CMPT 733: Final Course Project

Arash Vahdat

Fall 2015

The last few weeks of the course will be dedicated to projects. Below is a list of all necessary information related to the format, guidelines and a list of project ideas to get you started.

Timeline

- Friday, November 6th, 23:59: Project proposal submission deadline.
- Friday, December 4th, afternoon 14:00-17:00: Project poster presentation.
- Monday, December 7th, 23:59: Final code and report submission deadline.

Guidelines

- You can choose to work in pairs or individually.
- The project is worth 25% of the final grade.
- Your project has to be big data related, which means you should make sure you are using a large enough dataset. Discuss with me if you are not sure about the size of your data.
- Your solution should be *scalable*.
- Your project should involve training a machine learning model.

Deliverables

- Project proposal. (15%)
- Final poster presentation. (25%)
- Final report. (40%)
- Final code with a brief documentation. (20%)

Proposal It is important to start thinking about a topic ASAP. From now till the project proposal deadline, you are invited to talk to me about your project ideas during labs, or by e-mail. You can use this period to finalize your project and write your proposal. The document should be at most one page long and it should cover the following:

- The problem raised in your project and the motivation.

- The material you will use: dataset, libraries, etc.
- The method you are proposing to explore.
- The anticipated challenges you think you might encounter.

Poster Presentation We will hold a poster session on Friday afternoon, December 4th in Burnaby campus. Exact times will be confirmed during the term.

Final Report Your final report should contain the following:

- Problem Definition: What is the problem addressed? What are the challenges?
- Methodology: The proposed method, algorithmic details such as the computational and space complexity of your solution in terms of input size?
- Results and Discussion: What are the outcomes of the project, your conclusions?

Note that your report should not exceed **5 pages**. Try to use a good template for your report. For example you can use templates available at: https://nips.cc/Conferences/2015/PaperInformation/StyleFiles

Final Code

- Reproducibility: Your submission should contain the main functions used for your project.
- Documentation: Similar to your assignment reports, briefly explain which code was used for which task. Include a Readme.txt file clearly explaining how to run your code and reproduce your results.

Project Ideas The following is a list of suggested topic ideas to get you inspired. You can choose to pick a project from the list below as is, modify it to suit your taste, or propose your own idea. Be creative!

- **On-line Web-based Systems:** In our course, we considered different machine learning applications such as image classification, sentiment analysis, movie recommendation and anomaly detection. You could choose to create a web-based system that runs one of these applications on-line.
- **Course Applications with a New Solution:** You can try attacking one of the problems we reviewed during the course using a new approach. Naturally, it is expected that you compare your proposed solution with the solution studied in the course.
- **Other Big Data Libraries for Machine Learning:** You can aim at using other big data libraries such as Apache Mahout, H2O, and Cloudera Oryx for any of the applications seen during lectures or assignments.
- Check the following websites for more inspiration:
  - https://www.kaggle.com/
  - http://web.stanford.edu/class/cs224w/resources.html
  - http://cs229.stanford.edu/projects2012.html
  - http://courses.cms.caltech.edu/cs253/projects.html
  - http://ibmhadoop.devpost.com/details/data
  - http://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public