COMP-598: Applied Machine Learning

Final project: Machine Learning for Improving Healthcare

Due on December 2 (short oral presentation) and December 11 (written report)

Background:

There are significant opportunities for applying machine learning techniques to tackle challenging data analysis problem pertaining to healthcare. As an example, recent Kaggle competitions were pertinent to this theme: *https://www.kaggle.com/c/diabetic-retinopathy-detection, https://www.kaggle.com/c/seizure-prediction* Several workshops have also been held on this topic, e.g. *http://www.ml4chg.org*

The goal of this project is to investigate healthcare data. Begin by identifying an interesting prediction question that can be tackled using machine learning methods, and solve the problem using appropriate machine learning algorithms and methodology. You are not restricted to a specific dataset. You can incorporate data from several sources, or even collect additional data (e.g. new test set) if appropriate.

The choice of prediction task is open. Try to pick a prediction question that is relevant and important to improving health care delivery. Remember to design a prediction task that is well suited to your choice of dataset; and vice versa, pick the right data for tackling your prediction question. Avoid choosing a problem that you are already studying for your own research. You can use a dataset that has been previously analyzed (e.g. data from the kaggle competitions above), or aim to replicate results from an existing paper.

There are no mandatory algorithms to implement for this project. You can implement your own algorithms, or use any existing toolbox or software, as long as you reference it appropriately in your report.

The final project can be completed individually, or in a group of 2 or 3. You can work with any student in the class, including those with whom you worked on one of the previous mini-projects.

Requirements:

There are 2 components to submit:

1. Short oral presentation (3min / team) to be presented in class on December 2. One or all team members can speak. Slides can be used (pdf format only). Presentations will be timed precisely, so practice accordingly. The presentation should clearly identify the target prediction question, describe the data used, define the methods applied (or to be applied) and if possible, give preliminary results. Evaluation criteria: Importance of prediction question, pertinence of methods, discussion of results or expected results, quality of delivery, organization. Weight: 20% of final project grade.
2. Written report (1 / team). The report should have at most 8 pages. It should follow a standard conference format (e.g. see instructions from Mini-project #1: double-column, 10pt font, 1" margins or IEEE conference format). The report should clearly present the target prediction task, the data used, a clear and well-motivated methodology for analyzing the target question with the selected data, empirical results of the analysis and a discussion. When appropriate, related work (e.g. attempts to tackle a similar prediction question, or analyzes similar data) should be briefly discussed and referenced. Weight: 80% of final project grade.

Submission instructions:

Submission instructions will be posted at a later date in a revised version of these instructions.