Tutorial 3: Categorization

Please work through this tutorial sheet on your own time as much as possible before arriving in tutorial. We encourage you to work together and discuss your methods and solutions.

1 Three Models of Categorization

Download the following file, which contains the models that you will work with in this tutorial (it is linked from the Assignments page):

http://www.inf.ed.ac.uk/teaching/courses/ccs/tutorials/ccs_t03.tar.gz

Unpack this archive, which will result in three Matlab source files:

- DEMlnL.m implements the Generalized Context Model (GCM) and its deterministic version, the Deterministic Exemplar Model (DEM);
- GRTlnL.m implements the General Response Theory (GRT) model;
- catModels.m is the main program that runs the model on the data from Rouder and Ratcliff (2004) and graphs and quantifies model fit;
- fminsearchbnd.m, hessian.m, infoCriteria.m are auxiliary functions for performing bounded optimization, computing the Hessian matrix (to get standard errors on parameter estimates), and for computing the AIC and BIC.

The code is taken from Lewandowsky and Farrell (2011, ch. 7) and explained there in more details. Please read this chapter if you need help understanding the code.

Now run catModel.m, which will generate a plots of the model fits for the three models: GCM, GRT, and DEM. Each of the graphs contains the data of the six participants in blue (each participant has been fit separately). Model fit is graphed in green.

Question 1: Check that you are able to recreate the graphs given in the textbook (p. 276–278; also presented in the lecture).

Question 2: Based on the graphs, determine which of the three models provides the best overall fit to the data. Now compare the models separately for each of the six participants. Are there participants which all models fit well? Are there participants which the best overall model doesn’t fit well?

The main loop in catModel.m (lines 28–48) iterates over participants and computes the maximum likelihood parameter estimates for each participant for each model.

Question 3: Compare lines 30 and 36, which compute the parameters for GCM and DEM, respectively. Why are they both calling the function DEMlnL?

Now consider GRTlnL.m, which implements the function GRTlnL(theta, x, data, N). This function computes the predictions and the log-likelihood of the GRT model for the parameter vector theta and the data vector data.

Question 4: What is the role of the other two arguments of the function, x and N. Why are they required to compute the model predictions and the log likelihood?

Question 5: Explain the function of lines 19 and 20 GRTlnL.m and relate them to the mathematical definition of the model. What’s the role of line 23?
2 Model Comparison

In the lecture (and in ch. 6 of the textbook) to criteria for comparing models were introduced: Akaike’s Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

**Question 6:** The AIC and BIC trade of model fit against simplicity. Explain how this is formalized in each of the criteria.

The code in `catModels.m` not only produced graphs, but also computes tables with parameter values, standard errors, and negative log-likelihood values for the three models. These are stored in the structures `GCM`, `GRT`, and `DEM`. Furthermore, the code computes the AIC values and difference in AIC (in `AIC` and `AICd`) and the BIC values and differences in BIC (in `BIC` and `BICd`).

**Question 7:** Plot the negative log-likelihood values for the three models. How many parameters do each of these models have? Which model performs best based on negative log-likelihood? Then plot the AIC and BIC values for the three models. How does the pattern change?

3 Generalization

Measuring model fit on the training data is only one aspect of model evaluation. Normally, we are also interested in determining how well the model generalizes to new participants or experimental conditions. In the lecture, a number of ways of assessing the generalization ability of model were discussed.

**Question 8:** Describe how held-out data can be used to test whether a model generalizes. Explain how crossvalidation extends this approach. What is leave-one-out evaluation?

We will now investigate how the generalization ability of the model we saw in Section 2, the GCM, GRT, and DEM. To achieve this, we will split the experimental data into two parts: the first three participants will serve as training data, the other three participants will be used as test data. We train the models on the training data, and then compute the mean of the maximum likelihood parameters over the three participants in the training data. We then use these mean parameters derive model predictions and evaluate these predictions on the test data.

**Question 9:** Change `catModels.m` to implement the procedure described in the previous paragraph. Which AIC and BIC values do you get for the three participants on the test data?

References
