Student: Angela Zhao
Mentor: Kenny Zhang
Winter 2021

In this Direct Reading Program, we started from the logistic regression, which is a basic machine learning algorithm for classification. Two real-life examples are how the computer predicts whether an email is spam or not and how the machine concludes whether it is a benign or malignant tumor. One important property of logistic regression is that its range always locates between 0 and 1, which is convenient for us to classify. What we need to remember that the input for logistic regression is not the original data, but rather it should be the multiplication of the weight vector of the original data. As analyzing the hypothesis function for logistic regression, we often set the decision boundary to be 0.5: when the predict value is less than 0.5, we classify into the type corresponding to 0, otherwise into that corresponding to 1.

For the logistic regression, the cost function is not similar to that of linear regression because of the convexity of the cost function after plugging in the sigmoid equation into the mean squared error. What a convex function matters is that we can directly use gradient descent to find the global minimum. However, a non-convex function has several local minimum points, and the process for finding a good initial point for reaching the global minimum is time-consuming and complicated.

Then we moved on to the neural network and convolutional neural network. I learned the basic structure of the network and what each layer does. Because the operation of these networks is hard to be expressed in a mathematical formula, they resort to backpropagation to learn the parameters that make the model have the lowest loss function. We also talked about an important theory in machine learning called bias and variance tradeoff, and we are trying to find the optimal model complexity with the least total error, which is the sum of squared bias and variance. During model selection, both under-fitting and over-fitting model should be avoided. Surprisingly, modern research shows that after a certain flexibility, the total error decreases again, and this “magical” double descent actually makes sense in statistical perspective. Finally, we constructed a simple convolutional neural network in Google Colab to classify a large dataset of images of dogs and cats. After the model was trained 30 epoch, it achieved 95 percent of train accuracy and 78 percent of test accuracy, which can be applied for future classification.

This project really provides me an opportunity to get touch on deep learning. In fact, techniques of image classification are common used in many industries. What I have learned in this 10 weeks helps me to better understand the algorithms used in the internship.