CE888 Assignment 1

January 27, 2017

1 The assignment

The main aim of the two assignments is to analyse data and present that analysis both on-line and as if it was written for publication in an academic journal. Assignment 1 is simple - you need to identify a data science problem you will work on and write a report about it. You also need to identify the relevant journal that would publish that kind of analysis/research you will perform. The report should be split into the following sections:

1. Abstract: provide a short description of your work and try to convince the reader that is worth reading the rest of your paper!
2. Introduction: explains the purpose of your work and motivates it.
3. Background: description of similar efforts done in the past.
4. Methodology: describe what your analysis will achieve and what methods you will use to achieve your goals. Describe the dataset you are going to use and how the data was collected (or generated).
5. Experiments: outline any experiments you will perform and explain the reasons behind them.
6. Discussion: explain how you will evaluate the results and how you will gain insights from your experiments.
7. Conclusion: any concluding remarks you might have.
8. Plan: estimate of when you expect to implement in the experimental section, and how long it will take. Use dates or a gantt chart. Be realistic about what you can achieve.

The maximum page limit for the report is the page limit of the conference/journal you chose to (fictionally) publish this in. Note that this is a hard limit - you should never go above it. If no such limit exists, aim for 6 pages. Alongside your assignment you should implement a prototype of the work you are going to do. The prototype should at least be able to access the data you are going to use.

2 Deliverables

1. Report in Microsoft Word or PDF format, adhering to the journal standard of your choosing.
2. A link to a github project that contains the prototype code and the data you are going to use - it should be the same as the one you used for your labs. If the data used is massive, provide a link to it instead in your github README.md.
3. Complete project tree containing all files used in the project - basically a .zip file of your github project.