MATH 637 - Spring 2020
Mathematical Techniques in Data Science
Presentations
May 21, 2020

Time	Speaker(s)	Title
10:30 to 10:50	Charlie Hannum Mark Leadingham Jerome Roehm Dominick Sinopoli	Machine Learning Methods for Grade Prediction
10:55 to 11:15	Prabhat Kumar	Computer Vision, Convolutional Neural Networks and Breast Cancer Classification
11:20 to 11:40	Carlos Restrepo	Improvement of the Reconstruction Quality in Computed Tomography Through the Use of Convolutional Neural Networks
11:45 to 12:05	Melinda Kleczynski	Neural Differential Equation Model of a Simulated Oscillatory Epidemic
12:10 to 12:30	Prosper Kosi Anyidoho Grace Ashley Ibrahim Balogun Paul Steller	Optimizing Digit Recognition with the MNIST Dataset
12:35 to 12:55	Olivia Wanjeri Mwangi Desiderio Pilla Akshaya Ramesh	Twitter Sentiment Analysis
1:00 to 1:20	Tejas Dipak Patel Bilal Riaz	Spectral Clustering and Seriation: Theory and Algorithms

Abstracts

10:30–10:50 Charlie Hannum, Mark Leadingham, Jerome Roehm, Dominick Sinopoli
Machine Learning Methods for Grade Prediction.

Students often have difficulty determining how they will ultimately perform in a course at the add/drop deadline. Instructors may benefit from predictive tools that can be used in the advising process. To assist in this process, we obtained the anonymized gradebook data for over 1,600 MATH221 students over eight previous semesters. We applied machine learning methods to predict the final grade of students using quiz and exam scores that occur before the change of status deadline. When it came to determining pass or fail at the add/drop deadline an altered gradebook scoring and a nearest neighbors adaptation performed just as well as support vector machines and decision trees. However, the nearest neighbors algorithm was determined to be the best predictive tool overall and was able to predict student performance to within half of a letter grade. These results are promising for providing instructors and students with a new predictive tool to aid in withdrawal advisement.

10:55–11:15 Prabhat Kumar
Computer Vision, Convolutional Neural Networks and Breast Cancer Classification.

The first goal of this project was to explore computer vision and the research which built up the field, so we will discuss its history and look at the rise of modern convolutional neural networks (CNNs). Secondly, we look at the mathematics of CNNs and review its implementation on computers. Our main goal is to apply CNNs to the (binary) classification of invasive ductal carcinoma (IDC) breast cancer. So, we explore a few NN implementations: (1) the naive NN from the TensorFlow tutorial for a baseline comparison, (2) a CNN architecture made from scratch using generically accepted design principles and (3) state-of-the-art architectures like AlexNet and VGG, and we compare the results.

11:20–11:40 Carlos Restrepo
Improvement of the Reconstruction Quality in Computed Tomography Through the Use of Convolutional Neural Networks.

In an inverse problem like Computed Tomography (CT), we have x-ray projections taken from different perspectives of a sample. The idea is to reconstruct the original object from the observed measurements. It is important to notice that the solution to such a problem is ill-posed, and so, analytical solutions are unstable, i.e., they are very sensitive to noise. Because of that, iterative methods and compressed sensing strategies are typically used in this scenario. However, applying some of these techniques tends to be time-consuming, and may end up being impractical for real-time applications.

The purpose of this project is to take advantage of fast reconstruction techniques such as Filter Back Projection (FBP), which by itself may not be very accurate, and couple it with a Convolutional Neural Network to improve its performance. To that end, data augmentation techniques are used to generate enough samples to trained the model. Different architectures for the problem are tested, and the results show improvements that are typically between 2.30 and 4.45 dB in terms of PSNR.

11:45–12:05 Melinda Kleczynski
Neural Differential Equation Model of a Simulated Oscillatory Epidemic.

An epidemic model from the literature generates simulated data with intermittent spikes in the number of infectious individuals. A model involving neural networks learns a system of ordinary differential equations which describes the data. The learned model provides predictions regarding further time evolution of the data, and an alternate scenario with an initial vaccination campaign.

12:10–12:30 Prosper Kosi Anyidoho, Grace Ashley, Ibrahim Balogun, Paul Steller
Optimizing Digit Recognition with the MNIST Dataset.

The goal of our project was to optimize different models for digit recognition. Our models were trained on the popular MNIST dataset. Using the best model, we created an android app that would recognize user-drawn digits.

12:35–12:55 Olivia Wanjeri Mwangi, Desiderio Pilla, Akshaya Ramesh
Twitter Sentiment Analysis.

Social media has created a new way for individuals to express their thoughts and opinions. Sentiment analysis, or opinion analysis, is the process of retrieving textual information and discerning which emotions are exhibited by the author. The main goal of this project is to build and train a model that can detect the sentiment of large groups of tweets.

1:00–1:20 Tejas Dipak Patel, Bilal Riaz
Spectral Clustering and seriation: Theory and Algorithms.

In this project, we implemented spectral clustering and seriation algorithms and compared them to the different artificial data sets. The primary objective of this project is to understand seriation and its interconnection with clustering, more specifically spectral clustering. This project will be extended to explore new optimization schemes for seriation. In further work, seriation will be applied to physical data sets available online.