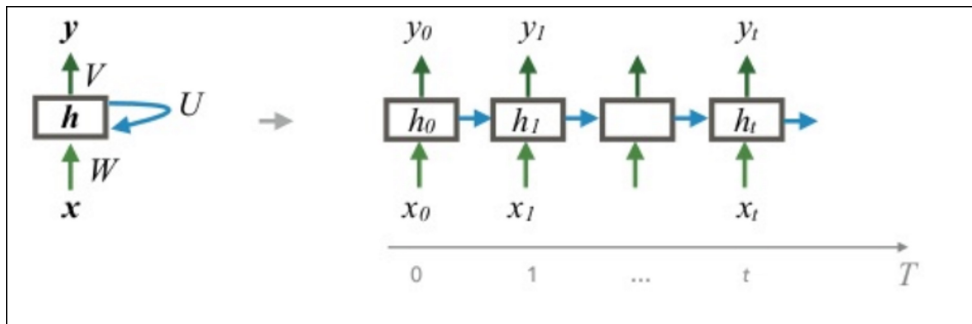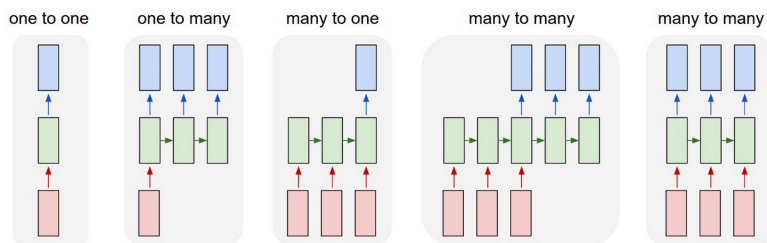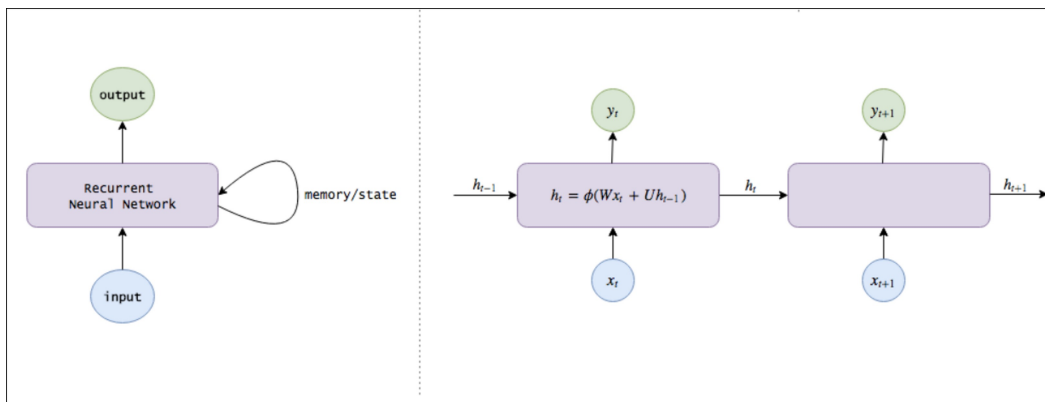# Recurrent Neural Network

27 May 2018   16:02

Recurrent Neural Network

$$h_t = \phi(WX_t + Uh_{t-1})$$
$$Y_t = Vh_t$$

The Unreasonable Effectiveness of Recurrent Neural Networks From <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

