

CSC 2515 Projects

1 General Guidelines

The idea of the final project is to give you some experience trying to do a piece of original research in machine learning and writing up your results in a paper style format. What we expect to see is an idea/task that you describe clearly, relate to existing work, implement and test on a dataset. To do this you will need to write code, run it on some data, make some figures, read a few background papers, collect some references, and write a few pages describing your task, the algorithm(s) you used and the results you obtained.

Students can work on projects individually or in pairs. The grade will depend on the ideas, how well you present them in the report, how clearly you position your work relative to existing literature, how illuminating your experiments are, and well-supported your conclusions are.

The paper should be in NIPS format. You can find the template here: <https://nips.cc/Conferences/2017/PaperInformation/StyleFiles>.

The idea is that this project report should be a manageable amount of work, but that if you want to turn your project into a paper, everything in the project report will need to be done anyway. If you feel that your project won't fit into this rubric, please talk to your instructor. There are many ways to make contributions to a field!

Your project should consist of an implementation of one or more learning algorithms, and their application to a dataset. Your project may be a comparison of several existing algorithms, or it may propose a new algorithm, which should then be compared to at least one other approach. Finding an application of machine learning – collecting an interesting and novel dataset and applying machine learning to it – can also be of interest. As a graduate student, you are free to pick a project of your own design.

You are free to use any third-party ideas or code that you wish as long as it is publicly available. You must properly provide references to any work that is not your own in the write-up. The project is not intended to be a stressful exercise; instead it is a chance for you to experiment, to think, to play and to hopefully have fun! Start with simple methods that work more or less out of the box and go from there.

2 Specific Requirements

Length: 4 to 8 pages, not including appendices. Don't be afraid to keep the text short and to the point, and to include large illustrative figures.

1. **Abstract (5 points):** a summary of the main idea of the project and its contributions.

- Should be understandable to anyone in the course.
- You don't need to say everything you did, just say what the main idea was and what were one or two takeaways.

Marking scheme:

- 5/5: A clear and concise abstract that gives the reader a clear idea of what the project is about and why it is interesting
- 3/5: The abstract is difficult to read and/or is very vague and/or doesn't sell the project as well as it might have

2. **Introduction (5 points):** a statement of the problem being addressed and why we might want to solve it.

Marking scheme:

- 5/5: A readable write-up that explains what the problem is and why it is of interest.
- 3/5: The write-up is difficult to read, somewhat vague, or doesn't make a really good case for why the problem is of interest.

3. **Illustration (10 points):** a figure or a diagram that illustrates the overall model or idea of your paper. The idea is to make your paper more accessible, especially to readers who are starting by skimming your paper.

- For the project, taking a picture of a hand-drawn diagram is fine, as long as it's legible. PowerPoint is another option. You will not be penalized for hand-drawn illustrations – you are graded on the design and illustrative power.
- For camera-ready diagrams, we recommend using Tikz, a LaTeX package.
- Be clear whether arrows indicate computational flow, or conditional dependencies, or both.

Marking scheme:

- 10/10: A thought-out, clear, and original illustration that makes the idea immediately clear.
- 8/10: An illustration that does the job, but is not particularly clear or original.
- 4/10: An illustration that is significantly lacking in some respect.

4. **Formal description (15 points):** a description of the model and/or loss function and/or conjecture and/or problem domain. Include at least one of:

- An algorithm box.
- Equations describing your model.
- A theorem or formally stated conjecture.
- A formal description of a problem domain.

Differentiate your work Highlight how your model is different from other approaches, or what the main relevant considerations are for the domain. This can be done by comparing it to an existing model, perhaps by using another diagram or in words. For example, if you are proposing a new algorithm that only changes one line in an existing algorithm, highlight that one line, or do a side-by-side comparison.

Marking scheme

- 15/15: A full and comprehensible description
- 13/15: A comprehensible description that is somewhat flawed
- 10/15: A somewhat incomplete description where details can be reconstructed with some effort
- 8/15: A flawed description: some things are just unclear
- 3/15: A very flawed description

5. Related work (20 points): A section describing related works and the bibliography.

- If your project builds on previous work, clearly distinguish what they did from what your new contribution is.
- Include a 1-2 sentence summary of other closely related papers. You might not know about all related papers (or have time to carefully read all related papers), and that's OK for this project. A rough guide is that you should be able to find 3-4 closely related papers, and another 3-4 papers that all those papers cite as foundational work. These foundational papers are often cited in the introduction.
- Using bibtex is annoying at first, but Google Scholar can give you the bibtex entries, and it will save you time in the long run.

