Exercise 1:

a) Take a look at the `spam` dataset (`?mlr3::mlr_tasks_spam`). Shortly describe what kind of classification problem this is and access the corresponding task predefined in `mlr3`.

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

Hint: Use `rpart.plot()` from the package `rpart.plot` to vizualize the trees. (You can access the model of a learner by its class attribute `model`)

c) Use the random forest learner `classif.ranger` to fit the model and state the oob-error.

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 an adequate variable importance filter as described in `https://mlr3filters.mlr-org.com/#variable-importance-filters`.

Exercise 2:

Generate an artificial dataset with the function call `mlbench.spirals(n = 500, sd = 0.1)`. (The function `mlbench.spirals` is part of the `mlbench` package.) Visualize the decision boundaries of a random forest using the `classif.ranger` learner from `mlr3learners`. Create plots with `plot_learner_prediction` from `mlr3viz` for an increasing number of trees. (Start with `num.trees = 1`) Explain what you see.

Exercise 3:

Given are the dataset

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

and the same dataset, but with the feature x log-transformed

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

Either manually compute the first split point that the CART algorithm would find for each dataset or implement your own CART split-point-finding algorithm with a few lines of code.

Exercise 4:

The fractions of the classes $k = 1, \dots, g$ in node $\mathcal{N}$ of a decision tree are $\pi_1^{(\mathcal{N})}, \dots, \pi_g^{(\mathcal{N})}$. Assume we replace the classification rule in node $\mathcal{N}$

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

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

For this setting, we want to estimate the misclassification rate in node $\mathcal{N}$, for data distributed like the training data. Assume independent observations therefor. (*Hint*: Then the observations and the estimator using the randomizing rule follow the same distribution) The misclassification rate is the fraction of the data where the observations and the corresponding estimators do not coincide. Compute the expectation of this misclassification rate. What do you (hopefully) recognize?