Title

Authors (*Sergio Latrofa, Federico Matteoni, Riccardo Parente*).
Master Degree in Artificial Intelligence.
s.latrofa@studenti.unipi.it, f.matteoni4@studenti.unipi.it, r.parente@studenti.unipi.it
Machine Learning course (654AA), Academic Year: 2021/22
Date: xx/yy/2022
Type of project: **A with ML**

Notes to you:

- Language: English (Italian as a special case, spelling and clarity are essential)
- Use word if you like, but latex is better. Anyway, send a PDF file
- **Font 11** at least! (here font 12 for clarity) ← Mandatory (else we will reject)
- 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, specify which ones)
  - **C** (ad hoc prj, refer our agreement email and date)

- Pages: up to **8** [2 students case], up to **10** pages for 3 students project [standard case]: use appendix for large or other tables or plots (if needed); the pages out of the limit will be not read nor evaluated!!! ← Important
- See instructions in the slides of the lecture for the project description!

Abstract

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

1 Introduction

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.
- 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.
  ○ 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.

- Preprocessing procedure (if any) [details may be postponed to 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) [details may be postponed to Section 3]
- Type of preliminary trials pursued (often summarized by text) [details may be postponed to Section 3]

Each figure/table should be referenced as in the following, see Fig 1. 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).
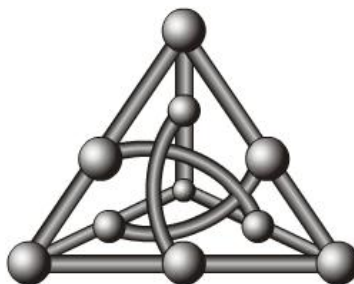

Figure 1: Do *not* insert a picture of your network (unless innovative).

3 Experiments

What you have found: The results should be the <u>larger part</u> of the report.

3.1 Monk 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.
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

Task	#Units, eta, lambda, ..	MSE (TR/TS)	Accuracy (TR/TS)(%)[i]
MONK1			
MONK2			
MONK3			
MONK3+reg.			

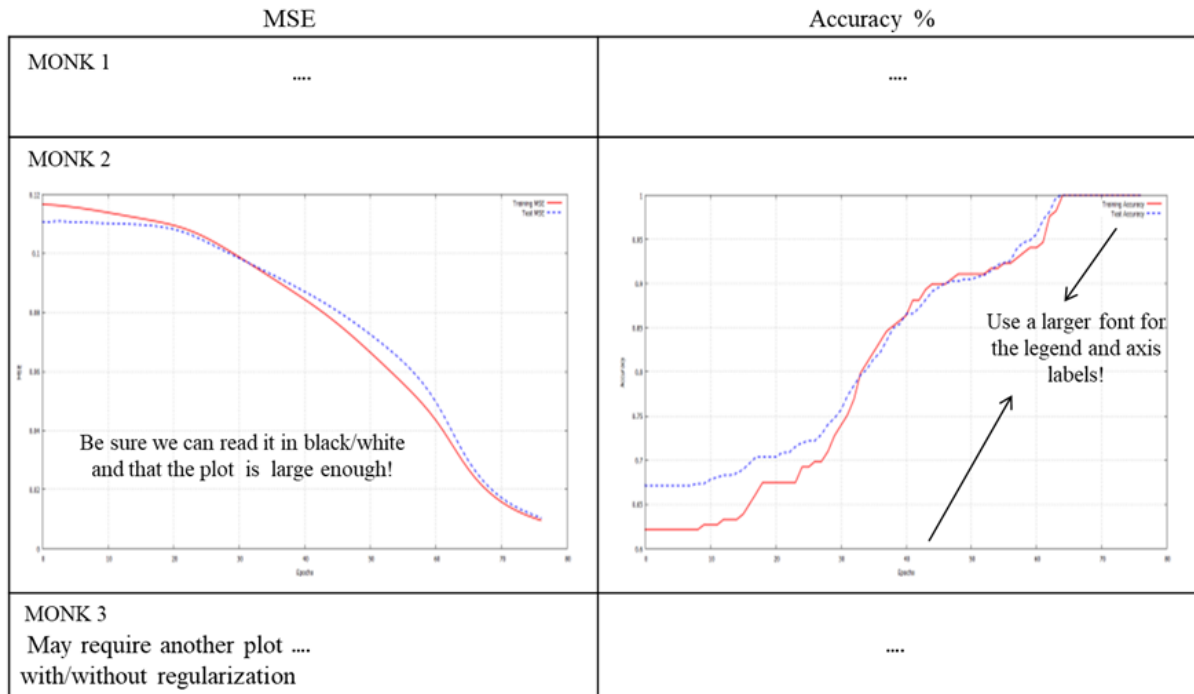Table 1: Average prediction results obtained for the MONK's tasks.

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 (after print)]

3.2 Cup Results (Detailed!)

Always include:

1. <u>Details of the Validation schema</u> (model selection and assessment schema): Splitting TR/VL/(internal)TS (% 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. Schema and <u>range</u> of explored hyper-parameters (values used for the grid search, possibly a table) [*]
4. Grid search: <u>TABLES of results</u> 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*** <u>FINAL model</u> used on the blind test set [*]. <u>Which is it among the candidat</u> Also write the hyper-param. of the final model [*].
8. ***Report*** for the <u>FINAL model</u> used on the blind test set the <u>TABLE with MEE</u> for <u>TR (training)</u>, <u>VL (validation)</u> and <u>TS (internal TS)</u>**iv** <u>in the original scale</u> [*][*]. Note again that you must have an internal test evaluation (see the note IV above)
9. Plot the <u>learning curve</u> TR/(VL)**v**/TS for the FINAL model [*] (the final model is unique)
10. Discussion (what you found more relevant in your experiments)

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

Further explanations and notes:

A good experimental report lists observed facts *supported by experimental evidence*:

○ Report of used setting (***replicability***), and of the procedure for model selection and validation/assessment
○ 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

4 Conclusion

What you have drawn and also:

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

Acknowledgments

If any.

References

From where you are taking information, where a reader can find details, credits, sources . . . 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.:

Appendix A

If needed, insert here (only) <u>TABLES, GRAPHS, **PLOTS**</u> 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 8/10 pages, without a specific limit (but should be reasonable!).

---

[i] Please see FAQs in the PRJ lecture: report the <u>mean</u> 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 <u>a point for</u> <u>each epoch</u> (defined as the total number of training data): if you are using on-line or <u>mini-batch</u> 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!