

CSCI 447/547: Proposal and Final Project

September 14, 2018

Proposal due Oct. 23. Project due Dec. 10.

1 Project

This course requires a final project (due on the course's final date), consisting of a paper in the scientific style (including abstract, introduction, methods, results, discussion, conclusions, expect 3–4 pages). Undergraduates may select one of the following options. Graduate students are restricted to option 2, and will be held to a somewhat higher standard of quality:

- 1. Book Report** Identify a source of primary literature (either an academic or industrial publication or a website implementing machine learning), and deliver a synopsis on what said literature is about or how it works.
- 2. Algorithm Application** Identify a machine learning algorithm or method and apply it to a data set. Either the algorithm or the data can be something that we covered in class, but not both.

A few things to note in no particular order. First, I strongly encourage ‘double-dipping’ in the sense that if this project can be useful towards your research or another project, then please pursue that. Second, I am very happy to work with you in the development of project ideas. Note that some excellent sources for large datasets include Kaggle, the UCI machine learning repository, and the University of Florida machine learning database. Some notable examples of previous projects include the development of a convolutional neural network for classifying tumors as benign or malignant, a naive Bayes classifier for fatal versus non-fatal traffic accidents, a report on face swapping in images, and a hidden Markov model for identifying matches in noisy genome sequences. Finally, the prospect of failure should not be an impediment to your project selection: it is well understood that much of Machine Learning is an art as much as a science, and researchers often work for years in an attempt to find good solutions to difficult problems. With that in mind, I will not grade based on classifier accuracy or any similar metric of model performance. Rather, I will grade on your demonstrated ability to understand a chosen algorithm or method, to reason about why that algorithm or method is a good choice for the problem you are approaching, and why the method ended up working well or not.