Answer the following:

1. [3 pts] Given the diagram on the right for an RNN, what are the recursive equations for $h(t)$ and $y(t)$ in terms of $\{x(t), x(t-1), h(t), h(t-1), y(t), y(t-1)\}$ assuming that we are using this model for regression? Make sure to include any weights and biases. Use g(·) as the activation function.

2. [1 pts] Which has a longer memory capacity a standard RNN or an LSTM?

3. [2 pts] In a sentence, What is the role of the forget gate in an LSTM cell?

4. [1 pts] What type of activation function is used for the forget gate?

5. [2 pts] In a sentence, What is the role of the input gate in an LSTM cell?

6. [1 pts] What type of activation function is used for the input gate?

y Output

h
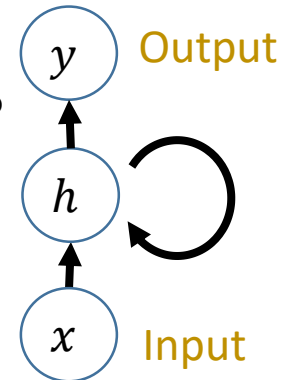
x Input

Answer

1. $h^{(t)} = g(W_h h^{t-1} + W_x x^t + b)$
   $y^{(t)} = V h^t + c$

2. LSTM

3. The forget gate generates a vector using the previous local state ($h^{(t-1)}$) and current data input (x^t). The vector is used in an element-wise product with the long-term state (c^{t-1}) to retain some features and forget other features.

4. Sigmoid

5. The input gate generates a vector using the previous local state ($h^{(t-1)}$) and current data input (x^t). The vector contains features which are to be added to the long-term state (c^{t-1}).

6. Sigmoid