Deep Learning - HW3 Report

Nadav Oved - 200689768
Aviv Sugarman - 305652729
Github Link: https://github.com/nadavo/DeepLearningCourse/tree/hw3

Model Architecture

6 layer network built according to the following structure:

```
nn.Sequential {
  [input -> (1) -> (2) -> (3) -> output]
  (1): nn.LookupTable
  (2): nn.Sequential {
    [input -> (1) -> (2) -> (3) -> (4) -> output]
    (1): nn.LSTM(212 -> 212, 424)
    (2): nn.Dropout(0.330000)
    (3): nn.LSTM(212 -> 212, 424)
    (4): nn.Dropout(0.330000)
  }
  (3): nn.TemporalModule {
    [input -> (1) -> output]
    (1): nn.Linear(212 -> 10000)
  }
}

Number of parameters:      2850800
```

We based our network model on Elad Hoffer's recurrent.torch implementation with 2 LSTM layers, a Dropout layer after each LSTM layer and a Linear layer before output.
We used the following parameters:

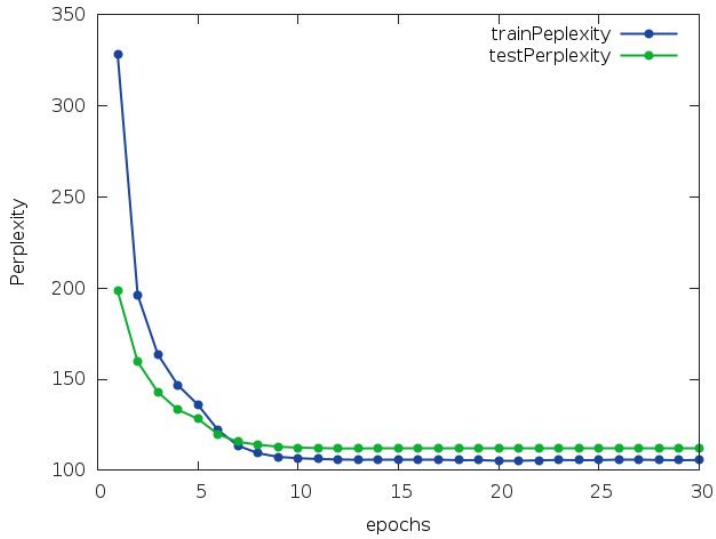
- LSTM hidden layer size = 212
- Dropout layer p-value = 0.33
- Batch Size = 50
- Learning rate = 0.0025
- Sequence Size = 10
- Loss Criterion = CrossEntropyCriterion
- Optimization Algorithm = RmsProp
- Total # of Epochs = 30

Training Procedure

After constructing the full network model according to the architecture described above, and defined the parameters noted above, we proceeded to run the following flow for 30 epochs:

1. Input the entire training set in batches of 50, were for each batch we did:
    a. Forward propagation of data samples into network
    b. Loss and Perplexity calculations according to our defined loss criterion
    c. Backward propagation through time to calculate gradients for network weights using RmsProp as our optimization algorithm.
2. Calculate loss and perplexity values for network predictions on validation and test sets respectively.
3. Sample 10 words from our network to complete the sentence "Buy low, sell high is the"
4. Decrease learning rate by ½ if validation loss didn't improve.

Perplexities Graph



Summary and Conclusion

In order to comply with this exercise's perplexity and number of parameters constraints, the main parameters we tuned by trial and error are described in the Model Architecture section. We achieved our final Test Perplexity of `111.97878265381` after 25 epochs and from the graph we can't see any major improvements after epoch 13, meaning the validation loss didn't improve and as a result the learning rate decreased exponentially.

Best Results + 5 Sentence Completions:

```
Epoch 29
Training Perplexity: 105.61737823486
Validation Perplexity: 119.69146728516
Sampled Text:
Buy low, sell high is the... volume of the fed trading they expects only his own
Test Perplexity: 111.97878265381
Learning Rate decreased to: 7.4505805969238e-11

Epoch 30
Training Perplexity: 105.6735458374
Validation Perplexity: 119.69146728516
Sampled Text:
Buy low, sell high is the... platform or high only its paper business that has been
Test Perplexity: 111.97878265381
Learning Rate decreased to: 3.7252902984619e-11

Best Iteration was 29, With a validation loss of: 4.7849174194595
Sentence 1:
Buy low, sell high is the... increase for example the help will be <unk> prices that
Sentence 2:
Buy low, sell high is the... differences of shares the bonds are n't slowing in s&p
Sentence 3:
Buy low, sell high is the... result of the fed who are virtually very major traders
said
Sentence 4:
Buy low, sell high is the... u.s. theory allow us to do if many investors have
Sentence 5:
Buy low, sell high is the... direct business there 's hard <unk> mr. roberti says
following
```