ECE 544NA Fall 2016 Assignment 2

Jifu Zhao Date: 09/29/2016

I. Pencil-and-Paper

1. Derivative of Softmax
Recall the softmax function:

$$\vec{y}[k] = \frac{e^{\vec{z}[k]}}{\sum_{l=1}^{C} e^{\vec{z}[l]}} \tag{1}$$

So, consider two cases, j = k and j != k.

When j = k:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = -\frac{e^{\vec{z}[k]} \cdot \sum_{l=1}^{C} e^{\vec{z}[l]} - e^{\vec{z}[k]} \cdot e^{\vec{z}[k]}}{(\sum_{l=1}^{C} e^{\vec{z}[l]})^2} = \vec{y}[k] - \vec{y}[k] \cdot \vec{y}[k] \tag{2}$$

When j != k:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = -\frac{e^{\vec{z}[k]} \cdot e^{\vec{z}[j]}}{(\sum_{l=1}^{C} e^{\vec{z}[l]})^2} = -\vec{y}[k] \cdot \vec{y}[j] \tag{3}$$

So, we can conclude that:

$$\frac{\partial \vec{y}[k]}{\partial \vec{z}[j]} = \begin{cases} -\vec{y}[k] \cdot \vec{y}[j] & \text{if } k! = j \\ \vec{y}[k] - \vec{y}[k] \cdot \vec{y}[k] & \text{if } k = j \end{cases} \tag{4}$$

2. Negative Log Likelihood loss for Multi-Class.
Recall the negative log likelihood:

$$L = -\sum_{i}^{N} \sum_{k}^{K} \mathbf{1}[y_i = k] \cdot log(\hat{\vec{y}}_i[k]) \tag{5}$$

In this way, we have:

$$\frac{\partial L}{\partial \hat{\vec{y}}_i[j]} = -\mathbf{1}[y_i = k] \cdot \frac{\partial log(\hat{\vec{y}}_i[j])}{\partial \hat{\vec{y}}_i[j]} = -\frac{\mathbf{1}[y_i = j]}{\hat{\vec{y}}_i[j]} \tag{6}$$

3. Avg-pooling (1D)
Recall Avg-pooling (1D) operation with window size W:

$$\vec{y}[i] = \frac{1}{W} \sum_{j=0}^{W} \vec{x}[i+j] \tag{7}$$

Then, we have:

$$\frac{\partial \vec{y}[i]}{\partial \vec{x}[j]} = \begin{cases} -\frac{1}{W} & \text{if } i \le j \le i + W \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

4. Max-pooling (1D)

Recall Max-pooling (1D) operation with window size W:

$$\vec{y}[i] = \max_{j=0}^{W} \vec{x}[i+j] \tag{9}$$

Then, we have:

$$\frac{\partial \vec{y}[i]}{\partial \vec{x}[j]} = \begin{cases} 1 & \text{if } \max_{k=0}^{W} \vec{x}[i+k] = \vec{x}[j] \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

5. Convolutional layer (1D)

Recall Convolution (1D) operation, assume $\vec{w}$ is length 3, and zero index at the center:

$$\vec{y}[i] = (\vec{w} * \vec{x})[i] = \sum_{j=-1}^{1} \vec{x}[i-j] \cdot \vec{x}[j] \tag{11}$$

From above equation, we have:

$$\frac{\vec{y}[i]}{\vec{x}[j]} = \begin{cases} \vec{w}[i-j] & \text{if } i-1 \leq j \leq i+1 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

$$\frac{\vec{y}[i]}{\vec{w}[j]} = \begin{cases} \vec{x}[i-j] & \text{if } j = -1 or 0 or 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

