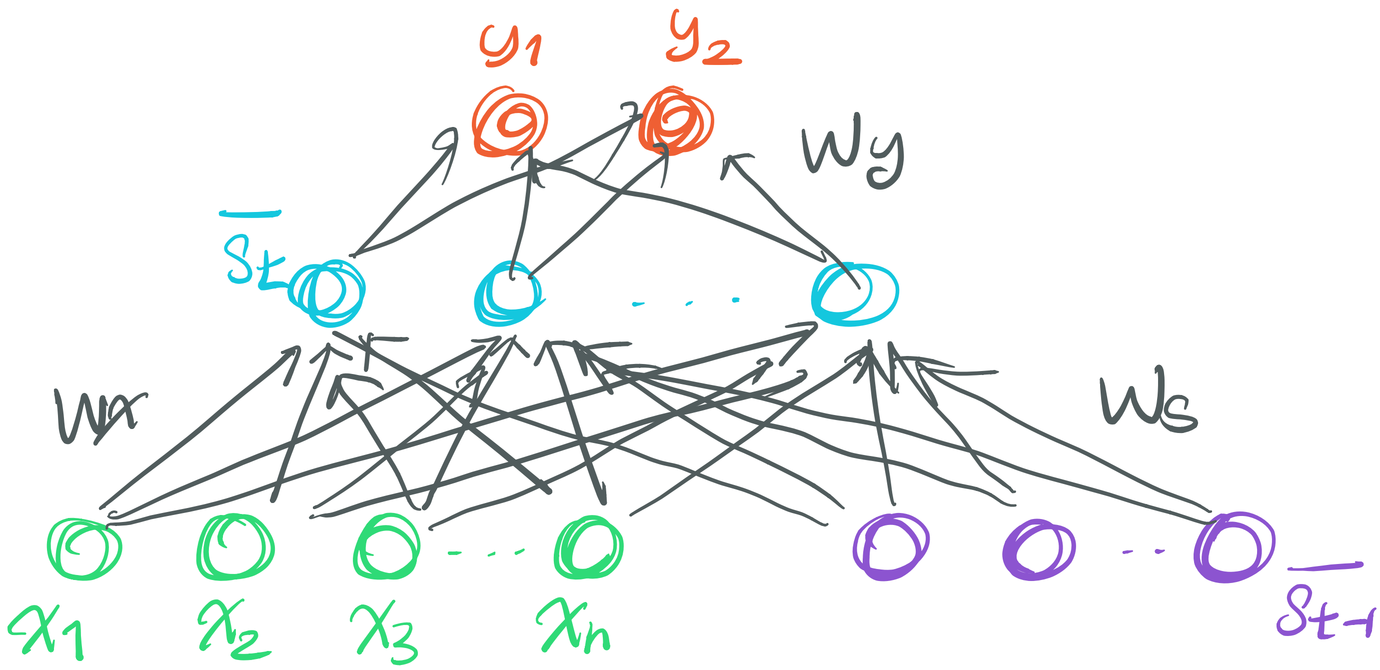
Nov 16, 2019 L3.17

# Backpropagation Through Time (BPTT)




$$\bar{y}_t = \sigma(\bar{s}_t \cdot W_y)$$

$$\bar{y}_t = \bar{s}_t \cdot W_y$$

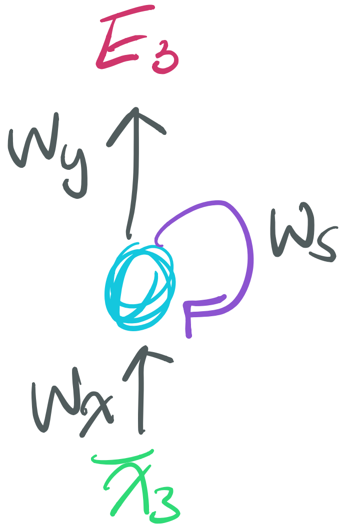
$$\bar{s}_t = \phi(\bar{x}_t \cdot W_x + \bar{s}_{t-1} \cdot W_s)$$

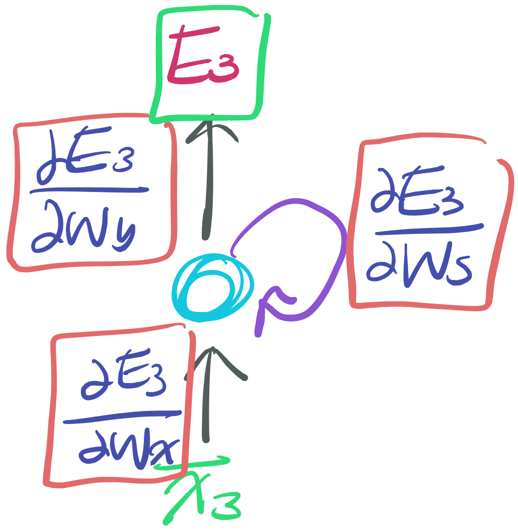
$$\bar{s}_t = \tanh(\bar{x}_t \cdot W_x + \bar{s}_{t-1} \cdot W_s)$$

