Q2

17/11/2021

Ans to the Qs. 1

In machine learning, the vanishing gradient problem is encountered when training artificial neural networks with gradient-based learning methods and backpropagation. The problem occurs because as more layers using certain activation functions are added to neural networks, the gradients of the loss function approaches 0 (zero) making the network harder to train, this effectively prevents the weight from changing its value. In the worst case, this may completely stop the neural network from further training. Traditional activation functions like the sigmoid function squishes a large input space between 0 and 1. Therefore, a large change in the input of the sigmoid function will cause a small change in the output. Hence, the derivative becomes small.

This vanishing gradient problem can be handled in LSTM.

$$c_t = f_t * c_{t-1} + i_t * c_t'$$

Here, f_t = forget gate controller
c_{t-1} = previous state info,
i_t = input gate controller
c_t' = new info. to process

An LSTM network has three gates that update and control the cell states - forget gate, input gate and output gate. The forget gate controller of LSTM mainly helps in solving the vanishing gradient problem. It controls what information in the cell state to forget, given new information that entered the network. The sigmoid function in LSTM keeps 0, 1 as switch. f_t changes C_{t-1} very slowly according to which information should be forgotten or remembered. When $f_t = 1$, it signals the passing of previous cell information to the next cell. When, $f_t = 0$, it means the previous cell information should be forgotten.