MISCADA Core IIA (3) Classification

A. Student

20 March 2020

Part I: Executive summary

Write your executive summary here.

In this part you should describe the modelling problem being addressed with this data set, the results of your analysis and any performance metrics in terms that would be easy for a non-expert to understand. This means you will need to *avoid* reporting figures such as AUC, true positive rate, etc and should instead provide tangible performance numbers and easily interpretable plots: imagine you are handing this to your line manager in a future job.

An important aspect of the executive summary is reflecting on what the real-world objective may be in relation to the data set you are using and how your analysis addresses it. For example, if it is a medical dataset, is raw accuracy really the best objective to optimise? If not, what is the real objective and how have you addressed it? Feel free to be creative here if necessary to define an interesting question (eg set economic values on different ground truth/prediction outcomes).

Do not exceed 2 A4 pages after knitting to PDF for this section.

Part II: Technical summary

Write the technical details here.

In this part you should describe how you approached the modelling problem and summarise the performance metrics for how well your final model performs in full technical detail. The following list provides for some pointers to things you may want to think about addressing in this part:

- initial data exploration;
- details of any data coding or feature engineering;
- any train/test/validate, cross-validation or bootstrap strategies employed;
- strategies used to address any missingness;
- the approach taken to model fitting, including any model design, early stopping criteria, hyperparameter selection or tuning, and algorithm choices;
- insights into improvements achieved through different architectures (deep learning), data augmentation approaches, regularisation methods, etc;
- details on the performance of the model, including calibration;
- reporting of loss function choices, any post-model analysis such as tuning true/false positive rates, and justification of alternative objective functions;
- any efforts at interpretability of models;
- supporting plots for any of the above points.

Do *not* include any code snippets in the written report, since code must be submitted separately anyhow.

Do not exceed 4 A4 pages after knitting to PDF for this section.

Do not exceed 6 A4 pages after knitting to PDF for the report as a whole.