Project Proposal
Data Mining

Luuk Arts, `s4396863`
Tristan de Boer `s1007313`

November 2017

1   Algorithm

The algorithm we will be implementing is an artificial Neural Network (NN). We will first implement this NN with one input layer with 28×28 nodes and one output layer of 1 node, which are connected by edges to which weights are assigned and a sigmoid function. The output of the initial NN will be determined by the following function.

$$output = input \cdot weights$$

The error is then calculated, and the weights are updated depending on the error using the following function.

$$weights+ = input \cdot (error \times \frac{\partial}{\partial x}(\frac{1}{1 + exp(-x)}))$$

Afterwards, we will experiment with different numbers of layers to see if that improves our results. The purpose of our implementation will be the classification of handwritten digits.

2   Data

The data we use to train and test our Neural Network is the MNIST handwritten digit dataset (http://yann.lecun.com/exdb/mnist/). This data consists of a collection of black and white images of the digits $0 - 9$ with a size of 28×28 pixels. We will normalize this data to values between 0 and 1 so we can easily use it in our neural network.

References

[1] Christopher M Bishop. Neural networks. In *Pattern recognition and machine learning*, pages 225–284. Springer, 2006.

[2] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[3] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962, 2003.