II. Code-from-Scratch

1. Methods

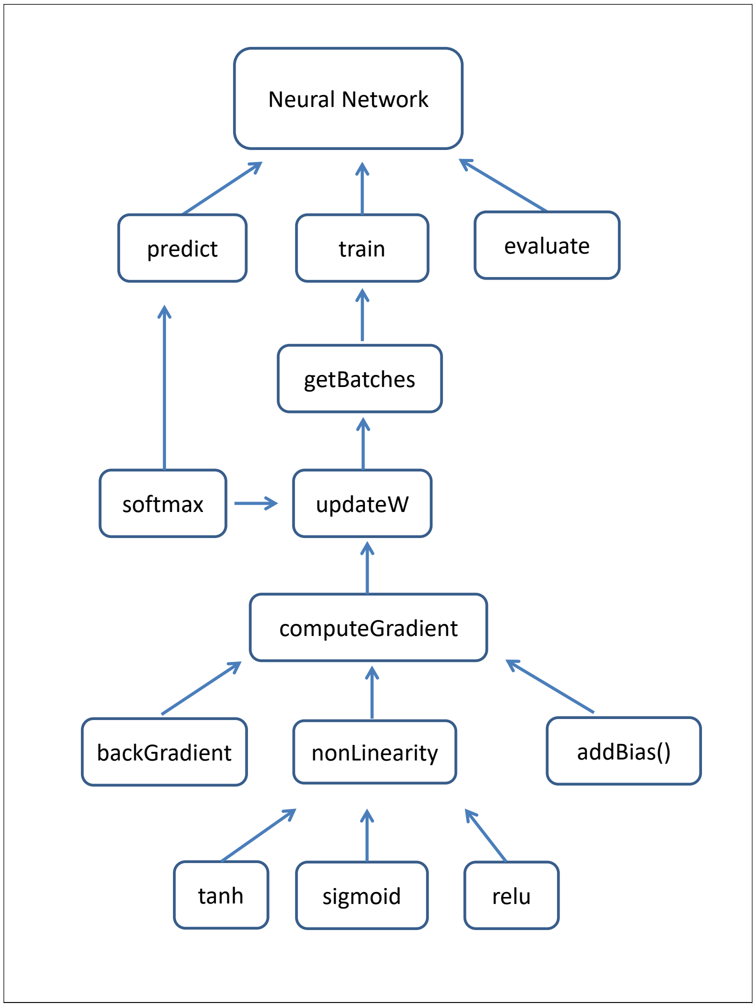

Figure 1: Algorithm Structure

2. Results

Table 1: Comparison of training accuracy

Hidden Nodes/Function	ReLU	Sigmoid	Tanh
10	66.42%	57.25%	63.16%
20	87.94%	76.72%	59.21%
30	60.71%	78.08%	75.35%
40	56.84%	83.69%	77.91%
50	69.90%	83.67%	79.72%

Table 2: Comparison of test accuracy

Hidden Nodes/Function	ReLU	Sigmoid	Tanh
10	39.38%	43.01%	41.03%
20	43.34%	46.09%	44.44%
30	43.23%	44.44%	44.11%
40	35.42%	45.10%	46.09%
50	40.70%	46.31%	43.01%

Table 3: Average time for one iteration (in seconds)

Hidden Nodes/Function	ReLU	Sigmoid	Tanh
10	0.00067	0.00309	0.00509
20	0.00073	0.00330	0.00535
30	0.00082	0.00352	0.00555
40	0.00092	0.00366	0.00582
50	0.00105	0.00383	0.00614

