Title: Statistical framework for detecting changes in high-dimensional data using neural networks in a transfer learning fashion

Mentors: The project will be supervised by Dr. Alon Kipnis

Description:
This project proposes a novel and relatively simple method to convert a trained neural network to an unsupervised discriminator between two or more new datasets. The only requirement is that the neural network can classify data closely related to the new datasets. Typically, the datasets involve images although other kinds of data come to mind. The idea is to combine many statistical "two-sample" tests based on the network's neurons in a novel manner to one number; large values of this number indicate that it is likely that a change between the two datasets has occurred while low values indicate the opposite. The challenges in this process revolve around deciding how to combine information from many statistical tests involving individual neurons or layers and exploring applications the resulting method has in real-world classification challenges. Intuition from previous works in the one sample setting (out-of-distribution detection https://arxiv.org/abs/2102.12967) is likely to help.

We anticipate from our would-be detector to be widely applicable in a host of unsupervised classification settings and transfer learning challenges. Examples include: classifying images with little to no training data, detecting unusual events in video footage, and discriminating artificially- from naturally-generated images. Depending on the progress, the anticipated deliverable can be a new software package and/or a research article.

Background:
Two sample tests are commonly used in statistical decision theory to indicate whether two datasets can be thought to be sampled from the same underlying generating mechanism – or not – without knowing/estimating/learning the actual distribution of any of the samples. In this project, we treat the response of each sample in each neuron or a group of neurons as a different feature and test for equality of the distributions of the two classes across the many features. We plan to use a two-sample t-test as our main working horse; we apply this test many times – once per neuron. We then combine the many p-values obtained from these tests using several possible methods for combining P-values. Different combination methods are useful for detecting different signals. Fisher's, Sims's, Bonferroni's, and Higher Criticism are several such methods. Finally, we characterize the combined values under a situation of "change" and "no change" and decide on the optimal threshold for balancing the two. Once calibrated, this threshold is expected to be useful in identifying differences between the two samples across many unsupervised discrimination challenges.

The one-sample setting of this problem was studied at https://arxiv.org/abs/2102.12967. The two-sample setting is anticipated to be much more useful in real applications since there is no need to calibrate the model per task.

Requirements: Students taking on these projects should have hands-on experience with neural networks for image classification and relevant programming packages like PyTorch. An advanced course in probability and statistics like “Statistics and Data Analysis” is highly recommended. At a minimum, the students should be motivated to learn new concepts in probability and statistics.