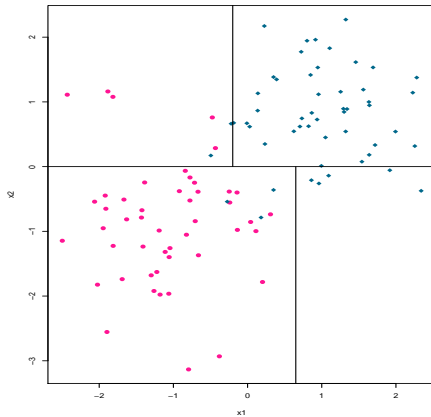


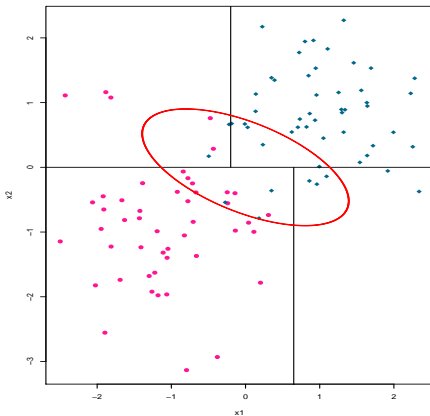
Characterizing Uncertainty in Decision Trees through Imprecise Splitting Rules

ISIPTA 2019

Malte Nalenz & Thomas Augustin

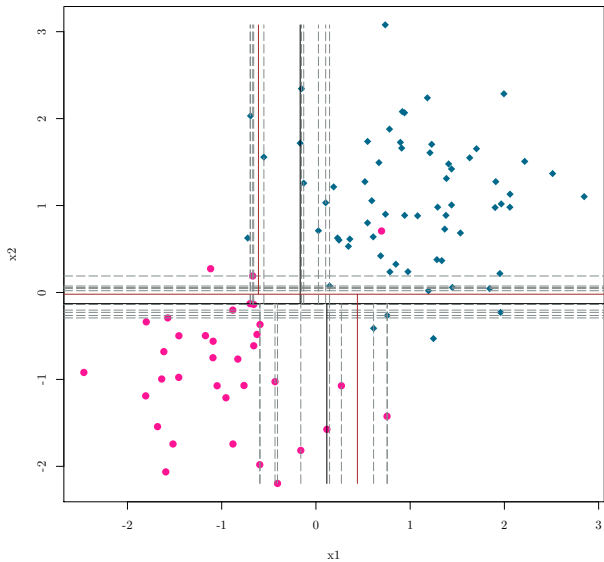
July 6, 2019





Oberservations close to the decision boundarie(s) more uncertain, as small perturbations lead to different prediction

Instead of using a single splitpoint $x \leq t_0$ we propose to consider a neighbourhood: $\mathcal{T} = \{t_{-k} = x_{-k}, \dots, t_0, \dots, t_k = x_k\}$, where t_0 is the candidate split and t_k and t_{-k} the k'th datapoints with higher and lower ordered covariate values as reasonable alternative splitting values.



For details and discussion please come to the poster!

Quantifying Uncertainty in Decision Trees through imprecise splitting Rules

Malte Nalenz, Thomas Augustin
malte.nalenz@uni-wuerzburg.de, thomas.augustin@uni-wuerzburg.de



Motivations

Decision tree (DT) induces uncertainty particularly in situations where the data are noisy and the model is overfitted. This uncertainty can be quantified by using imprecise splitting rules.

Basic Concepts

A node n of a DT induces uncertainty for the split τ if the uncertainty for the split is non-zero. This uncertainty can be quantified by using imprecise splitting rules. The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Imprecise Splitting Rules

Instead of using a single split rule to split the data, we use a set of split rules. The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Feasible Splitting Rules

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Non-binary Splitting

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Interpretation

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Acknowledgements

We thank the anonymous reviewers for their helpful comments and suggestions.

Illustration of Imprecise Splitting Rules

Consider the following data set D (see Table 1). The data are split into two classes, C_1 and C_2 . The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Prediction

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Single Prediction

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Prediction Interval

The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

Discussion

- The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .
- The uncertainty for the split τ is defined as the uncertainty for the split τ induced by the data D at node n .

References

[1] Malte Nalenz and Thomas Augustin, Imprecise Splitting Rules, in: Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI Press, 2017.

