**Final Course Project**

Project logistics:

**\*Pick one (1) project to work on, with a group.**

**\*Group size: 2-3 people per project**

Project possibilities

**PROJECT: [Data: mouseV1][Mentor Jennifer]**

Take the processed data (2 plans over 2 time points, before and after training) from this folder[Project](https://drive.google.com/drive/folders/1L7QC8iYN81VqHx-8oj265dnUgNR6MCfG?usp=sharing) . If you like, go through the matching so you get a sense of the data you will be processing, otherwise, the matched cell indices are in the matched folder.

**Dimension reduction**. Apply learned methods (PCA, NMF, or UMAP if you want to!) on V1data. Look at trajectories under different visual stimuli for each day. You can then try to compare the same neural populations between two different days (using matched pairs provided or do it yourself). Look at trials where the mouse is moving vs separately from trials where the mouse is still (run.matrix).

**Classification.** Apply learnt methods (e.g. LDA using the first 3 PCs) to compare the population decoding before and after training. Try varying time windows (e.g. 1s vs. 3s) to see what captures the difference between days, use regularization and cross validation to assess the outcome.

**PROJECT :[Data: Zebrafish][Mentor Mark]**

Download and pre-process at least five different fish. Pick several fish with 0-3 and several fish with different stimuli (beyond 0 to 3). For each fish, do **PCA** and **NMF** of cells within ten or more large regions; compare PCA and NMF trajectories. Which seems more meaningful?

How does swimming motion affect the trajectories/plots in phase space? Color trajectories where the fish swims L vs when it swims R when the fish is not swimming. Are there consistent deflections of neural trajectories related to swimming?

Do the same thing for eye convergence (preparation for hunting microbes)

Which pairs of regions seem to have similar trajectories in phase space during the experiment?

Study co-variation of trajectories between pairs of brain regions in each fish, by canonical correlation (week 2.) Identify pairs of regions that may share a communication sub-space with several dimensions (as Byron Yu presented) in each fish.

Start to investigate individual differences in fish (yes, there are!), by comparing neural trajectories of brain regions between two fish after the same transition. Compare communication subspaces connecting those regions that show the biggest individual differences.

**Project: [Data: Steinmetz][Mentor Pascal, Hadas, Mark (part-time)]**

**Preamble**:

This project represents a “full” neural data science stack. We will take it all the way from exploratory data analysis to classification.

We highly recommend using the “big data analysis cascade” introduced earlier in this course to organize your code - it will make your life a lot easier.

Make sure to use at least 2 recording sessions for your analyses - develop the models in one session and see if the findings generalize to other sessions (genuine replication). Perhaps characterize the model variance that way (if you use more than one session in the “test” set). Make sure to use a region (or regions) that show up in both sessions (so that results are comparable). Use at least 2 such regions (we recommend one “early” and “late” in cortical processing), so you can contrast the population response.

NOTE: This is a complex project by any metric - there is a lot going on in terms of the neural data, the behavioral data, and everything else that is going on. So let’s see if we can utilize it. Arguably, the authors themselves left *a lot* of data on the table in their original paper.

Will you be able to make sense of it? [Perhaps we should publish a canonical course paper?]

1. **Exploratory data analysis**: Characterize the neural activity of the population of neurons in the two regions you chose in terms of firing rates and tuning curves of individual neurons, as well as the population response as a whole. Is the population response normal, by any metric (meaning: *DISTRIBUTED* normally)? Make sure to determine neural activity in response to both visual stimulation as well as motor response. This will be important later. In fact, most of what you can determine later will already be implicitly contained in the results of this analysis. In particular, be on the lookout for “outliers” (extreme values) and be prepared to handle them later.
2. **Dimension reduction**. Perform **PCA** on neural responses of cells in both brain regions, using a sliding window of data (window and step size determined by you), count the number of components to account for, e.g., 80% of the variance (or some other cut-off, e.g. Scree-plot) to determine the “effective dimensionality” of this dataset. Make a time course plot of this effective dimensionality, time-locked to events in the experiment (stimulus onset vs. movement onset).. Do this separately for both brain regions and plot the effective dimensionality lines in the same plot, for maximal comparison.

Now repeat all of this with at least one other linear dimensionality reduction method and one non-linear one. Do you notice anything of interest?

1. **Regression**. Build the ultimate linear regression model! What we mean by this specifically is to build a model that predicts neural firing rates in your target region as well as possible (maximize R^2) *without* overfitting. There are all kinds of predictors available for this purpose (visual inputs, motor responses (eyes, face), and a plethora of other covariates (e.g. licks). Importantly, account for all of the issues we discussed, in particular multi-collinearity (hint: You might want to use the uncorrelated PCs from PCA from the previous part of the project for this purpose, or use the LASSO, or both). In any case, some kind of regularization, hyper-parameter tuning and cross-validation is probably needed. How well can you predict activity in early vs. late brain regions.
2. **Classification.** Do LDA \*and\* logistic regression in small time windows to produce a time series of windowed classification accuracy. Use different window sizes to see what integration window (and method) maximizes accuracy as a function or brain area (across sessions). Classify what? Behavioral choices of the animal. How well can you classify trials of different provenance (e.g. rewarded vs. unrewarded trials, left vs. right choices)?
3. **Extra credit**: Pick anything from the 2nd week (in particular some kind of network analysis) and do something interesting with it (that is not already specified above). If all else fails, you can do something with spike count correlations (I can help).

NOTES (BEWARE):

\*The SteinmetzSessionsRegions spreadsheet seems to be incomplete - there is more data available (from more regions) than denoted in that sheet (Sheet = areas Mark who created this sheet was interested in at the time, the mismatch is particularly large for visual areas).

\*Make sure you *develop* the model in an animal that has both data from an early (perceptual?) and late (motor?) region, but it is ok to mix and match the validation/test (in other words, you might be hard pressed to find an animal where both is matched), so you could find one with a match for the early area, and another one for the late one. As the animals are (presumably) independent, and independence is key, this should be fine.

\*Make sure to look at the Steinmetz “cheatsheet” (Data notes) if you are unsure of the structure of the dataset or the structure of the experiment. It’s a very rich ball of wax.

\*You *should* have editing privileges for the SteinmetzSessionsRegions file. Crowdsource replacing the “Xs” with the actual number of neurons available per session and animal. That would be very helpful.

Meeting notes 071225

* Give 3 projects (1 per dataset)
  + Choose sub-projects/ways of approaching data
  + Each prof assigned to dataset
  + Fish mark + hadas, mouse v1 jennifer, steinmetz pascal
* Profs write projects
* Update from last time, modify for conciseness and topic relevance
* Profs present projects in class a few minutes each and let them think
* Reshuffle teams
* Groups of variable sizes
* People pick their projects
* If more than 3 people pick, then we split
* They can choose how to split
* Pascal, Hadas, and Jennifer will work tonight to condense
* Mark is picking 2nd zebrafish project from last year
* Merging 3rd and 4th Steinmetz
* Jennifer wrote hers