Project Instructions (50.039 Deep Learning, Y2025)

Matthieu De Mari

1. Introduction

In this project, you may work on a topic of your own choice (assuming it fulfils a minimal list of requirements detailed below). You will have to design your custom model, and possibly create your own mock dataset (or better, look for one online). We advise to have a look at the typical dataset repository websites such as Kaggle et Google Data Search, as to not waste too much time on creating or processing a dataset (this is not the purpose of this project!).

Please assume that this project is to be submitted to a serious client (played by us professors) and take this project as an opportunity to work on your project delivery skills. More specifically, pay attention to the way you code and document your code.

Your submission will have to be uploaded for grading. Please only submit your code and notebooks on eDimension, not the dataset! (Otherwise, the submission will be too large). We also advise you to upload your final project on Github, to enrich your personal portfolio of projects. Shall the dataset be required to run your project, let us know where to download it (Kaggle or even better, upload it to our own Google Drive/Dropbox).

Finally, we expect you to present the project during one of the sessions of Week 13, time to be confirmed closer to the date.

2. TLDR

Groups of 2-3 people, will not allow people to work alone.

Free choice on project proposal, only requirement is that it should use numerical time series data.

Custom project proposal deadline and groups formation before February 23rd 2025, 11.59pm.

Use link for group formation: https://docs.google.com/spreadsheets/d/1EQ-85BTAK0Tn8yxeVHNp4dFdgLYBvauh7ku3x3VWlZA/edit?usp=sharing

Submission deadline for all project deliverables: April 19th, 11.59pm.

3. Objectives and guidelines for project proposal

You will need to submit the project proposal via email to me (matthieu_demari@sutd.edu.sg) by 23rd February 2025, 11.59pm, for approval.

The proposal should be a small PDF file, containing a brief description of the following elements:

- Topic, problem to be investigated,
- Expected inputs and outputs, dataset to be used,
- Architecture draft (briefly hinting at the type of architecture you plan to use, does not have to be extremely descriptive, and more importantly, it might change over the course of the project),
- Team members,
- What you are going to deliver,
- Final deliverable must be at least: a PDF report, code for training model from scratch, code for recreating the trained model from a file (ideally, the same one you used to show performance curves and other metrics in your report).

Project requirements: The only constraint we have decided to give for this project is the following:

The Deep Learning model should use some numerical time series data.

For instance, you could design a model that:

- receives historical data about the weather and attempts to predict the next type of weather (sunny, rainy, etc.)
- receives heart beats sound and attempts to classify if the patient has tachycardia,
- receives some numerical data about the stock market and attempts to do something with it (predict when to buy/sell, next values in the stock market graph, etc.)
- Etc.

At the moment, some of these concepts might feel a bit obscure, but you will see in the next coming weeks, how to use Neural Networks to process such data types, starting with Week 6 and Week 8 concepts. A quick note that a string of text is not considered a numerical time series data type.

Apart from these constraints, we are rather open to crazy topics!

4. List of expectations for projects

As mentioned before, you are expected to deliver a project, as if us professors were professional clients. As such, good practices will be critical, for instance:

- Train your model using a training set but evaluate your performance after training on a test set. Even better, use a train-test-validation split.
- A bit of hyperparameters tuning and searching never hurts.
- Give a quick comparison of your model performance against some state-of-the-art ones. You are not expected to beat the state-of-the-art models!
- You are expected to describe the different steps you have taken to create your architecture from scratch. A brief explanation should be given on how to recreate the model and train it from scratch.
- We also expect you to save your model to a file for reproducibility.
- Visualization of your model performance (e.g. accuracy and loss curves, performance of your system on some validation set images, etc.) are also expected. For each figure used in your report, there should be a clear description on how to recreate said figure. Being unable to do so, means that there is no way for us to confirm the results you are presenting!
- Please show examples of your model malfunctioning if any and discuss what might be the reason for such problems.
- Your report should contain everything we need to know to run your code (including the package dependencies).
- Put in your submission the group members and their contribution to the project.
- Your model should use PyTorch as a framework. Using any other framework (Keras, Tensorflow, etc.) will result in penalties being applied to your project score.

5. Project Delivery

Delivery details

Groups of 2-3 people, will not allow people to work alone.

Can form groups with members from both Class A (evening times and Thursdays) and B (morning times and Fridays). Keep in mind, however, that the presentations are meant to happen on the Week 13 time slots, and that it might lead to clashes. **It will not be your professor's responsibility to resolve these clashes, if any.**

Free choice on project proposal, only requirement is that it should use a numerical time series dataset.

Custom project proposal deadline and groups formation before February 23rd, 11.59pm.

Use link for group formation: https://docs.google.com/spreadsheets/d/1EQ-85BTAK0Tn8yxeVHNp4dFdgLYBvauh7ku3x3VWlZA/edit?usp=sharing

Submission deadline: April 19th, 11.59pm.

Recapitulative report

Your recapitulative report shall be submitted in a PDF format, along with your code.

You code may consists of .py files, .ipynb jupyter notebooks, or Google Colab notebooks. Your PDF report should explicitly mention what needs to be done to run your code (imports needed, where to find dataset, commands to run if any, etc.)

Properly documenting your notebooks/code files would be much appreciated, and you should practice it anyway (they are good practice!). Personally (Matt), I believe a presentable format is:

- some Jupyter Notebooks, combining MarkDown cells and code cells,
- along with .py files containing the largest parts of your code (you just have to import them later in your Jupyter Notebooks, to minimize the amount of code in your Notebook!)

But ultimately, the choice is yours!

Project deliverables

We strongly advise to upload your submission (code/notebooks + PDFs, but no dataset due to space restrictions) on a Github repository. You can then submit the link to your PUBLIC Github repository, during your submission on edimension.

Your Github repository for this project should contain your PDF report, your DOCUMENTED code/notebook files. It should also contain directions showing the required libraries and steps needed to re-train the model from scratch. And more importantly, it should also contain clear

directions on how to recreate the exact trained PyTorch model and its performance results you are presenting in the PDF, by loading some save PyTorch weights for your model.

This is essential, for reproducibility reasons and something that we, researchers, must do constantly for our papers. If the weights files are too heavy to be submitted on edimension or Github, you may upload them to a Google Drive/Dropbox instead.

Important note: you are nearing the end of your curriculum at SUTD, gathering your projects, and uploading them on your personal Github is much appreciated by recruiters as it allows them to immediately identify what your coding capabilities are. If you have not done so yet, please consider starting your project portfolio on Github!

Project presentation

Finally, we expect you to present the project during one of the sessions of Week 13. A small demo, along with some slides or a small video would be appreciated. Presentation times and details to be confirmed closer to the date.

Grading rubric

Project originality, quality of the report, quality of the code and presentation will be the main components of the final grade given to this project. More details will follow.