Assignment 2 Report

This report explains the basic logic of RNN in mathematical expression and shows the training result of sentiment analysis using RNN. The model is fine tuned to achieve best result. Units and layers are explored in finding such a model.

1. **RNN step forward**

Given input data at timestep t, hidden state from previous timestep , , , and b, we calculate by

1. **RNN step backward**

Given , gradient of loss with respect to next hidden state at time t, and with , we have

)

)

)

)

)

1. **RNN Forward**

To run a valila RNN forward on an entire sequence of data input, we just loop through T vectors for each input. From timestep t = 0 toward timestep t = T – 1, in each timestep/iteration, we have . is initialized to a zero matrix as there’s no previous h at t = 0.

1. **RNN Backward**

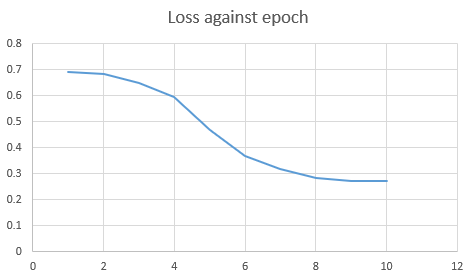
In RNN backward, for each parameter, its gradient is just aggregation of gradients across all positions:

1. **Sentiment Analysis**

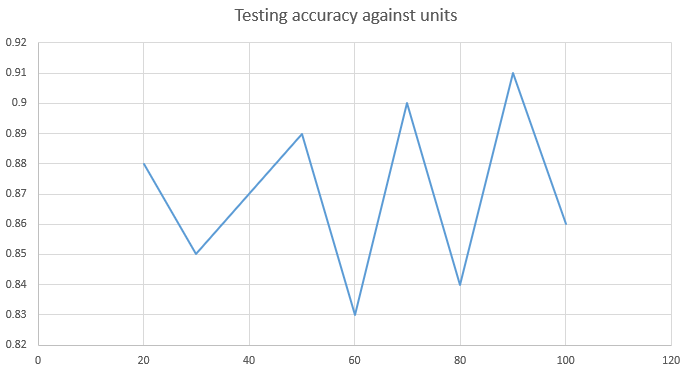
In the forward step, we pass forward data through five layers, namely, fully connected layer, bidirectional RNN layer, fully connected layer, temporal pooling layer and another fully connected layer. Loss is calculated by taking softmax and cross-entropy of the final result of forward step. In the backward step, we calculate gradients by going backward in the order inversely to forward process.

1. **Training Result**

The training batch is set as 20, valuation batch set as 100, test batch set as 100, epochs set as 10. The iteration loss is depicted as following.



1. **Model tuning**
2. Tuning units



The hidden units are tuned from 20 to 100. The testing accuracy against number of hidden unit result is given above. As shown in the chart, the highest testing accuracy is achieved at hidden units = 90. With hidden units increasing, the testing accuracy also demonstrates an increasing trend. Intuitively, this can be understood that more hidden units can capture more input information which will lead to a better testing result.

1. Tuning input features

Next, the hidden unit is fixed at 90, we tune input features from 200 to 600 in step of 50. The experiment result is shows below. It is observed that the testing accuracy peaked at 400. More features means that the model is pruned to overfitting; less feature means the model is pruned to under fitting.

