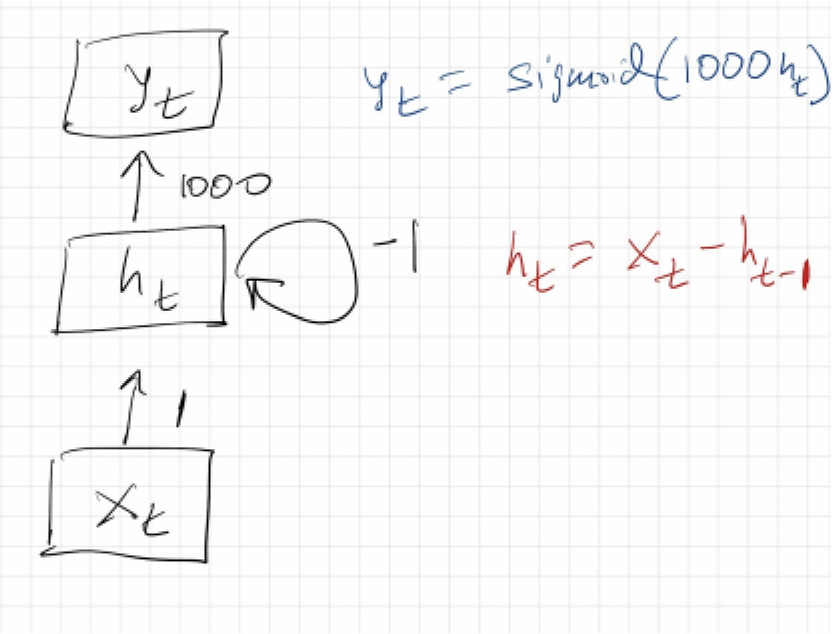
Deep Learning (Spring 2022) Homework 3

Assigned: Apr 14, 2022

Due: Apr 28, 2022

This homework is to be done individually. Please submit your solutions to the theory questions in a single PDF document.

1. **Recurrences using RNNs (30 points):** Consider the recurrent network architecture below in Figure 1. All inputs are integers, hidden states are scalars, all biases are zero, and all weights are indicated by the numbers on the edges. The output unit performs binary classification. Assume that the input sequence is of even length. What is computed by the output unit at the final time step? Be precise in your answer. It may help to write out the recurrence clearly.



2. **Encoder-Decoder Architecture for Translation(20 points):**
   1. Why do we use Encoder-Decoder Architecture rather than plain sequence-to-sequence RNNs?
   2. What is the attention mechanism? How does it improve the translation task?

3. **Embeddings for NLP applications (20 Points):**
   1. [10 points] State two differences between embeddings generated by BERT and Word2Vec.
   2. [10 points] Why do we need embeddings in the first place? Can’t we use one-hot encoding instead?

4. Faster Attention (30 points): Consider the standard dot-product self-attention mechanism that computes alignment scores between all pairs of input symbols; so if there are n tokens in a sequence this requires computing n^2 query-key dot products. In this problem, let us try to make this more efficient.

1. Consider autoregressive self-attention where every token only attends to its own position and all previous positions. Calculate how many dot-products are now required as a function of n.
2. Consider strided self-attention where every token attends to at most t positions prior to it, plus itself. Calculate how many dot-products are required as a function of n and t.
3. Consider windowed self-attention where the n tokens are partitioned into windows of size w (assume w divides n), and every token attends to all positions within its window and prior to it, plus itself.