Cognitive Modelling

Homework 4 – Modelling of face perception

This homework combines several of the topics that we have encountered in this course. For this homework, you will run a simple experiment in groups, collect and analyze data, and build a model.

This homework is inspired by the paper:

Fagertun, Jens, Tobias Andersen, and Rasmus Reinhold Paulsen. "Gender recognition using cognitive modeling." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012,

which is accessible on Inside.

The experiment in brief

During the experiment, the participant looks at a sequence of photos of faces, and for each they have to score them as Yes of No according to a binary variable that will be chosen (e.g., is the face smiling?). Several participants look at the same faces, and their binary answers are converted to a continuous scale combining the answer with the reaction time. These scores are the dependent variable in a linear model where the images themselves (or rather, their Principal Components' scores) are the predictors.

The homework step-by-step

Below you have a "recipe" of what needs to be done to complete this experiment. What you do and learn in the process will be the basis of your report. The steps give you some guidance, but you'll have to think and find the right way to proceed.

1. Setup the experiment
    a. Choose a face database to use as stimuli. You can download one of
        i. https://susanqq.github.io/UTKFace/ ("Aligned&Cropped" version)
        ii. https://kdef.se/index.html (requires registration)
        iii. https://fei.edu.br/~cet/facedatabase.html ("Frontal Images" version)
    b. Choose the variable to model
        i. Happy vs. not happy
        ii. Angry vs not angry
        iii. Masculine vs feminine looking
    c. Select your stimuli. Choose a subset of the stimuli in the dataset, at least 500 of them. For this part, you will have to manually review the files and/or use the tags set by the creators of the database (normally contained in the filenames). Ideally you want images that look similar (exclude those that have very different face position, lighting conditions, …) and have a roughly comparable number of stimuli in each of the categories for the variable you chose, with a variety of "intensity" levels. Place the chosen images in a folder on the computer where you will run the experiment.

    d. Download the provided Matlab code for stimulus presentation (in Inside, under Homework 4). Try to understand what it does, and set the parameters to the correct values for your computer.
    e. Run some pilot trials. Run the script and try to collect some data. Have a look to the results file (results.txt) to check that everything makes sense.
2. Run the experiment
    a. It's time to collect some data! Your group members are your test persons. Start the stimulus presentation script and let the test person score all the stimuli. You are welcome to collect data from more people.
    b. Remember to instruct the test person to answer without thinking too much. We need the reaction time, and if they spend a long time pondering about each picture, we cannot use it. It's better to make some mistakes than spending a long time deciding. Aim for below one second per answer.
    c. The answers from each test person, together with the response times, are automatically saved in the file "results.txt", as tab-separated values where the columns are:
       `subject_id, trial_number, stimulus_filename, answer, reaction_time`
3. Data pre-processing
    a. Load the data in your analysis environment of choice (Matlab, Python, …)
    b. Consider removing outliers for the reaction time (above several seconds, below 200ms).
    c. Convert the binary answers and the reaction times to a continuous scale.
    d. By aggregating across test persons, you will get a score for each face photo.
4. PCA and feature selection
    a. Run PCA on the images. The scores (i.e., the representation of the images in PCA space) will be the predictors in the model.
    b. Inspect the first few PCA components. What do they model? How much variance do they encode?
5. Select a subset of relevant components. They will be the predictors in your model. Not all the components model the chosen variable. Adopt a technique to select features. You can choose one of the following options:
    a. Matlab's sequentialfs. If you choose this one, you need to write a paragraph in the report explaining what it does.
    a. Homemade stepwise selection with cross validation.
6. Linear model
    a. Build a linear regression model that predicts the scores based on the PCA scores.
7. Generate samples from the model.
    a. The linear model expresses the scores on the continuous scale y as
       $y = Ax + b$
       where A is the vector of the coefficients for the predictors (components) and b is the intercept.
       How can you use the model to generate a synthetic face picture that corresponds to a given score?
    b. Use your model to generate synthetic face pictures with a given score.

Homework hand-in

The hand-in is a report about the experiment. You can structure it according to the points below.

1. Introduction to the experiment
2. Description of the dataset
3. Modelled variable
4. Execution of the experiment
    a. Describe also how many test persons you tested and how many stimuli were shown.
5. Description of the collected data
6. Processing of the data to obtain a continuous intensity scale. Describe how you combined the binary answers and the reaction time in order to obtain a continuous scale.
7. Principal Component Analysis
    a. Describe how PCA is applied
    b. Display the first few components and describe what they are modelling
8. Feature selection. Describe the method you used to select the components that are used as predictors.
9. Modelling
    a. Build a linear model that has the transformed answers (the continuous scale) as dependent variable and the chosen components as independent variable.
    b. Describe the statistics of your model
10. Use the model to generate faces. Display synthetic faces generated at different values along the scale.
    a. Moving along the continuous scale, choose 5-10 values and generate the corresponding faces, and represent them. These should look like different intensity values of your chosen variable.

Time plan

This homework has several parts, so make sure that you are on schedule. The hand-in deadline is Sunday December 1st.

A suggested time plan is:

- November 14 (in class): set up and run the experiment. At the end of the day you should have finished data collection.
- November 15-November 20: pre-process data, think of how you can produce a continuous scale, try running PCA on the images
- November 21: in class, we'll discuss the modelling part of the project (PCA and the linear model) and you'll have time to refine your model.
- November 22-November 27: finalize the model, work on the report
- November 28: work on the homework in class, ask questions. Finalize the work.