W4315 Final Project Instructions

The final project is meant to be a time-intensive, significant, team-based effort that demonstrates your knowledge of regression analysis techniques. Note that the final project takes the place of nearly 6 homeworks and a final exam. So, if you typically spend 3-4 hours per homework and would have spent 20-30 hours studying for the final examination then the total equivalent effort that you will be expected to expend on the final project will be, per person, something like 40-50 hours. In teams of up to 4, this means between that each project will be expected to benefit from between 160 and 200 man hours of labour, enough to get a significant amount of work done. This means that you will fail to complete a satisfactory final project if you wait until the last minute to begin and complete the project.

The final schedule is as follows:

Nov 10th – Final project groups are to be formed by this date. A project team may consist of 4 people maximum. Groups consisting of 2-3 people are preferred. Single member groups are strongly discouraged. Each group should submit a piece of paper with a group name and the names and CUID's of all group members by the end of class on the 10th.

Nov 17th – Final project proposals are due on this date. The proposal will be graded and will count as 25% of the final project grade. The proposal should consist of a 3-5 page, single space, full column-width, 10pt font paper describing the general outline of the intended final project. In particular, it should take the form of a research abstract short paper – title, authors, abstract, introduction, data, methods, conclusions, and references. Note that the only significant missing piece is the results. Other sections will grow in the final project report as well, but each should appear in the project proposal in reasonably complete form. Of particular interest are a description of the data to be used, a detailed description of how it was collected, a literature search of analyses already performed on the data and the conclusions drawn from those analyses, and an overview of the techniques you intend to use in analysing the data. Project proposals that do not outline a project with sufficiently high complexity will be given zero credit and groups that author such proposals will be required to resubmit project proposals on a weekly basis thereafter until the proposal reaches a sufficient level of complexity to warrant a passing grade.

Dec 8th & Dec 10th – Final project presentations. The presentation will be graded and will count as 25% of the final project grade. Each group will present their results in a short in-class presentation (single presenters are preferred). The day on which your group will give their presentation will be decided in class on Dec. 3rd by a random selection process. The duration of each presentation will depend on the total number of groups formed, but should not exceed 10 minutes. Slides for the presentations (plan on 1-2 per minute of presentation) must be delivered to the instructor 15 minutes before the start of class. The presentation should explain the domain of the regression analysis, highlight the methods used, and cover analysis results. Each group will be subject to 5-10 minutes of questioning from both the instruction and the other students in the class about all aspects of their project. Having backup slides (slides prepared, but not used in the presentation) for questions about methodology detail will be expected of all groups.

Dec 15th – Final project report. The final project report will be graded and will count as 50% of the final project grade. It should consist of a 10-15 page, single space, full column-width, 10pt font paper in the research paper style describing all aspects of the project in detail, in particular title, authors, abstract, introduction, data, methods, results, conclusions, and references. Of particular importance are the mathematical methods section which should be explicit and detailed and the results section which also should be very detailed. This should be thought, written, and formatted as a research paper and will be judged as such.

Comments and Project Ideas

As has been mentioned throughout class, there are (roughly speaking) two kinds of projects that will be considered to be above threshold in terms of complexity.

The first class of such projects are those that propose to use multiple regression analysis techniques to analyse a dataset of actual importance to either a member or members of a group or to the world at large. This kind of project will require a careful review of the data source, how it was collected, what kind of modelling assumptions can safely be made, etc. The point of such a project would be to produce an analysis of some dataset that is relevant to an application domain and the conclusions from which could be put into use by practitioners in the relevant domain (clinicians for medical data, traders and analysts for financial data, managers and executives for business data, etc.). These projects will be graded and evaluated on the validity and power of the domain specific conclusions drawn after taking into account the appropriateness of the assumptions made.

A second class of above threshold project is a methodological exploration project, i.e. a proposal to spend less time on data selection, data analysis, and the making of domain specific statistical conclusions and instead to implement and deploy one or more of the advanced regression methods we will cover later in the class (Bayesian regression, SVM regression, neural nets, etc.). For this kind of project, emphasis in both work product and in results will lean more towards producing and explaining either a new model (probably not actually new, but for the purposes of writing something up you can pretend that it is) or a new technical result. Comparisons of multiple regression models on a single dataset that consist of technically comparing and contrasting the pros and cons of each would be of interest.