Assignment 1

# Feedback Decision trees, Naive Bayes by Reineke Peter Waldsam Franz and Wimmer Jonatan

Lecture

**Machine Learning Algorithms**

at

**Chair of Information Technology**

**Written by:**
Daniel Wagermaier

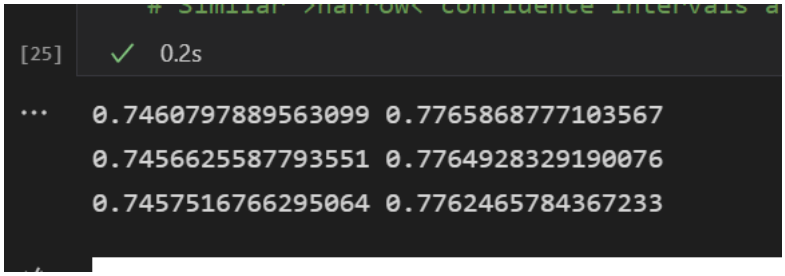
Leoben, 28.11.2022

*How clearly written and comprehensible is the submission in general?*

The submission, in general, is well-structured, comprehensive, and well-commented. However, numbered headings would have been nice to somehow keep track of where you are. Also, I would prefer to have written down beforehand what will be done in the following code block. Sometimes, this was done, and sometimes not. But since the code is very readable and filled with many comments, it is not a big deal. That is more of a question of who is supposed to read the notebook.

*How easy is it to understand the results described in the submission?*

Sometimes the output of a code block is not printed meaningfully. For example, the block after the heading "...confidence interval (95% choosen =standard)" prints out the upper and lower bound for the precession with no further information. If you are just skimming through the notebook to see the results, such outputs could be confusing.




Besides that, it is easy to understand the results of the submission.

*How convincing is the experimental setup?*

The setup seems to be convincing. Nevertheless, I would have decided to use more than 45% of the data for training. But since the evaluation of the created model will be done on 3000 randomly picked digits, this decision seems to make sense. Also, as experiments show, the Naïve Bayes classification is supposed to achieve better results with fewer training data.

*How do the authors justify their conclusion? Are you convinced? Which questions have not been addressed by the authors?*

To me, every question from the task description seems to be answered at least explicitly. However, it would be nice to have those answers printed out explicitly or written down in a markdown cell. Also, it would be nice if the formula for the upper and lower precision bound were written somewhere.

*Summarize the quality of the submission.*

The plots showing the label distribution have both the same heading. Therefore, it is hard to distinguish between the sets plotted there.

The split between markdown and comments in the code is a bit confusing. I think it would be better to write down what will be done beforehand. But since this is a Jupyter notebook, including the code, it is somehow hard to find the right amount of description. As code should be written "clean," meaning readable and self-explaining, it is hard to decide what should be explained and what not…