Week 7

Name: Kaley Nguyen Email: kaleynn2@illinois.edu

Question 1. What are the main messages you learned from this chapter?

- RNN is used to model longitudinal EHR data, clinical notes or continuous monitoring data. One-to-many (image to text), many-to-one (sequence to single prediction), many-to-many sequential prediction (sequence with same size input-output), and many-to-many seq2seq (unaligned size input-output) are the four different kinds of RNN structures. The vanishing gradient problem, however, occurs because the gradient is backpropagated and can get smaller and smaller.
- Several RNN variations, including LSTM, GRU, and bidirectional RNN, have been found to solve the vanishing gradient issue.

Question 2. Which part do you want to improve in this chapter?

Since the chapter is so much more advanced than what I can offer, I have no recommendations for it.

Question 3. What related resources (book, paper, blog, link) do you recommend your classmates to checkout?

I strongly suggest taking this amazing UC Berkeley course: https://fullstackdeeplearning.com

Question 4. What is the main difference between RNN and bidirectional RNN?

- RNN: RNN can only use information at time t-1, t-2, etc. at current time t; it cannot use information at time t + 1, t + 2, etc.
- Bidirectional RNN: uses to utilize all the event info before and after the current time t. With a bidirectional RNN, the output can be influenced by both recent and distant information.

Question 5. What is the main difference between LSTM and GRU?

- LSTM: introduces a new structure called cell state that is intended to retain important information over time and to forget unimportant information. LSTM has 3 gates: input gate, forget gate and output gate. Note that all the gates and the cell state are all embedding vectors.
- GRU: utilizes the hidden state as the cell state and output state while removing the cell state from the LSTM. GRU therefore only has 2 gates: an update gate and a reset gate.