The ways in which pre-neurons (known as neural progenitor cells, or NPCs) form communication networks is an active topic of research. Cell communication is often measured using the cross-correlation of their calcium concentration over time. However, this metric is arbitrary, and thus it is difficult to select a cutoff value to differentiate between “talking” and “not-talking” cells. In order to solve this problem, I attempted to use machine learning techniques to predict networks in a more robust manner, hopefully making conclusions based on such data more reliable.

I initially attempted to model correlated sets by generating artificial spike trains, supervised learning methods failed to accurately predict networks on real data, suggesting that I had insufficient skill to properly model these time series. With this in mind, I turned to using the K-Means clustering algorithm, with the cross-correlation used as a measure of distance, defined below.

$$argmax_x \frac{1}{n} \sum_x \frac{1}{\sigma_f \sigma_t} (f(x) - \bar{f})(t(x) - \bar{t})$$

Overall, simple modeling of calcium spike trains, even with gaussian noise, was insufficient to accurately capture the nature of real data. As such, my efforts to use supervised learning to solve this issue requires much more time to be spent on the modeling end, time which I did not have for this project.

However, my efforts to create a simple clustering algorithm to generate NPC networks seems to have worked quite well, attesting to the power cross-correlation has as a relative similarity metric.
References
