i$ using the observed values x_i. The marginal distribution function can be parametric or nonparametric
2: determine $u_i = \hat{F}_i(x_i)$, for $i = 1, \ldots, d$
3: calculate $z_i = \Phi^{-1}(u_i)$ where Φ is the cumulative standard Gaussian distribution function, for $i = 1, \ldots, d$
4: estimate the correlation matrix $\hat{\Sigma}$ for the random vector $(Z_1, \ldots, Z_d)$ using pseudo observations $(z_1, \ldots, z_d)$

Given that a Gaussian copula is a distribution function it is possible to simulate data from it. The main steps are the following: once a correlation matrix Σ is specified, a data set can be generated from a joint standard Gaussian distribution. The next step consists of transforming this data set using the cumulative distribution function Φ. For random vectors with a Gaussian copula associated to their joint distribution, the first step is to generate data from the copula and then determining their quantiles by means of their cumulative distribution functions. Algorithm 2 and Figures 1, 2 and 3 illustrate the sampling procedure for different correlations.

Algorithm 2 Pseudocode for generating data with Gaussian dependence structure

1: simulate observations $(z_1, \ldots, z_d)$ from a joint standard Gaussian distribution with matrix correlation Σ
2: calculate $u_i = \Phi(z_i)$ where Φ is the cumulative standard Gaussian distribution function, for $i = 1, \ldots, d$
3: determine x_i using quasi-inverse $F_i^{-1}(u_i)$, where F_i is a cumulative distribution function, for $i = 1, \ldots, d$

Figure 1-(a) shows 500 bivariate data with correlation $\rho = -0.5$ drawn from a bivariate standard Gaussian distribution (step 1, Algorithm 2). The histogram on the vertical axis and the histogram on the horizontal axis illustrate that both marginals are univariate standard Gaussian distributions. This data set is used to obtain a sample from a Gaussian copula, as shown in Figure 1-(b) (step 2, Algorithm 2). Both histograms illustrate that marginals are uniform, according to Definition 1. Figure 1-(c) shows a sample from a joint distribution with Gaussian copula and Beta marginals (step 3, Algorithm 2). This sample is obtained using the data set of Figure 1-(b). Figure 1-(d) shows a sample from a joint distribution with Gaussian copula, Student's t marginal distribution and exponential marginal distribution (step 3, Algorithm 2). This sample is also obtained from the data set of Figure 1-(b). In order to appreciate how the correlation parameter modifies the dependence structure, Figures 2 and 3 show the same information as Figure 1 with $\rho = -0.7$ and $\rho = -0.95$, respectively.

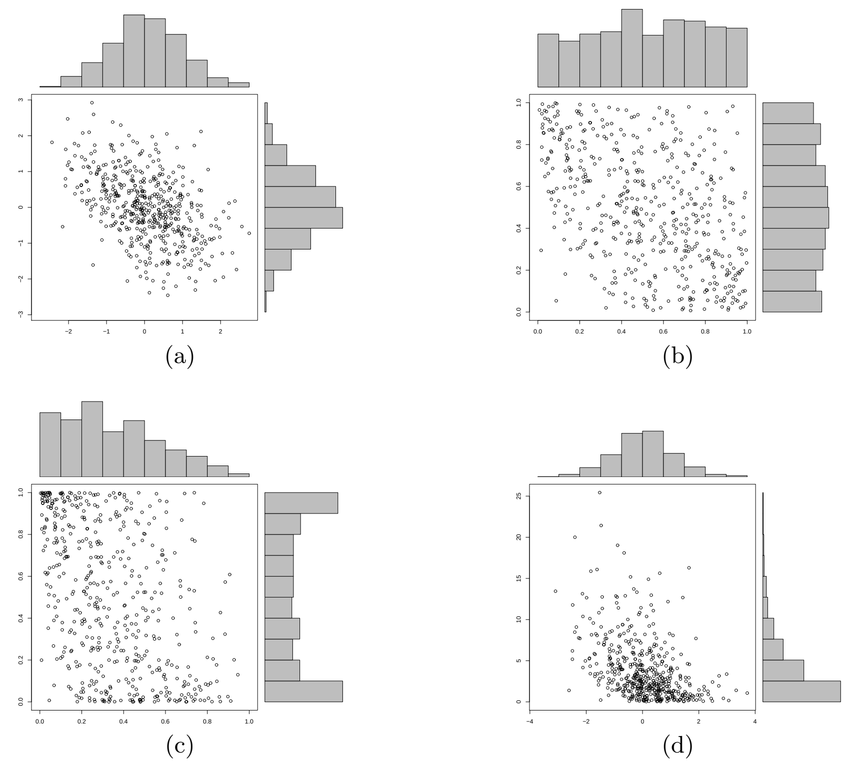


Fig. 1. (a) A sample of 500 points from a standard Gaussian distribution with parameter $\rho = -0.50$. (b) The corresponding sample for a Gaussian copula. (c) The associated sample for a joint distribution with marginal Beta distributions with parameters $(1, 2)$ (histogram on the horizontal axis) and $(0.5, 0.5)$ (histogram on the vertical axis). (d) The associated sample for a joint distribution with marginal t-Student distribution with 8 degrees of freedom (histogram on the horizontal axis) and marginal exponential distribution with mean 4 (histogram on the vertical axis).

