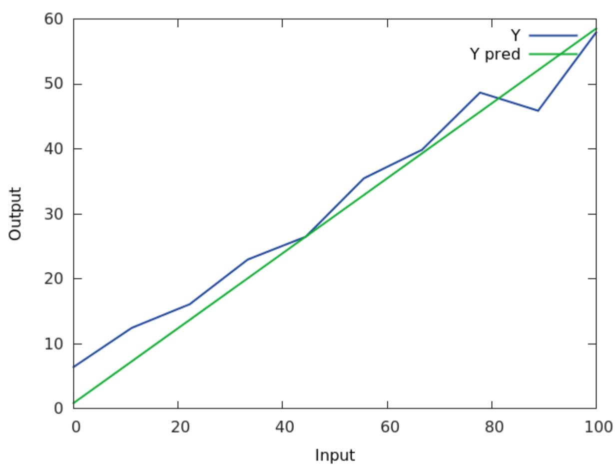
CS 8803 DL Assigment 1

Jingdao Chen

Jan 26

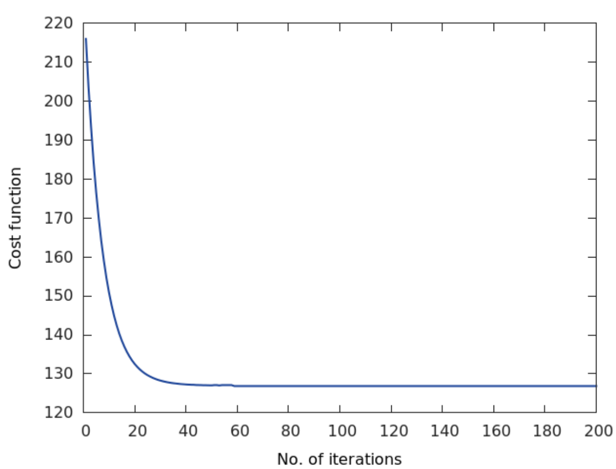
1 Problem 1

The plot of the estimated linear function is shown below:



]

The plot of cost function vs. no. of iterations is shown below:



]

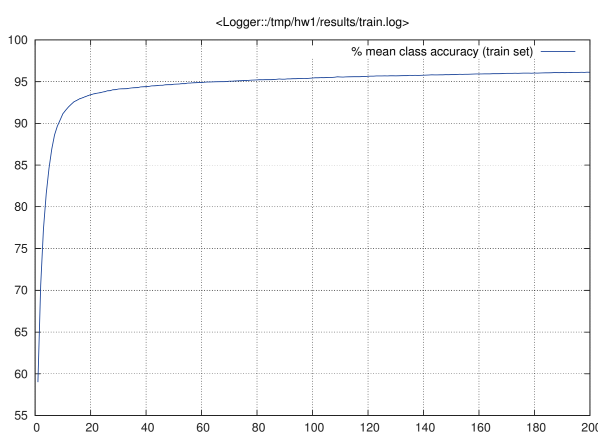
2 Problem 2

2.1 Linear Layer

Let the input be x_0 and the output be x_1. The linear layer is calculated as $x_1 = w_1^T x_0 + b_1$.

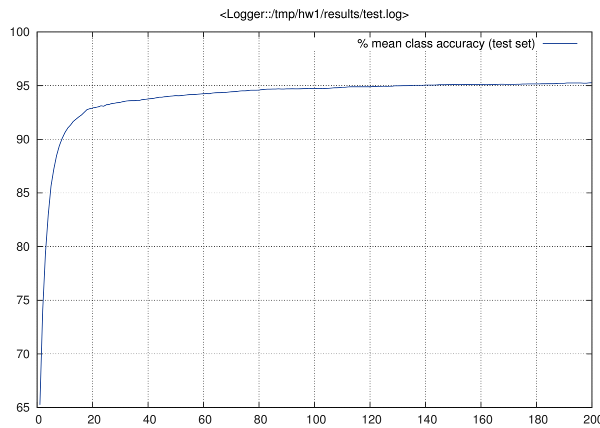
$$\begin{aligned}
\frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial w_1} \\
&= \frac{\partial E}{\partial x_1} x_0 \\
\frac{\partial E}{\partial b_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial b_1} \\
&= \frac{\partial E}{\partial x_1}\mathbf{1}
\end{aligned} \tag{1}$$

The plot of training accuracy vs. no. of epoch for is shown below:



]

The plot of testing accuracy vs. no. of epoch for is shown below:



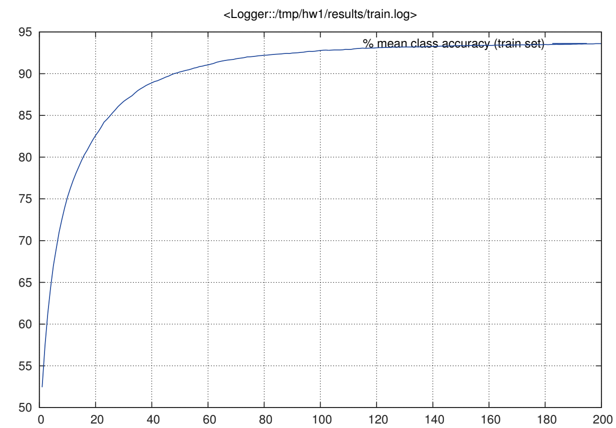
]

2.2 Linear-Sigmoid Layer

Let the input be x_0 and the output be x_2. The linear layer is calculated as $x_1 = w_1^T x_0 + b_1$. The sigmoid layer is calculated as $x_2 = \sigma(x_1)$.

