Udaikaran Singh and Wesley Kwan

Gary Cottrell

CSE 190

10/23/2018

**PA2: Mutli-Layer Neural Network on MNIST dataset**

**Udaikaran Singh Wesley Kwan**

Department of Computer Science Department of Computer Science

University of California, San Diego. University of California, San Diego

La Jolla, CA 92093. La Jolla, CA 92093

**Abstract**

[add abstract]

(a) Reading in MNST data

The hyperparameters we used for the neural network was training on all 50,000 training samples. The training type is stochastic gradient descent. Momentum was implemented based on method defined by Geoffrey Hinton of adding 0.5 to the

(b) Gradient Checker

We used epsilon = 0.01.

|  |  |  |
| --- | --- | --- |
|  | Gradient Approximation  E(w+e) – E(w-e) / 2e | Actual Gradient |
| Input to hidden weight 1 | 0.0 | -1.3771253495617533e-15 |
| Input to hidden weight 2 | -5.639302358417808e-08 | -2.194892411834406e-05 |
| Hidden bias weight | 0.0 | 1.3771253492539445e-15 |
| Hidden to output weight 1 | 0.0 | -1.3771253492539445e-15 |
| Hidden to output weight 2 | -5.639302358417808e-08 | -2.194892411834406e-05 |
| Output bias weight | -0.0005879615482928813 | -0.031112308600483817 |

All the gradients are approximately equal (within epsilon squared), except for the output bias weight gradient.

(c) Finding Optimal Number of Epochs

(d) Experimenting with Regularization

(e) Experiment with Activations

(f) Experiment with Network Topology

Individual Contributions

[add paragraph by each person of what they did]