Q1)

Q2) LSTM:

Forward loss:

$$F_t = \sigma(\omega_{Fi}[A_{t-1}, x_t])$$
$$\tilde{C}_t = \tanh(\omega_c \cdot [h_{t-1}, x_t])$$
$$i_t = \sigma(\omega_i \cdot [h_{t-1}, x_t])$$
$$O_t = \sigma(\omega_o \cdot [h_{t-1}, x_t])$$
$$h_t = O_t \otimes \tanh(i_t)$$
cell state $C_t = C_{t-1} \otimes F_t \otimes \tilde{C}_t \otimes i_t$

Let ξ_T be the error for time step $\leq T$

$$\frac{\partial \xi_T}{\partial \omega} = \frac{\partial \xi_T}{\partial C_T} \cdot \frac{\partial C_T}{\partial C_{T-1}} \times \frac{\partial C_{T-1}}{\partial C_{T-2}} \cdots \frac{\partial C_1}{\partial \omega}$$

$$\frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}}(C_t \otimes F_t \otimes \tilde{C}_t \otimes i_t)$$
$$= \frac{\partial F_t}{\partial C_{t-1}} \cdot C_{t-1} \times F_t + \frac{\partial i_t}{\partial C_{t-1}} \cdot \tilde{C}_t + i_t \frac{\partial \tilde{C}_t}{\partial C_{t-1}}$$

$$\frac{\partial C_t}{\partial C_{t-1}} = \underbrace{\frac{\partial}{\partial C_{t-1}}\left(\sigma(\omega_f[h_{t-1}, x_t])\right)}_{B_t} \cdot C_{t-1} + F_t \cdot A_t$$
$$+ \underbrace{\frac{\partial(\sigma(\omega_i[h_{t-1}, x_t]))}{\partial C_{t-1}}}_{C_t} \cdot \tilde{C}_t + \underbrace{\frac{\partial(\sigma[\omega_c[h_{t-1}, x_t])}{\partial C_{t-1}}}_{D_t} \cdot i_t$$

$$\frac{\partial \xi_T}{\partial \omega} = \frac{\partial \xi_T}{\partial C_T}\left(\prod_{t=2}^{t=T}(A_t + B_t + \Phi_t + \xi_t)\right)\frac{\partial C_t}{\partial \omega}$$

Vanishing gradient is less likely in LSTMs due to there being sum of product terms.