17A_Q4 RNN

1. state vector h_t

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

input sequence x_t

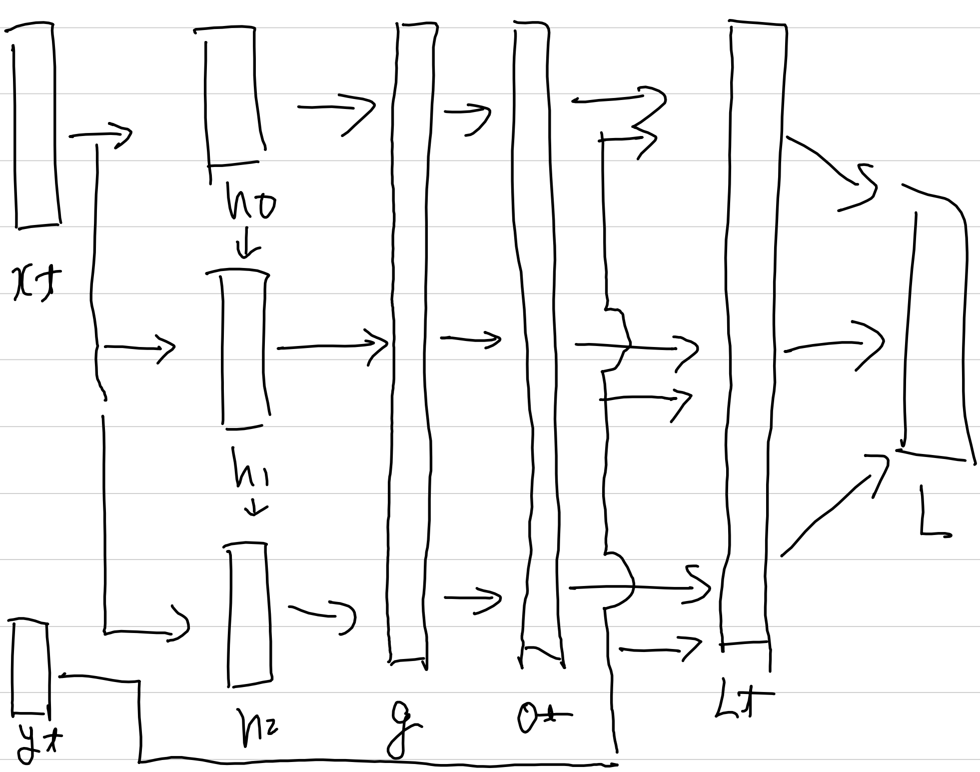
target sequence y_t

Loss $L_t = L(o_t, y_t)$

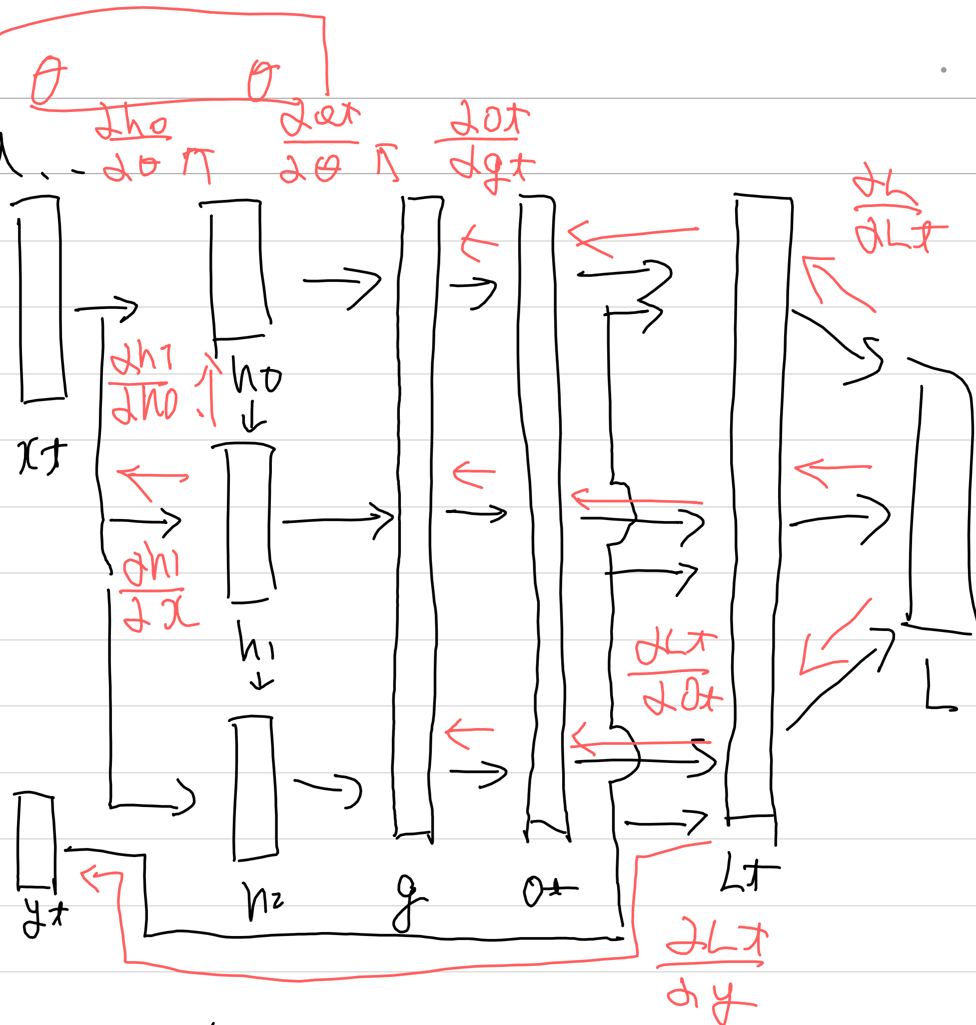
$h_t = f(h_{t-1}, x_t, \theta)$

Total loss of interests is $C = \sum_t L_t$

gradient propagation



Contd...



$$\frac{dC}{dL_t} = \frac{d}{dL_t}\left(L_1 + \cdots L_t + \cdots\right) = 1$$

$$\frac{dL_t}{dO_t} = \frac{d}{dO_t} L(O_t, y_t)$$

$$\frac{dL_t}{dy_t} = \frac{d}{dy_t} L(O_t, y_t)$$

$$\frac{dO_t}{d\theta} = \frac{d}{d\theta} g(h_T, \theta), \quad \frac{dO_t}{dh_t} = \frac{d}{dh_t} g(h_t, \theta)$$

