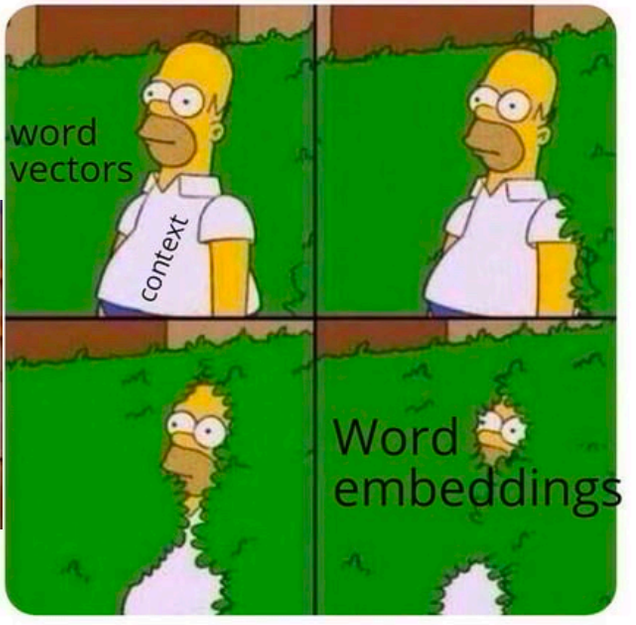
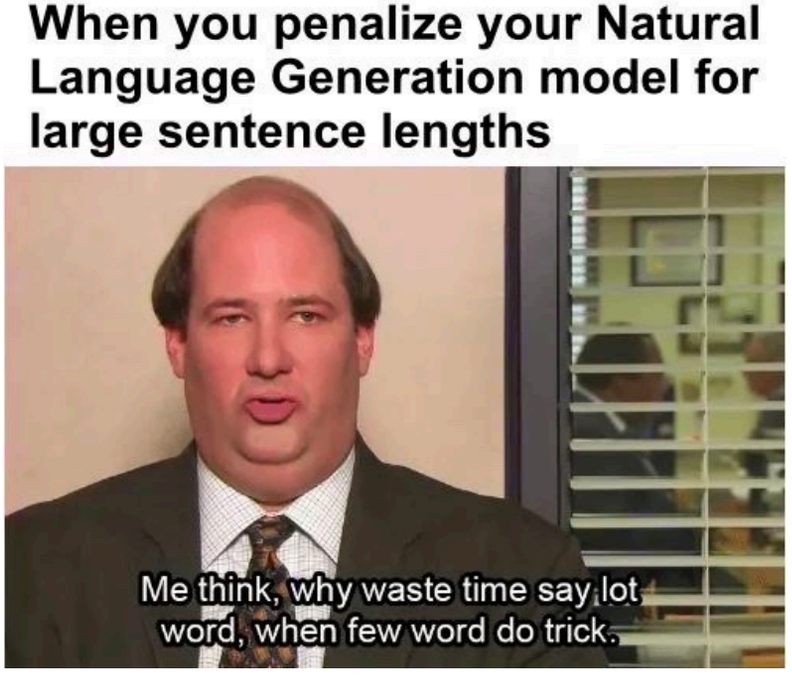
May 8, 2023.

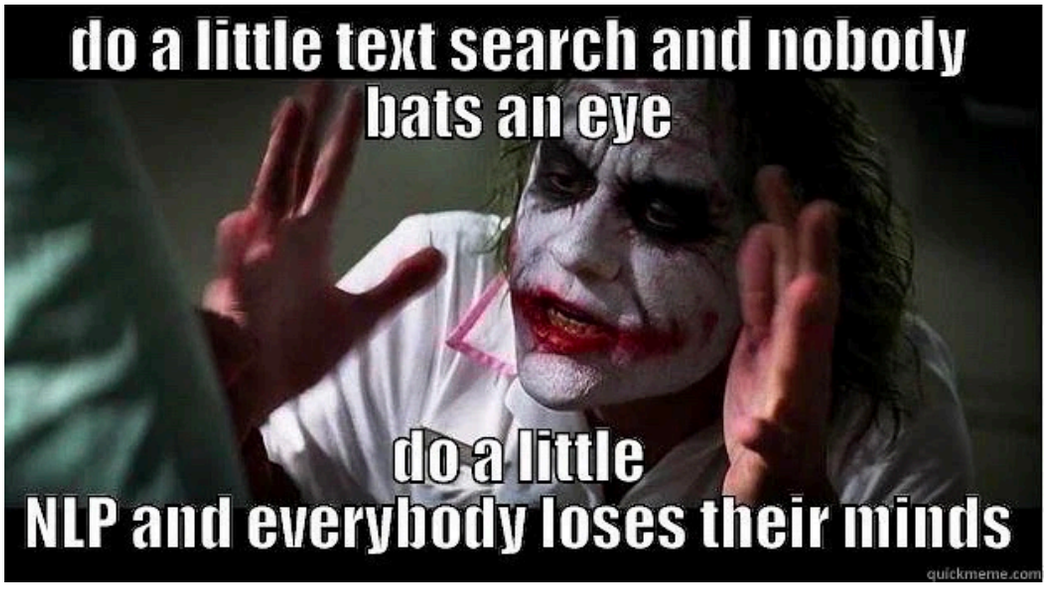
DSML : NLP module.

