"Machine learning for analyzing neuroimaging data from natural stimulus experiments"

Alex Huth. UC Berkeley. Estados Unidos.

Summary:

Natural stimuli provide the opportunity to examine many different hypotheses using the same dataset. In this course we will analyze two fMRI datasets, one collected while subjects watched natural videos, and another collected while subjects listened to natural stories. For each dataset multiple feature spaces can be extracted that represent different specific hypotheses about representations in the human cortex. We will look to machine learning methods for extracting useful feature space representations of the stimuli as well as fitting encoding models, which predict brain responses from stimulus features.

Program:

<u>Monday</u>

Assessing data quality: repeatability of responses to natural stimuli. One of the most important things in science is getting high quality data—garbage in, garbage out! This dat will start with a short lecture (~30 mins) about the various methods for computing repeatability, including simple (explainable variance, mean pairwise correlation), complex (coherence spectrum), and visualizations thereof. Students will then complete computational tutorials showing how to compute these metrics and create visualizations.

<u>Tuesday</u>

Constructing simple, categorical encoding models based on responses to natural stimuli. One of the easiest things to do with natural stimulus data is to create categorical models (e.g. is there a face in the scene or no?). We will explore these types of models and compare them (theoretically) to standard contrast-based fMRI experiments. This day will begin with a lecture (~45 mins), then students will complete a short tutorial, then there will be a discussion about the theoretical implications.

<u>Wednesday</u>

Linear regression methods for encoding models: ordinary least squares (OLS), ridge/tikhonov regression, and lasso. For simple models OLS is sufficient, but for complex models (with many features) OLS fails. This will begin with a longer lecture (~1 hour), then students will complete a short tutorial, and then there will be a discussion about the various regression methods and how they affect model interpretation, etc.

<u>Thursday</u>

Constructing complex high-dimensional encoding models. Here we will use the advanced regression methods discussed Wednesday to build high-dimensional encoding models of visual and language data. This will begin with a short lecture (~30 mins) followed by a longer tutorial in which the students will fit these models and investigate their properties.

<u>Friday</u>

Combining multiple feature spaces into single encoding models. Here students will extend what they learned Thursday to fitting models that combine multiple feature spaces in order to find which parts of the brain are well-explained by each model. This will begin with a short lecture (~30 mins) followed by a longer tutorial.