$$\frac{dh_t}{dh_{t-1}} = \frac{d}{dh_{t-1}} f(h_{t-1}, x, \theta) \qquad \frac{dh_t}{dx} = \frac{d}{dx} f(h_{t-1}, x, \theta)$$

$$\frac{dh_t}{d\theta} = \frac{d}{d\theta} f(h_{t-1}, x, \theta)$$

17A Q4-2

時間方向での展開

$$\frac{\partial C}{\partial \theta} = \frac{\partial}{\partial \theta}\sum_t L_t$$

$$= \frac{\partial}{\partial \theta}\sum_t L(o_t, y_t)$$

$$= \frac{\partial}{\partial \theta}\sum_t L(g(h_t, \theta), y_t)$$

θを含むのは

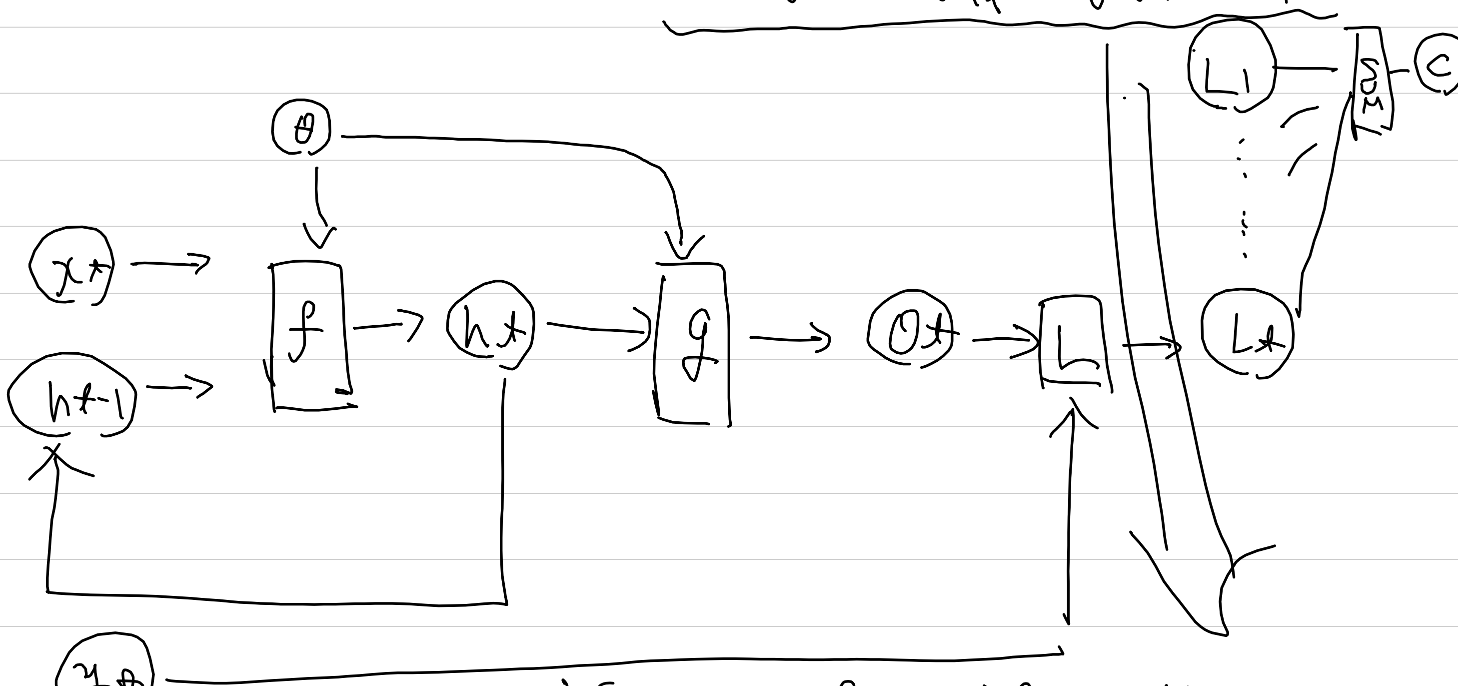
$\frac{\partial h_t}{\partial \theta}$ と $\frac{\partial o_t}{\partial \theta}$

