Convolutional Neural Networks for Classification of Data In High-Dimensional Datasets

Nishanth Bhaskara
Mentors: George Djorgovski, Ciro Donalek

Machine learning is an important technique for data analysis and pattern detection in complex or large data sets. Certain datasets are very high in dimensionality and so most existing machine learning techniques fail to scale in feasibility. Although methods such as feature selection or feature extraction exist, they fail to look at the data as a whole. Our research is mainly focused on using CNNs, a neural network which utilizes filters that convolve over the input vector, to try and provide a generic solution to the classification problem of any high dimensional data set. As a proof of concept, the research is geared towards video classification, namely fMRI clips of patients either testing positive or negative to autism. The fMRI clips are different from normal images as they add an extra spatial dimension (z-axis) and a temporal dimension. The CNN terminates with fully connected neural network which minimizes the loss function (cross-entropy in our case) over the training period, and outputs a probability that the patient has autism. The project is still ongoing, and the main obstacle is achieving the computational power to do our tests; we hope to have results soon.