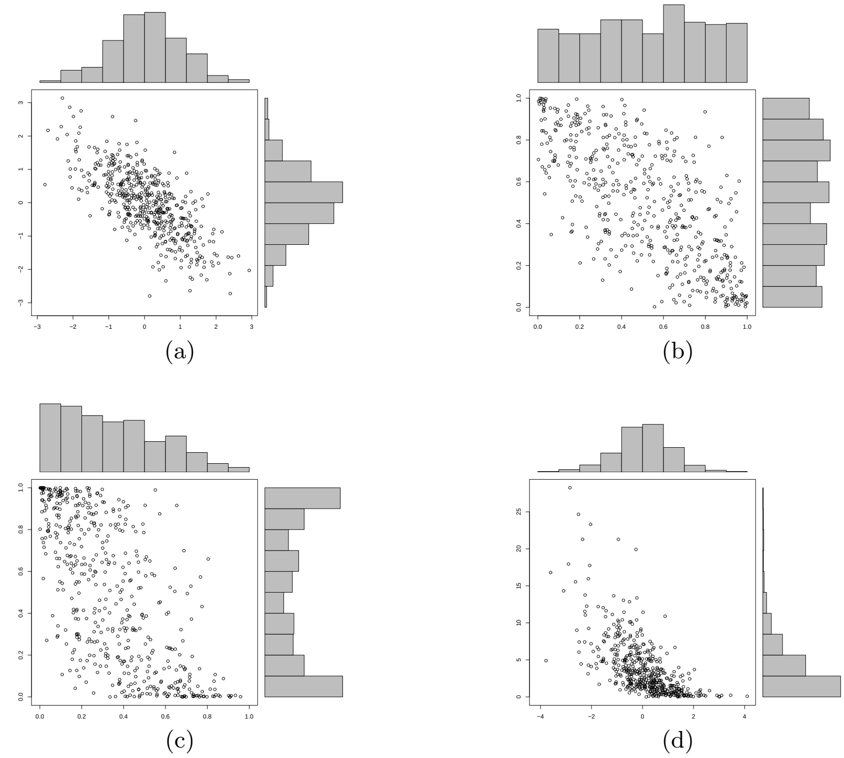


Fig. 2. (a) A sample of 500 points from a standard Gaussian distribution with parameter $\rho = -0.70$. (b) The corresponding sample for a Gaussian copula. (c) The associated sample for a joint distribution with marginal Beta distributions with parameters $(1, 2)$ (histogram on the horizontal axis) and $(0.5, 0.5)$ (histogram on the vertical axis). (d) The associated sample for a joint distribution with marginal t-Student distribution with 8 degrees of freedom (histogram on the horizontal axis) and marginal exponential distribution with mean 4 (histogram on the vertical axis).