$$\frac{\partial C}{\partial \theta} = \frac{\partial L_t}{\partial \theta}\frac{\partial C}{\partial L_t}$$

計算可能 = 1

$$= \frac{\partial o_t}{\partial \theta}\frac{\partial L_t}{\partial o_t}\frac{\partial C}{\partial L_t}$$

$$= \frac{\partial h_t}{\partial \theta}\frac{\partial o_t}{\partial h_t}\frac{\partial L_t}{\partial o_t}\frac{\partial C}{\partial L_t}$$



$$\frac{\partial C}{\partial \theta} = \frac{\partial f}{\partial \theta} \times \frac{\partial g}{\partial f} \times \frac{\partial L}{\partial g} \times 1$$

17A Q4 - 3

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

$$h_{t+1} = f(h_t, x_{t+1}, \theta)$$

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

$$\frac{\partial h_t}{\partial h_\tau} = \frac{\partial}{\partial h_\tau} f(h_t, x, \theta)$$

e.g. $h_t = h_{10}$
$h_\tau = h_7$

$$\frac{\partial h_{10}}{\partial h_7} = \frac{\partial h_{10}}{\partial h_9} \frac{\partial h_9}{\partial h_8} \frac{\partial h_8}{\partial h_7}$$

$$= \frac{\partial h_t}{\partial h_{t-1}} \cdots \frac{\partial \tau+1}{\partial \tau}$$

$$\frac{\partial h_t}{\partial h_\tau} = f'(h_{t-1}, x, \theta)(h_{t-1})'$$

$$= \frac{\partial}{\partial h_\tau} f(h_{t-1}, x, \theta) \frac{\partial h_{t-1}}{\partial h_\tau}$$

活性化関数の繰り返し適用は 異常な非線形挙動を引き起こしやすい

$$x_t \xrightarrow[\theta]{} h_t \xrightarrow[\theta]{} o_t \longrightarrow L_t \longleftarrow y_t$$

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

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

Same eigenvector wrt. $\frac{\partial h_t}{\partial h_\tau}$、

So、let consider f as linear function

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

$$= W^T h_{t-1} \quad \text{(just assume.)}$$

where $W = Q \Lambda Q^T$

収束
$\lambda < 1 \to 0$
$\lambda > 1 \to \infty$
発散

So、 $h_t = Q^T \Lambda^{t-\tau} Q h_\tau$

Then $\frac{\partial h_t}{\partial h_\tau} = \underline{Q^T \Lambda^{t-\tau} Q h_\tau}$

1回目のBackPropの勾配ベクトル

$g \to Jg \quad \cdots \quad \to J^n g$

J: hidden layerの重み行列
x: hidden layerに入ってくる特徴ベクトル

$$\text{where} \quad J_f = \frac{\partial f}{\partial x} \begin{pmatrix} \frac{\partial f_1}{\partial x} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

$$\text{where} \quad f = \begin{pmatrix} f_1(h_t, x_1 \cdots x_n, \theta) \\ \\ f_m(h_t, x_1 \cdots x_n, \theta) \end{pmatrix}$$

と定式化したとき gに $\delta \underset{\text{固有ベクトル}}{\underline{v}}$ で加えてみると

$$g + \delta v \to J(g + \delta v) \quad \cdots \quad \to J^n(g + \delta v)$$

その差分は $J^n(g+\delta v) - J^n g = J^n g + \underline{\delta |\lambda|^n}$

write down the Jacobian matrix $\frac{dh_t}{dh_\tau}$

$h_t =$

Contractive (縮退)

17A - Q4 - 4

$$x_t \to h_t \xrightarrow[\theta]{} O_t \to L_t \leftarrow y_t$$

$\frac{dC}{dL_t} = 1$ より

$$\frac{dC}{d\theta} = \frac{df}{d\theta}\,\frac{dg}{df}\,\frac{dL_t}{dg}\,\frac{dC}{dL_t}$$

$$= \frac{W h_t}{d\theta}$$

1