1. What’s the vanishing problem in RNN?

2. Explain the impact of different gates in LSTM?

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

4. Assume we have a stacked autoencoder with three hidden layers \( h_1, h_2, \) and \( h_3 \), in which each layer applies the following functions respectively, \( h_1 = f_1(x) \), \( h_2 = f_2(h_1) \), and \( h_3 = f_3(h_2) \), and the output of the network will be \( y = f_4(h_3) \). Do you think if it is a good autoencoder if it generates \( f_4(f_3(f_2(f_1(x)))) = x \) for all input instances \( x \). How can we improve it?

5. How does Gibbs sampling work? When do we need to use Gibbs sampling?

6. How do you tie weights in a stacked autoencoder? What is the point of doing so?