Trustworthy Machine Learning

HW 1

Shahar Sarfaty 305294498

Question #1 - Exploratory data analysis (EDA)

* How many samples, overall, are included in the dataset?
  + 28,911 samples overall.
* How many classes do the data belong to?
  + 4 classes.
* What are these classes?
  + The classes are tagged as {0,1,2,3}, and represent the actual handwritten digits 0-3.
* What is the dimensionality of the data?
  + Each datum is 28x28 pixels.
* What are the sizes of the training, validation, and test sets?
  + Training set size: 23,754 samples.
  + Validation set size: 1,000 samples.
  + Test set size: 4,157 samples.
* How many samples of each class are included in each set?
  + Training set: {0: 5713, 1: 6445, 2: 5721, 3: 5875}
  + Validation set: {0: 210, 1: 297, 2: 237, 3: 256}
  + Test set: {0: 980, 1: 1135, 2: 1032, 3: 1010}

Question #2 - Logistic regression

1. What is the test accuracy of each classifier?
   1. Binary Classifier – 97.93%
   2. Multiclass Classifier – 97.98%
2. Based on the visualizations, it seems that the classifiers work by applying a higher W values where it is probable to find the “ink” of the handwritten digit, and a vice versa. That means that both foreground and background features matter since the absence of “ink” in certain pixels can ease the job of classifying. For example, the existence of “ink” in the center of the image (represented by higher W values) is a good feature for classifying a digit as 1, While the absence of pixels (represented by lower W values) in the middle of the picture, is a good feature for classifying a digit as 0.

Furthermore, although learnt in a different process, the oddity W (Binary Classifier) seems to be an addition of the 1-W and 3-W, or a close approximation of it, where the center of the image is even more emphasized, as of being relevant to both 1 and 3.

Question #3 - Fully connected neural networks

1. What is the test accuracy of each classifier?
   1. Binary Classifier – 97.67%
   2. Multiclass Classifier – 98.20%