Department of Mathematics

February 7, 2022

Undergraduate Math Seminar

This week's seminar is **THURSDAY afternoon**. (The previously announced Tuesday seminar is cancelled.) Details for the seminar, including the title and abstract of the talk, can be found on the posters around Bailey Hall.

- **Thursday, February 10, 4:50-5:50 pm** in **Bailey 207.**

Pastries from Villa Italia will be available at the seminar!

Pieces from Thesis

Zhebin (Irene) Yin *wrote her senior thesis this past fall term under the direction of Professor Jue Wang.*

My thesis research was Diagnosing and Categorizing Lung Cancer using Deep Learning. We were trying to classify lung cancer types through the circulating tumor cells in liquid biopsy by using one of the most popular techniques in computer science — neural networks. This research is a perfect combination between Mathematics and Computer Science. I enjoyed using interdisciplinary learning to solve real-world complicated questions and challenges. This research experience with Professor Wang enhanced my enthusiasm for researching, exploring, and learning interdisciplinarily. It also triggered my desire to pursue further education in research and development.

In this research, we explored various neural networks models, including unsupervised and supervised learning models where each model is built on a lot of mathematical formulas. The models that we used are Convolutional Neural Network (CNN), Stochastic Gradient Descent (SGD), and U-Net architecture. The dataset for the model is extracted from the lung cancer patient's bloodstream by using liquid biopsy, and all in the format of RGB images (.tif). They are nothing but pixels. Each pixel is constructed by a vector of three numbers ranging from 0 to 255. To let the model better understand and recognize the object in the image, a lot of math formulas related to vectors and derivatives are needed. For example, in the SGD model, we used the chain rule and power rule to refine and revise the cost function to minimize the error. The derivative is one of the most popular methods for machine learning to optimize the problem.

The goal of this research is to see how the Deep Learning architectures would perform on classifying the types of lung cancers and see whether we could increase the accuracy and efficiency of classification biomedical images. We have two sets of datasets, and one of them contains images with and without segmentation. The accuracy of using CNN (Deep learning) is 95% while the accuracy of using SGD is 63.41% (note that the baseline is 25%). We found that CNN is particularly skilled at image classification for many datasets and classes. Although the SGD classifier needs less computational complexity and parameters, the accuracy of using SGD is lower than the accuracy of applying CNN on the dataset. In the future, we will try to use the U-Net model to get the segmentation of the biomedical images and to further explore and improve the accuracy of biomedical image classification.

I would highly encourage those who are interested in exploring to experience independent research with our highly decorated professors. I feel fortunate to have had this research experience which enhanced my skills of independent learning, researching, exploring, and asking questions. With the great help of Prof. Wang, who is resourceful, patient, motivated, and highly supportive, I realized the importance of self-learning, as well as my enthusiasm towards researching and exploring human-centric applications, which is my career goal.