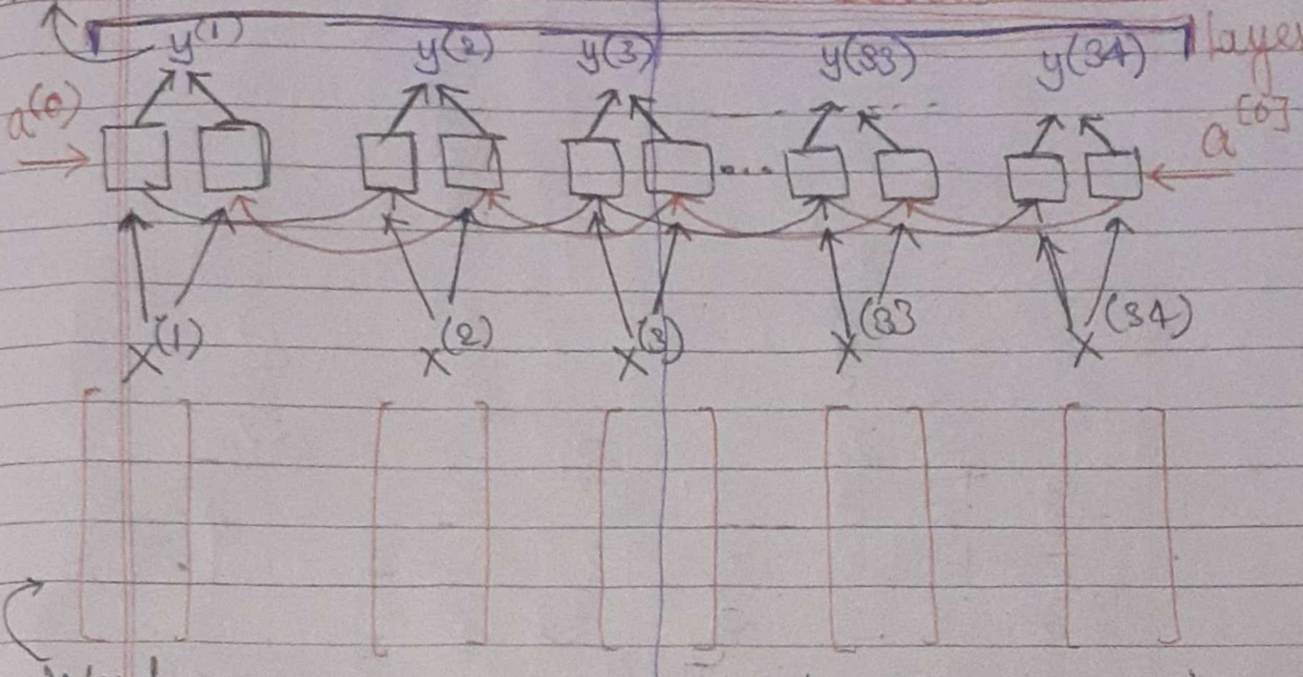
Project Work

34 LSTM (Cells)
One Bidirectional LSTM Layer

hidden state



Word embedding (300 dimentional)

Few Sentences are Padded so that Length of all Sentences goes upto (34) i.e Length of longest Sentence

ch $\quad A = \text{Bidirectional}(\text{LSTM}(n\_a, \text{return\_seq} = \text{True}))(X)$

This A will be the hidden state from
t function each of Bidirectional LSTM Cell

$y^{(1)}$ $y^{(2)}$ $y^{(3)}$ ——→ $y^{(33)}$ $y^{(34)}$

A =

of Number of training examples (m)

$m \times 34 \times (n\_a)$

n_a → dimention of each hidden state ($y^{(i)}$)

m → Number of training examples

34 → Number of hidden state required for each training examples

$A = m \times (34) \times (n\_a)$

(*) Now Size of each $y^{(i)}$ is n_a $\quad (n\_a = 32)$

Context Vector is Calculated as:

$$c^{[1]} = \sum_{i=1}^{34} \underbrace{\alpha^{(1,t_i)}} \cdot y^{(t_i)}$$

This is input to first Timestep of Post attention LSTM.

$c^{[1]}$ will have Same dimention as That of $y^{(t_i)}$

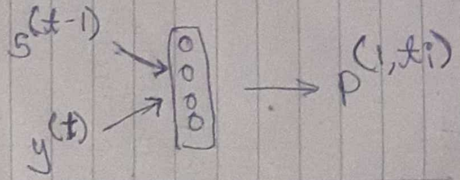
$\alpha^{(1,t_i)}$ is a Simple Number.

