

Recurrent Neural Networks

Setup: Sequence data-set $\{y_t, t=1, \dots, T\}$, $y_t \in \mathbb{R}^d$

$$y_t = (y^{(0)}, y^{(1)}, \dots, y^{(d)})$$

Goal: Predict the next sequence \hat{y}_t as a function of previously

observed sequences: $\hat{y}_t = \underline{f}(\underbrace{y_{t-1}, y_{t-2}, \dots, y_{t-L}}_{L: \# \text{ of lags}})$

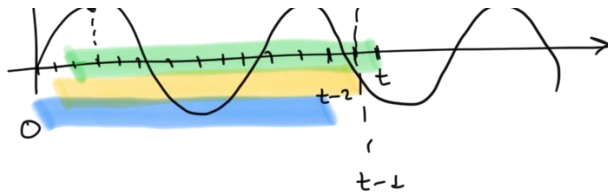
RNN architecture: $\hat{y}_t = \underline{f}_{\theta}(\underbrace{y_{t-1}, y_{t-2}}_{\text{inputs}})$

$$\begin{array}{c} \hat{y}_t \\ \uparrow v, c \\ h_t \\ \uparrow v, b \\ h_{t-1} \xrightarrow{w} h_t \\ \uparrow v, b \\ h_{t-2} \xrightarrow{w} h_{t-1} \\ \uparrow v, b \\ y_{t-2} \end{array} \quad \left\{ \begin{array}{l} \hat{y}_t = h_t V + c \\ h_t = \sigma(h_{t-1} W + y_{t-1} V + b) \\ h_{t-1} = \sigma(h_{t-2} W + y_{t-2} V + b) \\ h_{t-2} = 0 \end{array} \right.$$

$$\theta := \{V, b, W, V, c\}$$

Example: $y(t) = \sin(\pi t)$





$$y_t \in \mathbb{R}$$

Inputs : $X = (y_{t-1}, y_{t-2}, \dots, y_0)$
 $\underline{L \times N \times D}$

$$y_t = (y^{(2)}, y^{(3)}, \dots, y^{(t)})$$

Output : $Y = y_t$

$$y_{t-1} = (y^{(1)}, y^{(2)}, \dots, y^{(t-1)})$$

Forward pass :

$$N \times D$$

$$h_{t-2} = 0$$

$$y_{t-2} = (y^{(0)}, y^{(1)}, \dots, y^{(t-2)})$$

$$h_{t-1} = \sigma(h_{t-2}W + X[0, :, :])$$

$$h_t = \sigma(h_{t-1}W + X[1, :, :])$$

$$\hat{y}_t = h_t V + c$$

Long Short-Term Memory networks :

In vanilla RNNs : $h_t = \tanh(h_{t-1}W + y_{t-1}V + b)$

↑

$h_t = o_t \odot \tanh(s_t) \rightarrow$ output vector

$o_t = \sigma(h_{t-1}W_o + y_{t-1}V_o + b_o) \rightarrow$ output gate
sigmoid

$s_t = f_t \odot s_{t-1} + i_t \odot \tilde{s}_t \rightarrow$ cell state

$\tilde{s}_t = \tanh(h_{t-1}W_s + y_{t-1}V_s + b_s)$

$i_t = \sigma(h_{t-1}W_i + y_{t-1}V_i + b_i) \rightarrow$ external input gate

$f_t = \sigma(h_{t-1}W_f + y_{t-1}V_f + b_f) \rightarrow$ forget gate

$\theta := \{W, V, b, W_o, V_o, b_o, W_s, V_s, b_s, W_i, V_i, b_i, W_f, V_f, b_f\}$

↳ $(v_0, u_0, r_0, s_0, \dots)$

- Gate Recurrent Units (GRU cell)

inputs

-

-

-