Assignment 2
DASC521 Fall 2019
Introduction to Machine Learning
Erhan Tezcan 0070881
18.10.2019

1. TASK

We are given 1000 images (28×28 pixels $= 784$ pixels), each image belonging to one of 5 classes which are types of clothes. For the implementation, we mainly use methods specified in chapter 10.8 of the book "Introduction to Machine Learning" by Ethem Alpaydın. We will split the data in half, first 500 to be used for training and the remaining 500 to be used for testing.

2. IMPLEMENTATION

We are going to do Discriminaton by Regression for multiple classes. We will be using sigmoid function, which is defined as $\sigma : \mathbb{R} \rightarrow (0, 1)$

$$\sigma(a) = \frac{1}{1 + e^{-a}} \tag{1}$$

In this case, we will apply sigmoid every element in the result $X.W + W_0$, where X is the input matrix, where each row represents an image as an array of pixels, W is the weight matrix and W_0 is the bias.

It is important to note that we do One-Hot Encoding to our labels before we proceed, so that we can compare them to our predictions.

The error function we use is the sum of squared errors:

$$\mathrm{E}(w, w_0 | X) = \frac{1}{2} \sum_t (r^t - y^t)^2 \tag{2}$$

We want to minimize this function, thus by using Gradient Descent we get the values below for adjusting weights and biases:

$$\Delta W = \eta \sum_t (r_i^t - y_i^t) \times y_i^t \times (1 - y_i^t) \times x^t \tag{3}$$

$$\Delta W_0 = \eta \sum_t (r_i^t - y_i^t) \times y_i^t \times (1 - y_i^t) \tag{4}$$

As stated in the book, after certain amount of iterations the correct label should have a larger predicted value, compared to rest, which we then find by taking the maximum of the predictions per class. Our loop has two termination conditions:

(1) Maximum iteration count is reached.

(2) The total change of W and W_0 is less than some small ϵ.

We plot how the error changes by iteration, and we get the result below:
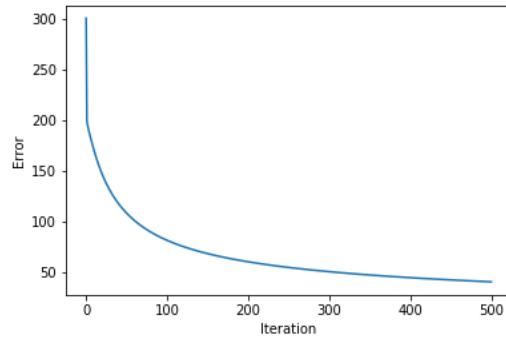
Overall, our accuracies are:

- Training Data: 94%
- Test Data: 93.2%