On the Accuracy of Copula-based Bayesian Classifiers: An Experimental Comparison with Neural Networks

Lukáš Slechan and Jan Górecki

Department of Informatics, SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic
{o130079,gorecki}@opf.slu.cz

Abstract. In this work, we compare three classifiers in terms of accuracy. The first is a copula-based Bayesian classifier based on elliptical and Archimedean copulas. The remaining two are Naive Bayes and Neural Networks. Such a comparison, particularly for the recently proposed Archimedean copula-based Bayesian classifiers, hasn't been reported in the literature. The results show that copula-based Bayesian classifiers are a viable alternative to Neural Networks in terms of accuracy while keeping the models relatively simple.

Keywords: copula · elliptical copula · Archimedean copula · Bayesian classification · Neural Networks

1 Introduction

In machine learning, classification is the problem of identifying to which of the set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known. Learning relationships between random variables is a decisive task in the field of knowledge discovery and data mining. The dependence between the observed variables can be studied by means of copulas. Copulas are distribution functions with standard uniform univariate margins and are widely used particularly for studying dependence between continuously distributed random variables; for more details on copulas, e.g., see [12]. The word copula comes from the latin copula, which means "a bond or a link" and was first used by Sklar [15]. Despite the fact that a large part of the success of copulas is attributed to finance [14], copulas are more and more adopted in data mining [9], [7], hydro-climatic and water-resources [4], [10], gene analysis [18] or cluster analysis [3].

In this paper, we consider copula-based Bayesian classifiers (CBCs) based on elliptical copulas (ECs) and Archimedean copulas (ACs). These classifiers are experimentally compared on 8 real-world datasets with other two commonly known classifiers - Neural Networks (NN), which are appreciated for their high accuracy but which however produce complex and thus low-understandable models, and Naive Bayes (NB), which, as a special case of CBCs assuming the independence copula for the variables, produces well-understandable however less-accurate models. The research reported in this paper follows the research presented in [7], where CBCs are compared with Classification and regression trees, Random forests and Support vector machines. Our comparison complements that work with a comparison of CBCs to NN.

The paper is structured as follows. Section 2 summarizes some needed theoretical concepts including ECs and ACs. Section 3 recalls CBCs. Section 4 presents the results of an experimental comparison of the above-mentioned classifiers based on real-world datasets and Section 5 concludes.

2 Preliminaries

2.1 Copulas

Definition 1 A d-dimensional copula C: $[0,1]^d \rightarrow [0,1]$ is a function which is a multivariate distribution function with uniform univariate margins on the interval $[0,1]$.

Copulas establish a connection between multivariate distributions and their univariate margins, see the following theorem.

Theorem 1 (Sklar's Theorem) Let F be a d-dimensional multivariate distribution function with univariate margins $F_1, \ldots, F_d$. Then there exists a copula C: $[0,1]^{\mathrm{d}} \rightarrow [0,1]$ such that

$$F(x_1, \ldots, x_d) = C\big(F_1(x_1), \ldots, F_d(x_d)\big) \tag{1}$$

holds for all $(x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Such a function C is uniquely determined, if $F_1, \ldots, F_d$ are all continuous. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F given by (1) is a multivariate distribution function with margins $F_1, \ldots, F_d$.

Hence, Sklar's Theorem allows to model the univariate margins and the copula of a multivariate distribution function separately, which allows to build very flexible class of multivariate distribution models. There are many parametric copula families available, which usually have parameters that control the strength of dependence. Some popular parametric copula models are recalled below.

2.2 Some parametric copula families

Basic information about copula families are presented, e.g., in [12]. In this paper, we use two families of ECs and three families of ACs.

