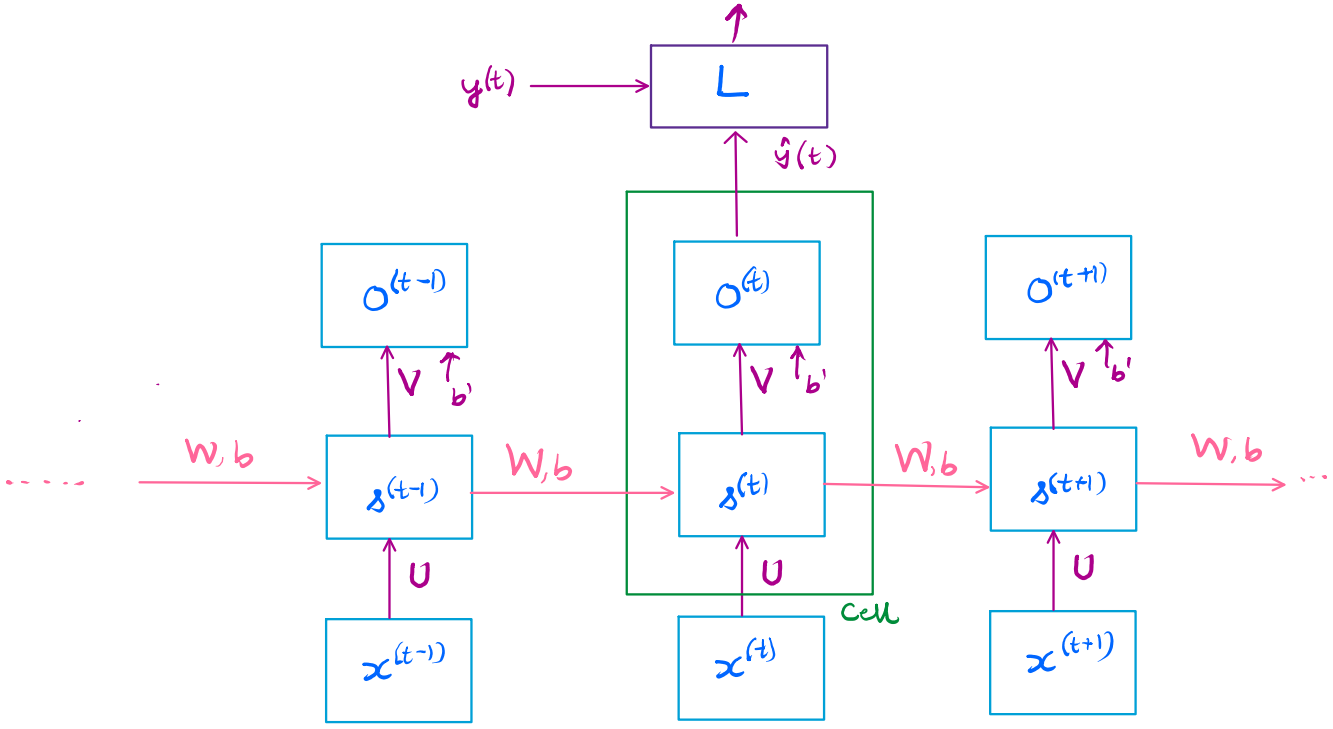
Lecture 13: Recurrent Neural Networks

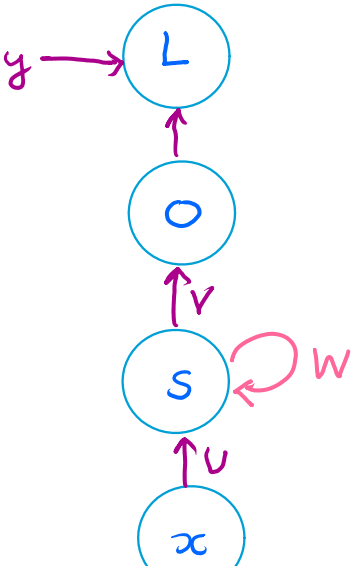
06 March 2024 10:57

In fit forward Neural Networks (FFNN), the size of data set is fixed. So, when data is sequential, (for eg: a sentence) we need to use more parameters, and there is a high chance of overfitting (due to huge redundancy) In such data, the ordering matters, and the length (size of data size) will be variable.

Examples i) autocomplete feature while typing
ii) Grammatic error correction
iii) Virtual assistance
iv) Language translation.

