Some Possible Projects

Imagine your name here!

Overview

These slides give you some *starter* ideas for projects.

You are not restricted to just these!

Many variations are possible, or use your creativity!

Project topic: Encoding of facial movements in V1

Scientific context: It has been known since Niell and Stryker, 2010 that sensory areas are modulated by locomotion. Recently it has emerged that V1 and the brain as a whole also encodes finer details of an animal's facial movements. Why would that be?

Specific question: Can you determine ***which*** patterns of facial movement are encoded? That might contain hints about the function of these signals.

Data set: Stringer with spontaneous behaviors.

Techniques: For tracking fiducial points on the face: DeepLabCut/APT/LEAP/DeepPoseKit. For dimension reduction on spatiotemporal point cloud data: PCA, HMM, Gaussian Process Factor Analysis (GPFA). For regression prediction: reduced rank regression, ridge regression, deep networks.

Controls: Cross validation and establish baselines based on Stringer, Pachitariu et al, Science 2019.

Difficulty: Medium. Prereqs: Linear Regression, Deep neural networks, Dim Reduction. Cross-ref to course content: W1D3+5, W3D4-5. Proposed by: Marius Pachitariu, Kristin Branson

Project topic: Reinforcement learning for foraging

Scientific context: Animals forage for food, integrating evidence and taking actions. Can this be explained by reinforcement learning problem.

Specific question: How should behavior on a foraging task depend on uncertainty, cost, dynamics?

Data set: None — theory project.

Techniques: Use Reinforcement learning or control to find optimal behavior on simple foraging tasks described in behavioral neuroscience literature. Model tasks by a Markov Decision Process. Reproduce published plots for animal behavior, but for optimal foraging.

Controls: Compare to simpler heuristic strategy.

Difficulty: Medium. Prereqs: probability, RL. Cross-ref to course content: W2D4-5. Proposed by: Xaq Pitkow

Project topic: Decoding X from Y

Scientific context: Many scientific questions ask whether some brain signals Y contain information about some task variables X. Decoding provides a way to estimate that: if you can decode X from Y using some technique, then you know that information is present (though not necessarily used).

Specific question: Pick a variable of interest, like orientation, frequency, running… Then pick a type of brain signal, like firing rate, oscillation amplitude, time series… How are these quantities related?

Data set: Any.

Techniques: Many, both linear and nonlinear. Linear: regression, Support Vector Machines, Canonical Correlation Analysis, Granger Causality. Nonlinear: Directed Information, Neural Networks...

Controls: Cross-validate for held-out data. Causality is harder to establish.

Difficulty: Medium. Prereqs: basic or advanced Machine Learning tools. Cross-ref to course content: W1D4-5. Proposed by: Xaq

Project topic: Biological learning rules for supervised and unsupervised tasks

Scientific context: Gradient descent is now the algorithm-of-choice for training artificial neural networks. However, gradient descent is well known to be implausible biologically. How well do existing biologically-plausible learning rules fare compared to gradient descent on a wide range of tasks?

Specific question: How does the performance of biologically learning rules compare with gradient descent? Can networks with biologically-inspired learning rules (Hebb's rule, STDP, etc) learn "good" representations of sensory-driven neural population activity for decoding by downstream circuits?

Data set: Can be just a pure theory project with no data, or could use the Stringer dataset of 10k cortical neurons responding to visual stimuli.

Techniques: Pick your favorite supervised/unsupervised/reinforcement learning task that isn't too difficult - for example decoding the visual stimulus from Stringer neural population data. Compare the performance of gradient descent-based algorithms with biological learning rules proposed for similar tasks. Do biological learning rules even exist for your tasks?

Controls: Be able to reproduce the performance of published results when using gradient-descent.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Guangyu Robert Yang, Cian O'Donnell

Project topic: State sequences in population dynamics across cortical areas

Scientific context: Metastable brain dynamics are characterized by abrupt, jump-like transitions where single-trial neural activity unfolds as a sequence of discrete 'states', modeled as network attractors. In this project, students will acquire familiarity with the theory and practice of state sequences in models and data; they will perform network simulations and analyze single-trial population activity using hidden Markov models, comparing latent dynamics across cortical areas. [video]

Specific question: In which brain areas is activity well captured by hidden Markov models (HMM)? Can we predict an animal's behavior or presented stimuli from state sequences? How are state transitions correlated across brain areas - which areas lead/follow?