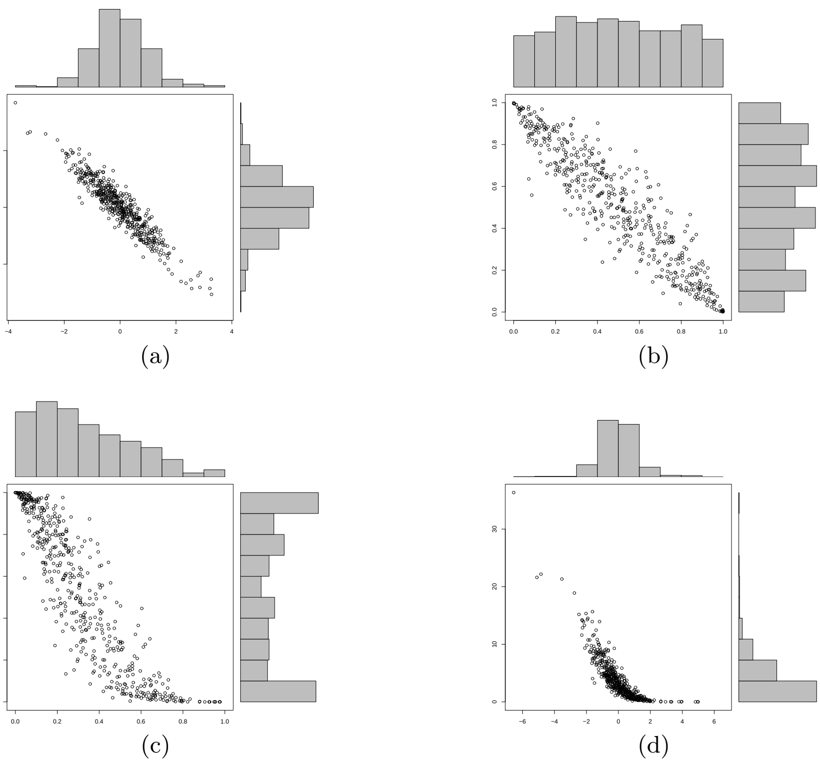


Fig. 3. (a) A sample of 500 points from a standard Gaussian distribution with parameter $\rho = -0.95$. (b) The corresponding sample for a Gaussian copula. (c) The associated sample for a joint distribution with marginal Beta distributions with parameters $(1, 2)$ (histogram on the horizontal axis) and $(0.5, 0.5)$ (histogram on the vertical axis). (d) The associated sample for a joint distribution with marginal t-Student distribution with 8 degrees of freedom (histogram on the horizontal axis) and marginal exponential distribution with mean 4 (histogram on the vertical axis).

Although the correlation is used to generate data from a Gaussian copula, it is not necessary the same for joint distributions with Gaussian copula and non-Gaussian marginals. However, the data sets in Figure 1 have the **same concordance value** measured in Kendall's τ. This important result for parametric bivariate copulas (see [18]) is explained through the equation:

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2; \theta) dC(u_1, u_2; \theta) - 1 \ , \tag{3}$$

which relates the dependence parameter θ of a copula and Kendall's τ. For a bivariate Gaussian copula, Equation (3) can be written as

$$\tau = \frac{2}{\pi} \arcsin(\rho) \ . \tag{4}$$

Given that is well established how to estimate correlation matrixes, evaluate densities, and calculate integrals for the multidimensional Gaussian distribution, the Gaussian copula function is relatively easy to implement.

3 The Probabilistic Classifier

As noted, the aim of this work is to introduce the use of Gaussian copula functions in supervised classification. According to Theorem 1, we can employ a copula function in a probabilistic classifier, such as a Bayessian classifier. In this section we present a three dimensional probabilistic model based on three empirical distribution functions and a trivariate dimensional Gaussian copula function.

The Bayes' theorem states the following:

$$P(K = k|E = e) = \frac{P(E = e|K = k) \times P(K = k)}{P(E = e)} \ , \tag{5}$$

where $P(K = k|E = e)$ is the posterior probability, $P(E = e|K = k)$ is the likelihood function, $P(K = k)$ is the prior probability and $P(E = e)$ is the data probability.

The Equation (5) has been used as a tool in supervised classification. A probabilistic classifier can be designed comparing the posterior probability that an object belongs to class K given its attributes E. The object is then assigned to the class with the highest posterior probability. For practical reasons, the data probability $P(E)$ does not need to be evaluated for comparing posterior probabilities. Furthermore, the prior probability $P(K)$ can be substituted by an uniform distribution if the user does not have an informative distribution.

3.1 The Probabilistic Classifier based on Gaussian Copula Function

For continuous attributes, a Gaussian copula function can be used for modeling the dependence structure in the likelihood function. In this case, the Bayes' theorem can be written as:

$$P(K = k|e) = \frac{c(F_1(e_1), \ldots, F_n(e_n)|k, \Sigma) \times \prod_{i=1}^{n} f_i(e_i|k) \times P(K = k)}{f(e_1, \ldots, e_n)} \quad , \tag{6}$$

where F_i and f_i are the marginal distribution functions and the marginal densities of attributes, respectively. The function c is a d-dimensional Gaussian copula density defined by Equation (2). As can be seen in Equation (6), each category determines a likelihood function.

3.2 The Probabilistic Classifier based on Independent Model

By considering conditional independence among the attributes in Equation (6), or equivalently, an independent structure in the likelihood function given a category, a probabilistic classifier can use the following expression in order to calculate posterior probabilities:

$$P(K = k|e) = \frac{\prod_{i=1}^{n} f_i(e_i|k) \times P(K = k)}{f(e_1, \ldots, e_n)} \quad . \tag{7}$$

Equation (7) uses an independent structure given by a copula density equals to one. This independent copula density can be also obtained by a Gaussian copula function when matrix Σ is the identity matrix I.

3.3 An Application Example

Consider the following specific classification problem: assign a pixel to a certain class according to its color attributes. If we have information about the color distribution of each class, then we can use this information and the Bayes' theorem in order to classify new pixels. This is an example of supervised classification. For a red-green-blue (RGB) color space and two classes, a Gaussian copula based classifier can be written as

$$P(k|r, g, b) = \frac{c(F_R(r), F_G(g), F_B(b)|k, \Sigma) f_R(r|k) f_G(g|k) f_B(b|k) \times P(k)}{f(r, g, b)} \quad , \tag{8}$$

where c is a trivariate Gaussian copula density.