Elliptical copulas

ECs are based on existing multivariate elliptical distributions and are derived directly using Sklar's Theorem. The Gaussian copula is based on the multivariate normal distribution and the Student's t-copula is based on the multivariate Student's t-distribution. Formally, a Gaussian copula is given by

$$C_{\Sigma}^{Ga}(u_1, \ldots, u_d) = \phi_{\Sigma}\left(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)\right), \tag{2}$$

where ϕ is the cumulative distribution function (CDF) of the univariate normal distribution and ϕ_Σ is the CDF of the multivariate normal distribution with a correlation matrix ∑. A Student's t-copula is given by

$$C^t_{v,\Sigma}(u_{1,\dots,}u_d) = t_{v,\Sigma}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d)), \tag{3}$$

where t_v is the CDF of the univariate Student's t-distribution with *v* degrees of freedom and $t_{v,\Sigma}$ is the CDF of the multivariate Student's t-distribution with a correlation matrix ∑ and *v* degrees of freedom.

Archimedean copulas

Definition 2 An *Archimedean generator* (simply, *generator*) is a continuous, non-increasing function : $[0,\infty] \to [0,1]$, which satisfies $\psi(0) = 1, \psi(\infty) = \lim_{t\to\infty} \psi(\mathrm{t}) = 0$ and which is strictly decreasing on $[0, \inf\{t \mid \psi(t) = 0\}]$.

Definition 3 Any *d*-copula *C* is called AC, if it admits the form

$$C(\mathbf{u}) \coloneqq C\,(\mathbf{u};\psi) \coloneqq \psi\big(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\big), \mathbf{u} \in [0,1]^d, \tag{4}$$

where ψ is a generator and $\psi^{-1}\colon [0,1] \to [0,\infty]$ is defined by $\psi^{-1}(\mathrm{s}) = \inf\{t\ \psi(t) = s\}, s \in [0,1]$.

To derive an explicit form of an AC, we need explicit generators. For the construction of CBCs described in Section 3, we use the three popular families of ACs presented in Table 1. For other families of ACs, e.g., see [9].

Table 1. The three considered one-parametric Archimedean copula families with the corresponding parameter ranges and forms.

Family	θ	$\psi\,(t)$
Clayton (C)	$(0,\infty)$	$(1+t)^{-1/\theta}$
Frank (F)	$(0,\infty)$	$-\log(1-\big(1-\mathrm{e}^{-\theta}\big)\exp(-t))/\theta$
Gumbel (G)	$[1,\infty)$	$\exp(-t^{1/\theta})$

3 Construction of copula-based Bayesian classifiers

We briefly recall some basics for Bayesian classifiers and a way how copulas could be integrated into them; see also [7], [14].

Let $\Omega = \{\omega_1, \dots, \omega_m\}$ be a finite set of m classes. The problem of classification is to assign each x from the variable space $\mathbb{R}^d$ a class from Ω. A Bayesian classifier is said to assign x to the class ω_i, if

$$g_i(x) > g_j(x) \qquad \text{for all } j \neq i, \tag{5}$$

where $g_i\colon [0,\infty)^d \to \mathbb{R}, i = 1, \dots, m$ are called *discriminant functions* that are defined by

$$g_i(x) = \mathbb{P}(\omega_i|x) = \frac{f\,(x|\omega_i)\mathbb{P}(\omega_i)}{\sum_{j=1}^{m} f\left(x|\omega_j\right)\mathbb{P}(\omega_j)}, \tag{6}$$

where $f\colon \mathbb{R}^d \to [0,\infty)$ is a probability density function and $\mathbb{P}(\omega_i),\ i = 1, \dots, m$ are the prior probabilities of the classes from Ω. Since any monotonically increasing function $Q : \mathbb{R} \to \mathbb{R}$ keeps the classification unaltered, the discriminant functions can be simplified by $g_i := Q \circ g_i$ with $Q(t) = \ln(t \sum_{j=1}^{m} f(\mathrm{x} \mid \omega_j)\mathbb{P}(\omega_j))$ from (6) to

$$g_i(x) = \ln f(x|\omega_i) + \ln \mathbb{P}(\omega_i)\,. \tag{7}$$

Provided *F* given by (1) is an absolutely continuous multivariate distribution function with margins $F_1, \dots, F_d$, the pdf *f* of *F* can be expressed by

