Cathy Zhuang

BIOL 450 Assignment 3: Dr. Sober

16 February 2022

*Per the syllabus: “Collaboration on assignments is acceptable (and, in fact, is encouraged!), but please list the names of everyone with whom you worked, and each student must turn-in their own, separately generated solutions.”*

People with whom I’ve worked: Leila May Pascual, Lauren Gao, Tony Wang

Bar Graph from In-Class Question 3

Chart, bar chart

Description automatically generated

1. Chart, line chart

Description automatically generated

2. Chart, line chart

Description automatically generated

3. For Neuron 1, as the time resolution increases (as the time bins decrease from 40 ms to 1 ms), the MI values increase, meaning more of the pitches are being predicted. The highest MI value is 0.4445, which corresponds with time bins of 1 ms, which is the largest time resolution. Thus, as we increase the time resolution, our predictions of the pitches get better. Neuron 1 is using a time code since we are only looking at the time scale.

For Neuron 2, as the time resolution increases (as the time bins decrease from 40 ms to 1 ms), we also see an increase in MI values, up to 1, where the most pitches are being predicted at the highest time resolution. The probabilities of the spikes given pitches are evenly distributed in these smaller time bins. We also see that overall, the MI values for Neuron 2 are larger than those of Neuron 1, even at larger time bins.

At a glance, to learn about behavior, we would want to use the temporal resolution of 1 ms time bins for each neuron, since both had the highest mutual information values. The higher the mutual information value, the more pitches are being predicted. For Neuron 2, one could also use 2 ms bins, as it had the same MI value as 1 ms bins and would also have a little less impact from an under-sampled regime. It seems like Neuron 2 uses a rate code as opposed to Neuron 1, since code words and rates are more predictive than simply timescale.

4. The possible values for 20 ms timebins would be 212 = 441, since there are two bins and 21 values each bin can take on (0-20). For 1 ms timebins, the plausible values in each bin would be 2 (0 and 1), so in total, with 40 bins, it becomes 240, or 1.0995116\* 1012.

5. Since we are working with 1003 observations, the only time bins where our number of samples is larger than the total number of possible outcomes would be 20 and 40 ms bins, as 20 ms bins has total possible values of 441. However, with larger bin sizes, temporal resolution is lost. Thus, we may prefer to rely on the estimates provided by the 20 ms time bins as opposed to the 40 ms time bins to compromise temporal resolution with the avoidance of the unsampled regime.

6. If we are working with an under-sampled regime, then every “word” may only occur a few times or even not occur at all. It would be difficult to distinguish between a chance coincidence and an actual pattern with so few occurrences of the “word”. The variance is very high. In our mutual information program, we are not treating any occurrences as coincidences, meaning we may ascribe value to something that is just chance. This produces an overestimation of mutual information, since we may be falsely linking patterns that are actually coincidences. We may have too much mutual information.