In order to classify a pixel, we use in Equation (8) a prior probability $P(K = k)$ based on the uniform distribution, nonparametric marginal densities $\hat{f}$ based on histograms to approximate $f_R(r|k)$, $f_G(g|k)$ and $f_B(b|k)$, and nonparametric marginal distributions $\hat{F}$ based on empirical cumulative distribution functions to approximate $F_R(r)$, $F_G(g)$ and $F_B(b)$. For modeling the dependence structure of the likelihood function $f(r, g, b|k)$ we present the use of a trivariate Gaussian copula function.

4 Experiments

We use two probabilistic models in order to classify pixels of 50 test images. The first model is an independent probabilistic model (I-M) based on the product of marginal distributions. The second model is a copula-based model (GC-M) that takes into account a dependence structure by means of a trivariate Gaussian copula. The image database was used in [3] and is available online [27]. This image database provides information about two classes: the foreground and the background. The training data and the test data are contained in the labelling-lasso files [27], whereas the correct classification is contained in the segmentation files. Figures 4, 5 and 6 show the description of three images from the database. Table 3 shows a description for each image. Although the database is used for segmentation purposes, the aim of this work is to introduce the use of the Gaussian copula function in supervised color pixel classification. We use the information for supervised color pixel classification, without taking into account the spatial information.

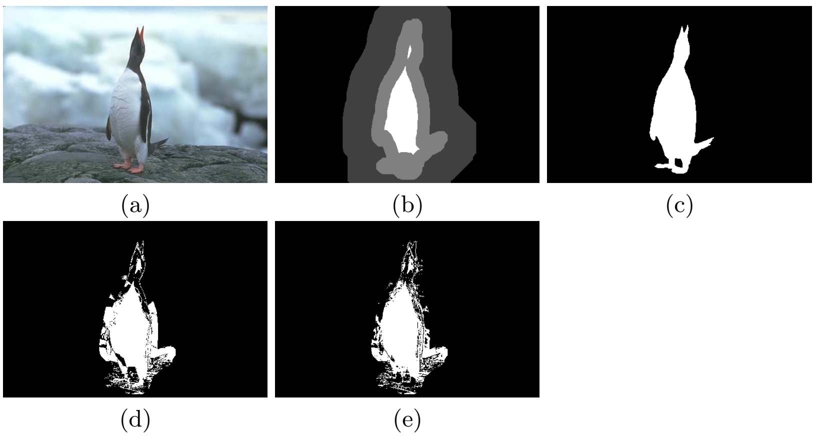
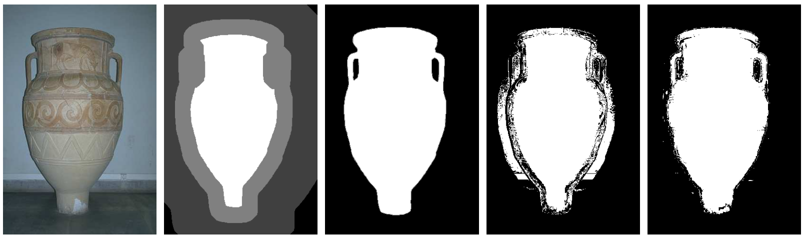


Fig. 4. (a) The color image. (b) The labelling-lasso image with the training data for background (dark gray), for foreground (white) and the test data (gray). (c) The correct classification with foreground (white) and background (black). (d) Classification made by I-M. (e) Classification made by GC-M.

Two evaluation measures are used in this work: *accuracy* and *Tanimoto coefficient*. The accuracy is described in Figure 7. We define the positive class as foreground and the negative class as background.

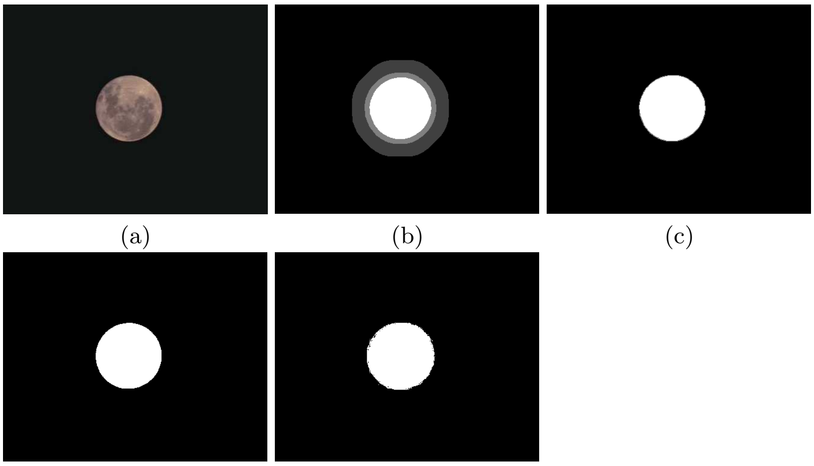
The Tanimoto coefficient (TC) is also known as *Jaccard similarity measure*. This measure, TC, is defined as:

$$TC(k) = \frac{V_{m\cap g}(k)}{V_{m\cup g}(k)} \ ,$$



(a) (b) (c) (d) (e)

Fig. 5. (a) The color image. (b) The labelling-lasso image with the training data for background (dark gray), for foreground (white) and the test data (gray). (c) The correct classification with foreground (white) and background (black). (d) Classification made by I-M. (e) Classification made by GC-M.



