Execution Tutorial

To conduct the experiments, navigate to the 'benchmarks' folder. From there, initiate the experiment with the simulated dataset (p=6) using the file 'run_benchmarks_simulated_data_multi_model_p=6.R'. The experiments with real-world data will be performed in the 'real world data' subfolder, utilizing the files 'run_benchmark_mtcars.R', 'run_benchmark_mushrooms.R' or 'run_benchmark_banknote.R'.

The basic functionality of the file has not changed compared to the original repository. The experiments are now simply launched directly from the files in this repository, for the reasons described in the paper, instead of from a parent file. This necessitates defining the seed, 'n,' 'N,' 'p,' and 'share_unlabeled' directly in the file. Additionally, in this version, the parameters of the alpha cut method must be specified. In addition to 'alpha,' the constant variance (sigma_priori), and the set of expected values for all priors (mu_priori_lower, mu_priori_upper) must be specified.

Furthermore, the dataset is modified as in the original repository. Before initiating all other methods in parallel, the Alpha Cut method is now launched. The file called 'benchmark-alpha-cut.R' (located in the R/Alpha_cut folder) functions almost the same as the other 'benchmark' files. The only two differences are that the additional parameters described above are passed into the method, and additional steps to monitor the process are executed. Also, the 'alpha_cut' method in the 'alpha_cut' file functions similarly to the other methods located in the 'R' folder. The main difference between the other methods and the Alpha Cut method is that the values for the decision function for each of the available pseudo-labeled data points are calculated in parallel (serial execution is also possible with 'parallel' = FALSE). From here, the individual functions described in the bachelor thesis are then invoked.

In the 'results' folder, you will now find, in addition to the results of individual methods and settings in the 'Results' subfolder, the results grouped in folders by experiment. Summaries of the experiments are also available as Excel files in the 'Final' (model at the end) and 'Maximum' (model at the maximum) folders.

With the 'analyze_alpha.R' in the 'analyze' folder, the performance of different alphas can be compared with each other in an otherwise identical setting.