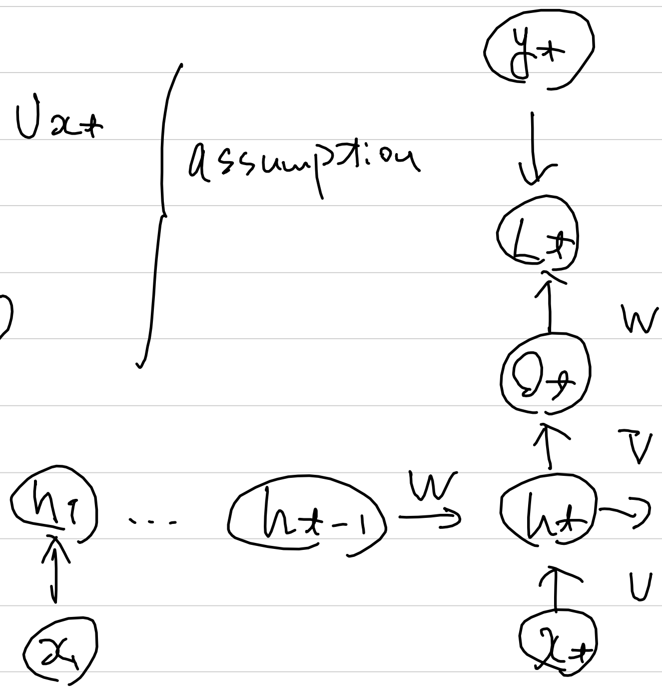
RNN Deep Learning Book 第10章

→ Recurrent Neural Networks
回帰結合型ニューラルネットワーク

$$a_t = b + W h_{t-1} + U x_t$$
$$h_t = \tanh(a_t)$$
$$o_t = c + V h_t$$
$$\hat{y}_t = \mathrm{softmax}(o_t)$$

assumption



$$h_t = f(h_{t-1}, x_t, \theta)$$

$$o_t = g(h_t, \theta)$$

$$h_t = L(o_t, y_t)$$

Lの内部で softmax(o_t)
↓
$\hat{y}_t$になる

$$(\log p(y|x_n))' = \frac{1}{p(y|x)}$$

xの系列と yの値の系列が与えられたとき、

L_tが $x_1 \cdots x_t$ で与えられた下での $y_1 \cdots y_t$ の

対数尤度と考えると、

$$L(\{x_1 \cdots x_\tau\}, \{y_1 \cdots y_\tau\})$$

$$= \sum_t L_t$$

$$= -\sum_t \log p_{model}(y_t \mid \{x_1 \cdots x_t\})$$

通時的誤差逆伝播法 (back-propagation through time : BPTT)

$$\frac{\partial L}{\partial L_t} = 1$$

時間 step tにおける出力の勾配 $(\nabla_{o_t} L)_i$

$$(\nabla_{o_t} L)_i = \frac{\partial L}{\partial o_{ti}} = \underbrace{\frac{\partial L}{\partial L_t}}_{=1} \frac{\partial L_t}{\partial o_{ti}} = \hat{y}_{ti} - 1_{i,y_t}$$

系列の最後 τ step から始める

$$\nabla_{o_\tau} L = V^T \nabla_{o_\tau} L$$

$$\nabla_{h_t} L = \left(\frac{\partial h_{t+1}}{\partial h_t}\right)^T (\nabla_{h_{t+1}} L) + \left(\frac{\partial o_t}{\partial h_t}\right)^T (\nabla_{o_t} L)$$

答 $\tanh(x)' = 1 - \tanh^2(x)$

$\tanh(x)$ の微分について。

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$$

$$= \frac{\frac{e^x - e^{-x}}{2}}{\frac{e^x + e^{-x}}{2}} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$= (e^x - e^{-x}) \times (e^x + e^{-x})^{-1}$$

$$= (e^x + e^{-x})(e^x + e^{-x})^{-1} + (e^x - e^{-x}) \times -1 \, (e^x + e^{-x})^{-2} \times (e^x + e^{-x})'$$

$$= 1 + (e^x - e^{-x}) \times (-1) \times (e^x - e^{-x}) \times (e^x + e^{-x})^2$$

$$= 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \tanh^2(x)$$