2.2.1 Overview

In this exercise, you will run two simple experiments. You will also create and validate a linear encoding model.

In Experiment 1, photos of faces are presented to the test participants. The participants must rate each face on a scale to indicate a facial feature. The facial feature can be the emotional expression of the face (e.g. happy, sad or angry) or it could be the perceived gender of the face. It could also be some other feature that select.

The participants' ratings are used to create a linear model in which the participants' rating is the dependent variable and the images, or rather, their principal component scores, are the predictors. You will use the linear model to create synthetic images of faces that correspond to given ratings. Some of the methods you will use are described in more detail in the Introduction to Machine Learning book by Herlau, Schmidt and Mørup [HMS21].

In Experiment 2 synthetic images of faces are presented to the test participants. The participants must rate each face on a scale similar to the one used in Experiment 1. The data from Experiment 2 are analysed using the psychometric function and ROC curves. The purpose of Experiment 2 is to validate the linear model experimentally.

This exercise is inspired by the papers by Fagertun et al. [FAP12; Fag+13]. You can find the papers on DTU Findit.

2.2.2 Instructions

Follow the instructions below to complete this exercise. Note that there are questions in bold in the instructions. Your report should consist of the answers to these questions. Please hand in your answers to the exercise as a pdf file. If you like, you can also submit your code and data (the images you used and the ratings).

1. Obtain photos of faces for your experiment. First, choose a database from which you will obtain the photos of face that you will use as stimuli. Also, decide on the facial feature that you want to use. It can be an emotional expression, perceived gender or some other feature that you would like to use.

• 'Aligned & Cropped' images from UTKFace (https://susanqq.github.io/UTKFace/)

• Karolinska Directed Emotional Faces (KDEF) (https://kdef.se/)

• FEI Face Database (https://fei.edu.br/ cet/facedatabase.html)

Then select the images that you want to use from the database you have selected. Choose a subset of at least 200 images from the database. For this part, you will have to manually review the files and/or use the tags set by the creators of the database (often contained in the filenames). Select the images based on the following two criteria

• The images should vary across the facial feature that you want to use. For instance, if the participants are rating the images on a scale ranging from happy to angry then you shouldmake sure to include an approximately equal number of faces that appear happy, neutral and angry.

• The images should not vary much in other features, such as the orientation of the face, lighting conditions and the age of the person in the photo.

Consider if you want to reduce the dimensionality of the images using cropping, down-sampling and gray-scaling.

What data base did you use and why? What facial feature did you select for Experiment 1. Describe in detail how you selected the images based on the feature of interest and on features not of interest.

2. Conduct Experiment 1

• Write a script that presents images and accept ratings from an observer. Importantly, the program must save the ratings and information relating each rating to image that was rated.

• Collect the data. Start the script and have the test person rate the stimuli. Your group members can be your test participants. You can also ask students from other groups, Friends or family to participate.

Describe the experiment in detail. How many test participants did you include? How many images did they rate? Did they all rate the same images?

3. Pool the (normalised) rating data

• Show a histogram of the ratings for each participant

• Do all your participants use the full range of the rating scale? In case they do not then you can normalise the ratings for each participant using min-max normalisation. Did you min-max normalise and why (not)?

4. PCA and dimension reduction

• Subtract the average image from all the images. Do not standardise (divide by the standard deviation).

• Run PCA on the images (with the average image subtracted). The scores (i.e., the representation of the images in PCA space) will be the predictors in the linear encoding model.

• Visually present the first few PCA components as images. Visualise a component by reconstructing the images that has the max/min score for that component. Compare these to the mean image. You can also create synthetic image using scores that lie somewhere between the max/min score. In this way you can visualise the effect that one principal component has on the images. Show figures of your visualisation. What image features do the PCs represent?

• Show a bar plot of the variance explained for all the PCs. Use it to select which PCs you want to include in your model. You should reduce the dimensionality significantly. How did you decide?

5. Build a linear regression model that predicts the (normalised) ratings from the selected PC scores.

• Use forward selection to select the relevant PCs for the model. You can use a routine from a toolbox or you can setup a homemade stepwise selection with cross validation. Some examples of routines from toolbox are

– sklearn.feature_selection.SequentialFeatureSelector (Python)

– sequentialfs (Matlab)

Explain how your feature selection method works (your own or a routine from a toolbox). Visualise the PCs that were selected in the same way as you did above.Do the selected PCs represent facial features relevant in the rating task?

• Finally you should fit the linear model using only the PCs that you selected using forward selection. What are the parameter values for this fit?

6. Generate synthetic images from the model

• Use your model fit to generate synthetic face images with a given rating using the method described in Section 2.1. You should generate one synthetic image for each possible rating (e.g. 7 images if your rating scale was 1-7). Also generate synthetic images that extends the rating scale in your data set (at least one image in each end of the range). Explain in detail how you generated the synthetic faces from the model. Show the synthetic images. Discuss if the images appear as expected and, if not, what the reason could be.

7. Conduct Experiment 2.

• Modify the script you used to present images in Experiment 1, so that it can present all the synthetic images that you have created at least 20 times in random order. Use the modified script to run the experiment. You can use the same test participants or some other participants. Describe the experiment in detail like you did for Experiment 1.

• Make a table with the response counts or response proportions for each test participant.

8. Analyse the data from Experiment 2.

• Analyse your results using ROC curves. Choose a baseline stimulus and fit and plot ROC curves for all the other stimuli with respect to the baseline stimulus. Also plot the data in the same plot. Do this for each test participant. Do the ROC curves appear like you would expect? Explain your reasoning.

• Analyse your results using psychometric functions. Fit and plot psychometric functions for each response criterion. Also plot the data in the same plot. Do the psychometric functions appear like you would expect? Explain your reasoning.