$$f(x_1, \dots, x_d) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \prod_{k=1}^{d} f_k(x_k), \tag{8}$$

where $c(u_1, \dots, u_d) = \frac{\partial^d C(u_1,\dots,u_d)}{\partial u_1 \dots \partial u_d}$ denotes the density of the copula $C(u_1, \dots, u_d)$ and f_k denotes the density of $F_k, k = 1, \dots, d.$ Using (8), $f(x|\omega_i)$ can be rewritten to

$$f(x|\omega_i) = c(F_1(x_1|\omega_i), \dots, F_d(x_d|\omega_i)|\omega_i) \prod_{k=1}^{d} f_k(x_k|\,\omega_i), \tag{9}$$

and thus (7) turns to

$$g_i(x) = \ln(c(F_1(x_1|\omega_i), \dots, (F_d(x_d|\omega_i)\,|\omega_i))) + \sum_{k=1}^{d} \ln(f_k(x_k|\omega_i)) + \ln(\mathbb{P}(\omega_i)). \tag{10}$$

Hence, the discriminant function g_i is composed of three ingredients: the conditional copula density, the conditional marginal densities and the prior probability of the class ω_i. Note that these ingredients do not impose any restrictions on each other.

4 Experiments

4.1 Design on the experiments

In this subsection, we evaluate the accuracy of 8 classifiers. Five of them, considering different underlying families of copulas, are CBCs, where two of them are based on ECs and three of them are based on ACs. More precisely, we consider:

- **Elliptical copula-based Bayesian classifiers**. For these classifiers, it is assumed that $\hat{C}(\cdot\,|\omega_i)$ is a Gaussian (denoted as ECBC(G)) or Student's t-copula (ECBC(t)), respectively. The computation $\hat{C}(\cdot\,|\omega_i)$ is implemented by Matlab's Statistics and machine learning toolbox function `copulafit` with the parameter `family` set to the value `Gaussian` or `t` copula, respectively.

- **Archimedean copula-based Bayesian classifiers**. For these classifiers, it is assumed that $\hat{C}(\cdot\,|\omega_i)$ is a Clayton (denoted as ACBC(C)), Gumbel (ACBC(G)) or Frank (ACBC(F)) copula, respectively. The copula parameter is estimated by inversion of pairwise Kendall's tau, e.g., see [4], [5], [7].

The estimates $\hat{F}_1(\cdot\,|\omega_i), \dots, \hat{F}_d(\cdot\,|\omega_i)$ of $F_1(\cdot\,|\omega_i), \dots, F_d(\cdot\,|\omega_i)$ in (10) are computed in the same way for all above-mentioned classifiers using the Kernel smoothing function `ksdensity` with the parameter `function` set to `cdf`.

These CBCs are compared with the following classifiers available in Matlab:

- **Naive Bayes** (denoted by NB). For the classifier $\hat{C}(\cdot\,|\omega_i), i = 1, \dots, m$ is independence copula. For the following tt is implemented by function `fitNaiveBayes` and it is training phase, we set the parameter `Distribution` to the value `normal` (this classifier is denoted by NAIVE(N)) or `kernel` (denoted by NAIVE(K)).

- **Neural Networks** (denoted by NN). It is a two-layer feed-forward network with hidden and softmax output neurons. The classifier is implemented by the function `patternnet` with the training function set to the default, i.e., the scaled conjugate gradient back-propagation is used. The parameter `hiddenLayerSize` is set to one of the values 5, 10 and 15 based on a 10-fold cross-validation.

Note that if the reader is interested in a comparison of CBCs to other types of classifiers, e.g., to Classification and regression trees, Random forests or Support vector machines, such a comparison is reported in [7].

