$\{1\} \quad \{2\} \quad \{3\} \quad \{4\} \quad \{6\}$

$1 \quad 4 \quad 9 \quad 16 \quad 36$

$$f_n = f_{n-1} + f_{n-2}$$

$$\begin{cases} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 2 & 3 & 5 & 8 & 13 \end{cases}$$

Let the data be $\quad S_1 \quad S_2 \quad S_3 \quad \cdots \quad S_{t-1} \quad S_t$

$S_t = f(S_{t-1}, \theta)$ when $\not\exists\, s_t = 3s_{t-1} + 5$ (θ)

↳ how do we discover this

Day	1	2	3	....	900	901	902	} stock market
	s_1	s_2	s_3		s_{900}	s_{901}	s_{902}	

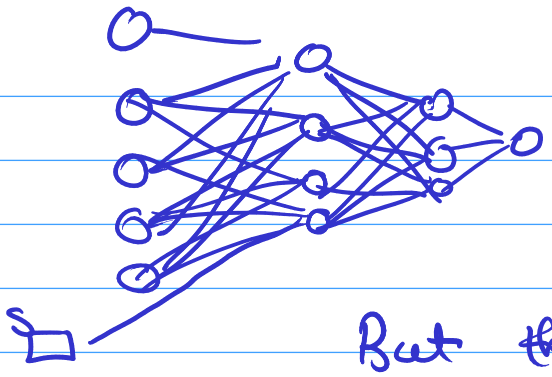
now lets say that $\not{f}(x_1, x_2 \cdots x_{50})$ be the feature vector for 1 day

$S_{903} = f(x_1, x_2 \cdots x_{50})$ is wrong

$S_{903} : f(S_{902}, x_1, x_2 \cdots x_{50})$

↳ consider this also as a feature

Develop a model



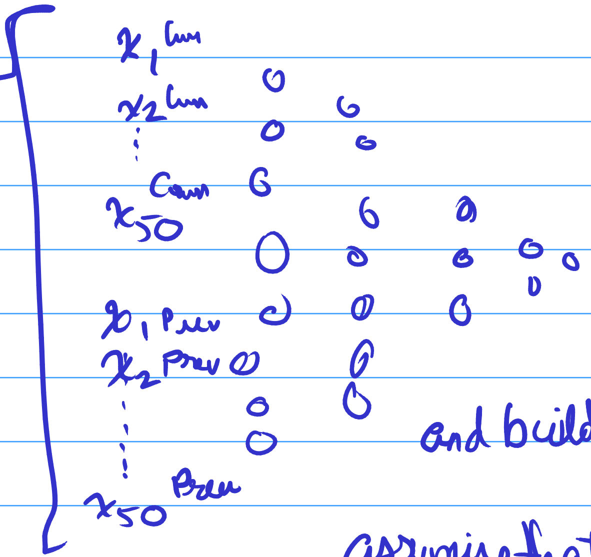
But that's stupid

S_{902} depends on the feature vector of the day 902

So we get

$g(x_1^{(902)}, x_2^{(902)}, x_3^{(902)} \cdots x_{50}^{(902)}; x_1^{(903)}, x_2^{(903)} \cdots x_{50}^{(903)})$

