Per Sentence.

$$S_t = \max[0, (Ux_t + WS_{t-1})] \Rightarrow S_t = \tanh(Ux_t + WS_{t-1})$$

$$\frac{d \tanh x}{dx} = 1 - \tanh^2 x \qquad \frac{d \max(0,x)}{dx} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\frac{\partial S_1}{\partial W} = \cancel{\frac{\partial(a + WS_0)}{\partial W}} = S_0$$

$$\frac{\partial S_1}{\partial W} = \frac{\partial \tanh(a + WS_0)}{\partial W} = (1 - S_1^2) \cdot \frac{\partial(a + WS_0)}{\partial W} = (1 - S_1^2) \cdot S_0$$

$$\frac{\partial S_2}{\partial W} = (1 - S_2^2) \cdot \frac{\partial(a + WS_1)}{\partial W}$$

$$= (1 - S_2^2) \cdot \left(S_1 + \frac{\partial S_1}{\partial W} \cdot W\right)$$

$$= (1 - S_2^2) \cdot (S_1 +$$

$$\frac{\partial S_3}{\partial W} = (1 - S_3^2) \cdot \frac{\partial(a + WS_2)}{\partial W} \Rightarrow \frac{\partial E_3}{\partial W} = \underbrace{\frac{\partial E_3}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3}}_{(\hat{y}_3 - y_3)} \cdot \underbrace{\frac{\partial z_3}{\partial S_3}}_{V} \cdot \frac{\partial S_3}{\partial W}$$

$$= (1 - S_3^2) \cdot \left(S_2 + \frac{\partial S_2}{\partial W} \cdot W\right)$$

$$=$$

in wild_ML.

$$\text{delta\_t} = (\hat{y}_3 - y_3) \cdot V \cdot (1 - S_3^2)$$

$$dW += \text{delta\_t} \otimes S_2$$