

Using neural decoding to understand how the brain processes information

A central goal of Neuroscience research is to understand how neural activity enables animals to perform complex behaviors. To study this, neuroscientists make different types of neural recordings, including functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) recordings from humans, and recording action potential of individual neurons. Once this data is recorded, neuroscientists then need to apply a data analyses to turn the recorded signals into information that can lead to an understanding of how the brain is processing information.

One method that has been shown to be very useful for interpreting neural activity is called *neural decoding* (the method is also referred to as *multi-voxel pattern analysis* by the fMRI community). The purpose of this tutorial is to explain what neural decoding is and how to run a decoding analysis in MATLAB.

An example experiment

To illustrate how neural decoding works, we will use data from an experiment run by Ying Zhang and Bob Desimone at MIT [1], which will be referred to as the Zhang-Desimone experiment. On each trial in this experiment, a monkey fixated on a dot for 500 ms, and then one of seven possible images was shown to the monkey for 500 ms (figure 1). While the monkeys viewed these images, Ying recorded the spiking activity from 132 neurons in the inferior temporal cortex (IT) which is a brain region thought to be involved with object recognition. Each experimental trial type was repeated ~60 times to collect a data set that had spiking activity from 420 trials (7 trials * 60 repetitions = 420 trials total). (The actual experiment was a bit more complex and consisted of showing the images at different locations and also an attentional manipulation but we will ignore this for now).

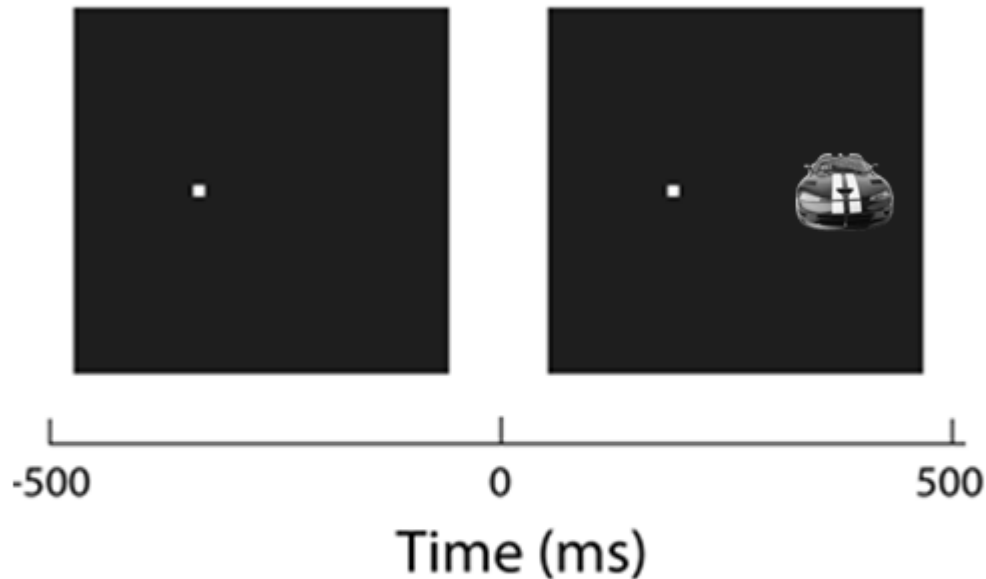
A**B**

Figure 1. An example experiment run by Ying Zhang and Bob Desimone at MIT. **A**, An example of an experimental trial where the monkey fixated for 500 ms and then was shown a stimulus for 500 ms. **B**, the seven different stimuli that could be shown to the monkey. Each stimulus was shown 60 times (20 times at each location) while the spiking activity of 132 neurons was recorded.

What is neural decoding?

Neural decoding is a data analysis method where a computer program, called a *pattern classifier*, uses neural activity to predict what experimental conditions are present on a particular experimental trial. The pattern classifier can predict what experimental conditions are present based on neural activity at a level that is above chance, this means that there is a reliable relationship between the neural activity and different experimental conditions, or put another way, there is information about the experimental conditions in the neural activity.