Data set: Steinmetz decisions. Stringer spontaneous.

Techniques: Matlab code for spiking network simulations and HMM fit from this repo. Python code from Scott Linderman's SSM repo for fitting generative models. Jupyter notebooks will be provided.

Controls: Spontaneous vs. stimulus evoked activity. Clustered vs. homogeneous networks.

Difficulty: Medium. Prereqs: Generative models, recurrent network modeling. Cross-ref to course content: W2D3. Proposed by: Luca Mazzucato

Project topic: Complexity of feedback loops

Scientific context: We see from the neural circuitry that lateral and top-down feedback play a significant role in vision. Evidence shows that feedback in higher-level areas such as V4, IT, or MT, with bigger and more complex receptive fields, can modify and shape V1 responses. We also know that feedback plays a critical role in behavioral regulation. Using a variety of representational frameworks, how can we characterize this complexity? Does an intelligent agent (or organism) require a minimal (or maximum) level of representational complexity to produce and act upon feedback?

Specific question: Track 1 - How can you model top-down feedback in an artificial neural network for vision? Track 2 - What is the tradeoff between feedback and performance given representational models at different levels of complexity?

Data set: Track 1- CIFAR-10, ImageNet. Track 2 -Mouse dataset (or EEG time-series) to train behavioral agent(s).

Techniques: Track 1- Recurrent neural networks; unsupervised learning; representation learning; autoencoders Track 2 - Bayesian model, Agent-based Model, or Reinforcement Learning.

Controls: Track 1 - Feed forward architectures (like traditional CNNs). Track 2 - No-representation feed-forward model of behavior.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Bradly Alicea, Edward Kim

Project topic: Understanding trained neural networks (RNNs and/or DNNs)

Scientific context: It is hypothesized that sensory neural representations are tuned to the statistics of the environment (eg. Barlow's redundancy reduction). Accordingly, properties of neural activity in V1 are explained by statistical properties of natural images (e.g. 1,2,3,4,5). In parallel work, deep CNNs accurately fit neural activity throughout visual cortex (e.g. 6,7). Understanding CNN hidden-layer statistics (e.g. 8,9) could bridge these fields and lead to new predictions.

Specific question: Are CNN hidden layer activations to natural images captured by existing models for V1-like filters? Which predictions of those models extend to higher visual cortex?

Data set: Images: 1 or 2. Pretrained CNN (eg. VGG16). MGSM toolbox.

Techniques: Compute CNN activations to images, compare statistics across layers to V1 literature (3). Fit Gaussian Scale Mixtures (GSM, MGSM), compare to Gaussian and Gaussian Mixtures. Formulate CNN-GSM predictions for mid-high visual cortex, compare to V1 literature (3,5).

Controls: Replace natural images with white noise and phase-scrambled images.

Difficulty: Medium. Prereqs: Probability, generative models, CNNs. Cross-ref to course content: W2D3. Proposed by: Srdjan Ostojic, Ruben Coen-Cagli

Project topic: Use CCA/RDM to compare brain areas

Scientific context: The brain consists of many different regions, that interact with one another to process sensory information and control behavior. While we know, at a coarse scale, the anatomical connectivity between brain regions (E.g., the mesoscale connectome published by the Allen Institute), we have not yet been able to study the dynamical interactions between areas in awake behaving animals..

Specific question: How do the different brain regions interact with one another?

Data set: Steinmetz has recorded simultaneously from large numbers of neurons in multiple regions of the mouse brain.

Techniques: We will use CCA (canonical correlation analysis), and RDM (representational dissimilarity matrix) analysis to study the similarity between neural representations in different pairs of brain areas, at different time lags.

Controls: Randomize the brain area labels associated with recorded neurons, and repeat the analysis.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Joel Zylberberg, Marieke Mur, Nick Steinmetz, Joshua Glaser

Project topic: Analyses on coding of cell assemblies

Scientific context: Broadly speaking, cell assemblies are any coalition of neurons that, by virtue of their temporary coordination, represent internal states or sensory information. While largely accepted as unitary “information token”, depending on the brain region, assemblies have different stereotypical activity patterns and encode different information.

Specific question: Does the brain use a rate or temporal code? How does activity pattern of assemblies contribute to the neural coding? Which information do they encode? Can assemblies mediate inter-regional interactions? Are assembles shared or re-organized across the spontaneous state and stimulus-evoked state?

