CSCI 6364 Machine Learning

Homework 1

Qinyang Li

For this homework, I used python to achieve KNN image classifier. Using existing libraries from Sklearn, Numpy, and Pandas. The Sklearn package have K Neighbors Classifier available, at the same time, this package also has scale function and train-test split algorithms.

First, I used 70% of the data to train my model and 30% of the train data to test my model. It was able to achieve 92% accuracy with K = 3. Then I decide to randomize my data so that the model I trained is more general. With 25% of data as test data, and split of data randomized, I was able to try different K value between 1 and 9 to find best accuracy of K. The result is shown follow. It shows that with this configuration, K=6 have best score of 84.7%. After select K = 6, I trained the model again and use it to predicate on test data. Then the results are saved in a csv file for submission.

|  |  |
| --- | --- |
| K | Score |
| 1 | 0.8015238095238095 |
| 2 | 0.8294285714285714 |
| 3 | 0.8308571428571428 |
| 4 | 0.8427619047619047 |
| 5 | 0.8414285714285714 |
| 6 | 0.8471428571428572 |
| 7 | 0.8423809523809523 |
| 8 | 0.8445714285714285 |
| 9 | 0.8433333333333334 |
|  |  |