Exercise 1:

a) Take a look at the `spam` dataset from the package `ElemStatLearn`. Shortly describe what kind of classification problem this is and create a task for `mlr`.

b) Use a decision tree to predit spam. Try refitting with different samples. How stable are the trees?

   Hint: You can use `getLearnerModel(model)` and `rpart.plot()` from the package `rpart.plot`.

c) Use a random forest to fit the model and plot the oob-error against the number of trees used.

d) Your boss wants to know which variables have the biggest influence on the prediction quality. Explain your approach in words as well as code.

   Hint: use `mlr::getFeatureImportance` and/or `randomForest::varImpPlot`.

Exercise 2:

Visualize the decision boundaries of a random forest using the package `randomForest` on the `mlbench.spirals` dataset. Create plots in which you start with a small number of trees and increase them. Explain what you see. Use `mlr` for visualization.

Exercise 3:

a) Try to manually compute the first split point that the CART algorithm would do on the following dataset, once using x as feature and once using $\log(x)$ as feature:

x	1	2	7.0	10	20
y	1	1	0.5	10	11

and the log transform of it

log(x)	0	0.7	1.9	2.3	3
y	1	1.0	0.5	10.0	11

b) Implement your own CART algorithm dealing with the above problem with a few lines of code.

Exercise 4:

The fractions of the classes $k = 1, \ldots, g$ in node $\mathcal{N}$ of a decision tree are $p(1|\mathcal{N}), \ldots, p(g|\mathcal{N})$. Assume we replace the classification rule in node $\mathcal{N}$

$$\hat{k}|\mathcal{N} = \arg\max_k p(k|\mathcal{N})$$

with a randomizing rule, in which we draw the classes in one node from their estimated probabilities.

Derive an estimator for the misclassification rate in node $\mathcal{N}$. What do you (hopefully) recognize?

Exercise 5:

Show that the variance of the bagging prediction depends on the correlation between trees.
Hint: compute $\text{Var}(\frac{1}{B}\sum_{b=1}^{B} f_b)$ when $\text{Var}(f_b) = \sigma^2$ and $\text{Corr}(f_i, f_j) = \rho$, where f_b is a single tree of the ensemble.