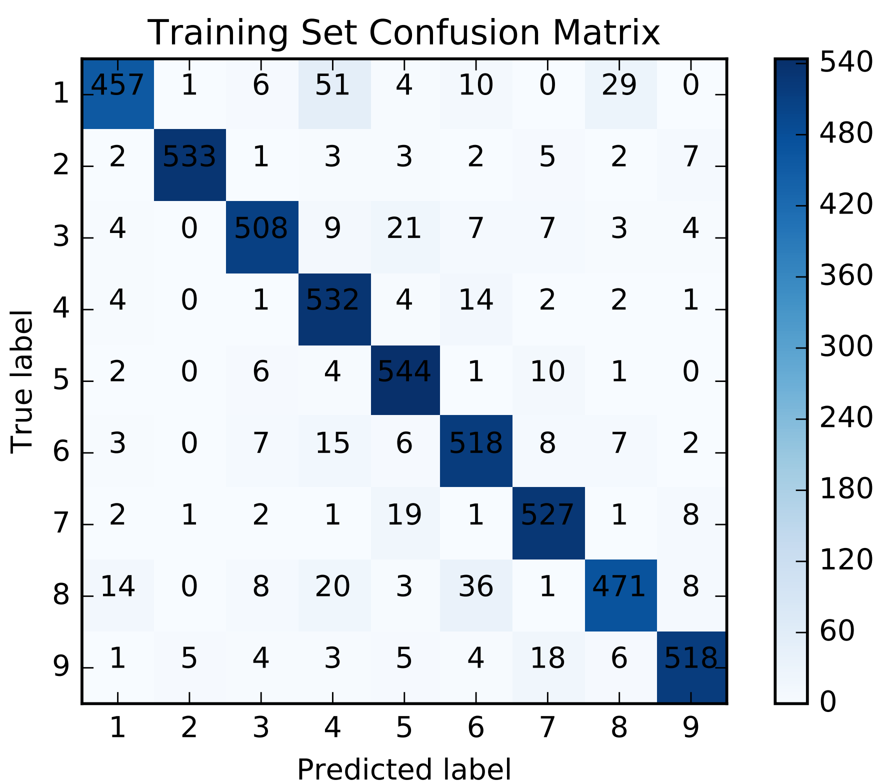

Figure 2: Training Set Confusion Matrix

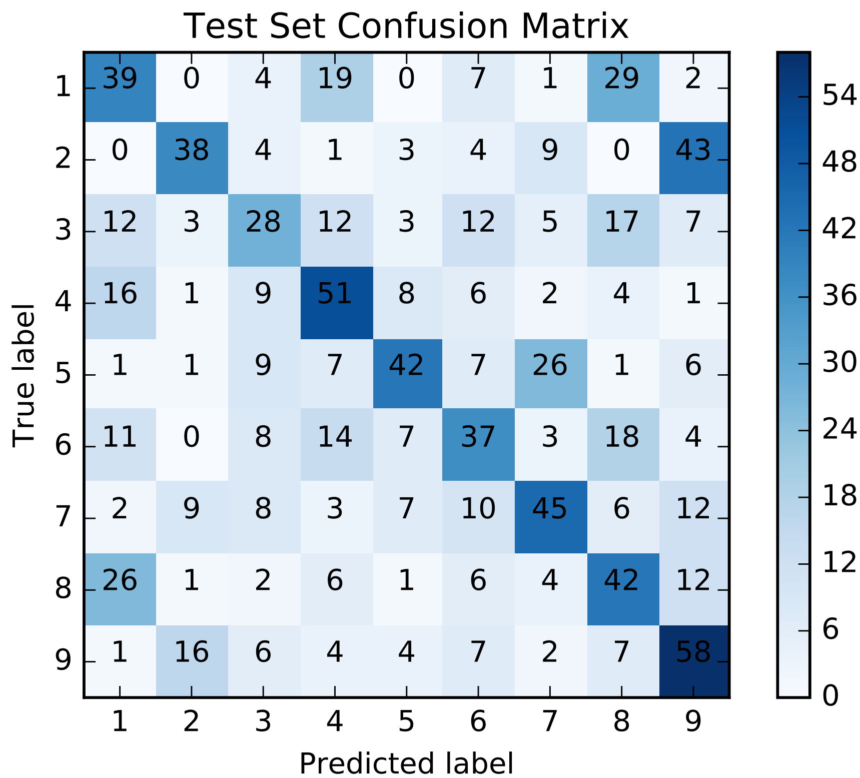

Figure 3: Test Set Confusion Matrix

III. TensorFlow

1. Methods

2. Results