P318 Computational Modeling
Spring 2014
Week 6 Homework
Due: Monday February 24th
(this is a longer assignment, over two weeks, worth double the points)

This homework assignment asks you to fit various special cases of the similarity-choice model to experimental data from an identification-confusion experiment. The main aim of this homework assignment is to use what you learned about quantitatively contrasting what are called “nested models.”

This problem starts with the code from last week’s assignment. You can either use the code you wrote or the solution uploaded to the web site.

Whatever you do, you need to make sure you send me ALL of the files needed to run your code. In other words, even if you use code that I uploaded, you still need to send me all the code I need to run your program. Consider compressing the files together into a ZIP file.

This assignment has several files online that you can use (homework6.zip).

For each of the models, find the maximum-likelihood parameter estimates. Do this by minimizing the negative log likelihood – recall that minimizing the negative of a fit measure is the same as maximizing the fit measure, and that minimizing the minus log likelihood gives you the same parameters as minimizing the minus likelihood.

In addition, report the sum of squared error (SSE) and the % variance accounted for. You don’t need to maximize the fit using these other measures, just report these measures after you maximize lnL.

While this assignment has many parts, some should involve a lot of copying and pasting, using my code if you wish, making a few changes here and there to the code to solve each problem.

I’ve given you the code that fits the basic similarity-choice model to the observed data by maximizing lnL. This is the same code you used in the last homework assignment, but uses lnL rather than SSE as the measure of fit. The code reports the maximum likelihood parameters found by using the Hooke and Jeeves hill-climbing algorithm. It also reports the fit values (lnL, SSE, and %Var).

- First, give the fit for the *saturated* perfect-fitting model (lnL, SSE, and %Var). This gives the upper bound on how well any model could fit the observed data. How many free parameters are there in the perfect-fitting models? Create an appropriately labeled plot of observed versus predicted identification frequencies. Note that you don’t need to do hill-climbing to solve this part.

- Does the full similarity-choice model fit significantly worse than the saturated model?

- Give the fit for the null model (lnL, SSE, and %Var). This does not really give the lower bound on how poorly any model could fit the data, but no candidate model would ever fit worse than this. How many free parameters are there in the null model? Create an appropriately labeled plot of observed versus predicted identification frequencies. You don't need to do hill-climbing to solve this part either.

- Now, fit the MDS-choice model to the observed data. Recall that in this model, rather than having free similarity parameters, similarity is defined as an exponentially-decreasing function of distances between points. The points represent the locations of each object in psychological space. For this example, assume an unconstrained two-dimensional space. That is, each stimulus is represented by an (x,y) position in that space.

  $$\eta_{ij} = \exp\left[-c \left(\sum_m \left| i_m - j_m \right|^r \right)^{1/r} \right]$$

  where i_m is the value of object i on dimension m and j_m is the value of object j on dimension m. You can set c equal to 1 without loss of generality. Assume r=1. Since we are estimating the i_m and j_m parameters, we do not include a w_m parameter. Convince yourself that we don't because w_m would be non-identifiable in this situation.

  Report the maximum likelihood parameters found by using the Hooke and Jeeves hill-climbing algorithm. Report the fit values (lnL, SSE, and %Var). How many free parameters are there? Does the MDS-choice model fit significantly worse than the similarity-choice model? Create an appropriately labeled plot of observed versus predicted identification frequencies.

- EXTRA CREDIT : I would like you to consider some kind of special case of the MDS-choice model that can significantly reduce the number of free parameters. Describe the model and justify your selection. Report the maximum likelihood parameters found by using the Hooke and Jeeves hill-climbing algorithm. Report the fit values (lnL, SSE, and %Var). How many free parameters are there? Does this special case fit significantly worse than the similarity-choice model? Create an appropriately labeled plot of observed versus predicted identification frequencies.

Feel free to work with other people on this assignment, at least so long as each person writes their own code by themselves, turns in their own solution, and contributes equally to the effort.