Data set: Stringer dataset.

Techniques: Many techniques can be used to obtain cell assemblies. Neurons change their activity differently according to the behavioral state (whisking/pupillometry could be used as proxy) or in reaction to stimuli, and thus can be divided into groups according to their coordinated modulation. One can divide units by, e.g., clustering their activities or by using more sophisticated techniques which allow detecting arbitrary assembly activity patterns and their temporal resolution (e.g., CAD, for Matlab *here* and python *here*). You can also use your own creativity and propose novel ways to detect assemblies. Once obtained, one can examine the various properties of these assembles, like the spatial-temporal structure, tuning properties, dimensionality, consistency of assemblies over time, and so on. Assignment of information content: t-test, regression, clustering.

Controls: Generating shuffled data; comparison across different brain regions, different brain states, cross-validate across different time periods.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Eleonora Russo, Xue-Xin Wei (with inputs from Cameron Smith)

Project topic: Communication within brain area

Scientific context: One of the goals of Neuroscience is to know how brain regions connect and communicate. This is often summarized in the figure of a connectome. Once having a connectome, one can easily simulate brain behaviour, therefore understanding cognitive processes and pathologies. However, is it possible to directly infer how the brain areas connect with each other using living data, i.e., while the subject is performing tasks?

Specific question: Can we infer the (functional) structure of the brain from neuron firing or EEG data?

Data set: Stringer Stringer et al V1 mouse data.

Techniques: Linear regression, PCA, cross-correlation, Granger causality, Information Theory.

Controls: Stringer et al V1 data without stimulus (or equivalent), or, if you're bold, theoretical simulation using Potjans-Diesmann model or similar.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W3D2+3. Proposed by: Matthijs van der Meer, Arthur Valencio, Bijan Pesaran

Project topic: State-dependence of stim+choice signals

Scientific context: Neural and behavioral responses to visual stimuli are strongly influenced by pre-stimulus activity (population firing rates, correlations, frequency content). The Steinmetz data set provides a novel way to test how strongly pre-stimulus activity is shared (or not) across multiple brain regions, and how it impacts stimulus responses and behavioral choices.

Specific question: Are various measures of pre-stimulus activity (e.g., low and high frequency LFP power, ensemble state) shared across visual and motor (cortical) areas? Does the degree of shared pre-stimulus activity predict variability of stimulus responses and/or behavioral responses?

Data set: Steinmetz (Neural: spikes and LFP; Behavioral: pupil, movement speed/accuracy)

Techniques: Linear regression, linear classification, spectral analysis

Controls: Do these factors simply reflect upcoming movements? Engagement / arousal?

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Matthijs van der Meer, Bilal Haider

Project topic: Spatial localization and dimensional characterization by frequency content

Scientific context: Cortical neurons show synchronous activities across a range of timescales (from milliseconds to minutes and hours). In principle, neurons can synchronize with different populations at different timescales. We will explore the spatial and dimensional structure of such timescale dependent synchrony.

Specific question (example): How does the frequency-resolved population coupling of neurons depend on their laminar position?

Data set: Calcium imaging of mouse V1 excitatory neurons during spontaneous behavior (Stinger et al., Science 2019)

Techniques: Spatial mapping, frequency domain analysis, dimensionality reduction (PCA, FA, GPFA). See (Okun et al., Cereb. Cortex 2019) as a potential starting point.

Controls: Internal controls, such as comparison across halves (in space or in time) of the data.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Takashi Kawashima, Michael Okun

Project topic: Intrinsic dimension of response manifold

Scientific context: How many dimensions does cortical activity have? Sparse coding theories predict high dimensional population activity compared to the number of neurons; dynamical systems approaches predict low dimensional activity. Finding an answer is tough: linear dimension reduction recovers only the dimensions in which the population activity is embedded, not the intrinsic dimensions of the activity itself.

Specific question: do population responses in V1 lie on a linear or non-linear manifold? How do the dimensions and their difference scale with number of neurons used? Can we recover the true dimensionality when it is known from the stimulus set (orientations = 1D).

Data set: Stringer orientations + external Stringer et al.

Techniques: linear dimension reduction (e.g. PCA) for the embedding dimensions; manifold estimation (e.g. correlation dimension). See Lehky et al (2014) Neural Computation, 26, 2135 for a starting point.

