Probabilistic models for neural data
Session 5: Generalized linear models #3 & dimensionality reduction

To do before this session:

- **Read Pillow et al. (2008) and prepare assigned presentation**
- **Complete the GLM exercise (it helps reading Pillow et al. (2008) before that)**
- **Read up on this session's statistical concepts**
- **Complete pre-session quiz**

In this session we will first discuss Pillow et al. (2008), where GLMs are used to model neural population activity, with a focus on correlations among neurons. After that, we will start looking at dimensionality reduction. Applied to neural data, dimensionality reduction hinges on the premise that neural population activity is a high-dimensional projection of some low-dimensional, unobservable (latent) states. The aim is then to identify these latent states, and with them gain a simplified representation of the high-dimensional data that might provide further insight. The different dimensionality reduction methods we will discuss in this section differ in which assumptions they make about the latent states and their relation to the high-dimensional projections. In terms of statistical concepts, we will discuss PCA, probabilistic PCA, and factor analysis.

Paper: Pillow et al. (2008). Spatio-temporal correlations and visual signaling in a complete neuronal population.

When reading the paper, please focus on the following:

- How do they use GLMs to model the simultaneous response of multiple neurons? Does the underlying theory change?
- How did they keep the number of model parameters low?
- Why was it important to maintain causality in the used filters? E.g., why could only past spikes impact current spiking?
- Why did the uncoupled model fail to capture the peaks in the neural response cross-correlation functions? What do we learn from this failure?
- How do they assess if the coupled model predicts neural data better than the uncoupled model?
- How do they assess how much information the model captures about the stimulus? Why does this allow them to estimate the information boost due to correlations?

Consider how the underlying graphical model changes with the introduction of multiple neurons.

Presentations:

1. Coupled GLM (Fig. 1a)
    a. Describe the different factors that the model assumes impact the activity of each individual neuron. How do the different factors interact?

    b. What is the purpose of the three different filters? What effects would the model be unable to capture if either of them would be removed?
    c. How do the different filter outputs interact to determine each neuron's spike rate? How did they preserve causality in their model?
    d. Describe how the model's components can be loosely compared to biophysical mechanisms.
2. Data and fitted filters (Fig. 1b-h)
    a. Describe the stimulus and the data was collected.
    b. How did they ensure that all RGCs in a local patch were recorded from?
    c. How did they formulate the stimulus filter? What assumptions does this formulation make about how neural activity is modulated by the stimulus?
    d. Describe and interpret the fitted filters in Fig. 1c-h.
3. Pairwise cross-correlation functions (CCFs; Fig. 2a-g)
    a. Describe what the shown pairwise CCFs measure. What is the interpretation of the different CCF shapes for different pairs of cell types?
    b. Which two model components are able to capture pairwise cross-correlations? Which of them is able to capture perfectly synchronous spikes? (Hint: think about how past spikes of other neurons can impact the instantaneous firing rate in the model)
    c. Which model component appears to best capture the CCF peaks around time zero? What does this suggest about their mechanistic origin?
4. Third-order CCFs (Fig. 2h-j)
    a. Describe and interpret Fig. 2h-i.
    b. Describe and interpret Fig. 2j.
5. Peri-stimulus time histogram (PSTH) vs. spike predictions (Fig. 3a-d)
    a. Describe the observed PSTHs and those predicted by the different models. How do you interpret the found match?
    b. Describe Fig. 3d. How was the spike prediction quality measured? How and why does the coupled model feature a higher spike prediction quality than the uncoupled model? Why does this difference not show up when comparing PSTHs?
6. The predictability of per-trial PSTHs (Fig. 3e-g)
    a. Describe how pre-trial PSTHs were computed. Do they differ from the true predictions, and if yes, why?
    b. How doe the authors quantify the prediction quality of single-trial PSTHs vs. true PSTHs in Fig. 3g. How do you interpret the shown results?
7. Assessing stimulus information (Fig. 4)
    a. Describe the different stimulus decoding models in Fig. 4b (you can ignore the maximum entropy model mentioned in the text).
    b. Describe how the decoding performance was measured in Fig. 4b. What is the interpretation of the results shown in Fig. 4b?

Statistical concepts: Dimensionality reduction

PCA and probabilistic PCA (PRML 12.0-12.2.3)

When reading up on the non-probabilistic version of PCA, make sure to understand

- Why it might be desirable to focus on linear subspaces that capture most variability in the data
- The benefits of projections into an orthonormal basis
- The equivalence between the variance-maximization and the reconstruction error minimization perspective of PCA

The application sections (PRML 12.1.3-12.1.4) provide additional intuition into how PCA operates.

In the section describing probabilistic PCA (PRML 12.2-12.2.3), make sure to understand

- The generative model underlying PCA
- The non-identifiability of the mapping W between the latent state and the observables
- The conditions under which the maximum likelihood solution to probabilistic PCA equals the PCA solution

You can skip the details of the math of the maximum likelihood solution, but should understand how this solution in general differs from the PCA solution, and why. You can also skip PRML 12.2.2 that describes the use of the EM algorithm for probabilistic PCA inference. (Advanced) The section on Bayesian inference for PCA (PRML 12.2.3) uses a few concepts we haven't previously discussed, but might nonetheless be an interesting read, as it describes how one might choose the latent space dimensionality without cross-validation by Automated Relevance Determination (ARD), which we have already encountered in Park & Pillow (2011).

Factor analysis (PRML 12.2.4)

For factor analysis, make sure to understand

- How the generative model for FA differs from that of PCA
- What the consequences of this difference are
- Why it shares the rotation-invariance of the latent space with PCA