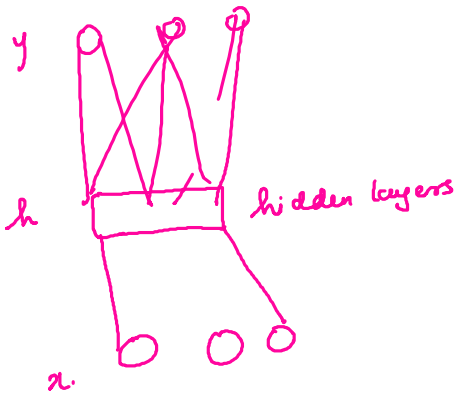
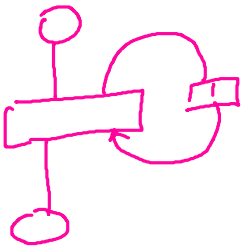
RNN

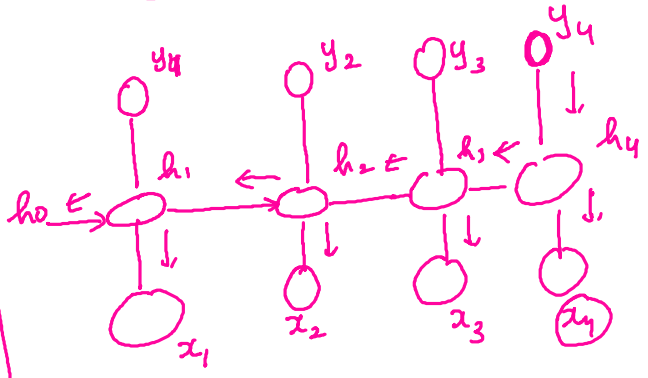


$$z_t = W_{hx}^T x + b_x$$

$$h_t = \sigma(z_t)$$

$$y = \text{softmax}(W_{hy} h_t + b_y)$$



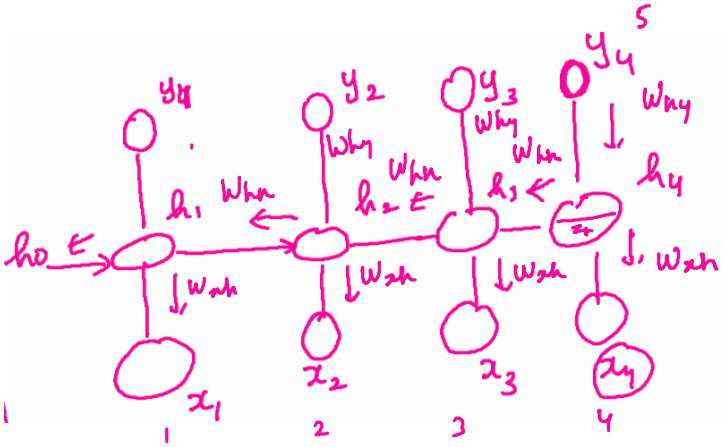


$$y_4 = f(h_3, x_4)$$

$$= f(f(h_2, x_3), x_4)$$

$$= f(f(f(h_1, x_2), x_3), x_4)$$

$$y_4 = f(f(f(f(h_0, x_1), x_2), x_3), x_4)$$

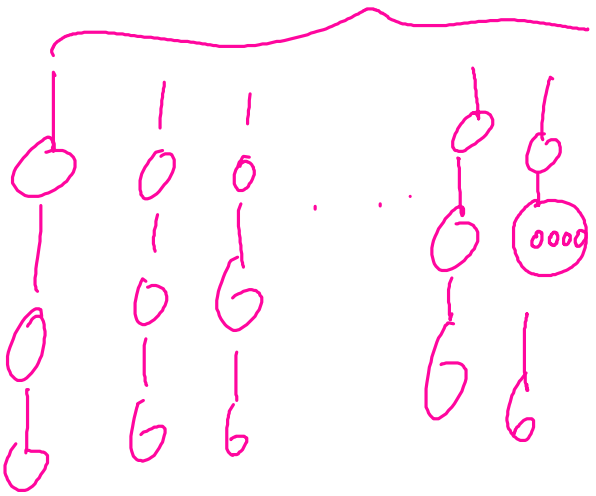


$$z_t = W_{xh} x + W_{hh} h_{t-1} + b_x$$

$$h_t = \tanh(z_t)$$

$$o_t = W_{hy} h_t + b_y$$

$$\hat{y} = \text{softmax}(o_t)$$



series = 1 2 3 . . . 700

seq. length = 10

features

1, 2, 3, ... 9

2, 3, ... 10

3, 4, 5, . . . 11

labels

10

11

12

1 50

seq. length

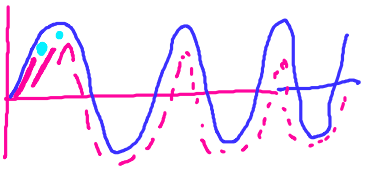
---

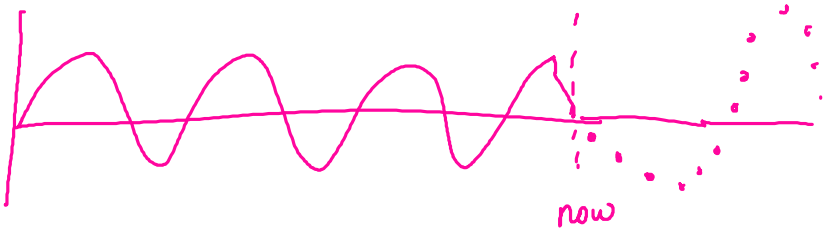
test x [0] ⇒ [0.5, 0.6, 0.7] [0.72] ✓

= [0.6, 0.7, 0.72] [0.7r]