Controls: for dimension estimation: cross-validation, shuffled data; for scaling, bootstrapping/resampling

Difficulty: Medium. Prereqs: Linear Regression, linear algebra. Cross-ref to course content: W2D3. Proposed by: Mark Humphries, Marius Pachitariu

Project topic: Beyond pairwise interactions in fMRI

Scientific context: In fMRI data analysis we mostly describe coordinated activity by means of pairwise correlations. Looking at multiplets of variables sharing common (redundant or synergistic) information might allow a better and more informative characterization of large scale brain dynamics.

Specific question: Which is the most recurrent dimensionality of informational multiplets in fMRI? How does the size and composition of the multiplets change in time and with task?

Data set: Kay/Gallant dataset of responses to natural images (but the technique is amenable to be used with other imaging modalities too)

Techniques: Information theory; conditional probabilities; autoregressive modelling

Controls: Compare the patterns associated with informational multiplets of higher dimension with those coming from pairwise interactions

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Daniele Marinazzo

Project topic: Impact of circadian rhythms on task performance

Scientific context: Circadian rhythms can have a big impact on behavior (e.g. here). Mice in IBL are trained at different times of the day, and only sometimes on a strict schedule.

Specific question: Do mice that are trained at fixed daily times learn the task faster, or slower than animals that train at variable times throughout the day? Does behavioral performance peak at a specific time of the day (and is this similar across animals, between male and female mice, and between institutions with a normal vs reverse light cycle)? Can you predict the optimal training schedule to achieve fast and reliable task learning? How much faster would animals learn under this schedule?

Data set: The International Brain Laboratory et al. 2020; https://data.internationalbrainlab.org/

Techniques: Regression, classification

Controls: Cross-validate on held-out data (between labs, or shuffled).

Difficulty: Beginner. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Anne Urai

Project topic: Trial history effects and dependencies

Scientific context: Trials matched for stimuli exhibit variability in both behavior and neural activity. Stimulus and trial outcome history (behavior and neural activity) may explain a portion of this variance.

Specific question: Can you determine how much, if any, variance in behavior or neural activity can be explained by trial history effects? Can past responses be classified using only activity from the current trial? Can we relate the behavioral effects to the neural data? Can the behavioral effects of trial history be understood within the framework of reinforcement learning?

Data set: Steinmetz (behavior and spikes); IBL (for behavior only)

Techniques: Regression, logistic regression, linear classification, Reinforcement Learning

Controls: Cross validate for held out data, baseline activity/behavior in absence of history.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Ian Greenhouse, Petr Znamenskiy

Project topic: Dimensionality within and between brain areas

Scientific context: It remains a big question how much of the information in one brain area is shared with other brain areas, and how this may change with behavior or internal states.

Specific question: How does the dimensionality of neural activity within brain regions, and shared across brain regions, change (or stay the same) during different epochs of the task? Is there information in the shared subspace between regions that is predictive of task performance?

Data set: Steinmetz neuropixels.

Techniques: Any dimensionality reduction technique incorporating one or multiple brain areas (PCA, factor analysis, isomap, kernel pca, autoencoder, CCA, kernel CCA, reduced rank regression).

Controls: Do these results stay consistent when changing the type of dimensionality reduction used (e.g. PCA vs. Factor Analysis vs. Isomap vs. Kernel PCA, or CCA vs. Kernel CCA), and the definition of dimensionality.

Difficulty: Medium. Prereqs: Regression, Dimensionality reduction. Cross-ref to course content:___. Proposed by: Joshua Glaser, Joel Zylberberg, Marieke Mur, Nick Steinmetz

Project topic: Task-related changes via dimensionality reduction

Scientific context: Ongoing (intrinsic/resting-state) activity in the brain may interact with task-related responses. Some suggest it's an additive response, whereas others suggest a more complex interaction where task input changes the intrinsic dynamics, which can be detected using higher order analytic methods.

Specific question: How does task-related activity interact with intrinsic activity of the brain?

Data set: HCP fMRI data or Kay/Gallant fMRI data

Techniques: Principal/Independent components analyses, regression, CCA/PLS

Controls: Resting state fMRI and resampling

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: Randy McIntosh

Project topic: Compare stimulus organisation in different brain areas

Scientific context: The visual cortices transform information from retinotopic to view invariant and from basic visual features to objects with meaning. This transformation happens along the ventral visual stream from V1 -> Rhinal cortices/Temporal pole.

