Mertcan Aşgün

63948

Homework 02: Discrimination by Regression

Our aim was to classify the handwritten letters, given in excel files. For this purpose, I started with reading those files using “pandas”. For each letter, I divided the data sets and split them as training set and test set using “sklearn”. Then, I combined each individual part to get a training set consisting of each letter data, training label set, test set and test label set.

After I have the data in the desired format, I defined sigmoid function, which I will use as score function, and defined methods for gradient w and w0. I also defined a safelog method, to avoid infinity when we work with 0. Additionally, I declared eta and epsilon values and initialized w and w0. Then in an iterative way, I updated w and w0 values. Finally, I plot the error-iteration graph and calculated the confusion matrix for the training data.

Having decided on w and w0 values in training set, I apply the results on test set. And finally, I printed the confusion matrix for the test set.





