Project Updates: Image Classification with CIFAR-10 Dataset Tim Lieb tlieb@wisc.edu

17-November-2020

Progress

Since the initial proposal, I have started the implementation of my code. First, I came up with a way to read in the data. I spent some time figuring out the formatting and organization of the data and determined how I would be able to utilize it. The data is already split into 5 training sets and 1 test set, but for cross validation, I will use all 6 sets as training and testing for different iterations. This will allow me to better compare all the different algorithms.

Then I began working on the Ridge Regression algorithm by creating a function to perform the calculations and analysis. I use 5 of the data sets as the training data, and the I split the one hold out set into two sets for the testing. The function first takes a set of lambdas (still need to figure out what set/ range to use) and finds the w for each. Then it chooses the lambda that gives the lowest error rate using the first test set. Then with the best lambda and corresponding w, the last test set is used to find the error rate and the squared error. While my method seems right, the code takes a long time to run, and so I will look into ways to speed it up.

In my project proposal, I proposed using K-means to classify the data. This was a mistake because K-means is used for unsupervised learning which is not what I want. I was thinking of a K-nearest Neighbors approach, but I'm not certain this will work. I will need look into the issue more and determine if it'll fit my needs. If not, I will need to choose another algorithm.

Action Items

- Figure out how to determine the set of lambdas used for Ridge Regression
- Look into approaches to help speed up the Ridge Regression calculations
- Check about using K-nearest neighbors—come up with an alternative algorithm
- Start implementing the second algorithm once finalized
- Work on Neural Networks aspect once we've overed it in class