Summary
On
Course5 - Week 3

Submitted to: Dr/Omar Nasr
Name: Mohamed Ahmad Nassar
Noha Ahmad Darwish

- Encoder-Decoder architecture are used in many wide spread applications like, Machine translation, image captioning, sentiment analysis and trigger word detection
- Encoder network used to summarize the input series into one output vector that could be interpreted by its decoder part
- To Select the most suitable output sequence an algorithm called Beam Search is used
- Beam Search is an algorithm that is used to Maximize the probability of of output sequence given input sequence
- As beam Width collapsed to 1, the algorithm turns into a greedy search algorithm, that so called “best path algorithm” which select the maximum output probability independently at each time step
- As beam width increases the beam search will be computationally very extensive, but it should produce a better result i.e better accuracy but bad performance
- a Log loss may be used instead of original loss in beam search to avoid very small numbers which can't be interpreted by computers
- Length normalization is a technique used in beam search algorithm which divide the loss function by sequence length which considerably improve its performance
- Blue score is an algorithm that used to solve the problem that when one input sequence could have more than one correct ground truth.
- Errors consist of two category,
    - RNN Rrror (when output a probability of correct sequence < probability of wrong seq)
    - B.S Error (When selecting a path with a lower probability than true label path)
- Encoder-Decoder networks aren’t good in remembering long-term dependencies, because of their limited capability in summarizing the input into one vector
- to overcome the forgetting problem, an Attention mechanism is used to allow the network to focus on a specific part of the input while generating the output
- attention mechanism is considered so important in seq-to-seq problem, but it’s computationally extensive