We use Recurrent Neural Network (RNN) as a way to get around this with, by changing the architecture

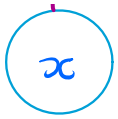




$$s^{(t)} = g(Ws^{(t-1)} + Ux^{(t)} + b) \quad g \in \sigma, \tanh, \text{ReLU}$$

$$O^{(t)} = Vs^{(t)} + b'$$

$$\hat{y}^{(t)} = \text{softmax}(O^{(t)})$$

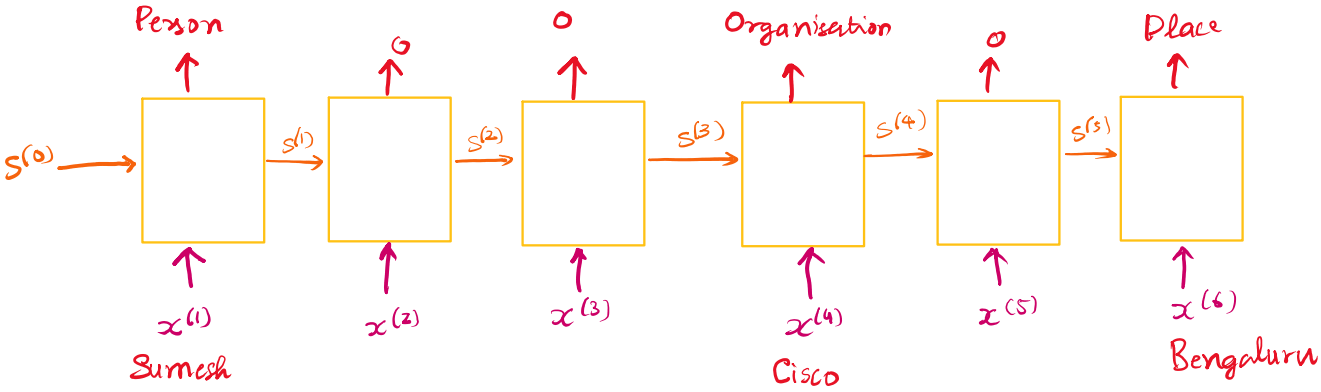


Example (Named Entity Recognition):

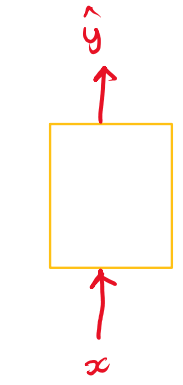
"Sumesh worked at Cisco in Bengaluru"

Person Organisation Place

We need to identify the "type" of each word.

