

Final Project Instructions

CMPSCI 589 - Spring 2017

Project Proposal: Due March 21, 2017 by 11:59pm. 5% of final grade.

Final Report: Due May 2, 2017 by 11:59pm. 30% of final grade.

Late days cannot be used for either the project proposal or the final report.

General Guidelines:

We expect to see a creative, well-researched, and well-executed project. Application projects should pose a novel and well-motivated problem and solve it using machine learning. Implementation projects should select non-trivial existing learning and/or prediction algorithms and implement them in novel computational contexts (on a smart phone, wearable device, embedded microcontroller, within a browser, or in a multi-core/parallel/distributed computing environment). Modeling and algorithms projects should propose new models and algorithms for existing problems. We expect most students to complete application projects, but proposals for other project types are welcome.

In general, we expect projects to involve a mix of activities including selecting or creating one or more data sets, creating learning and prediction pipelines by combining existing components or implementing new components, and designing and conducting experiments to evaluate properties of your pipeline or compare multiple alternatives. All projects require writing up your results in conference paper format.

Projects can be completed in groups of up to three students. The expectation is that groups of 2 or 3 will do 2 or 3 times more work than a single student could accomplish on their own. Students that are involved in research can tie this project to their existing work, but all of the work done for the course project must be new, and whatever you submit for the course project must be entirely your (or your group's) own work.

Specific Requirements for Project Proposals:

Your project proposal is a short description of what you plan to work on for your final course project. You should read the specific requirements for the final project report before starting to draft your proposal. You may change your mind about some of the details as you go, but you should contact us if you decide to completely change your project topic. The maximum page limit for the proposal is **1 page** plus an additional page for references, and a **1 page** collaboration plan for groups of two or more. Students working in groups should each submit a copy of the project proposal.

Your proposal should include:

1. **Title:** Select an informative title for your project
2. **Author(s):** List the names of all group members (your name if working alone).
3. **Problem:** Provide a clear description of the problem your project will address.

- Include sufficient context information to situate your project within the machine learning literature. Describe why the problem is important.
4. **Methodology:** A sketch of the methodology you plan to apply including what models or algorithms you will use or develop, what code libraries you will use or implement, and what hardware platform you will target.
 5. **Data Sets:** A brief description of the data sets you plan to use including a link to the data sets if possible or a clear description of how you will collect or get access to the data.
 6. **Experiments:** A description of what experiments you will perform to validate your approach, study its properties, and/or compare to other existing alternatives.
 7. **Related work and Novelty:** A statement of what others have done with the same or similar data/task/problem (including citations), and how your project will do something different or novel. All projects must identify and describe related work.
 8. **Collaboration Plan:** For groups of 2 or 3, provide an additional summary of which group members will do what work on the project (up to one additional page). There must be a clear separation of tasks among group members and the work must be equally distributed across the group.
 9. **References:** Provide a list of references to support your assessment of related work. You can provide up to one additional page of references.

Specific Requirements for Final Projects:

Your final project should follow standard machine learning paper structure including the following sections. The number of pages per section listed below should be taken as a rough guide, but there is a firm upper limit of **5/8/12 pages** excluding references for groups of 1/2/3 students. Your report should be prepared in [NIPS conference format](#). All group members should submit an identical copy of the final project report.

1. **Title:** Select an informative title for your project
2. **Author(s):** List the names of all group members (your name if working alone).
3. **Introduction (0.5-1 pages):** An introduction describing the problem you are solving, a discussion of why you think it's important or interesting, a statement of what is creative or novel about your project given the related work, and a summary of your solution and findings.
4. **Related Work (1-2 pages):** A related work section summarizing 3-5 pieces of prior research related to the problem your project addresses. You can use Google Scholar <http://scholar.google.com/> to help identify relevant papers. Look for papers in JMLR, MLJ, NIPS, ICML, UAI, AISTATS, AAI, IJCAI, or KDD.
5. **Methodology (1-4 pages):**
 - For application projects, this section will describe the pipeline you have built in detail. Your pipeline may include pre-processing components (missing data imputation, feature selection, feature learning, dimensionality reduction, etc.), one or more core models (classification, regression, dimensionality reduction, clustering), and hyper-parameter or model selection methods (cross validation etc.). You should include mathematical descriptions of the models used and indicate the considerations you took into account when selecting and evaluating components.

