Notes for the Report

This document is intended to provide you information on the content of the report that you will prepare in form of slides.
Do not write a report like this one, you have to just write and present at the oral session the slides. See the lecture of the project (ML-xx-PRJ-...pdf)

Notes to you:

- File name: Use the **surnames** of the group components
- Language: English (spelling and clarity are essential)
- Use any format that you like for the slide (within the constraints explained in the ML-xx-PRJ-...pdf). Anyway, you will send also a PDF file!
- Use **page numbers** for each slide ← Mandatory
- Type of project : can be A/B/(C)
  - **A** (new NN simulator with your code, in that case specify the programming language) or specify **A with CM**
  - **B** (comparing existing simulators or by LLM, specify which ones)
  - C (rare and ad hoc cases, refer our agreement, email and date)

- Time/Pages/Font: see ML-xx-PRJ-...pdf. Use appendix for large or other tables or plots (if needed); the slides out of the limit will be not necessarily read and therefore not necessarily evaluated!!!
- See the instructions in the slides of the lecture ML-xx-PRJ-...pdf for the project description!

ABSTRACT

Few rows (up to 5) summarizing the developed project (used model for the cup, validation technique, novelties if any) ….

1. INTRODUCTION AND OBJECTIVES

- Describe/define *your aim and the essential background* (**in short**):
  - Define your aims: To explore/improve what? What are you going to show?
  - Describe (just define) the used models, the used algorithms etc. (e.g. A Neural Network trained by … algorithm)
  - Describe your assumptions (if any significant).

2. METHOD (BE SCHEMATIC)

Briefly (short part) describe *what you developed and how*:

- The code (for type A implementation) or the used simulator(s) (for type B).
    - The used tools/libraries (if any)
    - Software overview and the software design choices (if interesting)
    - Implementation choices (a summary): e.g. architecture/s (and numb. of layers), type of activation function/s, type of training algorithm/s, batch/on-line/mb, initialization schema/s, regularization schema/s, stop condition/s. See also the *check list* in the notes of Section 3.2.
    - The novelties (if any) but not the standard approaches (*do not* describe the algorithms/models we already described in the lectures). Use references for the source of information. In case, it is important to stress what you have been investigated beyond ordinary techniques.
- Preprocessing procedure (if any) [details may be postponed, see Section 3]
- Validation schema (model selection and evaluation schema) for the Experimental part: report data splitting TR/VL/TS (% data for each set and/or the K values of the K-fold CV) [may be postponed see Section 3 and slides demo]
- Type of preliminary trials pursued (often summarized by text) [may be postponed see Section 3 and slides demo]

Each figure/table should be referenced as in the following, see Fig. 1. (*not need for slides*)
Do not use figure/table without a number. Do not write "see the next figure" (which one?).
Tables and plots have always a caption. All of the Figures and Tables should be cited in order, including those in the Appendix (which should be cited as, for example, Fig. A.1, and Table A.1).

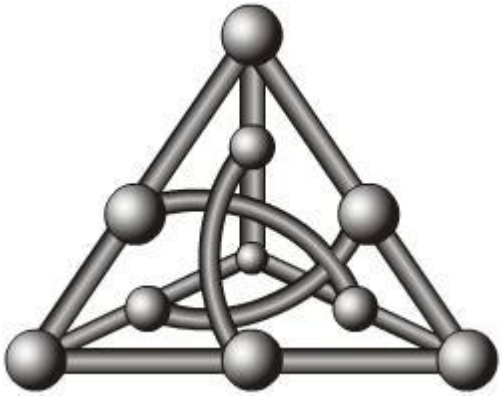

Figure 1. Do ***not*** insert a picture of your neural network (unless it is innovative)

3. EXPERIMENTS

What you have found: The results should be the larger part of the report.

3.1 MONK'S RESULTS (SCHEMATIC!)

In this subsection report only plots and tables, comment only if needed and very short.
[*] Lack of results in this part invalidates the report. Both for A and B prj (and for all the considered models of prj B).
Please report:

- Performance (see Table 1) and learning curves (MSE and accuracy plots for the 3 MONK's tasks, see Fig. 2 as an example filled with the plot for the MONK2).
- Used hyper-parameters.

Table 1. Average prediction results obtained for the MONK's tasks.