Usage
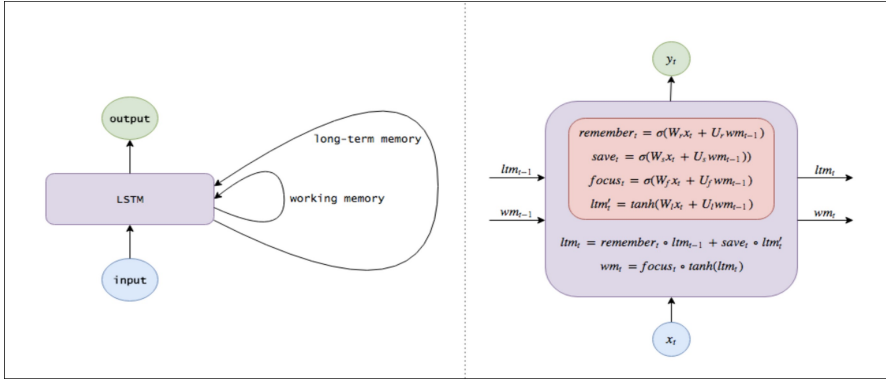
One to One -> Image Classification
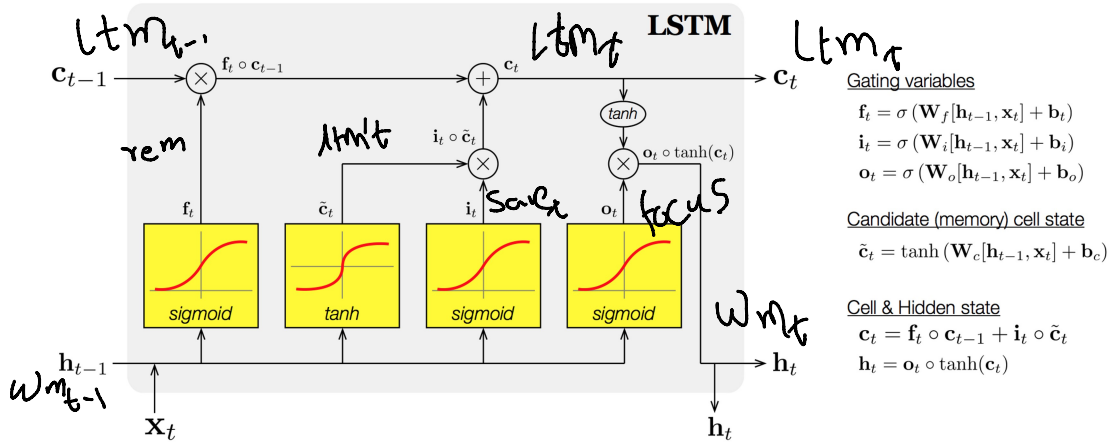One to Many -> Image Captioning
Many to One - > Sentiment Analysis
Many to Many (1) -> Machine Translation
Many to Many (2) -> Video Classification

LSTM (Long Short Term Memory)

LSTM Cell

Gating variables

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_t\right)$$
$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i\right)$$
$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o\right)$$

Candidate (memory) cell state

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c\right)$$

Cell & Hidden state

$$\mathbf{c}_t = \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \tilde{\mathbf{c}}_t$$
$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{c}_t)$$

Equations

$$remember_t = \sigma(W_r x_t + U_r wm_{t-1}) \quad \text{\# What to remember what to forget}$$

$$ltm'_t = \phi(W_l x_t + U_l wm_{t-1}) \quad \text{\# candidate addition to our long-term memory}$$

$$save_t = \sigma(W_s x_t + U_s wm_{t-1}) \quad \text{\# which parts of candidate are actually worth using and saving}$$

$$ltm_t = remember_t \circ ltm_{t-1} + save_t \circ ltm'_t \quad \text{\# updated long-term memory}$$

$$focus_t = \sigma(W_f x_t + U_f wm_{t-1}) \quad \text{\# focus our long-term memory into information that will be immediately useful}$$

$$wm_t = focus_t \circ \phi(ltm_t) \quad \text{\# Working Memory}$$

- The long-term memory, ltm_t, is usually called the **cell state**, denoted C_t.
- The working memory, wm_t, is usually called the **hidden state**, denoted h_t. This is analogous to the hidden state in vanilla RNNs.
- The remember vector, $remember_t$, is usually called the **forget gate** (despite the fact that a 1 in the forget gate still means to keep the memory and a 0 still means to forget it), denoted f_t
- The save vector, $save_t$, is usually called the **input gate** (as it determines how much of the input to let into the cell state), denoted i_t.
- The focus vector, $focus_t$, is usually called the **output gate**, denoted O_t