MATH 8820

Project 1

Forecasting/predicting is one of the most useful consequences of statistics. In this exercise you will be a forecaster, but note "A good forecaster is not smarter than everyone else, he/she merely has his/her ignorance better organized." The task, which is not for the faint of heart, will be to develop a spatio-temporal flu forecasting model. Forecasting other diseases would be acceptable, as long as I okay it. Your forecast should be at the monthly and county level or for a finer spatio-temporal granularity. Note, historical flu data is relatively easy to find from the Centers for Disease Control and Prevention.

- A paper outlining your approach/methodology will be prepared, this work should include the following: an abstract succinctly summarizing the objective of the paper, an introduction discussing any relevant information, a methodology section which outlines in detail the approach that you have taken to build your flu forecast, a data section which discusses where data was obtained, the relevance that it has, and how you will use it. The last section of your paper should present your forecast for the entire continental US. This paper should be no more than 10 pages and no fewer than 5, it should be single spaced, use standard margins, and 12 point font.

- I encourage you to look into what others have done previously to develop flu forecasts and adapt/adopt these techniques. In doing so, you should provide appropriate references. Note: Plagiarism is not acceptable, can easily be spotted, and will result in a failing grade.

- Data sources: You are free to use any data source that you so desire, but you must acknowledge where you collected the data in a formal fashion; i.e., you should formally reference your data sources. Further, you should hand your data files in with your paper at the prescribed due date. These files should be full annotated and fully described.

- Fully annotated code, which implements your analysis, should be appended to your paper. Your code does not add to the length of your paper. Note, the use of existing software packages is completely acceptable (and actually recommended). The goal here is that anybody should be able to take your code and rerun it to obtain the results that you present, and you should provide the necessary details on how this is done.

- I prefer that your paper be prepared in Latex, and I will provide a general template that you can use to do so.

Important: Do not procrastinate, get started on this as soon as possible. It is really easy to detect work that has been thrown together at the last minute and I expect more from you. If you need any assistance, as always I am willing to help, but much of this should be an independent study.

Grading: This project will be graded with respect to two criteria. First (50% of the grade) the approach and your write up, you will be graded on your approach and its appropriateness, your write up, and the general presentation of your results. Second (50% of the grade) the accuracy of your forecast, more on this in due course.