so we do



and build a model

assuming that today's condition doesn't depend on day before's state

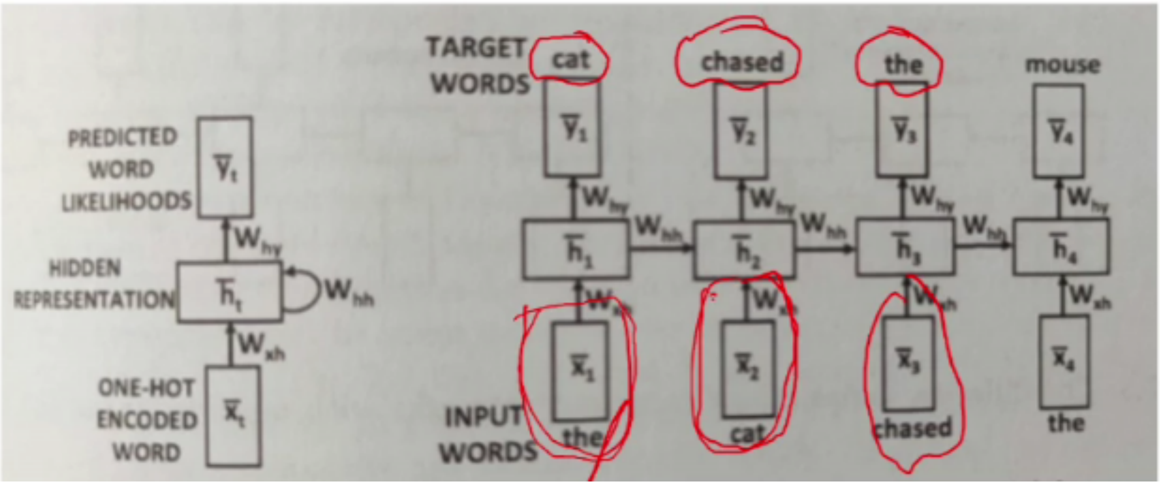
But as per diff. Financial experts, diff number of days are required!

$\Rightarrow$ each model for # of days! TF?

Another problem

The cat chased the . . . .

Now the issue is, the later words can influence the sentence



$\{\ \} + \text{chased} = \text{the}$

The essence of the previous input & the current inputs } that function remains the same over the sequence

f is the same

$f(h_0, x_1)$

$f(h_1, x_2)$

$f(h_2, x_3)$

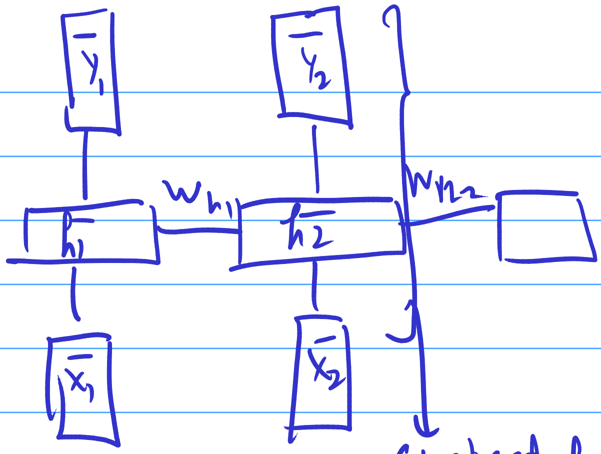
$f(h_3, x_4)$

if you believe that this f must be the same

almost..

f needn't depend only on position

So the output is independent of the position of input in the sequence by assuming that the function is same

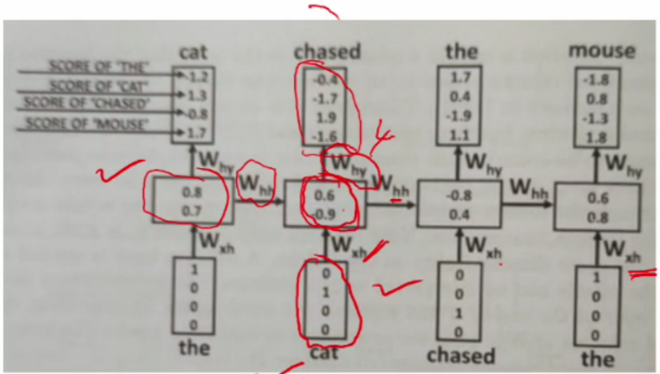


abstracted info
is sent to the next guy
(same dimension as X by the way)

So
our abstraction 'h' is as follows:

$$h_2 = \sigma(\underbrace{W_1 h_1 + W_2 X_2}_{\text{encode}})$$

$$\underbrace{W_{hy} h_2}_{\text{decode, get scores/word vector}} \Rightarrow \text{word}$$