(a) (b) (c)

(d) (e)

Fig. 6. (a) The color image. (b) The labelling-lasso image with the training data for background (dark gray), for foreground (white) and the test data (gray). (c) The correct classification with foreground (white) and background (black). (d) Classification made by I-M. (e) Classification made by GC-M.

		Truth	
		Positive	Negative
Model	Positive	tp	fp
	Negative	fn	tn

(a)

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn}$$

(b)

Fig. 7. (a) A confusion matrix for binary classification, where tp are true positive, fp false positive, fn false negative, and tn true negative counts. (b) Definition of accuracy used in this work.

where $V_{m\cap g}(k)$ denotes the number of pixels classified as class k by both the model and the ground truth and $V_{m\cup g}(k)$ denotes the number of pixels classified as class k by either the model or the ground truth.

4.1 Numerical Results

In Table 1 we summarize the measure values reached by the independent probabilistic model (I-M) and the copula-based model (GC-M). The information about the number of pixels well classified for each class is reported in Table 3. We include in Table 4 the performances of I-M and GC-M for each image.

Table 1. Descriptive results for all evaluation measures. BG stands for the background class and FG stands for the foreground class.

Measure	Minimum	Median	Mean	Maximum	Std. deviation
I-M					
Tanimoto coefficient – BG	0.369	0.690	0.695	0.955	0.143
Tanimoto coefficient – FG	0.341	0.633	0.638	0.953	0.168
Accuracy	0.571	0.792	0.795	0.976	0.107
GC-M					
Tanimoto coefficient – BG	0.450	0.816	0.797	0.976	0.118
Tanimoto coefficient – FG	0.375	0.780	0.758	0.972	0.141
Accuracy	0.587	0.889	0.871	0.987	0.083

To properly compare the performance of the probabilistic models, we conducted a hypothesis test based on a Bootstrap method for the differences between the means of accuracy and Tanimoto coefficients, for both probabilistic models. Table 2 shows the confidence interval for the means, and the corresponding p-value.

Table 2. Results for the difference between evaluation measure means in each model. A 95% confidence interval and a p-value are obtained through a Bootstrap technique. BG stands for the background class and FG stands for the foreground class.

Measure	95% Interval		p-value
Tanimoto coefficient – BG	-1.52E-01	-5.18E-02	2.67E-04
Tanimoto coefficient – FG	-1.80E-01	-5.96E-02	3.67E-04
Accuracy	-1.14E-01	-3.94E-02	2.67E-04

4.2 Discussion

According to Table 1, the GC-M shows the best behaviour for all evaluation measures. For instance, the mean accuracy for the I-M, 79.5%, is less than the mean accuracy for the GC-M, 87.1%. This means that using a I-M approximately has 8% more error rate than using a GC-M.

The average of the Tanimoto coefficient for the background class is greater than the average of the Tanimoto coefficient for the foreground class, for both I-M and GC-M (see Table 1). These coefficients are shown for each image in Figure 8-(a) and Figure 8-(b). Notice the Tanimoto coefficients for GC-M on the background and foreground are very similar, denoting a better classification than I-M. In most of the images and for each class, we can see in Figure 8-(c) and 8-(d) (also in Table 4) that GC-M outperforms I-M. In average, according to Table 1, the GC-M improves the I-M in both classes. For the foreground class from 63.8% to 75.8%, and for the background class from 69.5% to 79.7%.

