Maximum Margin Optimal Decision Trees

Federico D'Onofrio, Marta Monaci, Laura Palagi
Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG),
Sapienza University of Rome, Via Ariosto 25, Rome 00185, Italy
{federico.donofrio, marta.monaci, laura.palagi}@uniroma1.it

Giorgio Grani
Department of Statistical Sciences,
Sapienza University of Rome, Piazzale Aldo Moro 5, Rome 00185, Italy
{giorgio.grani}@uniroma1.it

Abstract. In recent years there has been a growing attention to interpretable machine learning models which are able to give explanatory insights on the decisions made by the algorithm. Especially in healthcare applications, it is of great interest to find a hierarchy on the subsets of features which mostly affect the outcome of a classification algorithm. Thanks to their interpretability, decision trees have been intensively studied for classification tasks, and, due to the remarkable advances in mixed-integer programming (MIP), various approaches have been proposed to formulate the Optimal Decision Tree (ODT) problem as a MIP model [1]. We present a novel MIP formulation for the ODT problem which exploits the generalization capabilities of support vector machines, see also [2, 3]. The Maximum **MARG**in **O**ptimal Decision **T**ree (MARGOT) selects at each node of the decision tree a maximum margin separating hyperplane using an ℓ_2-norm linear support vector machine. The resulting model combines such multivariate hyperplanes minimizing the global misclassification error. The formulation can also include feature selection constraints to improve interpretability. MARGOT has been tested on non-linearly separable synthetic datasets in a 2-features space to provide a graphical representation of the optimal hyperplanes. Finally, we evaluate our model on well-known datasets from the UCI repository.

Keywords: interpretable machine learning; mixed-integer programming; optimal decision trees; support vector machines

References

[1] Bertsimas, D. and Dunn, J. (2017). Optimal classification trees. Machine Learning, 106, 1039–1082.

[2] Bennett, K.P. and Blue, J.A. (1998). A support vector machine approach to decision trees. 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227), 3, 2396-2401 vol.3.

[3] Zhu, H., Murali, P., Phan, D., Nguyen, L.M., & Kalagnanam, J. (2020). A Scalable MIP-based Method for Learning Optimal Multivariate Decision Trees. ArXiv, abs/2011.03375.