Learning representations from time-series data with multi-step predictions

In neuroscience, understanding how brain activity drives behavior often involves analyzing high-dimensional neuronal time-series data. However, this data is inherently noisy and computationally challenging to interpret, making data dimensionality reduction techniques particularly useful for preserving both the temporal dynamics of neuronal activity and their relationship to behavioral states. Existing algorithms like BunDLe-Net address this by compressing neuronal signals into low-dimensional embeddings that preserve both behavioral and dynamical information. Preserving these multifaceted aspects of neuronal data is essential because these dynamics encode how brain activity relates to behavior. However, BunDLe-Net currently relies on a single-step prediction approach, forecasting only the immediate next state of neuronal activity, which limits its ability to model temporal dependencies over longer sequences. This limitation means it can miss behavioral motifs that emerge across multiple time steps and may amplify sensitivity to noise.

To address these limitations, this thesis will explore the integration of multi-step prediction techniques into BunDLe-Net. The aim is to enable the model to predict multiple future states at each time point. This approach is expected to capture richer information about the longer-term evolution of neuronal dynamics, allowing the model to distinguish between initially similar neuronal states that diverge into distinct behavioral motifs over time. Ultimately, the model could capture a broader view of how neuronal states evolve over time, learning meaningful low-dimensional abstractions of the underlying dynamics, enabling a more accurate simulation of neuronal activity in a given behavioral context.

The proposed research will begin with thorough data examination to confirm that the neuronal dataset contains sufficient future time-step information and to assess if such information can be effectively compressed into lower-dimensional embeddings. This will be followed by the development of simplified models to systematically evaluate several distinct multi-step prediction strategies and identify the most effective technique before its full implementation in the BunDLe-Net architecture. The approach must address several aspects, including mitigating the inherent noise in neuronal data, maintaining accuracy over extended prediction windows, ensuring generalization, and preserving meaningful representations of the underlying dynamical structures. The methods under consideration include a recursive method that iteratively predicts successive steps using prior outputs, a method that predicts each future state directly from the initial time point, and a method that forecasts the entire sequence of future states in a single forward pass. The optimal architecture, selected through empirical validation, will be integrated into BunDLe-Net. This integration will provide the framework with richer temporal abstraction, representing how neuronal states evolve over time. These low-dimensional representations are designed to encode the causal pathways governing neural dynamics and behaviors, capturing how past neuronal activity drives future dynamics and behaviors.

In conclusion, this research aims to better capture the dynamics of high-dimensional, noisy time-series data by integrating multi-step prediction techniques that capture longer-term temporal dependencies. By systematically evaluating multiple strategies, the study aims to identify an optimal approach that not only leads to more accurate prediction but also provides representations that are more interpretable and effective for understanding dynamical systems.