

## BIM 280: HOMEWORK 3

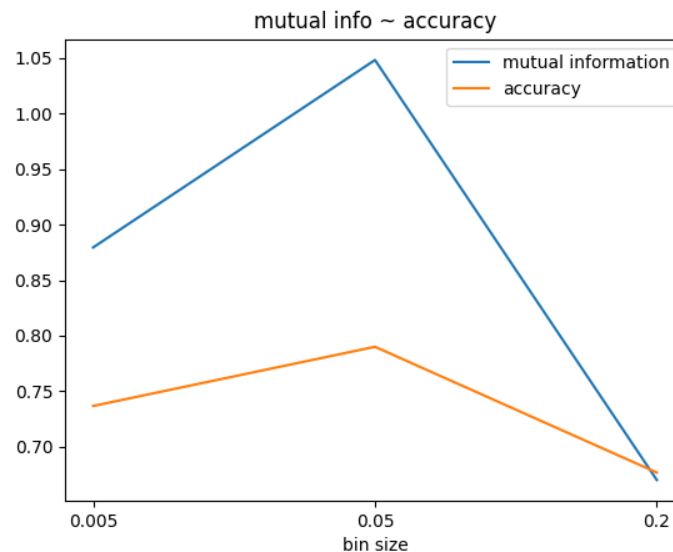
### Part 4: Conceptual Questions

1. **Compare and contrast the results across bin sizes. Which bin size is the most appropriate? Why do you think that bin size has the most information?**

The mutual information for 200 ms, 50 ms, and 5 ms bin sizes are 0.67, 1.05, and 0.88. Here, the most appropriate bin size is 50 ms because it provides the highest mutual information.

2. **Explain the relationship between the performance of the PSTH classifier and mutual information.**

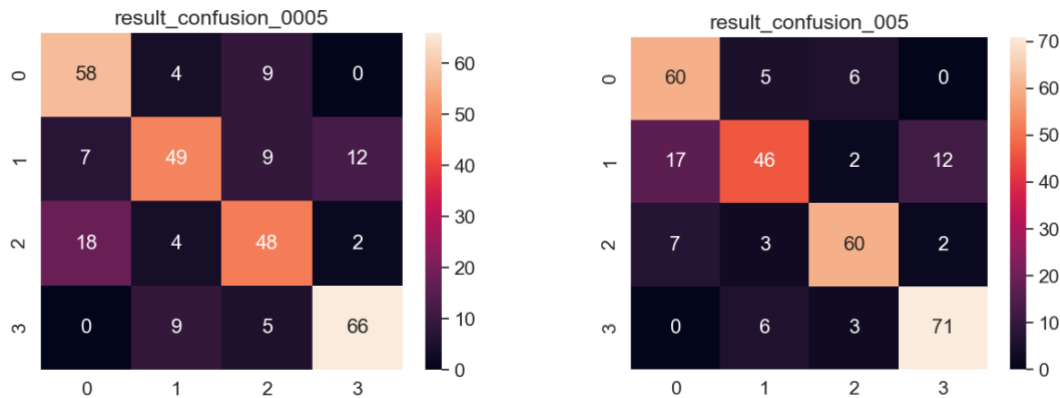
The PSTH classifier's performance and mutual information are positively correlated, therefore an improvement in mutual information results in an improvement in the PSTH classifier's performance. For example, the figure below shows that as accuracy of the classifier increases or decreases, the mutual information also changes in the same direction. Meaning, PSTH classifier performs better when the mutual information is higher.



3. **What does the diagonal in the confusion matrix represent? Do you want higher counts or lower counts along the diagonal? Why?**

The diagonal in a confusion matrix represents the number of correct predictions made by the classifier. We want to have higher counts along the diagonal of the confusion matrix. Because the number of counts indicate the performance of our

classifier. The higher counts mean that the classifier is able to make correct predictions. For instance, the classifier's performance for 5 and 50 ms bin sizes are respectively 0.74 and 0.79. Accordingly, the bin size for 50 ms (on the right) has higher counts along the diagonal in the heatmaps below.



**4. What do the off diagonals in the confusion matrix represent? Do you want higher counts or lower counts along the diagonal? Why?**

The off diagonal in the confusion matrix shows the number of incorrect predictions made by our classifier. For example, for 5 ms bin size above, out of 71 trials for event\_1, the classifier classified 4 of them as event\_2 and 9 of them as event\_3. We hope to have lower counts along the off diagonal. Because fewer off diagonal counts means that our classifier made less incorrect predictions ( $\hat{y}$ ) about the ground values ( $y$ ).