Marking scheme

- 20/20: An outstanding overview, with an insightful analysis of prior work and a clear connection between prior work and the proposed method
- 18/20: A comprehensive overview of prior work that gives the reader a clear idea of what's out there and how the proposed method is different
- 15/20: A fairly good overview of prior work, and some connection is made to the proposed method
- 12/20: An overview of several papers related to the proposed method, and some attempt is made to connect the prior work to the current method
- 8/20: An overview of several related papers, but not within a coherent conceptual framework
- 4/20: A bad attempt at describing prior work

6. Comparison or demonstration (20 points): Include at least one of:

- A demonstration of a theorem or conjecture. For example, an example or counter-example.

- A comparison of data generated by your model to a baseline model. Qualitative evaluation is OK for the project.
- An experiment demonstrating a property that your model has that a baseline model does not. Experiments should also include a description of how you prepared your datasets, how you trained your model, and any tricks you used to get it to work.
- If doing a review, include a table comparing the properties of the different approaches.

Marking scheme *The marking scheme will vary depending on the details of the project*

- 20/20: The experiments are comprehensive and clearly described. The author(s) have put thought into designing and running experiments that would test whether the idea works or not, and that would allow the author to draw conclusions about why and how the idea works or fails to work
- 16/20: The experiments fairly test out the idea, but they could be more comprehensive and/or better-designed.
- 12/20: The experiments are not comprehensive and/or there is a flaw in the design. The results are suggestive, but it's hard to draw firm conclusions from the experiments
- 8/20: The experiments are significantly flawed

7. Limitations (10 points): A section describing the limitations of your approach.

- Describe some settings in which we'd expect your approach to perform poorly, or where all existing models fail.
- Try to guess or explain why these limitations are the way they are.
- Give some examples of possible extensions, ways to address these limitations, or open problems.

Marking scheme:

- 10/10: An insightful and correct analysis.
- 8/10: A correct analysis that could be more complete and is not very insightful.
- 5/10: An incomplete or somewhat incorrect analysis.

8. Conclusions (5 points): A section describing your conclusions and ideas for future work.

- State the results achieved in relation to the problem described in the introduction.
- Repeat the main takeaways from your paper.

Marking scheme:

- 5/5: A clear and insightful summary of the paper, perhaps with interesting ideas for future work.
- 4/5: A summary of the experiments is given, but the conclusion is a mere summary. The ideas for future work are not interesting.
- 3/5: A flawed conclusion.

9. Depth (10 points): Marking scheme:

- 10/10: A novel and interesting idea for a technical method or an application; an interesting extension and/or combination of previous work in ML; a promising application area; Publication-worthy if born out by experiments. (Note: only one of the aspects needs to be excellent to score 10/10.)
- 9/10: A possibly-interesting technical idea and/or application area.
- 7/10: A reasonable idea: something that's worth trying out but isn't super-interesting
- 5/10: A technically flawed idea

2.1 Project proposal

You must turn in a brief project proposal. Your project proposal should describe the idea behind your self-defined project. You should also briefly describe software you will need to write, and papers (2-3) you plan to read. Please also say if you will have a partner, and if so, who it will be.

Include your email address on your proposal.

The proposal is to be submitted on MarkUs.

2.2 Project submission

Your final submission must include:

- Your research paper
- We encourage you to attach your code (or, better yet, a link to a github project), although this is not required. Code would be helpful to be able to better judge your work.

3 Friendly Advice

(For this project and for doing research in general)

- **Be honest!** You are not being marked on how “good” the results are. It doesn't matter if your method is better or worse than the ones you compare it to. What matters is that you clearly describe the problem, your method, what you did, and what the results were.
- **Be careful!** Don't do foolish things like test on your training data, set parameters by cheating, compare unfairly against other methods, include plots with unlabeled axes, use undefined symbols in equations, etc. Do sensible crosschecks like running your algorithms several times, leaving out small parts of your data, adding a few noisy points, etc. to make sure everything still works reasonably well. Make lots of pictures along the way.
- **Be curious and have fun!** Ideally, you'll stumble across an idea that you'll want to keep working on, as a project that would be interesting and useful to other people (and your career), or as a paper that you'll want to publish.