Figure 2 illustrates how this process works on one trial from the Zhang-Desimone experiment. On the trial shown in figure 2, a monkey was shown an image of a kiwi (Fig. 2A) while 132 neurons were recorded (Fig. 2B). The activity of these neurons was then fed into a pattern classifier (Fig. 2C) and the classifier predicted that a kiwi was present on this trial (Fig. 2D). Since a kiwi was in fact shown on this trial, the prediction made by the classifier is marked as being correct (Fig. 2E). This process is then repeated for many different trials, and the percentage of correct predictions is reported. If the pattern classifier was merely guessing which image was shown on each trial, then we would expect correct predictions to be made on approximately 1 out of every 7 trials, since there are 7 different possible images the classifier has to choose between. If the prediction accuracy is well above 1/7, then the classifier is unlikely to be guessing, which means there is a reliable relationship between the stimuli and the neural activity.

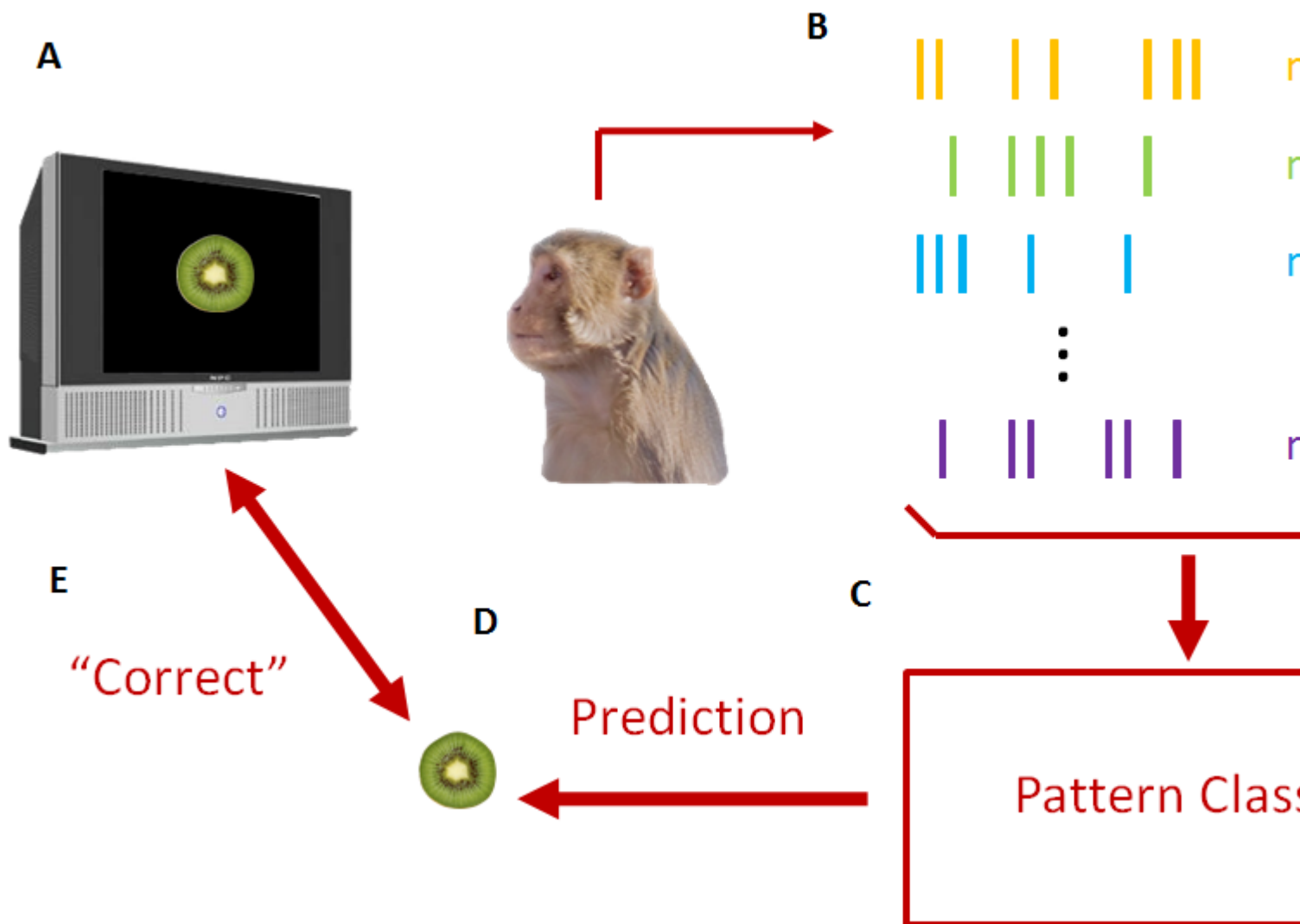


Figure 2. An example showing a decoding analysis for one trial from the Zhang-Desimone experiment. **A**, On this trial, the monkey is shown an image of a kiwi. **B**, While the monkey views the image, the spiking activity of 132 neurons is recorded. **C**, This neural activity is fed into a pattern classifier. **D**, The pattern classifier then makes a prediction about what image it thinks is present based on the neural activity. In this example, the classifier predicted the image was of a kiwi. **E**, The prediction of the classifier is then compared to the image that was actually shown to the monkey. In this example, the classifier's prediction was correct, so we make it a correct prediction. This process (steps A to E) is repeated over many trials, and the average prediction accuracy is reported as the final statistic from this analysis. If the average prediction accuracy is well above chance, then it means there is a reliable relationship between the neural activity and the different images, or in other words, the neural activity has information about the stimuli [2].

Training the classifier to make predictions

