Guideline for the Python coding assessment and the technical discussion

Python coding (20%):

Your coding assessment will be based on the "Bank Marketing Data Set" available via http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#. The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit.

The Python coding is again a group work and the group stays the same throughout the module, e.g. it is the same group as you have had for Tomasz Orpiszewski's assignment. The deadline will be one week before the technical discussion. These will be organized once all exams are scheduled.

In your Python coding task, you are supposed find the best model to predict client subscriptions. The model must be chosen from the models discussed during the lecture; other models will be ignored for the grading. In your code you can include several models and model specifications, but the aim is that at the end you propose one model specification that should be used for predicting if a client will subscribe or not.

To select your model, please generate a classification matrix and use this for your decision. For your prediction matrix please start with a threshold of 0.5, that is predict that a client will subscribe if the predicted probability is 0.5 or higher, e.g.

$$\hat{Y}_i = \begin{cases} 1 \text{ if } \hat{P}_i \geq 0.5 \\ 0 \text{ otherwise} \end{cases}$$

Where $\hat{P}_i$ is the predicted probability for client i. As part of your model selection, you can also discuss this threshold and choose it differently.

If you need to find a tuning parameter you can split your data 50/20/30 into training, validation, and test sample. For models without tuning parameter split your data 70/30 into training and test sample. The 30% test sample should be the same in all cases, e.g. you only have one test set for all specifications.

Please submit your work as a Python Jupyter notebook to Ruben.Seiberlich@zhaw.ch.

If you have questions, you can submit them any time to me via: Ruben.Seiberlich@zhaw.ch.

Please copy all group members cc whenever you have a question concerning the coding project. I will either organize a call or answer the mail directly. Please note that you can approach me as often as you want, however, if the support needed is significantly higher than those of the other groups, I will reflect this in the grading. The same holds for groups that require significantly less support than the others.

Technical Discussion (30%): The technical discussion will have the form of an expert talk. It will be about the methodologies and results of your Coding assignment. The technical discussion will be together with the whole group.