Introduction to Machine Learning

https://introduction-to-machine-learning.netlify.app/

Exercise sheet 7 WiSe 2020/2021

Exercise 1:

Given are the dataset

Х	1	2	7.0	10	20
У	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
У	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 2:

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

Compute the expectation of the misclassification rate in node \mathcal{N} , for data distributed like the training data, assuming independent observations. What do you notice? (*Hint*: The observations and the predictions using the randomizing rule follow the same distribution.)