▾ General Remarks

- **Collaboration**: you can work in teams of up to 3 people. Organisation in groups is welcome and appreciated, but the workload must be appropriate to the number of participants.
- **Honor code**: you can consult any papers, references, or available implementations for ideas and code, but you must clearly cite your sources in your code and in your report. The information leakage between groups is not allowed: originality of the work will be taken into account in the project assessment.

An idea of tasks to be carried out depending on the number of members in the group:

Image classification:

- 1-2 members: develop a CNN from scratch, a pre-trained CNN and compare the results
- 3 members: develop a CNN from scratch, two pre-trained CNN and compare the results

Time-series forecasting:

- 1-2 members: develop a RNN, a LSTM and compare the results
- 3 members: develop a RNN, a LSTM, implement a multi-step forecasting and compare the results

Project Goals

The general goals of the final project are the following:

- Identify a subject of your interest (make sure that the problem can be addressed in Colab) and motivate the relevance of Deep Learning techniques to address it.
- Frame the work w.r.t. the literature related to the problem and/or w.r.t. the inspiring implementations;

- Solve classification/regression/forecasting problems with pretrained and/or trained from-scratch Neural Networks. The project must deal with the topics covered during the lectures (e.g., use of CNN architectures in Computer Vision and/or use of CNN or RNN architectures for Sequence Analysis). In addition (and not as replacement), you can also explore topics not covered / partially covered in class (e.g. Generative Adversarial Networks, Image Captioning, Attention Mechanism, Time Series Forecasting ...).
- Demonstrate mastery of deep learning basic concepts (model selection, evaluation of underfitting / overfitting in the considered case study, application of regularization techniques).
- Adequately report and discuss the results obtained.
- Possibly,

    - evaluate/discuss the consequences of resorting to ensemble solutions, highlighting the accuracy gap between the base classifiers and the composite classifier.
    - evaluate/discuss aspects related to hyperparameter optimization.
    - evaluate/discuss aspects related to explainability.

Furthermore, it will be particularly appreciated:

- Creativity.
- Usage of algorithms, techniques and architectures that were not presented during the course.
- Adoption of techniques (e.g. architectures, preprocessing techniques, procedures) inspired by the literature. The inspiring works must be properly referenced.

Step 1: Project proposal

- Prepare a brief presentation (max 1 page), aimed at highlighting the following aspects:

    - Group members and MSc Degree;
    - Brief description of the problem;
    - Properties (or expected properties) of the dataset (ground-truth, metadata, dataset size ...). Describe the procedure of collection if you made, or plan to make, it.
    - Expected task(s) to be performed (e.g. image/text classification) and, possibly, describe the analysis you are proposing.

- Submit your proposal via e-mail:

    - approximately 4 weeks before the exam session which you intend to take (deadline could be less strict for the first session)
    - to:

        - beatrice.lazzerini@unipi.it
        - michele.baldassini@phd.unipi.it
        - put your group mates in CC if it is a group project.

    - mail subject: `"[CIDL_2022_2023] Project Proposal <SURNAME> [<SURNAME2> [<SURNAME3>]]"`

- Wait for project approval. The following aspects will be evaluated:

    - Relevance to the program of the course.
    - Level of detail and complexity of the proposed analysis

▾ Step 2: Project submission

Unless otherwise stated:

- The project must be submitted by sharing the Drive folder ("Visualization" mode) with the following accounts

    - beatrice.lazzerini@unipi.it
    - michele.baldassini@phd.unipi.it

- The project must be submitted few days (typically around 5) before the date of the oral exam.

The shared folder must be organized as follows:

```
|__Surname[_Surname2[_Surname3]]
    |____ notebook1.ipynb
    |____ notebook2.ipynb
    |____ ... (other .ipynb files)
```

```
|_____  Report.pdf
|_____  models
        |_____  store your models checkpoints here
```

In other words, you must include

- all the notebook files (ipynb) adequately described and commented. The cells output must reflect the results collected in the report.

- Network checkpoint models, stored in a dedicated folder.

- **Report.pdf**, including the following sections:

- **Introduction**. Describe the problem and the motivations of your work

- **Related works**. Discuss the state-of-the-art and the relationship of your project with existing works / implementations (if any).

- **Methods and experiments**. A detailed description of the proposed solutions, highlighting the motivation behind the performed choices, with commented figures and plots.

  - E.g.:

    - *simple model -> accuracy curve obtained on a validation set allows to observe underfitting*
    - *more complex model -> accuracy curve obtained on a validation set allows to observe overfitting*
    - *application of these regularization techniques: ...*
    - *performed these hyperparameters search: ...*
    - *hyperparameters chosen according to ...*

      - **Conclusion**. Summarize your results and the lessons learned. Possibly highlight weak aspects that would deserve further investigation. Assess whether the project has met the expectations discussed in the preliminary proposal phase and possibly highlight any criticalities.

This structure is to be considered indicative, you can use it as a starting point and develop it as you wish.

The report can be written either in English or in Italian.

Step 3: Project discussion

Unless otherwise stated:

- A part of the oral examination will be dedicated to the discussion of the project.
- All students in the group must participate in the project discussion. Example:
  - Let {*A*, *B*, *C*} be a group of students
  - *A* wants to take the exam during the first session, whereas *B* and *C* during the second session.
  - *A*, *B*, *C* will discuss the project during the first session; only project-related aspects will be discussed.

Ideas and examples

- **case studies** and **datasets**
  - Tensorflow datasets: [link](#), [list](#)
  - Kaggle datasets: [link](#)
  - your own data.
- **examples of past years students' projects** are available in the "FinalProject" directory. Please notice that:
  - we are sharing the projects just to give a coarse indication on the type of expected analysis and workload and on the overall organization.
  - **the documents represent student reports; as such, they should not be considered as peer-reviewed scientific publications**.
  - you are not required to analyze the same case study; you are not required to perform the same analysis.