MATH 442: Mathematical Modeling

Fall 2014

Instructor:	Prof. Wolfgang Bangerth Blocker Bldg., Room 507D bangerth@math.tamu.edu http://www.math.tamu.edu/~bangerth
Office hours:	Thursdays, 2:00–4:00pm or by appointment
Lecture:	Tuesdays + Thursdays, 8:00–9:15am Blocker Bldg., Room 133

Writing assignment 3

Assignment

Earlier this semester, we considered the change in population of a single species and came up with at least three different models: unrestricted exponential growth, the logistic model, and the Gompertz model. The latter two assume that there is a limit to how large a population can be supported long-term in the ecosystem (the carrying capacity).

From some reliable source, obtain as much population data as you can get for your home town, home state, home country, or some other entity you feel attached to. Your data should at least encompass 75 years. For example, for data about entities in the United States, the US Census website at census.gov provides this information, but other countries have similar data sets available.

Using this data, write a report that contains the following sections:

1. Introduction: State the population you consider; provide a scatter plot of the data set you will use; and give an outlook on the rest of the report.

2. Modeling the data: This section will be about fitting the parameters of various models to the data you have. Discuss the following topics:

   - Introduce briefly what the exponential growth, logistic growth, and Gompertz models represent.
   - As a baseline, compute the least-squares fits for a linear and quadratic model (these are not motivated by any considerations such as births and deaths, but are simple and can serve as a point of comparison for the more complicated models).
   - Compute the least-squares fits of the exponential growth, logistic growth, and Gompertz models to your data. These models are nonlinear in the parameters, and you may have to be creative in finding those parameters that minimize the misfit.
   - For each of the five models, present the values of the parameters that minimize the misfit and show the corresponding curves of all five best fits in one graph together with the actual data.
   - For each of the five models, list misfit and R^2 values in a table. Provide an ordering of the five models regarding their goodness of fit.

3. Predictions: Fitting models only uses past data, but we want to fit models because that may allow us to make predictions for the future. To this end, create a plot similar to the one above but that shows the curves out to the year 2200. In a table, state for each of the five models what population they predict for year 2025, 2050, 2100, 2200 and 2500.

4. Discussion: State which of the predictions you have obtained in the previous section seem realistic to you and why or why not. Elaborate *in detail* why the five models are not likely to produce any reasonable prediction beyond the immediate future. In other words, discuss why the models are not accurate descriptions of the effects that really affect *human* populations.

   Speculate which effects a more realistic model would have to take into account and how you think this would affect the ability to predict future populations in year 2025, 2050, 2100, 2200 and 2500. Correlate this with what you know about how population groth has changed over the past 200 years in various parts of the world, and what these changes may imply for the population you consider over the next 200 or 500 years. You may wish to research other sources of information for this.

5. Summary and conclusions.

There are no lower or upper bounds on your paper's length. However, you will need to address each of the points above in sufficient detail. (I anticipate that you will need around 10 pages to do so, but this is just a guess and not a recommendation.)

Specifics

- Write your report in the LyX word processor starting from the IEEEtran.lyx style file (or, if you want and are familiar with it, in straight LaTeX, for example using the `kile` editor or the http://www.writelatex.com service).
- You will do a peer review session during class on Tuesday 11/4/2014, so have a draft ready by that time.
- The final version of your assignment is due Friday 11/7/2014 5p.m. You need to submit it before this date through the eCampus "Turninit Assignments" function.

Grading

Grading will be based on the following rubric:

- *Correctness 40%:* Obviously, the data, models and fits you present must be correct and the best fit models must be accurate representations of the data. This part of the grade primarily concerns the *Modeling the data* and *Predictions* sections.
- *Insight 40%:* When discussing the validity and limitations of predictions, you will have to take into account a variety of sources outside the data you use to fit models. This part of the grading is about how well you take the *broader context* into account when making predictions of future population sizes. In particular, I will evaluate the comprehensiveness of your reasoning of why the models will/will not result in accurate predictions. I will also consider positively if you can provide more realistic models that take into account the distinctly *human* component of population change.
- *Structure, style and clarity 20%:* As a text, your report should be easy to read, i.e., it should be written with a reasonable subdivision into subsections, in proper English, and at an appropriate level of technicality suitable for your audience. It should include formulas where appropriate, show Maple commands if necessary for someone like you to repeat a computation, should have axes in graphs appropriately labeled, etc. Put yourself in the shoes of your reader if you think about whether to include a detail or not.

Why I'm assigning this project

Data was sparse 100 years ago when obtaining it involved either repeated manual measurements and copying things into a notebook, or going to the library and trying to find a book or paper that reproduced the data someone else had painstakingly recorded. On the other hand, incredibly large amounts of data are available today since we have cheap sensors for basically everything, and computers to record it with high frequency. To give just two examples, one big and one small: the New York Stock Exchange records prices for all traded stocks second by second, producing many gigabytes of data per day; and my bike computer records my speed, location, heart rate, cadence, altitude and other data every five seconds, producing several 100 kilobytes per bike ride. The challenge is to make sense of all of this data – a challenge so large that a whole new field of science, called *Big Data Sciences*, has emerged over the past 10 years.

This assignment works with some of the simpler things one can do with data: fitting models to it. It also asks you to use your fits to make predictions for the future, and in particular to critically think about the validity of such predictions. This last point is important since it has been shown over and over again that just using data is a poor predictor of the future – insight, critical thinking, and well reasoned inferences are equally important in projecting data into the future.