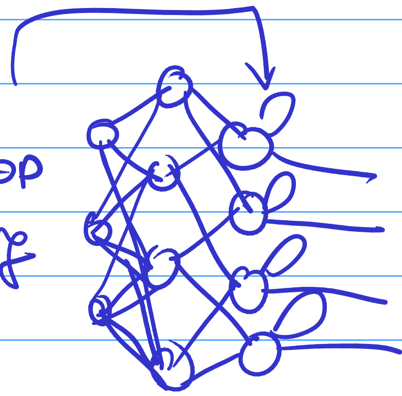


$$\bar{h}_t = \tanh(w_{xh}\bar{x}_t + w_{hh}\bar{h}_{t-1})$$

$$\bar{y}_t = w_{hy} h_t$$

BITS Pilani, Hyderabad C

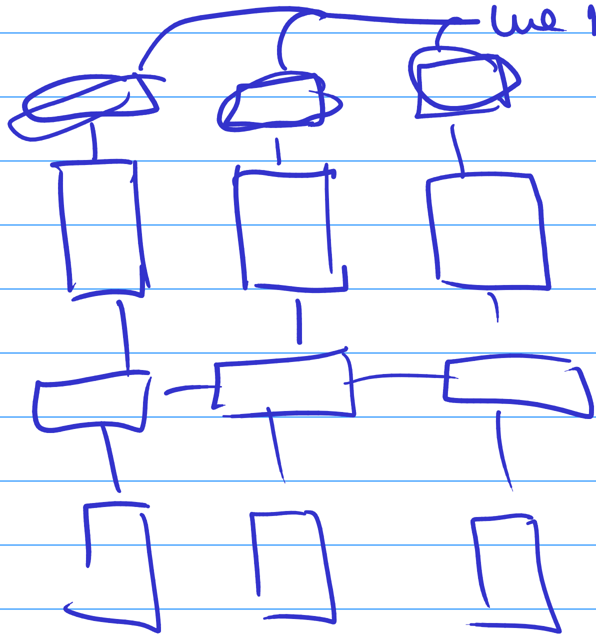
This is loop unrolling



$$\bar{h}_t = \tanh(w_{xh}\bar{x}_t + w_{hh}\bar{h}_{t-1})$$

now we know the ground truth whilst training

if ground truth is $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow$ then '1' has to be maximised.



we minimise the loss for entire sequence

The sequence length can vary but thats not a problem, since we built the model with that in mind as seen below

$$L = -\log(p_i^{y_i})$$

for gradient descent

$$\frac{\partial L}{\partial w_{xh}} \quad \frac{\partial L}{\partial w_{hh}} \quad \frac{\partial L}{\partial w_{hy}}$$

$$\begin{array}{cc} \begin{bmatrix} 1.8 \\ 1.3 \\ 0.8 \\ 1.7 \end{bmatrix} & \begin{array}{c} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{array} \end{array}$$

$\frac{\partial L}{\partial \hat{y}_1}$ we can find

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = W_{hy} h_1$$

$\frac{\partial L}{\partial W_{xh}}$: $g\, x$
↳ accumulated error

also $g^{(2)} x^{(2)}$ (so we get multiple gradients)
$g^{(3)} x^{(3)}$

$L = -\log(W_{xy} h h)$

how to maintain W_{xh} to be the same?

Similar to CNNs (Refer notes) $\frac{\partial L}{\partial \omega_{xh}^{(1)}}, \frac{\partial L}{\partial \omega_{xh}^{(2)}}, \frac{\partial L}{\partial \omega_{xh}^{(3)}}, \frac{\partial L}{\partial \omega_{xh}^{(4)}}$

$$\frac{\partial L}{\partial \omega_{xh}} = \frac{\partial L}{\partial \omega_{xh}} * \left(\frac{\partial \omega_{xh}}{\partial \omega_{xh}^{(1)}} + \frac{\partial W_{xh}}{\partial \omega_{xh}^{(2)}} + \frac{\partial W_{xh}}{\partial \omega_{xh}^{(3)}} \right)$$

$$\omega_{xh} = \omega_{xh}^{(1)} = \omega_{xh}^{(2)} = \omega_{xh}^{(3)} \Rightarrow \frac{\partial W_{xh}}{\partial \omega_{xh}^{(1)}} \Rightarrow 1$$

RNN → Recurrence