LSTM (Cell) Architecture

$$o_t = \sigma(x_t * U^o + h_{t-1} * W^o), \tag{1}$$

$$i_t = \sigma(x_t * U^i + h_{t-1} * W^i), \tag{2}$$

$$f_t = \sigma(x_t * U^f + h_{t-1} * W^f), \tag{3}$$

$$a_t = \bar{c_t} = \tanh(x_t * U^g + h_{t-1} * W^g), \tag{4}$$

$$c_t = f_t * c_{t-1} + i_t * \bar{c_t}, \tag{5}$$

$$h_t = \tanh(c_t) * o_t. \tag{6}$$

Loss of the Layer through Time

Since we're operating (with ∂) on functions over the space $\mathbb{R}^n$ (this instance $n = 3$) the composition operator for f, g is not multiplication any more, i.e., we use the inner product of the vector space $\mathbb{R}^3$.

$$p_t = \frac{e^{\hat{h}}}{\Sigma_i e^{h_i}}, \text{ where } i \leq dim(h_t), \text{ and } t \leq window_{time}, \tag{7}$$

$$\mathbf{L}(p_t) = -\ln(p_t), \tag{8}$$

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}, \tag{9}$$

$$f(\hat{h}) = e^{\hat{h}_t}, \ \frac{\partial f(\hat{h})}{\partial \hat{h_i}} = e^{\hat{h_i}} \tag{10}$$

$$g(\hat{h}) = \Sigma_i e^{h_i}, \ \frac{\partial g(\hat{h})}{\partial \hat{h_i}} = e^{h_i} \tag{11}$$

$$g(\hat{h})^2 = (e^{h_1} + e^{h_2} + e^{h_3})^2. \tag{12}$$

$$
\begin{aligned}
\frac{\partial f(\hat{h})}{\partial h_1} g(\hat{h}) - f(x)\frac{\partial g(\hat{h})}{\partial h_1} &= e^{\hat{h}_1} \times \Sigma_i e^{h_i} - e^{\hat{h}} \cdot e^{h_1} \\
&= [e^{h_1}, 0, 0] \times (e^{h_1} + e^{h_2} + e^{h_3}) - [e^{h_1}, e^{h_2}, e^{h_3}] \cdot [e^{h_1}, 0, 0] \\
&= e^{2h_1} + e^{h_1}e^{h_2} + e^{h_1}e^{h_3} - e^{2h_1} \\
&= e^{h_1}e^{h_2} + e^{h_1}e^{h_3}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial p_t(\hat{h})}{\partial h_1} &= \frac{e^{h_1}e^{h_2} + e^{h_1}e^{h_3}}{(e^{h_1} + e^{h_2} + e^{h_3})^2}, \\
\frac{\partial \mathbf{L}(\hat{p}_t)}{\partial h_1} = \frac{\partial \mathbf{L}(\hat{p}_t)}{\partial \hat{p}_t}\frac{\partial p_t(\hat{h})}{\partial h_1} &= -\frac{1}{p_t}\frac{e^{h_1}e^{h_2} + e^{h_1}e^{h_3}}{(e^{h_1} + e^{h_2} + e^{h_3})^2}.
\end{aligned}
$$

$$\frac{\partial p_t(\hat{h})}{\partial h_1} = \frac{e^{h_1}e^{h_2} + e^{h_1}e^{h_3}}{(e^{h_1} + e^{h_2} + e^{h_3})^2}, \tag{13}$$

$$\frac{\partial p_t(\hat{h})}{\partial h_2} = \frac{e^{h_2}e^{h_1} + e^{h_2}e^{h_3}}{(e^{h_1} + e^{h_2} + e^{h_3})^2}, \tag{14}$$

$$\frac{\partial p_t(\hat{h})}{\partial h_3} = \frac{e^{h_3}e^{h_1} + e^{h_3}e^{h_2}}{(e^{h_1} + e^{h_2} + e^{h_3})^2}. \tag{15}$$