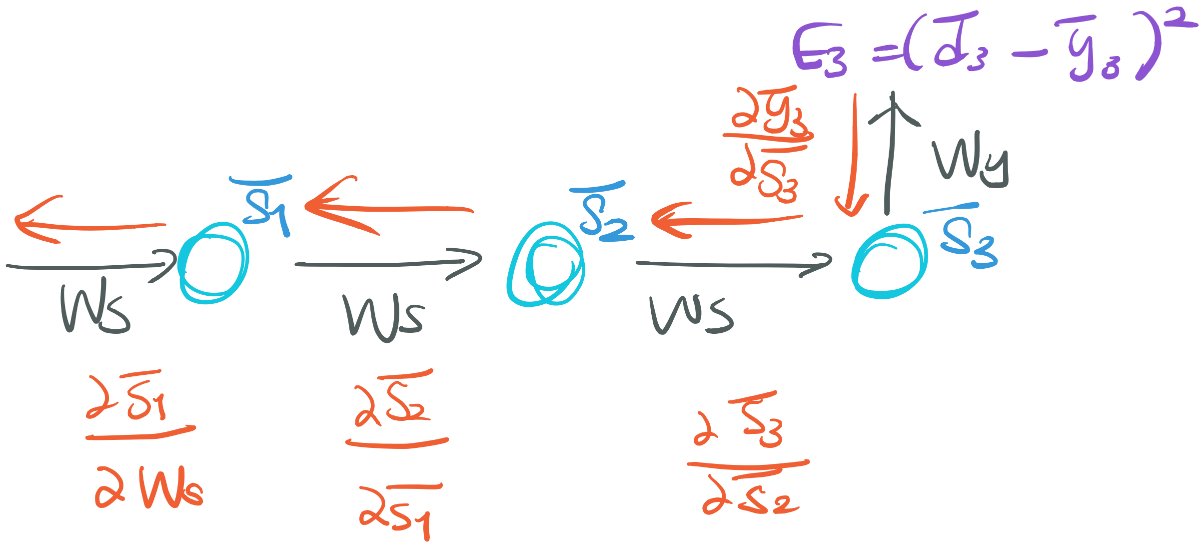
$$E_t = (\bar{d}_t - \bar{y}_t)^2$$

desired output

calculated output




Accumulative gradient @ timestep $t=3$.

$$\frac{\partial E_3}{\partial W_s} = \frac{\partial E_3}{\partial \bar{y}_3} \cdot \frac{\partial \bar{y}_3}{\partial \bar{s}_3} \cdot \frac{\partial \bar{s}_3}{\partial W_s}$$

$$+ \frac{\partial E_3}{\partial \bar{y}_3} \cdot \frac{\partial \bar{y}_3}{\partial \bar{s}_3} \cdot \frac{\partial \bar{s}_3}{\partial \bar{s}_2} \cdot \frac{\partial \bar{s}_2}{\partial W_s}$$

$$+ \frac{\partial E_3}{\partial \bar{y}_3} \cdot \frac{\partial \bar{y}_3}{\partial \bar{s}_3} \cdot \frac{\partial \bar{s}_3}{\partial \bar{s}_2} \cdot \frac{\partial \bar{s}_2}{\partial \bar{s}_1} \cdot \frac{\partial \bar{s}_1}{\partial W_s}$$

$$
\begin{tikzcd}
\bar{x} \arrow[r, "U"'] & \bar{s} \arrow[loop below, "W_1"] \arrow[r, "V_1"'] & \bar{z} \arrow[loop below, "W_2"] \arrow[r, "V_2"'] & \bar{y}
\end{tikzcd}
$$

Q1

Mathematical derivation of state Z @ time t.

$$\bar{z}_t = \phi(\bar{s}_t V_1 + \bar{z}_{t-1} W_2)$$

Q2

Error E. What's the update rule of weight mat V1 at time t, over a single timestep?

$$\Delta V_1 = -\alpha \frac{\partial E_t}{\partial V_1} = -\alpha \frac{\partial E_t \, \partial \bar{y}_t \, \partial \bar{z}_t}{\partial \bar{y}_t \, \partial \bar{z}_t \, \partial V_1}$$

Q3

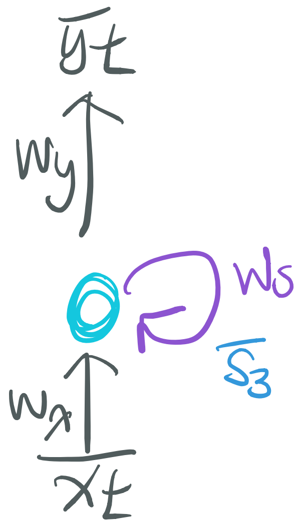
What is the update rule of weight mat. U @ time t+1 (over 2 timesteps)?

$$\frac{\partial E_{t+1}}{\partial U} = \frac{\partial E_{t+1}}{\partial \bar{y}_{t+1}} \frac{\partial \bar{y}_{t+1}}{\partial \bar{z}_{t+1}} \frac{\partial \bar{z}_{t+1}}{\partial \bar{s}_{t+1}} \frac{\partial \bar{s}_{t+1}}{\partial U}$$

$$+ \frac{\partial E_{t+1}}{\partial \bar{y}_{t+1}} \frac{\partial \bar{y}_{t+1}}{\partial \bar{z}_{t+1}} \frac{\partial \bar{z}_{t+1}}{\partial \bar{z}_t} \frac{\partial \bar{z}_t}{\partial \bar{s}_t} \frac{\partial \bar{s}_t}{\partial U}$$

$$+ \frac{\partial E_{t+1}}{\partial \bar{y}_{t+1}} \frac{\partial \bar{y}_{t+1}}{\partial \bar{z}_{t+1}} \frac{\partial \bar{z}_{t+1}}{\partial \bar{s}_{t+1}} \frac{\partial \bar{s}_{t+1}}{\partial \bar{s}_t} \frac{\partial \bar{s}_t}{\partial U}$$

'Folded' Model




'Unfolded' Model