Task	**#Units, eta, lambda, ..**	**MSE (TR/TS)**	**Accuracy (TR/TS) (%)**[i]
MONK 1		…	…
MONK 2		…	100%/100%
MONK3		…	…
MONK3 (reg.)		…	…

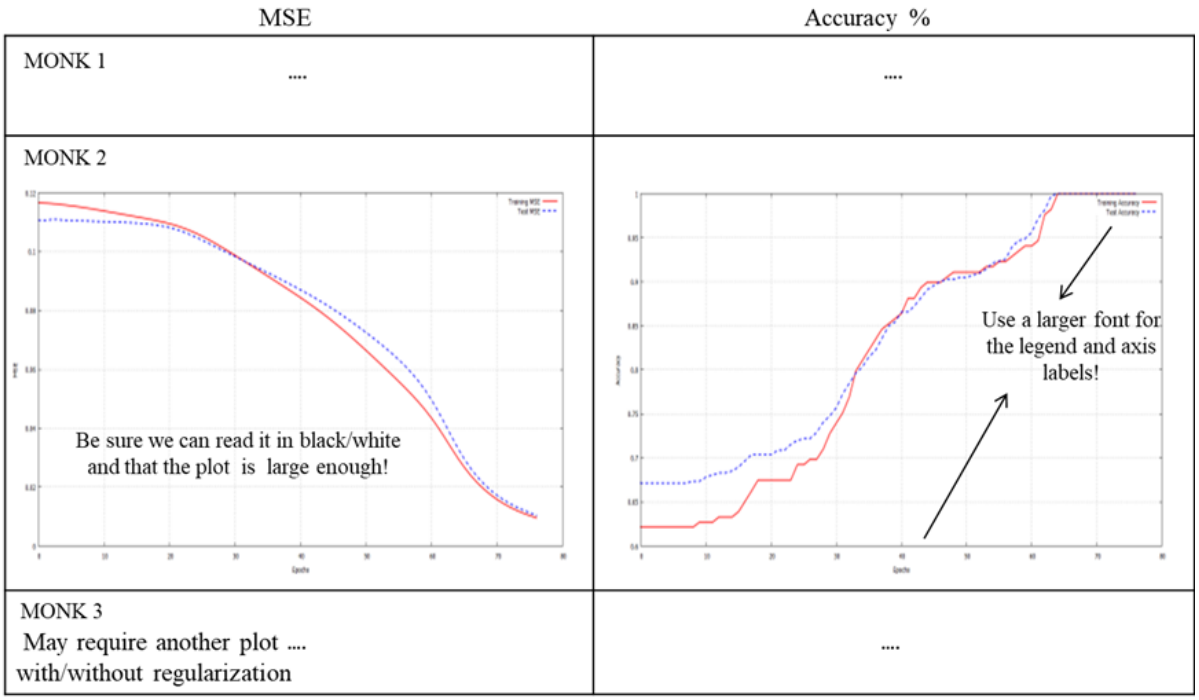

Figure 2. Plot[ii] of the MSE and accuracy for the 3 MONK's benchmarks.
[Be sure it will be easy to read them in Black&White and use a larger font for the legends and axis labels]

3.2 CUP RESULTS (DETAILED!)

***Always include**:*

1. Details of the Validation schema (model selection and assessment schema): Splitting TR/VL/(internal)TS (% of data for each set and/or the K values of the K-fold CV) [*]
2. Screening phase: Type of preliminary trials pursued (often summarized by text)
3. Grid search/model selection: Schema and range of explored hyper-parameters (values used for the grid search, possibly table/tables) [*]: see notes below

4. Grid search results: TABLES of results TR/VL/(TS)[iii] with MEE [*]. At least the most performant cases. Tables/Plots can be used also to show some relevant trends for specific hyper-parameters changes (if you think it is significant)

5. Provide an estimation of the (training) computing time (and of your HW resources)

6. Repeat and compare if you are comparing different models/algorithms/approaches (or different variants of the approach that you think are significant).

7. ***Define how you selected the* FINAL model** used on the blind test set [*]. Which is it among the candidates and why? Also write the hyper-param. of the final model [*].

8. ***Report*** for the FINAL model used on the blind test set the TABLE with MEE for TR (training), VL (validation) and TS ***(internal TS)***[iv] in the original scale [*][*]. Note again that you must have an internal test evaluation (see the note [iv] above)

9. ***Plot*** the learning curve TR/(VL)[v]/TS for the FINAL model [*] (the final model is unique)

10. **Discussion** (what you found more significant in your experiments, it is a relevant part!)

[*] = Lack of results in this part invalidates the report. Both for A and B prj.

