Lab assignment: Machine learning with representational similarity analysis.

By Ben Harvey

Krista Overvliet (k.e.overvliet@uu.nl) & Martijn van Ackooij (m.vanackooij@uu.nl).
December 4th and 11th.

Representational similarity analysis (RSA) is a data analysis method with broad applications to neuroscience. It is increasingly popular largely because it gives researchers the ability to compare patterns of responses across several types of neuroscience data: fMRI recordings, recordings from large numbers of individual neurons, behavioural responses, and the computational model unit responses.

The objective of this assignment is to implement a RSA and compare patterns of representational similarity across different types of neuroscience data. RSA is a good choice here because it is moderately complex and widely applicable to many types of data.

This assignment contains some explanation of representational similarity analysis and some tips to help you implement one. We expect you to use R, but this analysis is often implemented in Matlab, so you are also welcome to use Matlab.

The following text contains some explanation, some questions, and some instructions. Questions are clearly labelled and numbered: each group should submit a document answering these questions (with numbered answers), and it may help to include some of the images you make. Instructions are underlined to clarify what we want you need to actually do, as this is often mixed in with the explanation. The number of points available for each questions gives an idea of the expected depth of your answers.

Remote group work: A personal, interactive process
You should work in groups of 4 on these exercises, with help from a single teacher who is assigned to your group. We expect you to work on these exercises in class time so you can work with your group and teacher.

We suggest all group members doing these exercises on your individual computers simultaneously: this improves (student) learning and also makes it easier to find mistakes. Don't rely on other group members' answers if you don't understand why they are correct: this is meant to be an interactive collaboration with your group, so ask your group members to explain. If your group gets stuck on a question or different group members can't agree on an answer, ask for help from your teacher. Please share your video if bandwidth allows, at least when talking with your teacher.

When your group is happy with your answer, work together to finalise your answer in a document shared with the whole group. Google docs is an excellent platform for working together on a shared document. Show these answers to your teacher as you

work. You can share this document with the teacher too. It's often best to show your teacher answers for a few questions together, but you should always check your answers with your teacher after each few questions (so they can see you have completed them successfully). Your teacher will grade you as you work to monitor your progress and address problems. But we need a record of all your answers, submitted at the end of the assignment. Your teacher should be checking your final answers in this document as you work.

Many questions begin 'Discuss with your group, then describe to your teacher…'. In questions like this, it is generally best to start by asking every group member's opinion. Then work on a written answer together. Then explain your answer to your teacher. You can also ask your teacher to read what you wrote, but they will often ask questions. You may like to update this answer after talking with your teacher, but please tell your teacher what changes you made next time you talk.

Many questions build on previous questions being completed correctly, so you should be confident of your answer before using it in further questions: ask for help if you are unsure. But there are various points clearly marked (NEW TOPIC) where you do not rely on previous answers. If you get stuck and can't get help immediately, you can move on to the next topic until your teacher can help.

Starting data

Here we will start with the simulated response amplitudes of 100 fMRI recording sites in human inferior temporal cortex (columns) in response to 92 images (rows). This simulated data follows patterns of responses common in human inferior temporal cortex. This data is stored in the file 'NeuralResponses'. Load this text table into R using read.table.

This simulated data lacks the noise that is common in real experimental data, which is important if we want to understand how experimental data is used.
First, we will simulate the type of responses we would actually expect to see for a group of 12 subjects in an experiment.
To simulate data for each subject, add normally-distributed noise to the starting data (using rnorm), with a mean of zero and a standard deviation of one. Repeat this 12 times to simulate 12 'subjects'. Keep the original data and the data from each of the 12 'subjects'.

The basic RSA

Do questions 1-3 (below) and show your teacher all of the results together.
For questions where you generate a plot, table, or number, you should copy this in your answer document (maybe using a screenshot).

There are various ways to quantify dissimilarity. In class I described using a Euclidian distance and 1-correlation, which both have advantages and disadvantages. Here, you should use 1-correlation, where 'correlation' is Pearson's correlation coefficient (confusingly called 'r').

