Machine Learning with R at LRZ: Introduction to batchtools

Exercise 1

Take the benchmark from the second `mlr` exercise sheet and run it on the cluster. The spam dataset has the OpenML `data.id` 44.

1) Run it on the *serial* cluster. Each slurm job should do a full 10-fold CV.

2) Run it on the *mpp2* cluster using 2-level parallelization with `parallelMap`. Each slurm job is still a 10-fold CV but is parallelized on a node using the multicore backend of `parallelMap`.

3) If you want to replicate the experiment with multiple different 10-fold crossvalidations, how do you need to change the code?

Exercise 2

Have a look at the *Best Practices* slides.

1) Take the demos (`demos/batchtools_mlr_classif.R`, `demos/batchtools_mlr_tuning.R`) and try to apply these tips to improve the code.