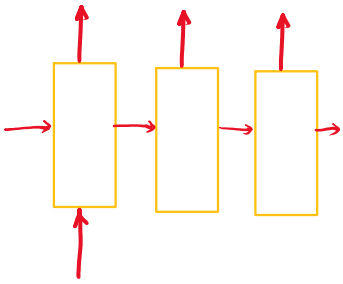


Possible Structures of RNN



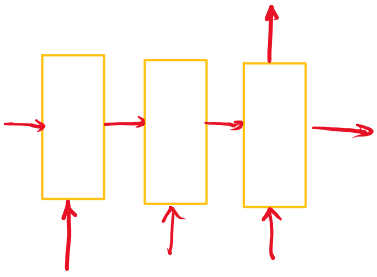
FFNN

One - many RNN



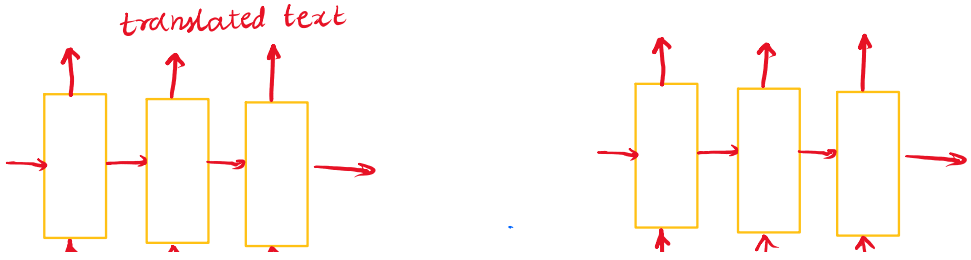
Eg: Image captions

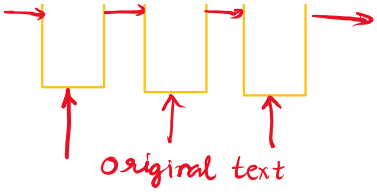
Many - one RNN



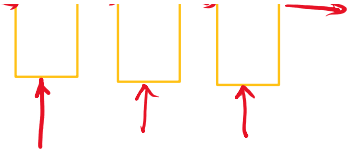
Eg: Rating

Many - Many RNN





Staggered : $len(x) \neq len(\hat{y})$
Eg : translation

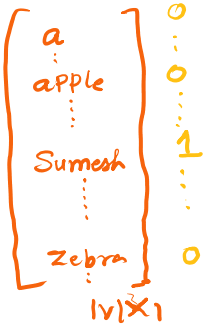


Synchronized
Eg : Named Entity Recognition

Coming back to our example :

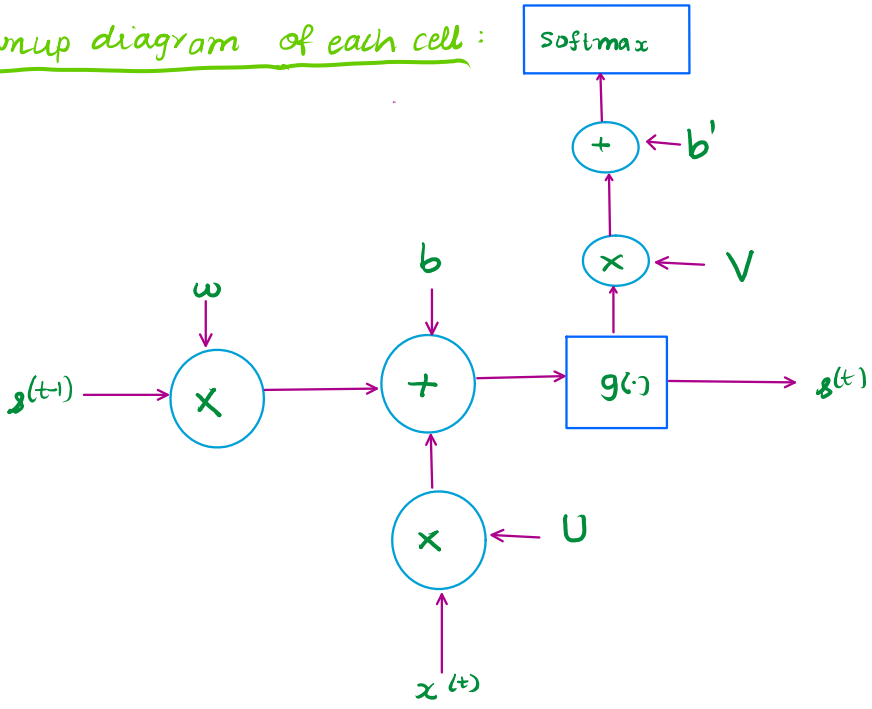
"Sumesh worked at Cisco in Bengaluru"
Person organisation Place

We have a vocabulary, V. The input is given as a vector with length |V|, using 1-hot representation
See eg of x_1 shown. Similarly, $x_2, \ldots x_6$ are also obtained as vectors in one hot representati

