EECS545 Lecture 13 Quiz Solutions

March 7, 2024

1. What are the benefits of using an RNN over applying CNN over time (i.e., 1D CNN)? (**Choose all options that apply**)

   (a) RNNs require less inference time than CNNs in Language modeling.

   (b) RNNs can better understand the sequential dependencies.

   (c) RNNs can better handle sequences with unknown lengths.

   (d) Training RNNs are easier than CNNs as RNNs are less likely to have gradients explode/vanish issue.

   Solution: (b) and (c)
   (a): RNNs need to infer each word one by one during inference time (See Slide 69).
   (d): RNNs are more likely to suffer gradient explode/vanish issues.

2. Which tricks can help address the exploding/vanishing gradient problem? (**Choose all options that apply**)

   (a) Use LSTM instead of vanilla RNN.

   (b) Use sigmoid activation instead of tanh in vanilla RNN.

   (c) Use orthogonal initialization.

   (d) Use gradient clipping.

   Solution: (a) (c) (d)
   Please revisit the slides from page 38). (b) would still suffer exploding/vanishing gradient issues; tanh can be written as $2\sigma(2x) - 1$ if sigmoid is $\sigma(x)$.

3. (True/False) The vanishing gradient in RNN could cause the parameters to be biased to capture short- term dependencies.

   Solution: True. Long-term dependencies in the sequence become affected as the gradient vanishes in the middle (Slide 42)