Name: Connor Raymond Stewart
ID: 20673233

CS 886 Homework Three:

Question 1 Calculate the gradient $g := \frac{\partial \ell(x,y)}{\partial w}$ given the data (x,y)

We know that: $\ell(x,y) = |-yw^T x + \epsilon \|w\|_{\frac{p}{p-1}}$

thus, we make the following simplification:

$$g = \frac{\partial}{\partial w}\left(|-yw^T x + \epsilon \|w\|_{\frac{p}{p-1}}\right)$$

$$= -yx + \epsilon \frac{w \circ |w|^{p-2}}{\|w\|_p^{p-1}}$$

Note:

$$\frac{\partial}{\partial w}(aw^T) = a \text{ so } \frac{\partial}{\partial w}(yw^T x) = yx$$

$$\frac{\partial}{\partial w}(\|w\|_p) = \frac{w \circ |w|^{p-2}}{\|w\|_p^{p-1}}$$

For Example:

When $p=1$:

$$g = -yx + \epsilon \frac{w \circ |w|^{-1}}{\|w\|_1^0} = -yx + \epsilon(w \circ |w|^{-1}) = -yx + \epsilon \frac{w}{|w|}$$

When $p=2$:

$$g = -yx + \epsilon \frac{w \circ |w|^0}{\|w\|_2^1} = -yx + \epsilon \frac{w}{\|w\|}$$

When $p=3$:

$$g = -yx + \epsilon \frac{w \circ |w|}{\|w\|_3^2}$$

Name: Connor Raymond Stewart
ID: 20673233

Question 2:

The accuracies of the three norm values are as follows:

p-norm	Accuracy Values for Norms
1-norm	99.1016548463357%
2-norm	100%
3-norm	100%

Edge cases are treated by taking all label predictions less than or equal to zero and treating them as cases aligned to negative one on the hyperplanes margin. Label predictions greater than zero are treated as cases aligned along positive one on the hyperplanes margin. Therefore, label predictions that are exactly zero or less are treated as cases of the digit zero from the MNIST dataset, whereas all cases are treated as cases of the digit one from the MNIST dataset. Finally, we notice that all cases greater than one or less the negative one still fit into the margin of one or negative one along the hyperplanes margin respectively.

Natural accuracies are good for all three cases, with the training dataset showing accuracies of one hundred percent for the two and three norm values, and approximately 99.1% for the one norm values.

MNIST datasets are downloaded from the classes drop-box as was the case for assignment two. An epsilon value of 0.01, an epoch value of 1000, a learning rate of 0.001, and norms ranging from one to three are used. Gradient descent is used along with the SVM linear classifier to classify numbers sampled from the MNIST dataset as was generated by the skeletal code provided by course instructor.

The data was tested on a desktop PC using a GTX1080 GPU, 16GB of RAM, and an intel i7-4770k CPU. Time taken to complete the execution of the code was roughly five minutes.