Implementation of a copula-based classifier

1 Description

Classifiers based on probabilistic graphical models (e.g. those based on Gaussian and Bayesian networks) are extensively applied in many areas. Recently, several methods based on Copulas have been proposed in this domain [1, 2, 3, 4, 5, 6, 7]. Copulas provide a flexible tool to build multivariate distributions from marginals wich are modeled by different distributions, and a function (the copula) that links these marginals. Classifiers based on copulas can be especially suited to deal with problems with a continuous representation in which there are intricate relationships among the problem variables.

2 Objectives

The goal of the project is to implement a copula-based classifier in Python. The student should: 1) Select one of the copula-based models proposed in the literature, or propose a new one, and implement it. 2) Identify which datasets are relevant for evaluating the classifier. 3) Validate the implementation using the datasets.

As in other projects, a report should describe the characteristics of the design, implementation, and results. A Jupyter notebook should include calls to the implemented function that illustrate the way it works.

3 Suggestions

- The copulalib Python library provides implementation of several relevant functions for learning copula from data `https://pypi.python.org/pypi/copulalib`.
- Also the pyvine library provides more sophisticated functions for learning and sampling vine copulas `https://pypi.python.org/pypi/pyvine/0.5.0`.
- One straightforward approach to build classifiers with copulas is using a Gaussian Naive Bayes approach using copulas instead of gaussian models. See `https://machinelearningmastery.com/naive-bayes-classifier/-scratch-python/` for an implementation of Naive Bayes from scratch.
- Alternatively, any of the classifiers proposed in the literature could be implemented [1, 2, 3, 4, 5, 6, 7].

References

[1] Diana Carrera, Roberto Santana, and Jose A Lozano. Vine copula classifiers for the mind reading problem. *Progress in Artificial Intelligence*, pages 1–17, 2016.

[2] Gal Elidan. Copula Bayesian networks. In *Advances in neural information processing systems*, pages 559–567, 2010.

[3] Gal Elidan. Copula network classifiers (CNSs). In *Artificial Intelligence and Statistics*, pages 346–354, 2012.

[4] Gal Elidan. Copulas in machine learning. In *Copulae in mathematical and quantitative finance*, pages 39–60. Springer, 2013.

[5] Rogelio Salinas-Gutiérrez, Arturo Hernández-Aguirre, Mariano JJ Rivera-Meraz, and Enrique R Villa-Diharce. Using Gaussian copulas in supervised probabilistic classification. In *Soft Computing for Intelligent Control and Mobile Robotics*, pages 355–372. Springer, 2010.

[6] Saket Sathe. A novel bayesian classifier using copula functions. *CoRR*, abs/cs/0611150, 2006.

[7] Lukáš Slechan and Jan Górecki. On the accuracy of copula-based Bayesian classifiers: an experimental comparison with neural networks. In *Computational Collective Intelligence*, pages 485–493. Springer, 2015.