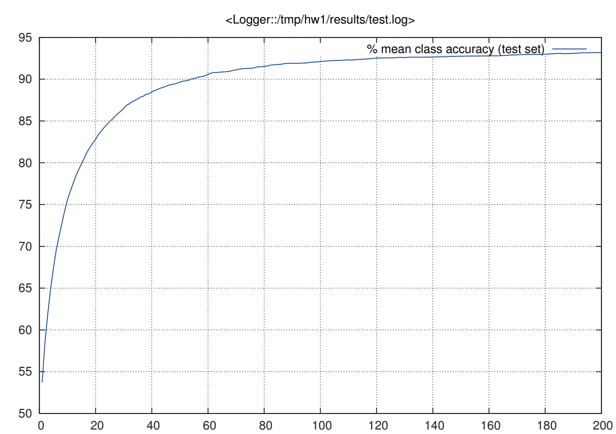
$$\begin{aligned}
\frac{\partial E}{\partial x_1} &= \frac{\partial E}{\partial x_2}\frac{\partial x_2}{\partial x_1} \\
&= \frac{\partial E}{\partial x_2}(x_2)(1 - x_2) \\
\frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial w_1} \\
&= \frac{\partial E}{\partial x_1}x_0 \\
\frac{\partial E}{\partial b_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial b_1} \\
&= \frac{\partial E}{\partial x_1}\mathbf{1}
\end{aligned} \tag{2}$$

The plot of training accuracy vs. no. of epoch for is shown below:



]

The plot of testing accuracy vs. no. of epoch for is shown below:



]

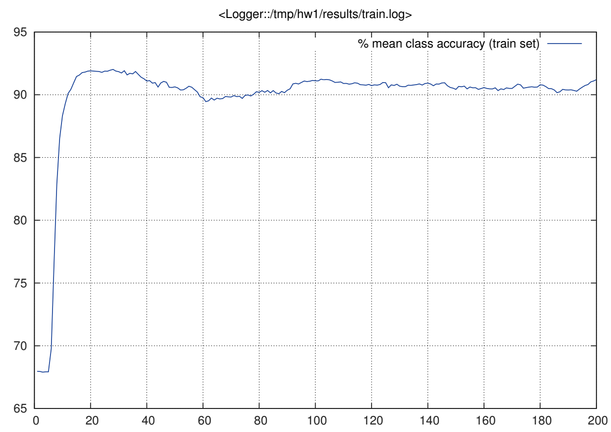
2.3 Linear-Sigmoid-Linear Layer

Let the input be x_0 and the output be x_3. The first linear layer is calculated as $x_1 = w_1^T x_0 + b_1$. The sigmoid layer is calculated as $x_2 = \sigma(x_1)$. The second linear layer is calculated as

$x_3 = w_2^T x_2 + b_2$.

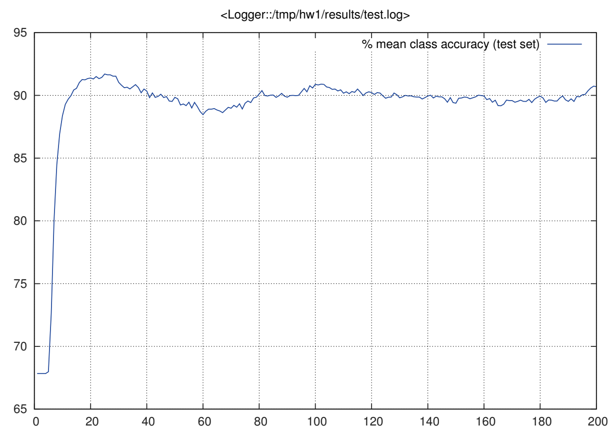
$$
\begin{aligned}
\frac{\partial E}{\partial w_2} &= \frac{\partial E}{\partial x_3}\frac{\partial x_3}{\partial w_2} \\
&= \frac{\partial E}{\partial x_3} x_2 \\
\frac{\partial E}{\partial b_2} &= \frac{\partial E}{\partial x_3}\frac{\partial x_3}{\partial b_2} \\
&= \frac{\partial E}{\partial x_3}\mathbf{1} \\
\frac{\partial E}{\partial x_2} &= \frac{\partial E}{\partial x_3}\frac{\partial x_3}{\partial x_2} \\
&= \frac{\partial E}{\partial x_3} w_2 \\
\frac{\partial E}{\partial x_1} &= \frac{\partial E}{\partial x_2}\frac{\partial x_2}{\partial x_1} \\
&= \frac{\partial E}{\partial x_2}(x_2)(1 - x_2) \\
\frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial w_1} \\
&= \frac{\partial E}{\partial x_1} x_0 \\
\frac{\partial E}{\partial b_1} &= \frac{\partial E}{\partial x_1}\frac{\partial x_1}{\partial b_1} \\
&= \frac{\partial E}{\partial x_1}\mathbf{1}
\end{aligned}
\tag{3}
$$

The plot of training accuracy vs. no. of epoch for is shown below:



]

The plot of testing accuracy vs. no. of epoch for is shown below:



]

3 Gradient check

One-sided finite difference approximation is calculated by:

$$\nabla J = \frac{J(\theta + \epsilon) - J(\theta)}{\epsilon} \tag{4}$$

Two-sided finite difference approximation is calculated by:

$$\nabla J = \frac{J(\theta + \epsilon) - J(\theta - \epsilon)}{2\epsilon} \tag{5}$$

The two-sided version is more accurate because taking the linear approximation of $J(\theta)$

on both sides of θ evens out the error.