Date : _____

DL   CS 583 A   Fall 2021

Quiz 3

December 1, 2021

Instructor : Jia Xu

Student name : Shreeya Kokate

Student ID : 20005256

Student email address :   skokate@stevens.edu
                           shreeya.csai@gmail.com

Q)

Ans Comparing attention models & RNNs we can say that,

Advantages

1. Instead of RNN which have maximum path length across time propostional to the number of time steps, Attention based models have constant path length between the encoder input & decoder hidden state. Because of this learning becomes easicar.

2. Attention based models achieve efficient content based addressing at the cost of recomputing context vectors at each time step.

3. RNN architecture limits the parallelism potential for longer sequences whereas attention models supports this & are best at it.

4. Unlike RNN encoders, the attention

encoder output do not depend on the order of inputs.

Disadvantage

1. RNNs, can detach the encoder & decoder so they have different lengths unlike attention model.

Q/ Choose the correct one.

1) In practice, what is the most accurate description for activation function used in neural networks?

   1. They must be differentiable
   2. They can be non-differentiable, but only for a small number of points.
   3. They can be any continuous function
   4. They must be non-linear to be learnable.

Ans   (B) or (2)

2) Given a neural network with N input nodes, no hidden layers, one output node, with entropy loss & sigmoid activation function, which algorithm can be used to find the global optimum.

1. Stochastic Gradient Descent
2. Batch & Gradient Descent
3. Mini-Batch Gradient Descent
4. All of the above

Ans   All of the above (4)

3) You want to train a neural network to predict the next 30 daily prices using the previous 30 daily prices as inputs. Which model selection & explanation makes the most sense?

1. A fully connected deep feed forward network because it considers all input prices in the hidden layers to make the best decision.

2. A single one-directional RNN because it considers the order of the prices & the output length is the same as the input length.

3. A bidirectional RNN because the prediction benefits from future labels.

4. A one-directional encoder-decoder architecture can generate a sequence of future prices based on all historical input prices

Ans   (4) A 1D encoder-decoder architecture.

Q// Draw a computational graph of a one-hidden layer feed forward NN & write derivaties of each variable in the backpropagation
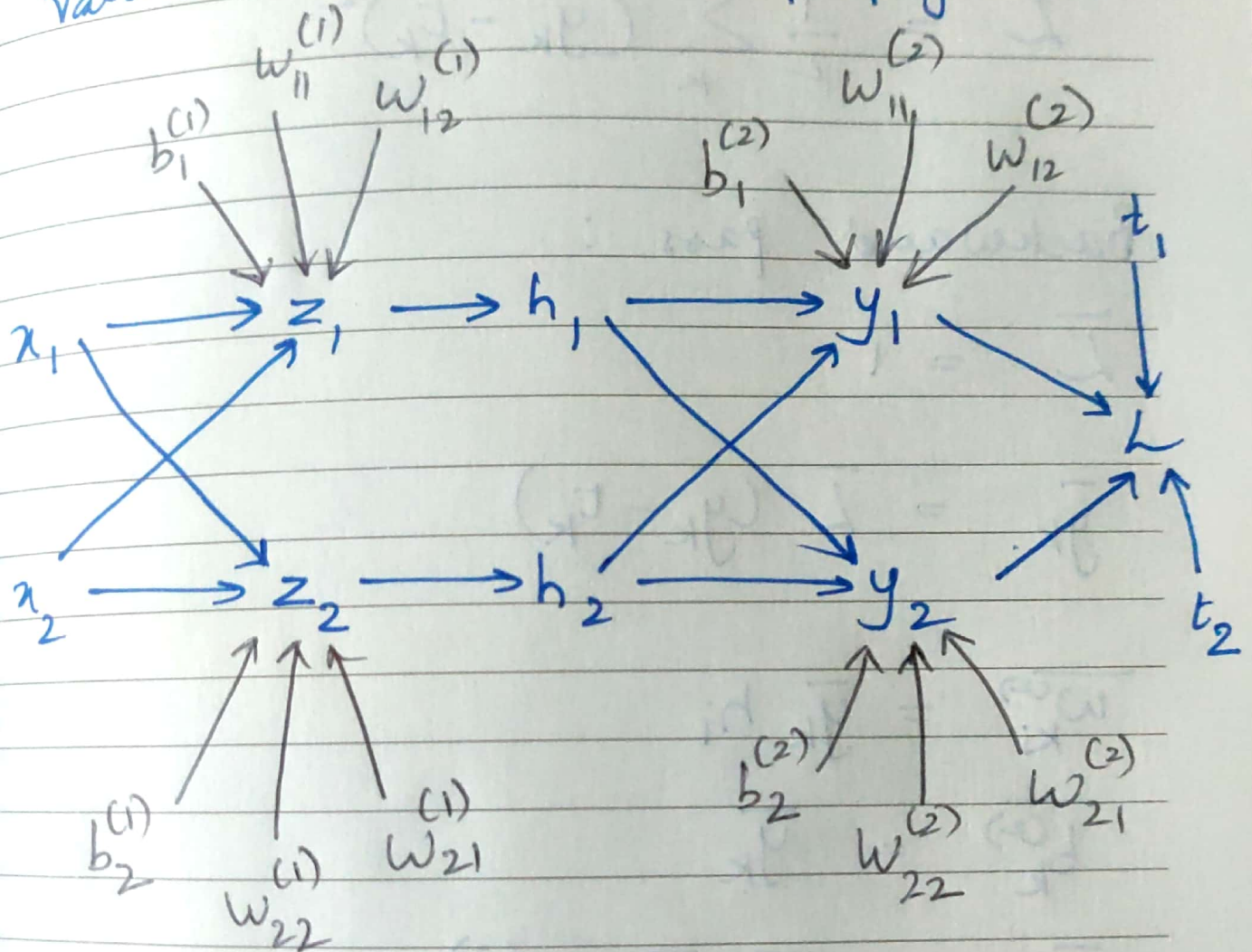
Forward pass :

$$z_i = \sum_j w_{ij}^{(1)} x_j + b_i^{(1)}$$

$$h_i = \sigma(z_i)$$

$$y_k = \sum_i w_{ki}^{(2)} h_i + b_k^{(2)}$$

$$L = \frac{1}{2} \sum_k (y_k - t_k)^2$$

Backward pass :

$$\bar{L} = 1$$

$$\bar{y_k} = L(y_k - t_k)$$

$$\overline{w_{ki}^{(2)}} = \bar{y_k} h_i$$

$$\overline{b_k^{(2)}} = \bar{y_k}$$

$$\bar{h_i} = \sum_k \bar{y_k} w_{ki}^{(2)}$$

$$\bar{z_i} = \bar{h_i} \sigma'(z_i)$$

$$\overline{w_{ij}^{(1)}} = \bar{z_i} x_j$$

$$\overline{b_i^{(1)}} = \bar{z_i}$$