Course Project

CMSC 678 — Introduction to Machine Learning

<table>
<thead>
<tr>
<th>Item</th>
<th>Summary</th>
<th>Length</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal</td>
<td>Monday March 12th, 2018</td>
<td>1-2 pages</td>
<td>30</td>
</tr>
<tr>
<td>Progress Report</td>
<td>Monday April 9th, 2018</td>
<td>2-3 pages</td>
<td>70</td>
</tr>
<tr>
<td>Final Report and Code</td>
<td>Wednesday May 23rd, 2018</td>
<td>4-6 pages</td>
<td>200</td>
</tr>
</tbody>
</table>

In this project, you will have the opportunity to explore an aspect of machine learning that interests you. Teams of up to four people are allowed: that is, you may work alone, or with one, two, or three other classmates. You may choose any aspect of ML; that aspect does not have to be one that we cover in class, though it’ll probably be easier if it is.

You have a lot of freedom in this project. The main requirement is that there is some novel experimentation, implementation, analysis, or problem formulation. This could include, but is not limited to:

1. reimplementing a prior effort and applying it to a new source of data, but of the same type for which it was initially developed, e.g., reimplementing a CV image recognition system and train/test it on new classes of objects;
2. reimplementing a prior effort and applying it to a different type of data, e.g., reimplementing a CV (NLP) system and applying it to language (vision) data;
3. formulate, implement and compare a novel solution to an existing problem;
4. formulate, acquire data for, and apply a sufficient baseline for a novel task.

Your project must be well motivated. This is especially true if you are reimplementing an existing system: you must explain the hypothesis for why that model might do well on the novel data.

**Deliverables**

This project has three main deliverables:

**Progress Proposal** A proposal is due Monday, March 12th by 11:59 AM. You should use the ACL style format (see below). The report should be at least one page, not including references; while there is not an upper bound, a two page report is reasonable. The report should include

- a title,
- authors (members of the group),
- a brief (at most 200 word) abstract,
- a description and motivation of the problem,
- a description of your proposed solution,
- a (brief) description of how your solution fits in with what has been previously done on the problem (or related problems),
- any blocks you’ve experienced (or anticipate experiencing).

Be sure to cite any references.

**Progress Report** A progress report is due Monday, April 9th by 11:59 AM. You should use the ACL style format (see below). The report should be at least two pages; while there is not an upper bound, any report longer than three (not including references) should be justified with particular questions and confusions (supported by data/results). The update can be an extended version of the proposal: you can use what you wrote for the proposal but extend (or adjust) what is already written and include the [progress] components.

- a title,
- authors (members of the group),
- a brief (at most 200 word) abstract,
- a description and motivation of the problem,
- a description of your proposed solution,
- a (brief) description of how your solution fits in with what has been previously done on the problem (or related problems),
- [progress] the progress you’ve made,
- [progress] what remains to be done, and
- any new blocks you’ve experienced (or anticipate experiencing).

Be sure to cite any references.

**Final Report and Code** A final report and all of your code, data, and final models are due Wednesday, May 23rd by 11:59 AM. If you acquired new data, or added new annotations to existing data, please either include this data or (if the dataset is large) make arrangements with the course instructors ahead of time.

The report should be at least four pages, and no more than six, not including references. In writing the final report, you may reuse what you wrote for the progress report, but be sure to specifically and thoroughly address the [final] components. The report should include

- a title,
- authors (members of the group),
- a brief (at most 200 word) abstract,
- a description and motivation of the problem,
- a description of your proposed solution,
[final] a description (longer than in the progress update) of how your solution fits in with what has been previously done on the problem (or related problems),
[final] evidence of proper experimental methodology,
[final] a clear description and analysis of your results,
[final] an analysis of the limitations of your work, and
[final] potential follow-up work, both short- and long-term.

You should use the ACL style format (see below).
Do not write just to fill up space. While four pages (the minimum) may seem like a lot at first, you can fill it up very quickly.

**Code Requirements**

Your code should be compatible with standard Linux machines, and please make it easy to run ☺. You may use any external libraries you wish to use. Please do not assume access to something hardware dependent like CUDA; if you think you’ll need CUDA to complete the project, please consult the instructors to work out what resources you need.

**Report Requirements**

Both reports should be in the ACL format. Please use the ACL 2017 style guide. I have downloaded and made available both \LaTeX and Microsoft Word (docx) versions: https://www.csee.umbc.edu/courses/graduate/678/spring18/materials/sty

Be sure to cite appropriately and follow all academic honesty standards. You may include figures (your own, reproductions, or copies of existing figures); be sure to provide appropriate credit for the figures. However, make the figures count: do not include them simply to pad the paper.

Do not consider just “recent” papers; try to find papers from the past 25 years. You may be surprised at the recurrence of ideas.

**Where to Start**

Google Scholar is an easy way to find linked and cited papers. The [AAAI digital library] also offers an extensive listing of AI-based conferences and proceedings. Of particular relevance are the flagship AAAI, ICML (International Conference on Machine Learning), and KDD (Knowledge Discovery and Data Mining) proceedings. Papers from [NIPS (Neural Information Processing Systems)] often tend to the more theoretical, but with a decided focus on neural networks.

You are welcome and encouraged to come talk with me, either during office hours, over email, or by appointment to discuss topics, advice on finding relevant papers, and the direction of your project.