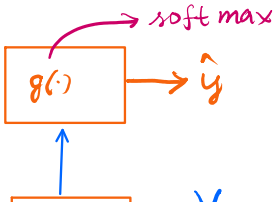


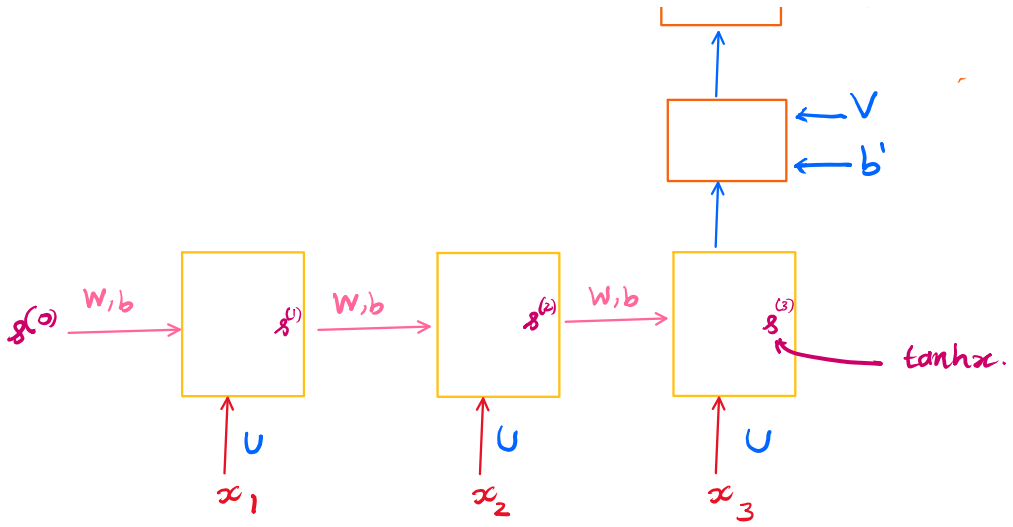
Eg: x_1

Blownup diagram of each cell :



Training RNN's





$$\text{Loss}(y,\hat{y}) = -\sum y_k \log \hat{y}_k = L \qquad \frac{\exp\{w^T \dots\}}{\sum \exp\{\dots\}}$$

$$= -\sum_{k=1}^{K} \mathbb{I}\{y=k\} \log(\hat{y}_k)$$

Categorical loss entropy fn

We adopt Backpropagation through time BPTT to train RNN.

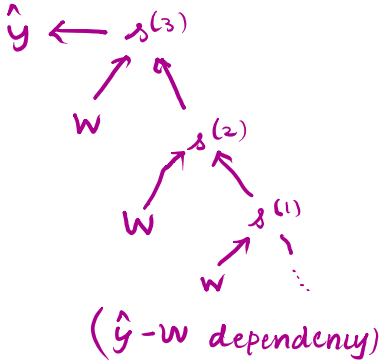
(a) Forward Pass: Compute $s^{(1)}, s^{(2)}, \dots$, and $\hat{y}$.

Here, $s^{(t)} = \tanh\left(W s^{(t-1)} + U x^{(t)} + b\right)$

$\hat{y} = \text{softmax}\left(V s^{(t)} + b'\right)$

(b) Backward Pass: We need to compute $\frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}, \frac{\partial L}{\partial U}, \frac{\partial L}{\partial V}, \frac{\partial L}{\partial b'}$.

- Observe $\frac{\partial L}{\partial V} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$ $\quad \frac{\partial L}{\partial \hat{y}} = \frac{\partial L}{\partial \hat{y}}\left(-\sum y_k \log \hat{y}_k\right)$ $\quad \frac{\partial \hat{y}}{\partial V}$ can be easily calculated
- Similarly we can calculate $\frac{\partial L}{\partial b'}$



- $\frac{\partial L}{\partial W} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W}$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s^{(3)}} \cdot \frac{\partial s^{(3)}}{\partial W} + \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s^{(3)}} \cdot \frac{\partial s^{(3)}}{\partial s^{(2)}} \cdot \frac{\partial s^{(2)}}{\partial W} + \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s^{(3)}} \cdot \frac{\partial s^{(3)}}{\partial s^{(2)}} \cdot \frac{\partial s^{(2)}}{\partial s^{(1)}} \cdot \frac{\partial s^{(1)}}{\partial W}$$

$$= \sum^{T} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s^{(i)}}$$

$$= \sum_{i=1}^{T} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s^{(i)}} \frac{\partial s^{(i)}}{\partial w}$$

- $\frac{\partial L}{\partial u}$ can also be calculated in a similar way as the dependents of $\hat{y}$ and u is very similar to W.

Disadvantages:

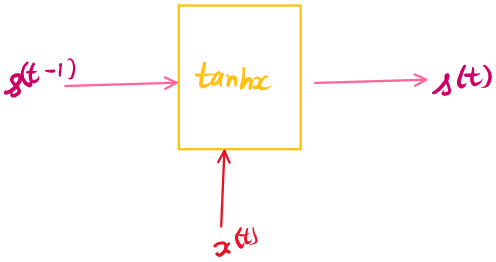
* (Vanishing Gradient Problem): In traditional RNN, the grammar checker won't be able to detect as error as below: (because gradient vanishes for long inputs)

Isha went for a vacation ..... He was overjoyed

long text

To mitigate this, a modification to the cell structure is done.

Cell before modification



Cell after modification.