The main difference between 1-correlation and Euclidian distance is that responses to one object may be larger overall than responses to another object (i.e. their mean responses may differ). The 1-correlation measure is not affected by these differences in mean response, while they can have a large effect on Euclidian distance. Discuss with your group, then describe to your teacher, some reasons why each method may be more suitable for different data types or research questions. In other words, in which situations is it best to ignore differences in mean response, and when are they important to consider? Give examples of such situations. (Question 1, 4 points)

Now calculate the representational dissimilarity of each object pair, constructing a representational dissimilarity matrix (RDM). Take the responses of each object (across all recording sites) and calculate the dissimilarity from the responses to each other object. You will need to repeat this for every possible pair to produce an RDM. You can calculate Pearson's correlation coefficient for a single pair using the function 'cor', though you may be able to find a more efficient approach to compare all pairs of recordings. Remember that the RDM should contain measures of **dis**similarity (i.e. 1-correlation) rather than similarity (i.e. correlation).

After you have calculated RDMs for each of the 12 'subjects' and for the original data, make an image of the RDM for a single example subject and for the original data. Include a legend to show the scale. Discuss with your group, then describe to your teacher: how do these two RDMs differ? (Question 2, 5 points)

Researchers often group data across many separate measurements, for example separate experimental subjects, to reduce the noise level in the data. This is also possible with RDMs. Average together the RDMs from all 12 'subjects' and compare these averaged RDMs to the RDMs from a single example subject and from the original data. Discuss with your group, then describe to your teacher: How do these differ, and why? (Question 3, 2 points). Keep the resulting 'average subject' RDM for later.

Hypothesis testing
NEW TOPIC

Do questions 4-10 (below) and show your teacher all of the results together.

The displayed images have been categorized and ordered by the type of object they contain. 'CategoryVectors' is a table containing binary classifications of the images into various categories. 'CategoryLabels' contains descriptive names for each category. For example, objects 1-48 are animate while 49-92 are inanimate, so element 1 of 'CategoryLabels' is "animate" and column 1 of 'CategoryVectors' has the value 1 (true) for objects 1-48 and value zero (false) for objects 49-92. Load both tables.

We can use then use each column of CategoryVectors to construct RDMs representing specific hypotheses about the data. Each of these RDMs gives the pattern of dissimilarity predicted by a hypothesised response to the category

described by the column. For example, the RDM from column 1 (animate/inanimate) predicts high dissimilarity of responses when one object is animate and the other is inanimate, but low dissimilarity when both objects are animate or both are inanimate.

Note here that the value in this RDM reflects whether the pair have the same animacy state, NOT whether the pair are both animate: if both objects are inanimate, they share an animacy state and should have a similar representation. This is a very important principle of RSA: similarity reflects *the similarity of* the stimulus category, not the stimulus category itself.

But are objects with similar animacy represented more similarly than objects with different animacy?

For column 1 of CategoryVectors (i.e. labels for animate objects), compare each pair of values, filling the RDM with '0' where and object pair has the same state of animacy and '1' where the two objects have a different state of animacy. This forms a mask you can now use to select different groups of pairs. Now, use a two-sided unpaired t-test to compare the dissimilarity of pairs with the same animacy against the dissimilarity of pairs with different animacy. Remember to exclude pairs where an object is compared to itself. Also remember that each dissimilarity measure is included twice in the RDM, but should only be included once in the t-test.

What values of the t-statistic and probability (p) do you find for an effect of animacy on the **original data**'s pattern of representational dissimilarity? (Question 4, 1 point)

Compare the t and p values you find for an example **individual subject** and for the '**average subject**' RDM. Write these values in your answers document. Discuss with your group, then describe to your teacher: how do they differ, and why? (Question 5, 3 points)

Now let's look at whether responses to faces differ from responses to other animate objects. Column 6 of the CategoryVectors shows whether each object was a face or not. Following the same principles, make an RDM that predicts the pattern of dissimilarity predicted by a hypothesised response to faces. Then test whether the 'face-ness' of objects affects the **original data**'s pattern of representational dissimilarity. What values do you find for t and p? (Question 6, 1 point)

Because we have already determined that animacy has a highly significant effect, we may expect that face-ness may only have an effect among animate objects: inanimate objects cannot be faces, and the difference between a face and an inanimate object is likely to follow the difference in animacy. Now repeat this test for effects of face-ness on response similarity using only the part of the RDM where animate objects are compared. What values do you find for t and p? Discuss with your group, then describe to your teacher: how do these compare to the values over the whole RDM, and why? (Question 7, 3 points)