= [0.7, 0.72, 0.7r] [0.8]

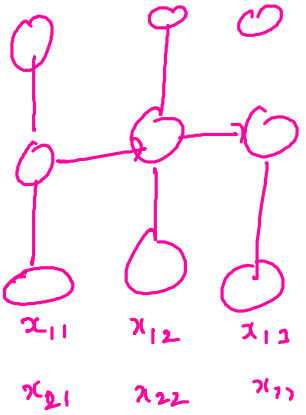
= [0.72, 0.7r, 0.8] [ ]





→ [[0.48, 0.611, 0.68]] [0.71]

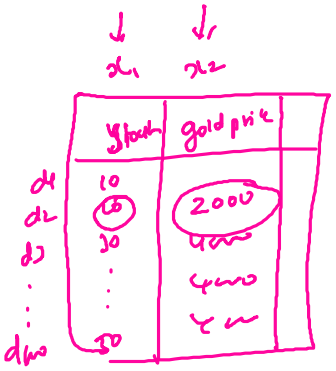
→ [[0.611, 0.65, 0.71]] [ ]



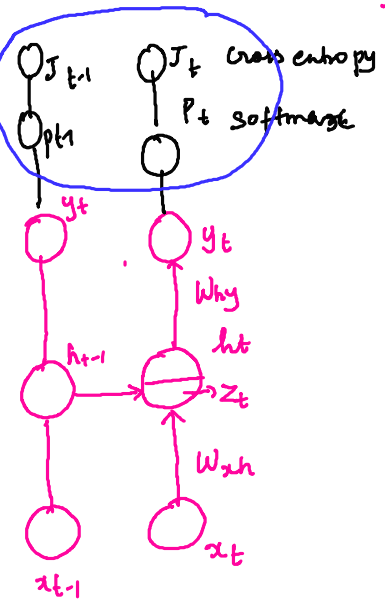
679, 20

679, 1, 20

679, 2, 20



Forward & Backward Pass RNN



Forward pass

$$z_t = W^{xh} x + W^{hh} h_{t-1}$$

$$h_t = \tanh(z_t)$$

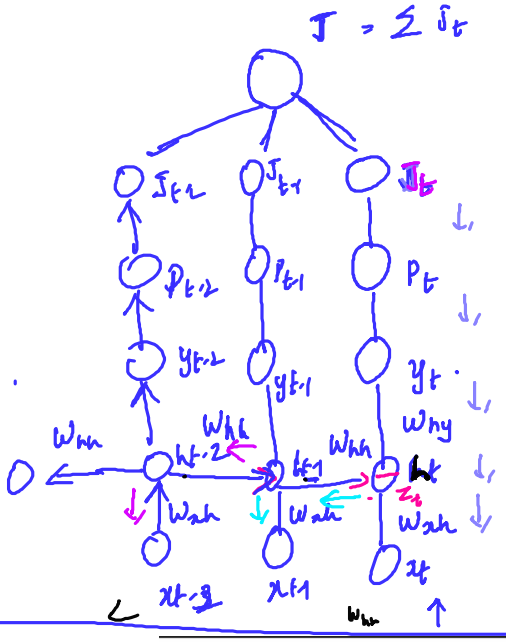
$$y_t = W^{hy} h_t$$

$$P_t = \text{softmax}(y_t)$$

$$J_t = \text{crossentropy}(P_t, \text{labels}_t)$$

&

$$J = (W^{xh}, W^{hh}, W^{hy})$$



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

Back Propagation in Time

$$\frac{\partial J_t}{\partial W^{hy}} = \frac{\partial J_t}{\partial P_t} \cdot \frac{\partial P_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial W^{hy}}$$

$$\frac{\partial J_t}{\partial P_t} = \frac{\partial}{\partial P_t}(\text{cross entropy}(P_t, \text{labels}))$$

$$\frac{\partial P_t}{\partial y_t} = \frac{\partial}{\partial y_t}(\text{Softmax}(y_t))$$

$$\frac{\partial y_t}{\partial W^{hy}} = \frac{\partial}{\partial W^{hy}}(W^{hy} h_t)$$

Forward pass

$$z_t = W^{xh} x + W^{hh} h_{t-1}$$

$$h_t = \tanh(z_t)$$

$$y_t = W^{hy} h_t$$

$$P_t = \text{softmax}(y_t)$$

$$J_t = \text{crossentropy}(P_t, \text{labels}_t)$$

&

$$J = (W^{xh}, W^{hh}, W^{hy})$$

$$\frac{\partial J_t}{\partial W^{hh}} = \frac{\partial J_t}{\partial P_t} \cdot \frac{\partial P_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial W^{hh}} +$$

$$\frac{\partial J_t}{\partial P_t} \cdot \frac{\partial P_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial z_{t-1}} \cdot \frac{\partial z_{t-1}}{\partial W^{hh}} +$$

$$\frac{\partial J_t}{\partial P_t} \cdot \frac{\partial P_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial z_t} \cdot \cdots \cdot \frac{\partial z_{t-1}}{\partial h_{t-2}} \cdot \frac{\partial h_{t-2}}{\partial z_{t-2}} \cdot \frac{\partial z_{t-2}}{\partial W^{hh}}$$

$\frac{\partial J_t}{\partial W^{ah}} = ???$ Exercise

Derivative of mse

sigmoid

softmax

cross entropy

---

features

[Th, i, course, on]

[i, coose, on, natural]

labels

[natural] [50)

[language] [20]