

15.097 Prediction: Machine Learning and Statistics Project Suggestions

Here are some ideas for the final project. Of course, you are also encouraged to come up with your own ideas that do not necessarily fall into any of these categories.

I. Study and prepare lecture notes on a set of topics not covered in class

Our goal was to give you an overview of the major topics in machine learning and statistics, but here are some topics we did not have time to cover:

- a. Expectation-Maximization and PageRank (the two top 10 algorithms not covered)
- b. One-class classification (anomaly detection) and multi-class classification
- c. Recommender systems and collaborative filtering
 - i. You see this on online sites that recommend products as you make purchases
 - ii. Implement an algorithm for low-rank matrix factorization and test it on data similar to that from the Netflix competition (<http://www.netflixprize.com/>).
 - iii. Investigate other matrix completion methods as well.
- d. Neural networks
- e. Use of first-order optimization methods in large scale machine learning

II. Address a problem from a competition

There are a variety of data mining competitions. Try to obtain their data and come up with solutions to their tasks. Possible competitions include the following (there are others; just do a Google search):

- a. KDD Cup (<http://www.sigkdd.org/kddcup/index.php>)
- b. Kaggle (<http://www.kaggle.com/competitions>) – Sign up for an account to download the data. You might want to take a look at “What Do You Know?” or “Photo Quality Prediction.”
- c. Kaggle in Class (<http://inclass.kaggle.com/competitions>)
- d. Computer Cooking Contest (<http://liris.cnrs.fr/ccc/cc2011/doku.php?id=welcome>) – Figure out a way to predict missing ingredients of recipes. Association rules might be a good place to start.

III. Design a new algorithm to solve a machine learning problem

Those of you who have training in optimization methods may put it to good use in designing new algorithms and showing how they compare against conventional methods. Think of alternative techniques, such as:

- a. First-order methods
- b. Linear programming
- c. Integer programming

For example, you might be able to use math programming for clustering.

IV. Acquire an interesting dataset and learn something insightful from it

Data are all around us. See if you can use machine learning techniques to tell you something interesting. For instance:

- a. Crawl Amazon for product reviews and determine associations between reviews.
- b. Find statistics on NCAA teams and develop a model to predict the Final Four.

MIT OpenCourseWare
<http://ocw.mit.edu>

15.097 Prediction: Machine Learning and Statistics
Spring 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.