Machine Learning Practical
Group Project Guide 2022/2023

This document contains guidelines and ideas for designing a semester 2 project for the MLP course. The listed ideas are non-exhaustive, and in no way do they represent the space of possible projects. They are merely a tool provided to you, so you can create your own project, by simply combining the provided tips and ideas. You can see this project as an opportunity to apply what you learned in the class to a problem of your interest.

Choosing your group

Semester two will be based on group projects.

- Each group should consist 2-3 students. Note that it is not allowed to have more than 3 people and not recommended to do the project alone. We are expecting projects to have the amount of work from a 2-3 person group while marking. Also interacting with your team is an important experience.

- You can discuss any aspects of the assignment with your group and divide up the tasks any way you like. Note that best projects happens when the team collaborates on each part.

- You can use Piazza "Search for Teammates" to help form a group – but make sure you use the MLP Piazza for Semester 2 (piazza.com/ed.ac.uk/spring2023/infr11132).

- You can use the MLP Student Group 2022-23 team on MS Teams to help form a group (in any existing or new channel that you would like to set up). Recommended that you use the existing formatting for such groups: "[Looking for Group] - Insert Group Theme/Topic".

- We have set up an Office 365 spreadsheet (accessible to University accounts) here (you need to be logged in with your University account). This has 3 sheets:
  - "MLP Group Setup" is for specifying your project group (deadline is Wednesday, 18 January 2023 at 12.00). If you cannot find any group and do not fill this in, your group will be randomly allocated. This sheet will also act as a reference for your tutors when tutorials start (which is why we include columns for topic, keywords, and online document).
  - "Group Formation" is for indicating interest in working in a specific area. Use this to find colleagues to discuss and potentially form groups with. Simply write your name and UUN under the column representing your area(s) of interest. These are also the lists that will be used for random allocation (so if you are listed in multiple columns, consider removing all but one as the deadline approaches). Remember that you can find your colleagues by UUN on the Teams app.
  - "Propose New Topic Area" is for proposing new columns for the "Group Formation" sheet. As long as 5 students 'sign-off' on the title, it will be added (discretion is advised).

- Note that you will self-select the tutorial hours and will not be automatically assigned to tutorials. Once the groups are formed, we will share an additional sheet for this purpose.

Choosing your project

Your first task is to pick a project topic. If you don't have an interesting project idea yet where you can apply deep learning techniques, you might want to have a look at the publications from previous years and example projects from other courses. Here are some examples:

- Machine learning conferences: ICLR, NIPS, ICML.

- Computer vision conferences: CVPR, ICCV.

- NLP conferences: ACL, EMNLP

- Speech conferences: ICASSP, Interspeech,

- Challenges: Kaggle,

- Projects from other courses: Stanford CS229, CS231.

To get a better feeling for what we expect from MLP group projects, we encourage to take a look at the example project reports from previous years:

- G008 - Audio Super-Resolution with Generative Adversarial Networks

- G045 - Image Deconvolution with Deep Image Prior

- G056 - Does adversarial training generalize across adversaries?

- G060 - Generative Audio Inpainting in the Frequency Domain

- G065 - Forging the Perfect Recognition

- G085 - 3D Shape Reconstruction from Two View Images

- G109 - Reinforcement Learning with Neuroevolution

Most students do one of three kinds of project.

Application project.

This is by far the most common type. It involves applying an existing method to an interesting application. Ideally the existing method has not already been tested on the application by prior work.

1. Choose a dataset with an interesting task in mind (e.g. graffiti classification, predicting Eurovision contest winner based on audio, wine quality prediction based on pH and sulphates, polymer type classification, predicting chemical reactions),

2. Identify a state-of-the-art method for your task (see the conference links above) and apply it to the dataset, evaluate its performance,

3. This method is now your baseline. Investigate its architecture/setup and aim to improve its performance in terms of an appropriate metric (e.g. classification accuracy, BLEU score, convergence speed, memory consumption). This can be done via exploration of architectures, data augmentation strategies, loss functions, optimization techniques and/or reqularization methods.

Here are some example abstracts from the previous year's project reports that fall into this group:

- Speech/audio project example: Audio super-resolution with Generative Adversarial Networks.

- Computer vision project example: 3D Shape Reconstruction from Two View Images.

- NLP project example: Forging the Perfect Recognition.

Analysis project.

It involves analyzing certain aspects of a family of methods which has not been explored before. This requires critical understanding of existing methods and usually a rigorous experimentation to analyse the influence of several factors in terms of performance.

1. Choose an interesting method/learning paradigm,

2. Read related papers from recent conferences (tip: once a paper is picked, use google scholar to find related papers; other recommended alternatives are Semantic Scholar, Papers with Code, and Connected Papers),

3. Identify the strengths/weaknesses of the existing methods,

4. See whether the existing methods have been compared to each other on a fair basis (using similar experimental procedure, similar network architecture, hyperparameters are well-tuned for all the methods), whether the comparison still holds when applied to different datasets (indoor/outdoor images, English/French documents) or when the data is varied (e.g. low/high resolution images) or when a more meaningful evaluation metric is chosen (BLEU/CIDEr).

5. Choose the appropriate research question,

6. Compare the methods in a rigorous experimental setting and analyze them.

Analysis project example: Does adversarial training generalize across adversaries?

Algorithm project.

It involves developing a new learning method or a novel extension to an existing method and experimentally demonstrating that the proposed method is superior to the existing method. This is typically *significantly harder* than the other project types and requires good knowledge of the field and critical understanding of existing methods.

1. Choose an interesting method/learning paradigm,

2. Read related papers from recent conferences (tip: once a paper is picked, use google scholar to find related papers),

3. Identify the strengths/weaknesses of the existing methods and try to propose a new idea that can address the weakness of a method while maintaining its strengths.

