CS 224D Spring 2016 Assignment #3

SUNet ID: 06106557

Name: Gregory Lambert

1 RNN's (Recursive Neural Network)

(a)

$$
\begin{aligned}
\hat{y} &= softmax(\theta) \\
z^{(1)} &= [h_{left}^{(1)}, h_{right}^{(2)}]W^{(1)} + b^{(1)} \\
ReLU(z^{(1)}) &= max(z^{(1)}, 0)
\end{aligned}
$$

$Node$ 1

$$
\begin{aligned}
\delta_3 &= \hat{y} - y \\
\delta_2 &= U^T\delta_3 \circ ReLU^{'}(z^{(1)}) \\
\delta_{below} &= {W^{(1)}}^T\delta_2 \\
\frac{\partial J}{\partial U} &= \delta_3 {h^{(1)}}^T \\
\frac{\partial J}{\partial b^{(s)}} &= \delta_3 \\
\frac{\partial J}{\partial W^{(1)}} &= \delta_2 \left[h_{left}^{(1)} h_{right}^{(1)}\right]^T \\
\frac{\partial J}{\partial b^{(1)}} &= \delta_2
\end{aligned}
$$

$Nodes$ 2 & 3

$$
\begin{aligned}
\delta_3 &= \hat{y} - y \\
\delta_2 &= \delta_{above} \circ ReLU^{'}(z^{(1)}) \\
\delta_{below} &= {W^{(1)}}^T\delta_2 \\
\frac{\partial J}{\partial w^{(1)}} &= \delta_2 \left[h_{left}^{(1)} h_{right}^{(1)}\right]^T \\
\frac{\partial J}{\partial b^{(1)}} &= \delta_2 \\
\frac{\partial J}{\partial b^{(s)}} &= \delta_3 \\
\frac{\partial J}{\partial L} &= {W^{(1)}}^T\delta_2
\end{aligned}
$$

(b)

(a,b,c) see *rnn.py*

(d) Final training accuracy 1.0 and final validation accuracy 0.77. Best validation accuracy for all 30 epochs was 0.79 with default hyperparameters.

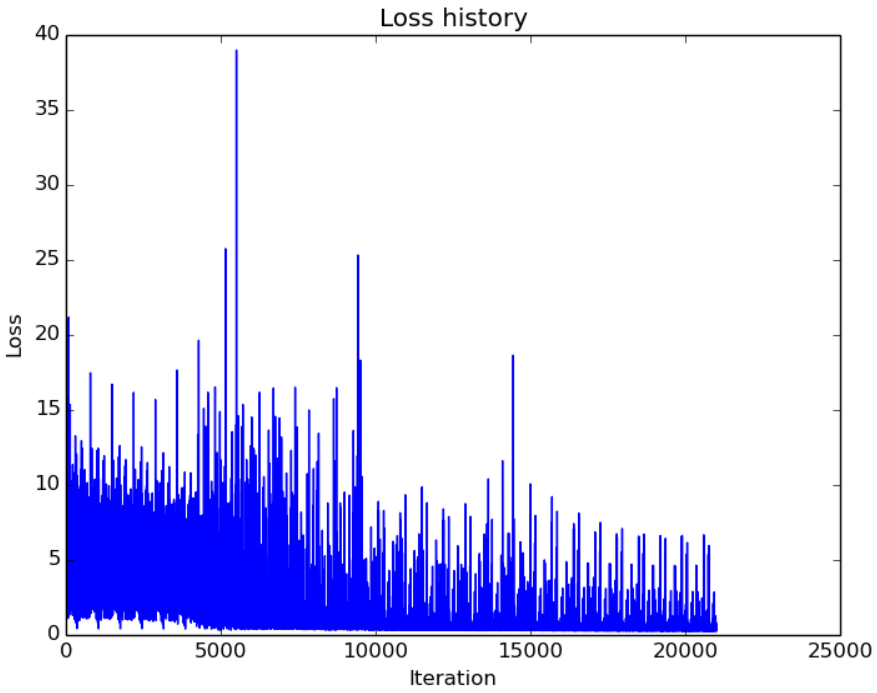

Figure 1: RNN loss plot default settings

embed_size = 35

label_size = 2

early_stopping = 2

anneal_threshold = 0.99

anneal_by = 1.5

max_epochs = 30

lr = 0.01

l2 = 0.02

(e) We will tune two hyperparameters for dealing with overfitting, dropout and L2 regularization in an attempt to improve the generalizability of our RNN model and increase validation accuracy. Our default L2 hyperparameter is 0.02, taking small incremental changes (e.g. 0.025,0.030,0.035, 0.040) we did see some improvement (final and best validation accuracy 0.82) with a regularization strength, $\lambda = 0.030$. Next we experimented with applying dropout to the output of our hidden layer using default hyperparameters and L2 regularization ($\lambda = 0.030$) in an attempt to remove overfitting of the model to the training data. Our final model with L2 regularization ($\lambda = 0.030$) and dropout achieved a best validation accuracy of 0.86.