In summary, we evaluate these 8 classifiers on 8 commonly known datasets from UCI-dataset repository [2], namely on Iris (4 variables, 3 classes), BankNote (4 variables, 2 classes), Seeds (7 variables, 3 classes) and BreastTissue (9 variables, 4 classes), Wine (13 variables, 3 classes), and two datasets from the KEEL-dataset repository[1], namely Hayes-Roth (5 variables, 3 classes) and Appendicitis (7 variables, 2 classes). The eighth dataset, which is a results of a recent real-world application in catalysis [11], we the Catalysis dataset. We selected these datasets in order to all considered classifiers could be applicable for each of them.

The accuracy computation for a given classifier and dataset is based on a 10-fold cross-validation and repeated 10 times. All computations were performed in Matlab on a PC with Intel Core i3-3220 CPU @ 3.30 GHz, 8GB RAM.

4.2 Results of the experiments

The accuracy of the classifiers computed for the selected datasets is shown in Fig. 1.

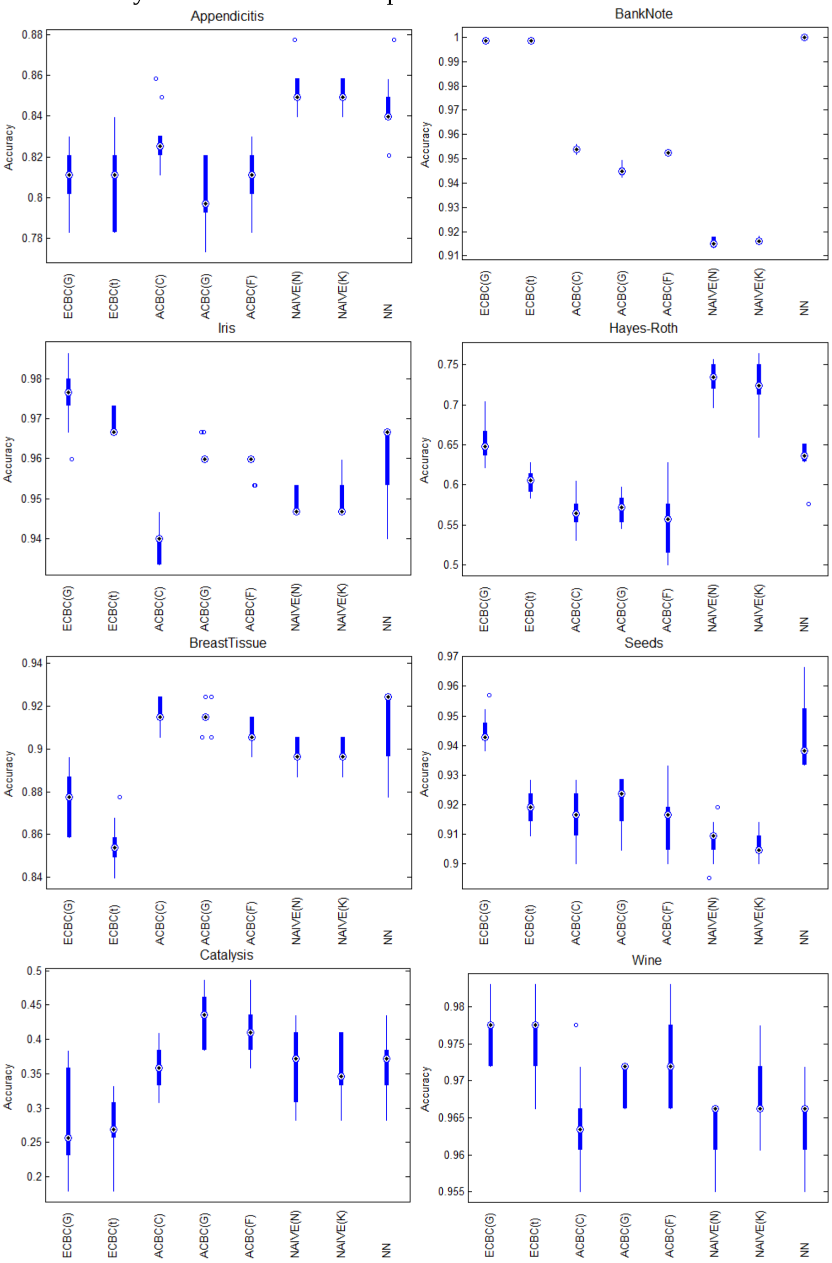

Fig. 1. The accuracy (boxplots) of the classifiers computed on the Appendicitis, BankNote, Iris, Hayes-Roth, BreastTissue, Seeds, Catalysis and Wine datasets.

It can be observed in Fig. 1, that there is not a top winning classifier on all chosen datasets. It thus confirms the "No Free Lunch Theorem" [17]. However, one can observe that there are classifiers which scored higher more often than others. This observation is supported by rankings of the classifiers shown in Table 2. Each of the classifiers is ranked according to its averaged accuracy (1 - the highest, 8 – the lowest).

Table 2. The rankings of the classifiers according to the averaged accuracy on a given dataset.

Classifier	ECBC(G)	ECBC(t)	ACBC(C)	ACBC(G)	ACBC(F)	NAIVE(N)	NAIVE(K)	NN
Appendicitis	5	7	4	8	5	1	2	3
BankNote	3	2	4	6	5	8	7	1
Iris	1	2	8	4	5	7	6	3
Hayes-Roth	3	5	7	6	8	1	2	4
BreastTissue	7	8	2	3	4	5	5	1
Seeds	1	4	5	3	6	7	8	2
Catalysis	8	7	5	1	2	3	6	4
Wine	1	2	8	4	3	7	5	6
average rank	3.625	4.625	5.375	4.375	4.75	4.875	5.125	3

Observing the averages of the ranks – the average rank row in Table 2 - three groups of classifiers can be distinguished:

- The highest-ranked classifiers – NN (average rank = 3) and ECBC(G) (3.625);
