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

(b) Gradient Checker

We used epsilon = 0.1.

|  |  |  |
| --- | --- | --- |
|  | Gradient Approximation  E(w+e) – E(w-e) / 2e | Actual Gradient |
| Input to hidden weight 1 | 0.0 | 3.836095938522184e-17 |
| Input to hidden weight 2 | 7.227379809904289e-07 | -1.5189384418679225e-06 |
| Hidden bias weight | 0.0 | -3.836095938522184e-17 |
| Hidden to output weight 1 | 6.245004513516506e-16 | -1.6111442405766055e-16 |
| Hidden to output weight 2 | -5.551115123125783e-16 | 5.46151463875278e-16 |
| Output bias weight | 3.366672782868785e-05 | -6.40231002820909e-05 |

All the gradients are approximately equal (within epsilon squared).

(c) Finding Optimal Number of Epochs

(d) Experimenting with Regularization

(e) Experiment with Activations

Note: found that sigmoid takes considerably longer to train than tanh

(f) Experiment with Network Topology

Individual Contributions

[add paragraph by each person of what they did]