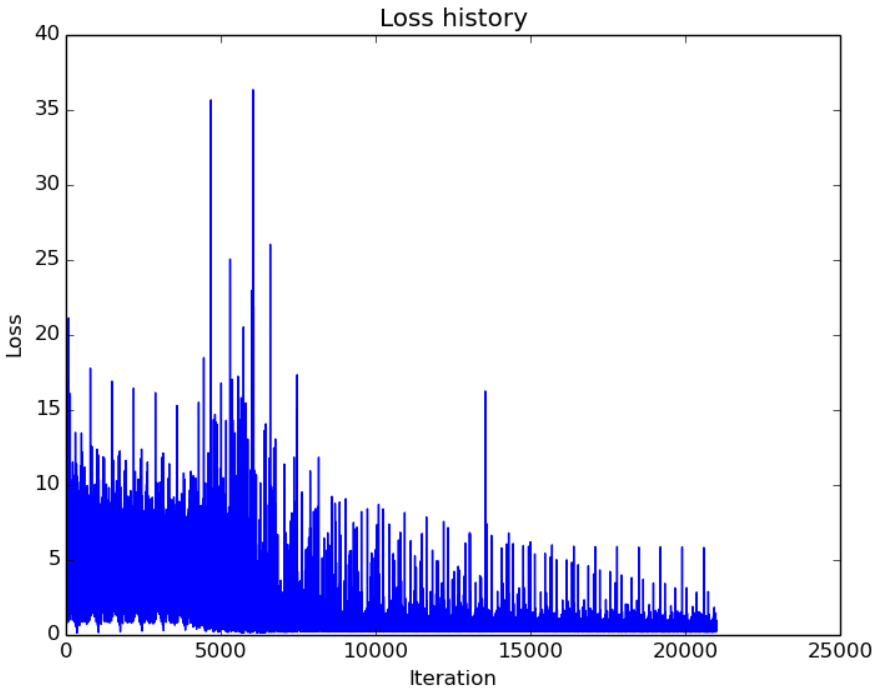

Figure 2: Regularization strength ($\lambda = 0.03$)

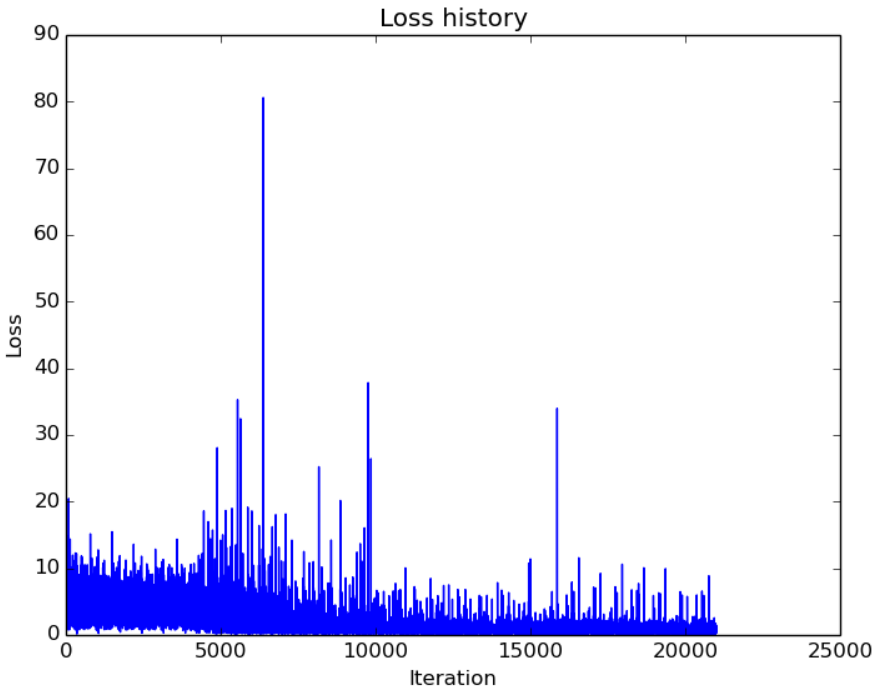

Figure 3: Regularization strength ($\lambda = 0.03$) with dropout (0.80)