- The middle-ranked classifiers – ACBC(G) (4.375), ECBC(t) (4.625), ACBC(F) (4.75) and NAIVE(N) (4,875);
- The lowest-ranked classifiers – NAIVE(K) (5.125) and ACBC(C) (5.375).

These results show that CBCs, particularly ECBC(G), could be considered competitive with the best in average performing NN. It should be also mentioned that, even if the CBCs based on ACs perform in average worse than the highest-ranked classifiers (this result corresponds to [7]), these classifiers, namely ACBC(G) and ACBC(F) are the best performing classifiers on the Catalysis dataset and the second best performing classifiers on the BreastTissue dataset, and hence could also be considered as a viable and simple alternative to NN.

5 Conclusion

In this work, elliptical and Archimedean copula-based Bayesian classifiers are experimentally compared to Neural Networks and Naive Bayes in terms of accuracy for the first time. The results based on 8 real-world datasets have shown that copula-based

Bayesian classifiers are, in terms of accuracy, a viable alternative to highly accurate Neural Networks while keeping the models relatively simple.

In further research, we would like to extend the research presented here by involving other copula-based Bayesian classifiers based on, e.g., hierarchical Archimedean copulas, pair copulas, etc. These families of copulas, for which serious researching effort can be recently observed, e.g., see [6], [8], [16], overcome some restrictions of elliptical and Archimedean copulas, are flexible but more computationally demanding, see, e.g., the discussion concerning the computation of hierarchical Archimedean copula density functions in high dimensions in [7]. Nevertheless, in our opinion, bringing these families into a play could substantially increase the accuracy of copula-based Bayesian classifiers while still keeping the models relatively simple.

Also, the considered copula-based Bayesian classifiers, despite ranked lower than the neural network based classifier in the averaged accuracy, outperform it on several datasets. In the light of this fact, it would be desirable to consider, e.g., the relation between the datasets features and the ranks, or which are the subclasses of problems for which the copula-based classifiers perform better than other classifiers.