Specific question: Can you show this change in organisation across brain areas with fMRI

Data set: Kay et al.

Techniques: Multivariate fMRI with your favourite clustering algorithm/unsupervised learning algorithm. Take the pre-prepared responses to the images. Extract regions of interest around key brain areas (eg V1, Lateral Occipital Complex, Perirhinal cortex). How does the representational structure change between areas?

Controls: If you are brave you could compare these results to your favourite DeepNet. If you get stuff that looks like the rhinal cortices, you are cleverer than anyone else so far :)

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: XXXXX

Project topic: Natural movie processing - hierarchical timescales.

Scientific context: Understanding stuff is a hierarchical process. Some parts of your brain process incoming sensory events millisecond-by-millisecond. Other parts retain information about the current context for many seconds (or years!).

Specific question: Can you show this for subjects watching a movie. For example, early visual cortices should change with the visual stimulus, but association cortices (such as medial frontal and parietal cortex) may hold information over a whole scene. The representational structure should change fast in some areas but slow in others. (eg Baldassano et al Neuron 2017)

Data set: http://studyforrest.org/access.html or http://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179

Techniques: You could start by simply showing the autocorrelation structure of representations in different areas. Visualising medial parietal and early visual cortices should be revealing. Then you can try to quantify this either by modelling the average autocorrelation structure, or by making your favourite clustering technique. Chris Baldassano's HMM version is here: https://github.com/intelpni/brainiak.

Difficulty: Medium. Prereqs: Linear Regression. Cross-ref to course content: W2D3. Proposed by: XXXXX

Project topic: Natural movie processing - dimensionality of representation.

Scientific context: The previous slide says that sensory cortices change their representations faster than association cortices, but do they move in a low d manifold, or are they actually higher dimensional than association cortices?

Specific question: Measure the dimensionality of fMRI representations whilst subjects are watching a movie. Are early visual areas higher dimensional than association areas. Is this only true for a movie with temporal structure, and not for a series of independent images?

Data set: http://studyforrest.org/access.html or http://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179 compare with Kay data.

Techniques: Get ROIs from different brain regions (eg with FSL). Use your favourite dimensionality estimation tool (PCA would be a good start).

Project topic: Natural movie processing - representational connectivity.

Scientific context: This project is intended for students who would like to irritate Konrad Kording. To do this it will be important to refer to the correlations you observe as "connectivity" in your final report.

Specific question: Does the "connectivity" between brain regions depend on the representations. We know that voxels fluctuate together in networks, can we go deeper. Different representations exist within brain regions. Do some representations in brain region A selectively co-fluctuate with particular representations in brain region B?

Data set: http://studyforrest.org/access.html

Techniques: Get ROIs from different brain regions (eg with FSL). Dig out different representations in each brain region (eg top 10 PCA components). Is there selectivity so that some components in region A only correlate with some in region B? Can you cross-validate components from different halves of the data?

Controls: Why is this happening? Is it because both regions represent the same visual stimulus (e.g. character), and emerge simultaneously whenever the stimulus appears? Or is it a more basic phenomenon? Do the same representations co-fluctuate during the retinotopic mapping experiment, where completely different visual stimuli are presented? (this requires you to know you to align images across experiments). .

Project topic: Communication between brain areas with engagement

Scientific context: Fast-timescale changes in synaptic efficacy (“dynamic gain control”, “routing”) are thought to be an important building block for cognition. However, empirical support has been piecemeal. Recent Neuropixels data with hundreds of neurons in multiple, connected brain areas offer new opportunities to examine how such rapid changes in connectivity occur and how they relate to behavior.

Specific question: Fit models to infer inter-area state variables based on LFP and/or spiking. What is the relationship between the inferred variables and behavioral engagement (speed, accuracy)?

Data set: Steinmetz Neuropixels data. Will need to perform initial screen (localizer) to identify most promising area pairs/networks. Reflect on if we can test for communication without causal manipulations.

Techniques: Multiple regression, cross-correlation, signal processing, causality, information theory.

Controls: Dependence on number of neurons used for state estimation. Resampling methods.

Difficulty: Medium. Prereqs: Linear Regression, Wrangling Experimental Data. Cross-ref to course content: W3D2+3. Proposed by: Matthijs van der Meer, Arthur Valencio, Bijan Pesaran