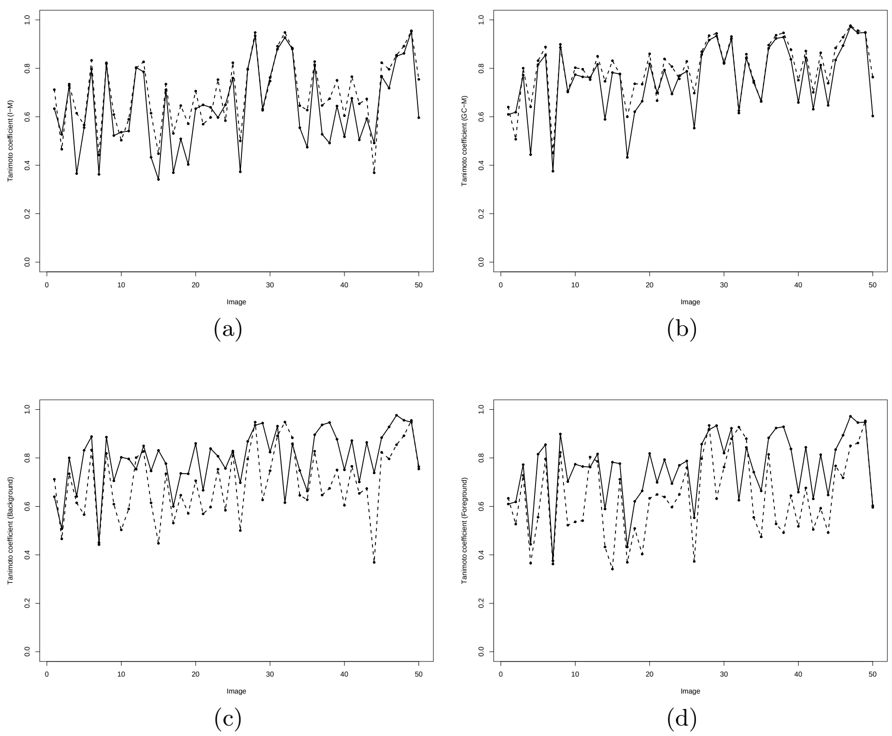


Fig. 8. Tanimoto coefficient for the supervised pixel classification on the 50 images of [3]. Image order is the same as in Table 4. (a) Background (dashed line) and foreground (solid line) for the I-M. (b) Background (dashed line) and foreground (solid line) for the GC-M. (c) I-M (dashed line) and GC-M (solid line) for the background class. (d) I-M (dashed line) and GC-M (solid line) for the foreground class.

Table 1 also shows information about the standard deviations for each evaluation measure. For all cases, the standard deviation indicates that using a GC-M in pixel classification is more consistent than using an I-M.

In order to statistically compare the performance of the probabilistic models, Table 2 shows confidence intervals and p-values that confirm differences between the models. None of confidence intervals include the 0 value and all p-values are less than $\alpha = 0.05$.

5 Conclusions

In this work we introduce the use of Gaussian copulas in supervised pixel classification. According to numerical experiments the selection of a Gaussian copula for modeling structure dependence can help achieve better classification results. An specific example is the image *227092*, which appears in Figure 5, its accuracy for the I-M classifier is 57.1%, whereas its accuracy for the GC-M classifier is 89.5%. For this image, the Gaussian copula improves its accuracy.

Although we model the dependence structure for each image with the same copula function, this is not necessary. There are many copula functions and the Gaussian copula has been chosen due to its practical usefulness and easy implementation. However, having more than one copula at hand may improve the performance of the copula-based classifier. In such case, a copula selection procedure is necessary. The evaluation results are the consequence of the selected dependence structure and marginals. For instance, on the image *106024*, Figure 4, the performance of the I-M classifier is 57.6% accurate (accuracy), whereas the GC-M classifier is 58.7% accurate. For most applications better results can be obtained by selecting the best fitted copula function from a set of available copulas. For example, in the experiment reported, the performance of the I-M classifier is better than GC-M for image *fullmoon*, Figure 6. However, the GC-M is expected to improve the performance of the I-M classifier if we used the proper copula.

Acknowledgments. The first author acknowledges support from the National Council of Science and Technology of México (CONACyT) through a scholarship to pursue graduate studies in the Department of Computer Science at the Center for Research in Mathematics.

References

1. Bacigál, T., Komorníková, M.: Fitting Archimedean copulas to bivariate geodetic data. In: Rizzi, A., Vichi, M. (eds.) Compstat 2006 Proceedings in Computational Statistics, pp. 649–656. Physica-Verlag HD, Heidelberg (2006)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2007)
3. Blake, A., Rother, C., Brown, M., Perez, P., Torr, P.: Interactive Image Segmentation using an adaptive GMMRF model. In: Proc. European Conference in Computer Vision (ECCV), Springer-Verlag (2004)