- For implementation projects, this section will describe the models and/or algorithms that you decided to implement including mathematical descriptions, a description of the hardware/software platform/programming language you decided to implement the models and/or algorithms in, your design methodology or implementation architecture, and any libraries or existing resources you leveraged for your implementation.
 - For models and algorithms projects, this section should describe your proposed model and algorithms in detail, including mathematical descriptions for the model, and detailed derivations for learning/inference/prediction.
- 6. Data Set(s) (0.5-1 Pages):** Describe your data including where it was obtained from or how you collected it. Describe the number of data cases, the number of features, what the features represent and what their data types are, etc. For data sets with large numbers of features, you should provide a summary of the features and not an exhaustive listing (include a reference to a published paper or website that describe the data in more detail for large data sets if possible).
 - 7. Experiments and Results (1.5-4 Pages):** You must perform experiments to explore some aspect of your solution and compare it to one or more alternatives. Typical experiments may involve the use of cross validation to optimize hyperparameters, and testing several different models to determine which works best on your data in terms of speed, accuracy, energy use, etc. You must carefully describe the experiments you perform, justify your methodological choices, and report the results using suitable figures (your report must contain at least 2/4/6 results figures or tables for groups of 1/2/3).
 - 8. Discussion and Conclusions (0.25-1 Pages):** Discuss the results of your experiments. How do your results relate to what has been reported in the literature previously? What seemed to work well and what didn't? Did you run into any particular problems? What else would you have done if you had more time?
 - 9. References:** Provide a list of references to support your assessment of related work. You can provide up to one additional page of references.

Final Project Marking Scheme:

The following criteria will be taken into account when marking final project reports:

- 1. Creativity and Novelty [15%]:** Did you use an existing data set to pose a new problem? Did you engineer or learn new features or representations for the data? Did you leverage multiple data sets in a novel way? Did you collect a new data set? Did you combine methods in a novel way or propose a new method? Did you investigate methods or hardware platforms not covered in assignments or in class?
- 2. Clarity [15%]:** Did you clearly describe the problem you are trying to solve, the proposed solution, your evaluation methodology, and your results and conclusions? Is your report easily readable? Are the figures and tables properly labeled?
- 3. Related Work [20%]:** Did you give a detailed discussion of the relationship between your problem and previous work? Did you include references and use citations appropriately? All projects **must** identify and describe related work.
- 4. Technical Correctness [20%]:** Are your descriptions of models, methods, and experimental procedures correct? Are the conclusions you draw valid?

5. **Experiments and Results [20%]:** Are your experiments well designed? Did you select hyperparameters in a valid way? Did you compare methods in a valid way? Did you present results using appropriately selected graphs?
6. **Reproducibility and Code Quality [10%]:** Did you design your experimental code so that your results are reproducible? Is your code well documented?

General Project Advice:

- **Be selective!** Don't choose a project that has nothing to do with machine learning. Don't attempt to address a problem that is irrelevant, ill-defined or unsolvable.

- **Be creative.** Implementation and Models and Algorithms projects will generally satisfy the creativity and novelty criteria. Application projects that apply machine learning methods covered in class to previously-defined problems on highly curated data will result in a creativity score of **zero**. Some suggestions for improving the Creativity and Novelty of application projects are:

- **Use data in a novel way:** Think about new and creative ways to formulate machine learning problems over existing data sets that may have been used for other purposes in the past. Don't be afraid to use messier data sets that have not already been curated for use with machine learning algorithms. You can also collect a lot of interesting data on the web, but if you do, make sure you follow the acceptable use policy of any web sites you collect data from, and allow sufficient time for data collection.
- **Explore advanced methods:** Explore advanced topics not covered in the class including dealing with missing data, noise, and outliers in supervised learning; use of structured prediction for sequences, time series or spatial data; learning from weakly labeled data; active learning; reinforcement learning; deep learning; etc.
- **Be careful.** Don't make basic mistakes like testing on your training data, setting parameters by cheating, comparing unfairly against other methods, using undefined symbols in equations, etc. Do sensible crosschecks like running your algorithms several times.
- **Be honest.** You are not being marked on how good the results are. What matters is that you try something creative (but sensible), clearly describe the problem, your method and experiments, and what the results were.
- **Use time well.** Don't wait until the end of the semester to start seriously working on your project. Don't underestimate the time needed to identify or collect data, understand new methods, get code running, run experiments, produce figures and tables, ensure reproducibility, and write a clear report.
- **Have fun!:** If you pick something you think is cool, that will make getting it to work less painful and writing up your results less boring.