In order for the pattern classifier to be able to make correct predictions, the classifier must be able to *learn* the relationship between different patterns of neural activity and different experimental conditions. To do this, the neuron recordings from different trials are divided into two non-overlapping sets of data. One set of data is called the *training set* and is used by the pattern classifier to learn the relationship between the neural recordings and experimental conditions. The second set of data is called the *test set*, and the classifier must take this test set and try to correctly predict which experimental

conditions were present on all of the trials in this set. Again, if the classifier can make correct predictions on the test set, it means there is a reliable relationship between the neural activity and the different experimental conditions. In a typical decoding analysis, the process of dividing the data into different training and test sets is repeated many times, and the final decoding accuracy is averaged over many different training and test splits.

Downloading the Neural Decoding Toolbox (NDT) to run a decoding analysis

In order to illustrate how to run a decoding analyses on the Zhang-Desimone data, we will use MATLAB and the Neural Decoding Toolbox (NDT). To get the toolbox, go to <http://www.readout.info/wp-login.php?action=register> and register for an account. Once you have an account, you should log into the www.readout.info website. You can then download the toolbox from <http://www.readout.info/downloads/download-the-toolbox/> and you can download the data from Zhang-Desimone experiment from <http://www.readout.info/downloads/datasets/zhang-desimone-7-object-dataset/> .

Once these files are downloaded, unzip them and place the two folders in a directory where you would like to do your analyses. Then open MATLAB, and set your MATLAB current folder to be the folder that contains the NDT and the Zhang-Desimone data.

How decoding works

cross-validation, etc.

Basic decoding analyses

Explain Zhang 7 object experiment and how to analyze data from it

Testing position invariance

Additional analyses

References

[1] Zhang Y, Meyers E, Bichot N, Serre T, Poggio T, and Desimone R (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences*, 108:8850-8855.

[2] Quiñones-Curioga R, and Panzerri S (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173-85.

[3] Meyers E (2013). The Neural Decoding Toolbox. *Frontiers in Neuroinformatics*, 7:8.