

1. Your answers to Question 1 go here. In all answers, please adhere to the instructions in **organization.pdf**. Keep in mind the one-page limit (excluding potential references) and don't forget to fill in your names and group number.
2. One way to detect overfitting is to compare model performance against a validation data set. By varying, for example, tree depth when fitting a random forest classifier, performance gain can be compared in both data sets against depth of the model. As decision trees often suffer from overfitting when model depth is large, the optimal tree depth can be determined this way. As can be seen in Fig 1, the validation score stabilises only after increasing the tree depth to about 20. At this level the training data already gives a score of almost 1.0. This means that not overfitting of the training data is happening as increasing the tree depth does not further increase the training score, while also not further decreasing the validation score
3. Your answers to Question 3 go here.
4. Your answers to Question 4 go here.
5. Your answers to Question 5 go here.
6. Your answers to Question 6 go here.

Appendix

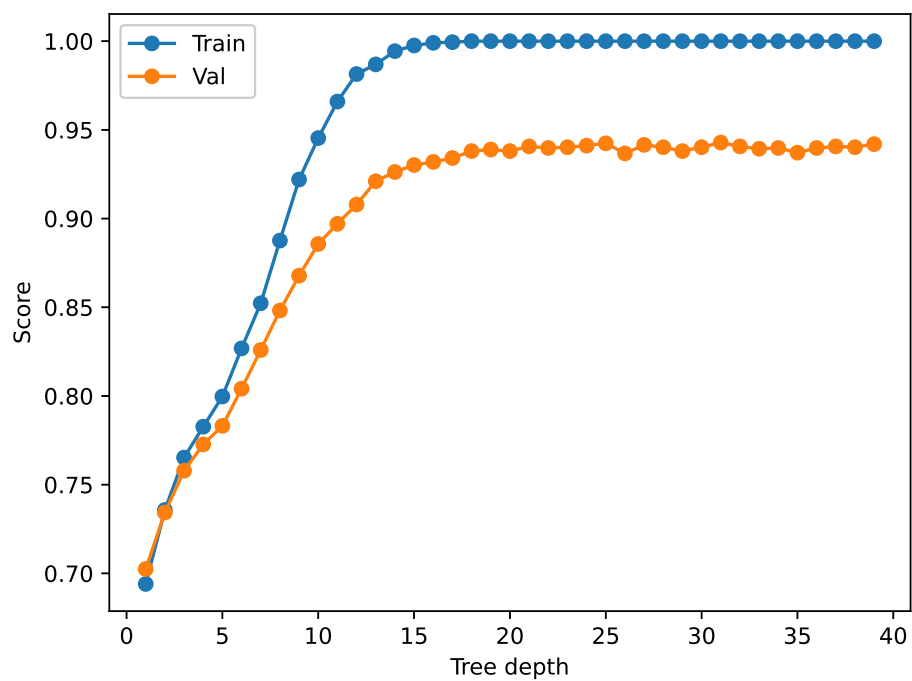


Figure 1: Score vs Tree-depth

References

Your references go here. Feel free to use any packages you need to streamline your referencing.