Neuroscience 299: Computing with High-Dimensional Vectors
Assignment 7: Connections to information theory
Due October 20, 1pm

Reminder: Please do *either* the writing assignment *or* the programming assignment. Expected length for the writing assignment is approximately 250-500 words, but there is no strict minimum or maximum.

Writing assignment:

Elaborate why is it important to understand the capacity of distributed representations. Pick a capacity result from either the focus paper or the extended reading list and explain what was proven. Finally, discuss how a capacity result could be useful for certain applications (i.e. a HD computing/VSA system designed for a certain task).

Programming assignment:

You will implement the analytical calculations from the capacity theory and validate them using empirical simulations using one of the HD computing/VSA models.

1. For this assignment, you can choose one HD computing/VSA model. Based on the previous assignment you might already have implementations for the following ones:
   - Multiply-Add-Permute (Assignment 1);
   - Holographic Reduced Representations (Assignment 2);
   - Fourier Holographic Reduced Representations (Assignment 2).

You may use another model (e.g., Binary Spatter Codes), but in any case make clear *which* implementation you are using.

2. Once the model is chosen, please implement the transformation of sequences into distributed representations with permutation and superposition operations (see slide #11 in lecture for Module 4). Here is a high-level description of the process:
   - Specify the dimensionality of HD vectors (denoted as N), the size of the "alphabet" (denoted as D), and the length of a sequence (denoted as M) as parameters of the transformation. The notations are chosen to be consistent with the focus paper.
   - For each symbol in the alphabet, assign a random N-dimensional HD vector and store it in the item memory, which in this case is an [N x D] matrix.
   - Generate a sequence of length M randomly using D symbols of the alphabet.
   - For each symbol in the sequence, pick the corresponding HD vector from the item memory and permute it according to its absolute position in the sequence.
   - Form an HD vector of the whole sequence by superimposing all the permuted HD vectors in individual symbols. (Do not forget to normalize the resultant HD vector if this is required by the chosen HD computing/VSA model.)

3. Implement the recovery/decoding procedure to reconstruct the sequence given the item memory and the HD vector representing the sequence. The process of decoding a symbol in the given position is very simple. Apply the inverse permutation in this position to sequence's HD vector and use the resulting HD vector as an input to the item memory. Return the symbol corresponding to the result of the nearest neighbor search as the prediction.

4. Implement the calculations required to analytically compute the accuracy of correctly recovering sequences' symbols (p_{corr}). The main equation to be implemented is (2.12). Note that the values of N and M, as well as the choice of HD computing/VSA model, affect the statistical moments used when calculating the sensitivity in (2.13), so pay attention to these details in your computations.

5. Use the primitives implemented in steps 2 and 3 to perform empirical simulations measuring the accuracy of correctly recovering sequences' symbols for the fixed N, M, and D. Average the results over several initializations of item memory and sequences. Compute the analytical accuracy (step 4) for the same values N, M, and D. Perform this step for several fixed combinations of M and D and for a range of N (e.g., 64, 128, …, 2048). Plot the obtained empirical and analytical curves in a figure (N is x-axis; Accuracy is y-axis).

The next steps ask you to elaborate on the following aspects of your experiments. Explain your answers in about a paragraph per question.

6. Do the empirical accuracy curves follow the analytical curves? (Hint: they should, so double-check your implementation in case they do not.)

7. How do N, M, and D affect the accuracy? How are their effects on the accuracy accounted for in equation (2.12)?

8. What are the key assumptions of the analytical equation (2.12)? Provide an example of the modification of the setup for the experiment above where some of the assumptions would be violated.