Acknowledgement

The research of this work was funded by project SGS/21/2014 – Advanced methods for knowledge discovery from data and their application in expert systems, Czech Republic. This work has been originally published in Lecture Notes in Computer Science (vol. 9329), see http://link.springer.com/chapter/10.1007/978-3-319-24069-5_46. For a BibTex citation, see http://suzelly.opf.slu.cz/~gorecki/en/research.php.

References

1. Alcalá, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing 17, 255–287 (2010)
2. Bache, K., Lichman, M.: UCI machine learning repository (2013). URL http://archive.ics.uci.edu/ml
3. Cuvelier, E., Noirhomme-Fraitur, M.: Clayton copula and mixture decomposition. In: Applied Stochastic Models and Data Analysis, ASMDA'05. Brest (2005).
4. Genest, C., Favre, A.: Everything you always wanted to know about copula modeling but were afraid to ask. Hydrol. Eng. 12, 347–368 (2007)
5. Górecki, J., Holeňa, M. (2014). Structure determination and estimation of hierarchical Archimedean copulas based on Kendall correlation matrix. In: A. Appice, M. Ceci, C. Loglisci, G. Manco, E. Masciari, Z.W. Ras (eds.) New Frontiers in Mining Complex Patterns, Lecture Notes in Computer Science, pp. 132–147. Springer International Publishing.
6. Górecki, J., Holeňa. M. (2013). An alternative approach to the structure determination of hierarchical Archimedean copulas. Proceedings of the 31st International Conference on Mathematical Methods in Economics (MME 2013), pp. 201 – 206, Jihlava, Czech Republic.

7. Górecki, J., Hofert, M., Holeňa, M. (2015). An Approach to Structure Determination and Estimation of Hierarchical Archimedean Copulas and Its Application in Bayesian Classification. Journal of Intelligent Information Systems, pp. 1 - 39, Springer Science+Business Media New York.
8. Górecki, J., Hofert, M., Holeňa, M. (2014). On the consistency of an estimator for hierarchical Archimedean copulas. Proceedings of the 32nd International Conference on Mathematical Methods in Economics, pp. 239 – 234, Olomouc, Czech Republic, 2014.
9. Holeňa, M., Ščavnický, M. (2013). Application of copulas to data mining based on observational logic. In ITAT: Information Technologies Applications and Theory Workshops, Posters, and Tutorials, North Charleston: CreateSpace Independent Publishing Platform, Donovaly, Slovakia.
10. Kao, S.C., Govindaraju, R.S.: Trivariate statistical analysis of extreme rainfall events via plackett family of copulas. Water Resour. Res. 44 (2008)
11. Moehmel, S., Steinfeldt, N., Engelschalt, S., Holena, M., Kolf, S., Baerns, M., Dingerdissen, U., Wolf, D., Weber, R., Bewersdorf, M.: New catalytic materials for the high-temperature synthesis of hydrocyanic acid from methane and ammonia by high-throughput approach. Applied Catalysis A: General 334(1), 73–83 (2008)
12. Nelsen, R. B.: An Introduction to Copulas. 2nd ed. New York: Springer (2006)
13. Rank, J.: Copulas: From Theory to Application in Finance. London: Risk Books (2006)
14. Sathe, S.: A novel Bayesian classifier using copula functions. arXiv preprint cs/0611150 (2006)
15. Sklar, A.: Fonctions de répartition a n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231 (1959)
16. Stöber, J., Joe, H., Czado, C. (2013). Simplified pair copula constructions - limitations and extensions. Journal of Multivariate Analysis, 119, 101-118.
17. Wolpert, D.H.: The supervised learning no-free-lunch theorems. In: Soft Computing and Industry, pp. 25–42. Springer (2002)
18. Yuan, A., Chen, G., Zhou, Z.C., Bonney, G., Rotimi, C.: Gene copy number analysis for family data using semiparametric copula model. Bioinform Biol Insights. 2., 343–355 (2008)