Repeat this procedure to test for an effect of whether an object was human (column 3 of the CategoryVectors). Does humanity have an effect on the similarity of

response patterns when all data are objects are included in the comparison? Is this also true when only animate objects are included? What t and p values do you find for this effect of humanity in each case? Discuss with your group, then describe to your teacher how you interpret these answers. (Question 8, 2 points)

So it seems that both animacy and face-ness affect the patterns of responses. As these effects of animacy and face-ness seem separate, we might think of a response model where animacy and face-ness both have effects, with face effects only occurring among animate objects. We can use the predicted RDMs from Question 4 (for animacy) and Question 7 (for face-ness among animate-animate object pairs: the dissimilarity of all inanimate-inanimate and animate-inanimate pairs can be set to 0 here), assuming that both have some contribution.

Use these two predictors separately to make either an ANOVA or a general linear model to test the hypothesis that both animacy and face-ness both have effects. In this combined model, do animacy and face-ness both have effects, and what statistical values and probabilities are found for each? (Question 9, 3 points)

What are the relative effect sizes for these two effects? (Question 10, 2 points). These effect sizes are 'eta' values for an ANOVA or 'beta' values for a general linear model, though you are welcome to take another approach to determine effect sizes.

Comparisons between different types of data
NEW TOPIC

Do questions 11-15 (below) and show your teacher all of the results together.

So far we have used fMRI response amplitudes to construct and test RDMs. However, the same approach can be applied to recordings from a group of single neurons. The table 'NeuroRDM' contains an RDM constructed from responses to the same images in a group of individual neurons in the inferior temporal cortex of the macaque monkey, a common animal model of human visual object processing. Load this table.

Determine the correlation between this macaque neuron data RDM and the '**average subject**' RDM we made earlier, using Pearson's correlation. Are these RDMs significantly correlated? What correlation coefficient (r) and probability (p) do you find? (Question 11, 2 points). Remember to exclude pairs where an object is compared to itself. Also remember that each pair's dissimilarity measure is included twice in each RDM, but should only be included once in the correlation.

Now make a scatter plot of the dissimilarity values from the 'average subject' RDM against the dissimilarity values from the macaque neuron data RDM. Note that these form two major clusters, which correspond to objects with the same animacy state (low dissimilarity) and different animacy states (high dissimilarity). See also that each of these clusters also has some structure within it.

Now take the part of both RDMs where all pairs are both animate (i.e. the upper left quarter of the RDM). Following the same principle as Question 11, determine whether the patterns of representational dissimilarity among these animate objects are still significantly correlated. What does this tell us? (Question 12, 3 points).

And take the part of both RDMs where the pairs are both INANIMATE. Are the patterns of dissimilarity significantly correlated here? Discuss with your group, then describe to your teacher: Why might this be different from the result seen in the previous question? (Question 13, 2 points).

The same approach can also be applied to behavioural similarity ratings. The table 'BehaviourRDM' contains an RDM constructed from responses of human observers asked to arrange the objects by their similarity. Load this table and make an image of this RDM. You may see that some patterns are like those in the neural data, while some are not found in the neural data. This suggests that humans use other information, together with neural responses from inferior temporal cortex, when judging the similarity between objects.

Nevertheless, at least some of our behaviour seems to be predicted by these neural responses. How well does the behavioural RDM correlate to our '**average subject**' RDM? How about when using only animate objects? And when using only inanimate objects? Discuss with your group, then describe to your teacher: What does this tell us, and can you think of a possible explanation for this result? (Question 14, 3 points)

Finally, we can compare our neural data to the responses of a set of computational model units in a particular model of visual object processing, called HMAX. You can find a summary of how HMAX works at http://maxlab.neuro.georgetown.edu/hmax.html.
The table 'HmaxRDM' contains an RDM constructed from the responses of HMAX model units to the same images. Load this table and take a look at this RDM, noting what is similar and what is different from the 'average subject' RDM. How well does the HMAX RDM correlate to our 'average subject' RDM? How about when using only animate objects? And when using only inanimate objects? Discuss with your group, then describe to your teacher: What does this tell us about the limitations of HMAX as a model of human object processing? (Question 15, 3 points)