

Large Scale Machine Learning and Deep Learning

Review Questions 7

1. Explain how can we use RNN for language modeling?

Answer: the RNN slides, pages 27-28

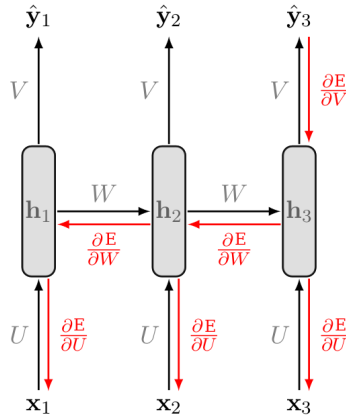
2. What's the vanishing problem in RNN?

Answer: the RNN slides, page 66, and training feedforward network slides, pages 40-43

3. Explain the impact of different gates in LSTM?

Answer: the RNN slides, pages 70-73

4. Assume the error of the following network is $E = E^{(1)} + E^{(2)}$, then compute the $\frac{\partial E}{\partial u}$.



Answer:

$$E_{\text{total}} = E^{(1)} + E^{(2)}$$

$$\frac{\partial E^{(1)}}{\partial u} = \frac{\partial E^{(1)}}{\partial \hat{y}^{(1)}} \frac{\partial \hat{y}^{(1)}}{\partial z^{(1)}} \frac{\partial z^{(1)}}{\partial h^{(1)}} \frac{\partial h^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial u}$$

$$\frac{\partial E^{(2)}}{\partial u} = \frac{\partial E^{(2)}}{\partial \hat{y}^{(2)}} \frac{\partial \hat{y}^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial s^{(2)}} \frac{\partial s^{(2)}}{\partial u} + \frac{\partial E^{(2)}}{\partial \hat{y}^{(2)}} \frac{\partial \hat{y}^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial s^{(2)}} \frac{\partial h^{(1)}}{\partial s^{(1)}} \frac{\partial s^{(1)}}{\partial u}$$

$$\frac{\partial E_{\text{total}}}{\partial u} = \frac{\partial E^{(1)}}{\partial u} + \frac{\partial E^{(2)}}{\partial u}$$