

Artificial Neural Networks

5th Assignment - Shahid Beheshti University - Bachelor's Program

December 16, 2022

I hope all is well with you. This is the fifth series of homework for the Artificial Neural Networks course. The deadline for this assignment is **due date January 1, Sunday**. All students are expected to submit their homework on time. Feel free to **ask questions regarding the exercises in the course Telegram group** if needed. As part of your assignment, you are required to write a detailed report.

Exercise 1

What's the difference between Stateful RNN vs Stateless RNN? What are their pros and cons?

Exercise 2

Explain the main differences between Encoder-Decoder RNNs vs plain sequence-to-sequence RNNs.

- Which one do you prefer for automatic translation and why?

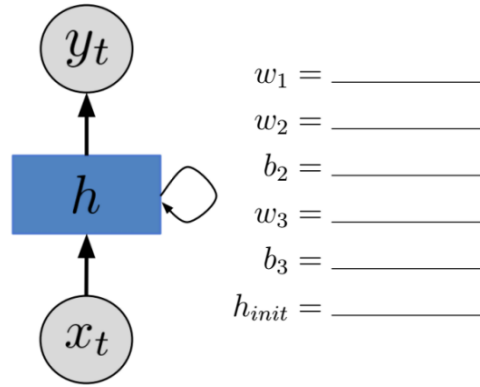
Exercise 3

Suppose we want to build a gated RNN cell that sums its inputs over time. What should be the gating values be? To focus on the gating aspect, your design can change the activation function of the RNN cell (e.g., replace tanh by linear).

- For the LSTM architecture, what should be the value of the input gate and the forget gate?
 - For the GRU architecture, what should be the value of the reset gate and the update gate?
- You can use this link to find out more about the GRU architecture.

Exercise 4

Extra bonus point question: Consider the following architecture:



where the model is defined as follows:

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b_2)$$

$$y_t = g(w_3 h_t + b_3)$$

where the activation functions f and g are given as

$$f(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$g(x) = x.$$

This RNN is supposed to output 0 as long as no 1 occurs in the sequence. As soon as a 1 shows up in the sequence, it is supposed to only output 1s. For example, if the input sequence is 00011001, the output sequence should be 00011111. Find values for the weights, biases, and the initial value of the hidden state that generates this output (+10 pts).

Exercise 5

Create a Persian language model using RNNs or LSTM cells for sequence prediction. You should train your data on the Persian Wikipedia Dataset in order to predict the entire sentence by giving the first 3 to 5 words of a sentence in the corpus. Note that the architecture of your model is your choice.

- You have to calculate the perplexity of your model (a lower perplexity score indicates better generalization performance).
- An extra mark will be awarded to the student who achieves the lowest perplexity on the test data in the class (+10 pts).