Q4: Vanishing gradient problem : occurs when the update of RNN network involves unrolling the network for each input time step, in effect creating a very deep network that requires weight updates. When a network has too many layers, it becomes untrainable and that leads a system to not converge towards the right solution.

LSTMs solve the problem using a unique additive gradient structure that includes direct acess to the forget gate's activations, enabling the network to encourrage desired behaviour from the error gradient using frequent gates update on every time step of learning process