Further explanations and notes:

A good experimental report contains a list of observed facts *supported by experimental evidence:*

- Report of used setting (***replicability***), and of the procedure for model selection and validation/assessment. The synthesis of the hyperparameters setting must be reported on the slides, while a complete description can be inserted in a document within the code package (call it "hyperparameters-setting") and in the Appendix slides (selecting what is more relevant). See the *check list* note below on "Model selection"
- Report accurate results in numerical and graphical form (clear tables/plots with measures of training, validation and test errors)
- Make a selection of the evidences that you think are important but for those show with graphs / tables etc, e.g. by plots varying the hyperparameters values
- **Make critical remarks** (+/-) on the effect of your choices (and, if necessary, their agreement with the theory of ML or report and discuss unusual evidences)
- Note that the relative position between the different cases is interesting (not the absolute performance)

- Learning curve to show the TR, Val/Test error with the progress of training: Curve with LMS (loss used for training) and measure of final error curve (e.g. accuracy for classification) can be different.
- ***Check list*** of the necessary features for the model implementation and the CUP application, info to be reported in the "Your contributions" slides:
  - Models/Architecture/s (and number of layers): which architectures/models have you implemented/used and compared? And for NN, how many layers?
  - Type of activation function/s (for NN): which activation function/s have you implemented and compared? E.g. Sigmoid, TanH, ReLU, Leaky-ReLU, …)
  - Type of training algorithm/s: which training algorithm/s have implemented/used and compared?
  - Batch/on-line/mb, initialization schema/s: which did you implement /use?
  - Initialization of NN: which approaches have been used? Were multiple trials performed with different initial random weights during the model selection phase? Does the results report the standard deviation for different initialization?
  - Regularization technique/s: which did you implement /use? Rember that L2 (Tikhonov) is a must, other schema are optional
  - Stop condition/s: which stop condition have been implemented and used? E.g. Fixed number of 1000 epochs, number of epochs selected in the grid search or other criteria, Early stopping, % of training error decreases, … etc. But see the slides of the course for the discussion of the best approaches!
- ***Check list*** of info to be reported in the "Model selection" slide (further note):
  - Grid search: did you use an unique or a multi/nested grid search? How many hyperparameters configurations have been considered? Possibly use table/tables. The synthesis of the hyperparameters setting must be reported on the slides, while a complete description can be inserted in a document within the code package (call it "hyperparameters-setting") and in the Appendix slides (selecting what is more relevant).

- ***Check list*** for the results part (further note):
  - **See the all the [*] points above**
  - How the final model have been built or selected among the competitors? Using which part of the data or validation schema?
  - Initialization of NN: Finally, important, what is the criteria to set up the last model (used for the blind test results) among different possible initializations?
  - **Discussion**: have you discussed the effects of your proposals and, for instance, the most significant findings and/or hyperparameters effects? (see the slide of the PRJ lecture and the slide demo for more details)

4. CONCLUSIONS

What you have drawn and what you learned and also:

BLIND TEST RESULTS: name of the result files and your nickname

ACKNOWLEDGEMENTS

If any.

We agree to the disclosure and publication of my name, and of the results with preliminary and final ranking.

REFERENCES

From where you are taking information, where a reader can find details, credits, sources
For the style see any paper bibliography to take examples. The items here should be numbered and the number should be used in the text. Double check the bibliography!!!

In particular, always include (with a uniform style):

Authors, Title, Journal/Proceedings/Editor, Volume, Pages, Year (URL if needed)

E.g. see the following references:

[1] A. Micheli, A. Sperduti, A. Starita, A.M. Bianucci: Analysis of the Internal Representations Developed by Neural Networks for Structures Applied to Quantitative Structure–Activity Relationship Studies of Benzodiazepines. *J. Chem. Inf. Comput. Sci.*, *41*, 202–218, **2001**.

[2] A. Micheli, A. Sperduti, A. Starita, A.M. Bianucci: Design of New Biologically Active Molecules by Recursive Neural Networks, IEEE *Proc. Int. Joint Conference on Neural Networks*, *4*, 2732–2737, **2001**.

APPENDIX. A.

If needed, insert here (only) TABLES, GRAPHS, **PLOTS** that are **not** essential in the main sections, or that are too large.
Take care: The report can be read without reading the appendix.
These parts may be out of the maximum number of slides, without a specific limit (but it should be reasonable!).

---

[i] Please see FAQs in the PRJ lecture: report the mean result over the different initialization, not just the best result. Moreover, remember that the MSE does not include the penalty term.

[ii] General note on plots in the report: to be comparable, use a point for each epoch (defined as the total number of training data): if you are using on-line or mini-batch provide the mean over an epoch for each point in the plot.

[iii] The internal test set (see the note IV) results can be or can be not present during the grid search according to your decisions on the data splitting/cross-validation approach. Hence, I'm using the notation (TS) with brackets therein.

[iv] The "***internal test set***" is a set of data drawn by yourself from the provided data for the model development (according to your TR/VL/TS data splitting/cross-validation approach). It is needed to use an internal test set to compare your results with the blind test results for our didactics aims. The test set results can be or can be not obtained in this phase, depending on the cross-validation schema that you use. In any case report the value also in this table as a summary.

[v] The validation set plot can be available or not for the final training depending on your usage of data. Hence, I'm using the notation (VL) with brackets therein. However, the results for VL must be provided in the tables for the final model according to the results obtained by the model selection phase. See point 8!

Note: please **do not** use footnotes in your report!