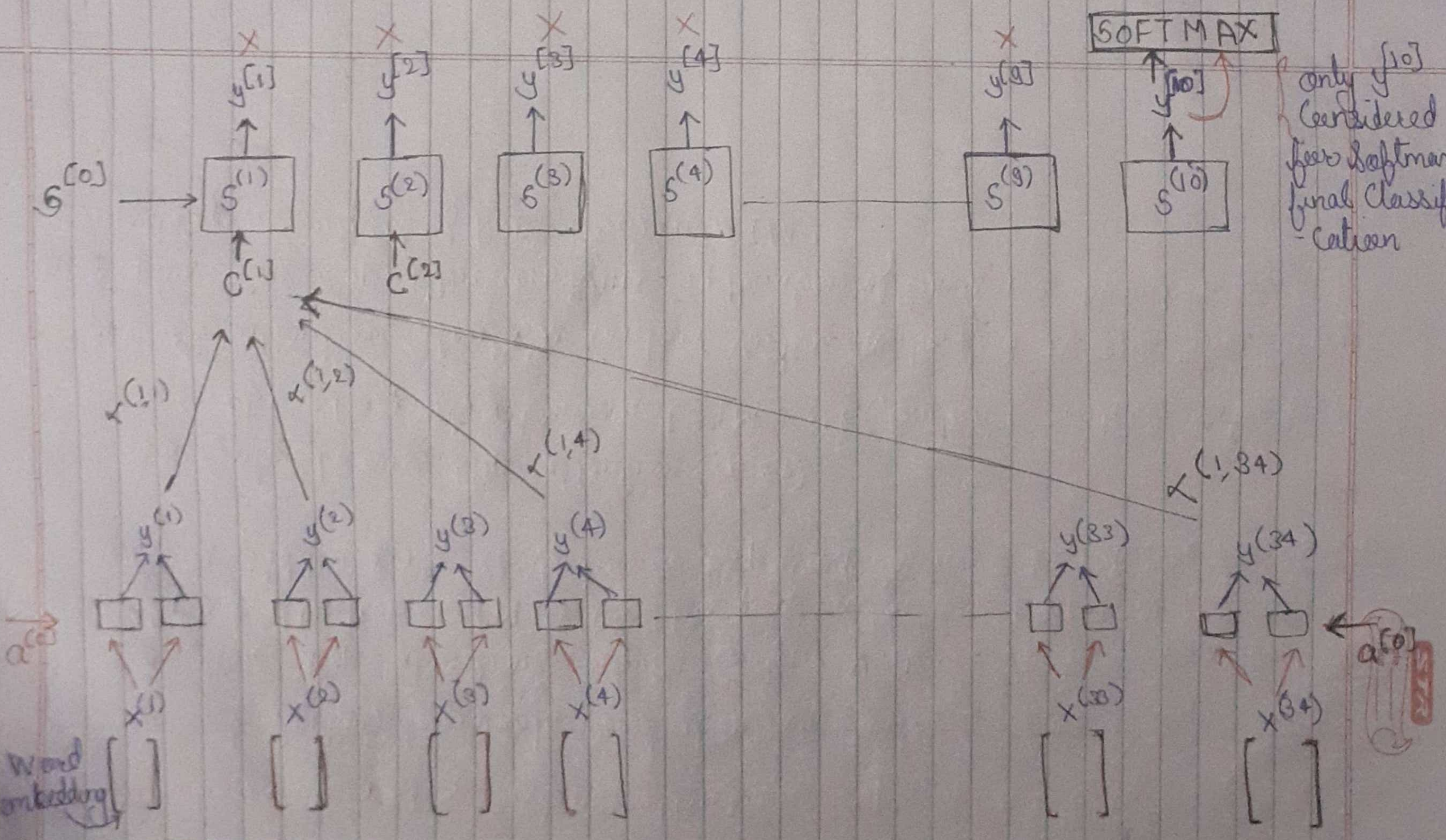
↓

Amount of attention $y^{(t_i)}$ should give While decoding 1st Word.

$$\alpha^{(1,t_i)} = \frac{e^{[p^{(1,t_i)}]}}{\sum_{j=1}^{34} e^{[p^{(1,t_j)}]}}$$

Now this $p^{(1,t_i)}$ is Calculated using a Small NN





we have created a Custom attention Layer, which Calculates Context Vector for each Post attention (2^{nd} Layer LSTM) input <u>Timestamp</u>

One-step-attention outputs Context Vector, to be feed as input to 1 timestamp ~~to~~ post-attention-LSTM-Cell. This post-attention-LSTM Cell further outputs hidden state & Vector to be passed to next LSTM-Cell timestamp

Context1 = one-step-attention(a, si)

s1, _, c1 = post-attention-LSTM-Cell(Context1, initial-state=[si, ci])

one-step-attention Calculates Context Vector to be feed to one timestamp, and This One-step-attention runs Ty Number of times i.e Number of timestamp in output of Post attention LSTM.