Data Classification Using Neural Networks: Configuring the hyperparameters of a multilayer perceptron to maximise learning speed and accuracy when classifying handwritten digits.

Key Words:

"Neural Network", "Multilayer Perceptron", "Data Classification", "MLP", "Activation Function" "Learning Rate", "MNIST", "Handwritten Digits", "Rectified Linear Unit","ReLU", "Leaky ReLU", "Sigmoid", "Hyperbolic Tangent".

Abstract:

There has been a growing societal interest in neural networks recently because of the many potential applications such as in image recognition, a specific case being recognising handwritten digits. Neural networks such as the multilayer perceptron (MLP) have many hyperparameters, such as the learning rate and activation function used. These hyperparameters are altered to increase the speed at which neural networks learn and their accuracy at a given task. This research project's goal is to find the best learning rate and activation function, from a given set of functions, for a multilayer perceptron that has the goal of classifying handwritten digits.

The MLPs used in experimentation each had four layers, the first layer being the input layer with 784 neurons, the two intermediate layers having 200 each, and the output layer having 10. The weights and biases for the MLPs were initialised randomly. The MLPs were trained on the MNIST training dataset for three epochs, with inputs rescaled to take values between 0.01 and 1.0. The activation functions selected for the comparison were the sigmoid, hyperbolic tangent, rectified linear unit (ReLU) and the Leaky ReLU. For each activation function, the learning rate which gave the highest accuracy on the MNIST testing dataset was selected for comparison on how accuracy improved with the number of iterations.

From the experiments conducted, the learning rates that gave the best accuracy for the sigmoid, hyperbolic tangent, ReLU and Leaky ReLU were 0.3, 0.0084, 0.0072 and 0.012, respectively. With these learning rates, LeakyReLU and Sigmoid appeared to gain a high accuracy on the MNIST test dataset with, after a sharp initial increase, a steady increase of accuracy, ReLU had jumps of accuracy every few thousand iterations, and the accuracy oscillated with hyperbolic tangent. The four MLPs compared had an accuracy approaching approximately 97% as the number of iterations increased. Further research could investigate the effect of different rescaling of inputs on the accuracy of the MLP.