4. Brunel, N., Pieczynski, W., Derrode, S.: Copulas in vectorial hidden Markov chains for multicomponent image segmentation. In: Proceedings of the 2005 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 717–720 (2005)
5. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. Wiley, Chichester (2004)
6. De-Waal, D.J., Van-Gelder, P.H.A.J.M.: Modelling of extreme wave heights and periods through copulas. Extremes 8(4), 345–356 (2005)
7. Dowd, K.: Copulas in Macroeconomics. Journal of International and Global Economic Studies, 1(1), 1–26 (2008)
8. Flitti, F., Collet, C., Joannic-Chardin A.: Unsupervised Multiband Image Segmentation using Hidden Markov Quadtree and Copulas. In: IEEE International Conference on Image Processing (2005)
9. Frees, E. W., Valdez, E. A.: Understanding relationships using copulas. North American Actuarial Journal, 2(1), 1–25 (1998)
10. Genest, C., Favre, A.C.: Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. Journal of Hydrologic Engineering, 12(4), 347–368 (2007)
11. Grigoriu, M.: Multivariate distributions with specified marginals: Applications to Wind Engineering. Journal of Engineering Mechanics, 133(2), 174–184 (2007)
12. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer (2009)
13. Jajuga, K., Papla, D.: Copula Functions in Model Based Clustering. In: Proceedings of the 29th Annual Conference of the Gesellschaft für Klassifikation e.V., pp. 606–613. Springer (2005)
14. Joe, H.: Multivariate models and dependence concepts. Chapman and Hall, London (1997)
15. Mackay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge (2008)
16. Mercier, G., Bouchemakh, L., Smara, Y.: The Use of Multidimensional Copulas to Describe Amplitude Distribution of Polarimetric SAR Data. In: IGARSS 07 (2007)
17. Monjardin, P.E.: Análisis de dependencia en tiempo de falla. Master's thesis, Centro de Investigación en Matemáticas. Guanajuato, México (2007) (in Spanish)
18. Nelsen, R. B.: An Introduction to Copulas. Springer, Heidelberg (2006)
19. Sakji-Nsibi, S., Benazza-Benyahia, A.: Multivariate indexing of multichannel images based on the copula theory. In: IPTA08 (2008)
20. Salinas-Gutiérrez, R., Hernández-Aguirre, A., Villa-Diharce, E.R.: Using Copulas in Estimation of Distribution Algorithms. In: MICAI 2009: Advances in Artificial Intelligence, pp. 658–668. Springer (2009)
21. Salinas-Gutiérrez, R., Hernández-Aguirre, A., Villa-Diharce, E.R.: D-vine EDA: a new Estimation of Distribution Algorithm based on Regular Vines. In: Genetic and Evolutionary Conference (GECCO-2010). Accepted for publication (2010)
22. Schölzel, C., Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. Nonlinear Processes in Geophysics, 15(5), 761–772 (2008)
23. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231 (1959)
24. Stitou, Y., Lasmar, N., Berthoumieu, Y.: Copulas based multivariate gamma modeling for texture classification. In: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1045–1048 (2009)
25. Trivedi, P.K., Zimmer, D.M.: Copula Modeling: An Introduction for Practitioners. In: vol. 1 of Foundations and Trends® in Econometrics Now Publishers (2007)

26. Venter, G., Barnett, J., Kreps, R., Major, J.: Multivariate Copulas for Financial Modeling. Variance, 1(1), 103–119 (2007)
27. Image and Video Editing, `http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm`

Appendix

Table 3: Description of images used in this work. BG stands for the background class and FG stands for the foreground class. Columns 3 and 4 give the size of test data. The last 4 columns give the number of pixels well classified for each class and for each probabilistic classifier.

Image name	Image size	Test pixels		I-M		GC-M	
		BG	FG	BG	FG	BG	FG
21077	321×481	4322	3353	3648	2551	3144	2763
24077	321×481	11529	10983	6577	8399	6348	10000
37073	321×481	8260	6642	6404	6187	7115	6014
65019	321×481	9853	8398	9181	3317	9099	4061
69020	321×481	25203	22634	17561	16813	22798	20421
86016	321×481	3271	2215	2765	2166	2937	2179
106024	321×481	9093	7368	5528	3961	5574	4087
124080	321×481	18286	18773	16487	16924	16307	18653
153077	321×481	13851	12098	11072	7774	10806	10638
153093	321×481	12027	11809	7617	8699	11414	9615
181079	481×321	23845	23110	18650	15320	22494	18705
189080	481×321	23363	23523	20726	21020	19722	20707
208001	481×321	10227	9530	9994	7669	10064	7914
209070	321×481	6696	4075	5117	2447	5894	2874
227092	481×321	19656	17321	12869	8229	19129	13966
271008	321×481	10909	9216	8934	7967	8800	8795
304074	481×321	7239	4794	5017	2591	5534	2810
326038	321×481	10781	7680	8730	4952	9488	5571
376043	481×321	13654	13485	12022	6094	13072	9343
388016	481×321	17800	15592	15633	11248	17596	12929
banana1	480×640	29983	24052	17120	23964	20285	23601
banana2	480×640	27433	21518	17063	20378	25373	18698
banana3	480×640	26205	20164	25588	12405	26035	14115
book	480×640	26087	21474	15689	20699	19852	21325
bool	450×520	20123	16850	19500	13279	18726	14373
bush	600×450	32513	22099	21072	12504	27734	14870
ceramic	480×640	30549	25709	24809	25069	27328	24791
cross	600×450	34602	25733	32824	25703	32918	25132
doll	549×462	18866	15106	12976	13269	17947	14960
elefant	480×640	27858	22787	20918	22656	23158	22540
flower	450×600	16125	13246	14612	12977	15036	13225
fullmoon	350×442	1580	1043	1498	1043	983	1026
grave	600×450	12294	12832	11977	11567	12219	10889

