## RNN
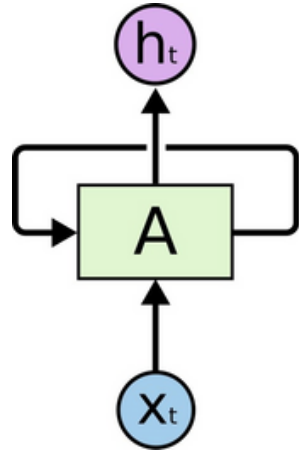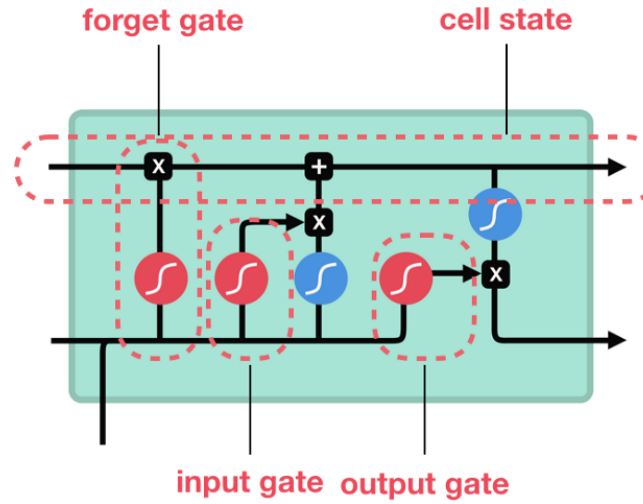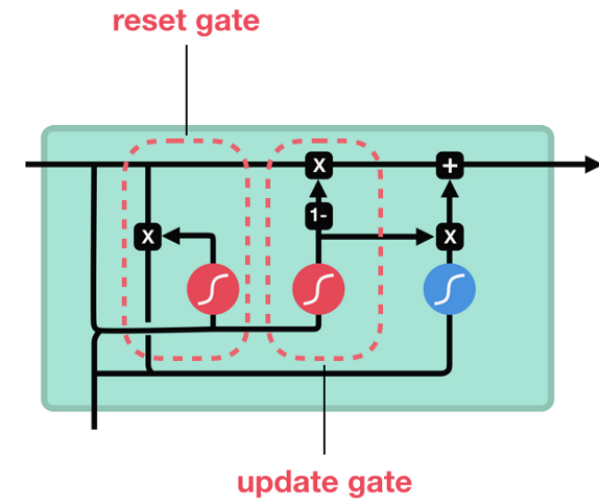


Input features: [batch_size, timesteps, dimension]

Input shape: [timesteps, dimension]
   * In case our timesteps vary, we can input the
     timesteps as "None", that way the RNN will be unrolled dynamically

Cons:
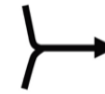   - Exploding/ Vanishing Gradient
   - Short term memory

## LSTM



forget gate

cell state

input gate    output gate

## GRU



reset gate

update gate



**sigmoid**

**tanh**

**pointwise multiplication**

**pointwise addition**

**vector concatenation**