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

- In deep learning applications, accuracy and interpretability are more crucial than just accuracy. More knowledge about the operation of the Autoencoder mechanism is provided by the deep learning model's capacity to be explained in terms of the attention weights.
- Attention mechanism based on Encoder-Decoder RNN model

> 1. *Annotation vectors through an encoder: encoding of the input sequence and we can use any variation of RNN as the encoder.*
> 2. *Alignment model: measures the correspondence between input and output sequence pairs. The alignment model is simply the dot product of the annotation vectors and hidden states at time t.*
> 3. *Attention weight: refers to how much emphasis is placed on the input words when creating the output words. The attention weights are the normalized vector from the output of the alignment model.*
> 4. *Context vector: dynamic summary of the input depending on the current output position. The weighted total of the annotation vectors and attention weights is the context vector.*
> 5. *Decoder: The output hidden state $$s_i$$ can be generated as a function of previous hidden state $$s_{i-1}$$, previous output $$y_{i-1}$$ and current time context vector $$c_i$$. The the output $$y_i$$ will be generated from the output hidden state $$s_i$$.*

- Understanding the connection between clinical events and results can be aided by understanding attention mechanisms in healthcare. There are several models that aim to understand the EHR data such as RETAIN, which uses a two-level attention mechanism to explain the risk scores associated with some of the medical diagnosis codes and medications before the heart failure onset event. Another application of the attention mechanism in healthcare is the representation learning from the EHR data to explain the medical ontology DAG graph. When the training data sizes were raised from 10% to 40%, GRAM considerably outperformed RNN, simple roll up, and roll up rare.

Question 2. What related resources (book, paper, blog, link) do you recommend your classmates to checkout?

I took a class at corise and they have a useful class to understand deep learning for a beginner level that covers all the basic essentials.

https://corise.com/course/deep-learning-essentials

Question 3. Which part do you want to improve in this chapter?

There are so many healthcare applications from the attention mechanism from the chapter and I do not have any recommendation regarding the chapter.

Question 4. What is the difference between sequence-to-sequence model with RNN and that with attention?

- seq2seq with RNN: the encoder RNN maps the input sequence into a **fixed-length** context vector $$c$$. The context vector $$c$$ will be passed into each step of the decoder RNN to generate the output sequence.
- seq2seq with attention RNN: Instead of compressing the entire context variable $$c$$ into a static variable, the attention mechanism allows a **dynamic** context variable $$c$$ to provide the positioning and translation between the input / output. The alignment helps to find the relevancy of the input of sequence with respect to the output of the sequence.