Continued on next page

Table 3 – continued from previous page

Image name	Image size	Test pixels		I-M		GC-M	
		BG	FG	BG	FG	BG	FG
llama	371 × 513	8930	8445	7783	5322	7547	7287
memorial	600 × 450	14853	12598	12900	6902	10936	10964
music	480 × 640	23945	19494	20457	18723	21794	19112
person1	450 × 600	19092	16384	16452	10041	18831	15372
person2	450 × 600	12796	9595	11492	5358	12465	9219
person3	600 × 450	14649	11450	13494	8122	14022	10112
person4	450 × 600	19250	16631	15230	10691	18197	11653
person5	600 × 450	13990	11332	13009	8327	13025	10377
person6	600 × 450	19015	15645	16753	9038	16071	11732
person7	600 × 450	12110	9634	9934	6998	11795	8093
person8	480 × 640	16684	12741	6740	11157	14690	9534
scissors	480 × 640	30768	23335	28152	19910	30181	19960
sheep	600 × 450	5331	3733	4750	3098	5243	3415
stone1	480 × 640	18716	15635	16087	15525	18376	15528
stone2	480 × 640	22002	18489	21556	16315	21788	17692
teddy	398 × 284	13892	13739	13790	13191	13426	13466
tennis	472 × 500	19471	13129	18054	8673	18322	8613

Table 4: Evaluation measures for each image. TC stands for the Tanimoto coefficient, BG stands for the background class and FG stands for the foreground class. Columns 2, 3 and 4 give the results for the independent probabilistic model. The last 3 columns give the results for the Gaussian copula-based probabilistic model.

Image name	I-M			GC-M		
	TC-BG	TC-FG	Accuracy	TC-BG	TC-FG	Accuracy
21077	0.712	0.633	0.808	0.640	0.610	0.770
24077	0.466	0.527	0.665	0.507	0.619	0.726
37073	0.735	0.728	0.845	0.801	0.772	0.881
65019	0.615	0.366	0.685	0.641	0.444	0.721
69020	0.566	0.555	0.719	0.832	0.816	0.903
86016	0.833	0.796	0.899	0.888	0.855	0.933
106024	0.442	0.362	0.576	0.450	0.375	0.587
124080	0.819	0.823	0.902	0.886	0.899	0.943
153077	0.609	0.523	0.726	0.706	0.703	0.826
153093	0.503	0.536	0.685	0.803	0.774	0.882
181079	0.590	0.541	0.723	0.796	0.765	0.877
189080	0.801	0.804	0.890	0.753	0.762	0.862
208001	0.827	0.786	0.894	0.850	0.816	0.910
209070	0.615	0.433	0.702	0.746	0.589	0.814
227092	0.448	0.341	0.571	0.831	0.782	0.895
271008	0.735	0.712	0.840	0.777	0.777	0.874
304074	0.531	0.369	0.632	0.600	0.432	0.693

Continued on next page

Table 4 – continued from previous page

Image name	I-M			GC-M		
	TC-BG	TC-FG	Accuracy	TC-BG	TC-FG	Accuracy
326038	0.646	0.509	0.741	0.736	0.621	0.816
376043	0.571	0.403	0.668	0.735	0.664	0.826
388016	0.706	0.633	0.805	0.860	0.818	0.914
banana1	0.569	0.649	0.760	0.667	0.699	0.812
banana2	0.597	0.639	0.765	0.839	0.793	0.900
banana3	0.753	0.597	0.819	0.807	0.694	0.866
book	0.584	0.649	0.765	0.757	0.770	0.866
bool	0.823	0.760	0.887	0.829	0.788	0.895
bush	0.500	0.373	0.615	0.698	0.553	0.780
ceramic	0.795	0.797	0.887	0.868	0.857	0.926
cross	0.948	0.934	0.970	0.935	0.917	0.962
doll	0.627	0.632	0.773	0.944	0.934	0.969
elefant	0.747	0.762	0.860	0.824	0.820	0.902
flower	0.891	0.879	0.939	0.931	0.923	0.962
fullmoon	0.948	0.927	0.969	0.616	0.626	0.766
grave	0.883	0.880	0.937	0.858	0.844	0.920
llama	0.646	0.555	0.754	0.748	0.741	0.854
memorial	0.628	0.474	0.721	0.663	0.664	0.798
music	0.828	0.815	0.902	0.896	0.883	0.942
person1	0.647	0.528	0.747	0.937	0.924	0.964
person2	0.675	0.492	0.753	0.946	0.929	0.968
person3	0.751	0.644	0.828	0.877	0.837	0.925
person4	0.605	0.518	0.722	0.751	0.659	0.832
person5	0.765	0.676	0.843	0.872	0.844	0.924
person6	0.654	0.505	0.744	0.701	0.631	0.802
person7	0.674	0.593	0.779	0.864	0.813	0.915
person8	0.369	0.492	0.608	0.739	0.647	0.823
scissors	0.823	0.767	0.888	0.884	0.834	0.927
sheep	0.796	0.718	0.866	0.928	0.894	0.955
stone1	0.855	0.850	0.920	0.976	0.972	0.987
stone2	0.892	0.862	0.935	0.956	0.946	0.975
teddy	0.955	0.953	0.976	0.948	0.948	0.973
tennis	0.755	0.596	0.820	0.764	0.603	0.826