# CPS 360 Final Project

**Important Dates:**

November 15 (Thursday): Proposal Due.
November 19 (Monday): Last day to have meeting with the professor discussing your proposal
December 10: Presentations and Q/A in class
December 12: Project due on Canvas at 11:59 PM (**no late submissions!**)

The final project is an opportunity to spend time on a project that you will enjoy working on, and that has some significance to you. You will work individually on a project of reasonable size. You will experiment with either an approach or an application beyond what we discussed in class (or both!). This project will be a chance for you to apply machine learning to a problem that interests you or to dig deeper into a topic that intrigues you. You will produce a paper that presents your results and give a presentation to your peers. This project should be different from any of the homework assignments. *The complexity of your program should be on par with 1.5-2 homework assignments.* At any point during this process I welcome you to consult with me for feedback.

The project should be a research project. It could be some problems you identified in your other classes or research fields which need to be solved by machine learning algorithms. It could be an empirical study (comparison of the performance) of several machine learning algorithms on some datasets you are interested in. It could also be possible improvements (you designed) to certain machine learning algorithms.

The datasets you will use in this study could come from your other research projects. You may generate your own dataset of a problem by extracting features which you think are important. You are also allowed to use some public datasets of your research fields for problems you are interested in. If you are experimenting with an improvement to a machine learning algorithm, you may use the datasets provided by UC Irvine, which has a repository that could be useful for you project: http://www.ics.uci.edu/~mlearn/MLRepository.html. Sam Roweis also has a link to several datasets out there: http://www.cs.toronto.edu/~roweis/data.html.

Here are some possible project ideas.

1. Application of regression algorithm to some real life problems. For example,
    - Weather prediction problem. This dataset includes 45 years of daily precipitation data from the Northwest of the US:http://www.jisao.washington.edu/data_sets/widmann/. How you split the datasets? How much history data you use to predict the next few days' weather?How to evaluate your predictions?
    - Stock market prediction, etc.

2. Implementation and comparison of some machine learning algorithms and compare their performances on several datasets. For example, you may implement UPGMA, Neighbor-Joining algorithms and compare their performances with K-means on image segmentation.

3. Use artificial neural network to solve non-linear regression problems and compare its performances with linear regression approach to non-linear problems.

4. Use artificial neural network to compress and decompress digital images. What is percentage is compression rate, how well can you restored these images, how log does it take to compress and

decompress images, etc.

5. Use evolutionary algorithm (genetic algorithm) to solve an optimization problem. The problem could be a difficult problem in Computer Science, Math, Physics, Finance, Economics, or other fields.

6. Use evolutionary algorithm to train the weights of neural networks and compare its performances with that of back-propagation.

7. Empirical study of some machine learning algorithm by comparing the performance with various parameter settings. For example,
    - You may compare the performance (i.e., how many iterations till it converges, what is the average best solution found, etc) of genetic algorithms by using various crossover and mutation strategies in solving problems such as the traveling salesman problem. Of course, for stochastic algorithms such as genetic algorithm, you may run the algorithm multiple times for each setting.

**Proposal (2 point):** Your proposal document should be uploaded to Canvas. Your document should be around 1-2 pages long. In the proposal, it should contain the following information: 1) topic of the project; 2) your name; 3) background and motivation of this study; 4) datasets; 5) simple description of the method(s) you will use; 6) possible experiment design; 7) expected outcome; 8) timeline of the project (when to finish coding, experiment, result analysis, report drafting, etc). Think about setting key goal posts and how long you expect it will take to reach them (implementing major pieces of functionality, testing and documentation, etc.).

I will then read your proposal and write comments to provide you with suggestions about your final project. If the project is too simple, then I may ask you to elaborate on it. If the project seems too ambitious, I may help you set more modest goals. The proposal must be submitted by November 15. You will need to meet with me in person to discuss your proposal by 1:00 PM, November 19.

**Project Report (5 points):** This writeup should be submitted as a PDF file, along with your project code. It should be around 4-6 pages. It must contain:

- The title of this study
- Abstract
- Introduction (background and motivation of this study).
- Datasets and method description: Write simple introduction of the datasets used in your study. Explain the method you have implemented/used.
- Experimental design and results analysis: What experiments have been done to test such algorithms or methods, e.g., cross-validations? What statistics (e.g., error function, TP, FP, TN, FN, sensitivity, specificity, accuracy, MCC, etc.) have been used to evaluate the method? What are the experiment results?

Conclusion and discussion of this study: What are the major contributions of this study? What inspirations can it give us? Are there any shortcomings with this approach? What other possible improvements can you think of? What is the possible future work or direction? Can this approach or method be applied to other problems you can think of?

**Project Code (5 points):** all the programs implemented in your study

- Write a readme file or documentation file describing the steps in reproducing your study. Your

project may contain several programs. Please explain how to run these programs.
- For each program you wrote, document it well.
- All your programs should run.

**Final Presentation and Q/A (3 points):** You will briefly present your project/study to the class, discuss your design decisions, challenges and your results, and answer questions. Note that the presentations are scheduled before the final due date. Your project should be near to completion by the presentation date (i.e., pretty much finish the coding and experiments and at least have some initial experiment results). The remaining days should be spent on polish, testing, debugging, and documentation, report writing, *not* on implementing core functionality.

**Bonus (1-2 points):** If your project goes beyond the basic requirements and shows significant additional complexity and/or imagination, you will receive 1-2 bonus points depending on the originality of the study.

The completed project is due by 11:59 pm, December 12. NO LATE SUBMISSIONS WILL BE ACCEPTED. You are *highly* encouraged to submit *before* this deadline to ensure that you get credit for your work! Double check your submitted files and make sure that all are complete and the correct versions. Remember that all necessary files, including the project report, all programs and documentation files, and result files.