4. Evaluate its performance and compare to the existing methods.

[Algorithm project example: Image deconvolution with deep image prior](#)

Note that some projects can also combine elements of different types.

Tutorials

Next semester we will have weekly lectures on advanced topics in deep learning by the course instructors and also by the industry people. Differently from the first semester, we will run weekly tutorials where you can discuss the progress of your project with a tutor. Each project group will be assigned to a tutor, who will discuss and review progress. We strongly encourage each project group to set up an Office 365 Doc for the group (shared with the tutors and instructors). This will be used as a communication channel between your group and tutor to report progress and experimental results, give plans, raise questions.

We understand that you might find it hard to choose a project and overwhelmed by the possibility of too many interesting topics or deciding whether the topic is good enough or too ambitious for the MLP project. We strongly encourage you to discuss about designing or evaluating the difficulty/feasibility of a project with your appointed tutor.

In addition, we will also use a new Piazza for next term (`www.piazza.com/ed.ac.uk/spring2023/infr11132`). Please keep using Piazza to ask and answer questions, search for teammates, etc. Note that there will be no scheduled labs in next semester.

Prize event

IBM UK kindly sponsored a prize for the best project in MLP as in the previous years. The selection process for the best projects will be done as follows. Once the final reports are marked, a committee including the instructors, TAs, 2 external academics from Informatics will shortlist the best 6-7 projects and select a winner. We'll announce the short list of projects for the IBM MLP Prize and have a prize giving event in late April. At the event each of the short-listed groups will be invited to give a short talk about their work and IBM UK will present the award (and also tell us why IBM are interested in this).

You can check last years winning submissions on the MLP Learn page, under [IBM](#).

Frequently asked questions

- **When should I start working on the project?** You are not expected to start work on your project until the second semester but it is a good idea to think of what project topic you would like to work on and to find group members.

- **Is it ok to have my group project in the same topic with my dissertation project?** The group project needs to be different such that there is no issue of work done in the MLP project appearing as a contribution in your dissertation and you must also cite the group project report (maybe add your project report to the appendix of your dissertation).

- **What are the deadlines for the coursework in next semester?** The deadlines for cw3 and cw4 are at 12:00 on Friday 10 February 2023 and Friday 24 March 2023 respectively:

  - Cw3 (Interim report) will not be marked and will be only for feedback purposes. A good interim report should include motivation and introduction to the project, research goals, details of dataset and task, first phase of experiments, any interim conclusions, plan for the remainder of the project, including discussion of risks, backup plans.
  - Cw4 (Final report) will be marked and worth 50% of your total mark. A good report should include a brief introduction, including a reprise of the aims and objectives, details of the data and the task, experiments, methodology, results, discussion and interpretation, conclusions with respect to aims and objectives, research questions.

- **What are the computational resources that I can use?** We will use the MLP server in next semester that hosts several Nvidia GTX1060's which has 6GB memory. While these GPUs are suitable for many deep learning projects, many state-of-the-art deep methods use very deep and wide neural networks (e.g. ResNet101) and are trained on big datasets (e.g. ImageNet). There are no per-account disk space limits. We will have a lecture and provide a detailed guide about the MLP server early next semester.

- **Are we going to get more Google credits?** No.

- **Do we have to use PyTorch or TensorFlow?** No, you can use your favorite deep learning library such as Keras, MXNet, OpenAI Gym, etc.

- **Are there any good resources to learn how to write good papers?** Yes, see the links below:

  - https://old.reddit.com/r/MachineLearning/comments/85cwiu/d_wellwritten_paper_examples/
  - https://cs.stanford.edu/people/widom/paper-writing.html
  - http://karpathy.github.io/2016/09/07/phd/

- **Does my project have to use deep learning?** Yes.

- **Can I use a company dataset which is not publicly available?** Depends. There should be no problem with using company data as long as (1) the company is happy for them to use the data, and to write about it in their report; (2) the data is available (make sure that you can get access to the data timely, experience tells me that the estimated duration between a company promising data and actually providing the data can be unbounded); and (3) you need to consider ethical issues if the data contains any personal information. For (3) you have to go through the ethics procedure if either the data contains any sort of personal information or if the date involves employing people in any stage of the process (e.g. subjective evaluation).

- **Will each member of my group get the same mark?** Yes.

- **The members of my group did not contribute to the project as much as I did and I had to do more work to cover them. Can my mark be evaluated differently than the other members?** No.

- **I am having problems with my team. What should I do?** Discuss first with your tutor. If problems persist, or at your request, the tutor will inform the lecturers who will be in contact with you.

- **Things are not going well, and I would like to leave my team. May I, and what will happen?** Yes, but only before the coursework 3 submission deadline. You and your former teammates will have to provide each other with your work so far (regardless of who worked on it). You are then in a 1 person team. Obviously this is not advised but in very rare occasions it is preferable, and it is best to know the option is there. There is no further effect, though you may request to change your tutorial allocation.

- **I am afraid to come forward with my team problems because I think my teachers might see this in a bad light.** Please don't be. We only evaluate the final submission. No communication is taken into account in marking. More than that, it is best to discuss problems early on when we can still help. It is not a bother to ask for help; it is a sign of maturity.

General Notes

1. It is ok to have difficulties at first in sharing the team work. Discuss early on what each team member's expectations are and what they are comfortable contributing. You should aim to keep shares equal, but that does not mean you cannot be smart about how you time the work. Talk to your tutor and ask them for help in coordinating your work.

2. You will not generally know exactly what you will do at the start. Your research question will evolve as you learn and implement solutions. Keep detailed notes of what you try and why. Try out different things, and start small. Get a baseline solution to understand your problem and any technical difficulties. Work up from there.