Named Entity Recognition.

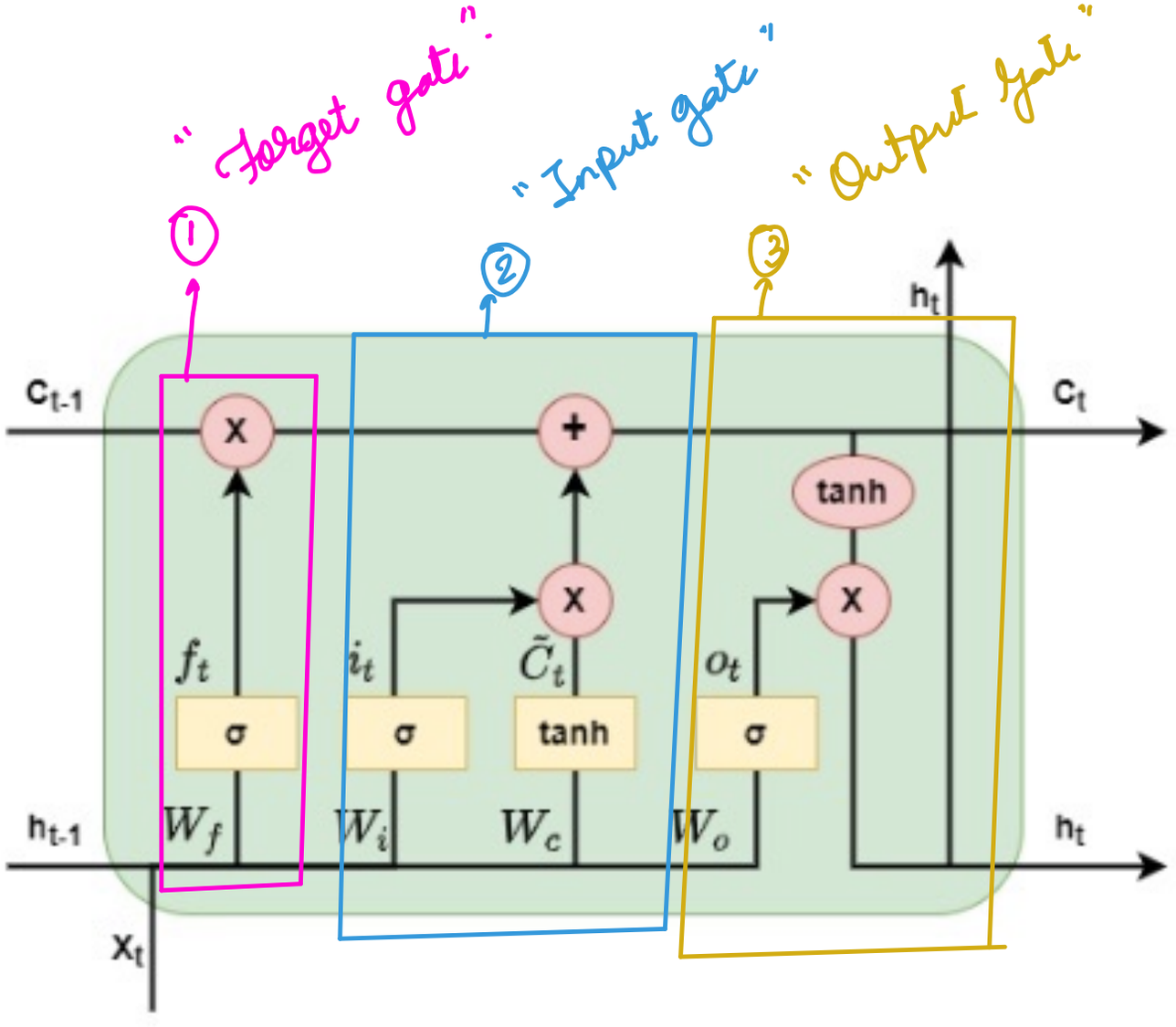
Class starts @ 9:05







Recap: Long Short-term Memory.

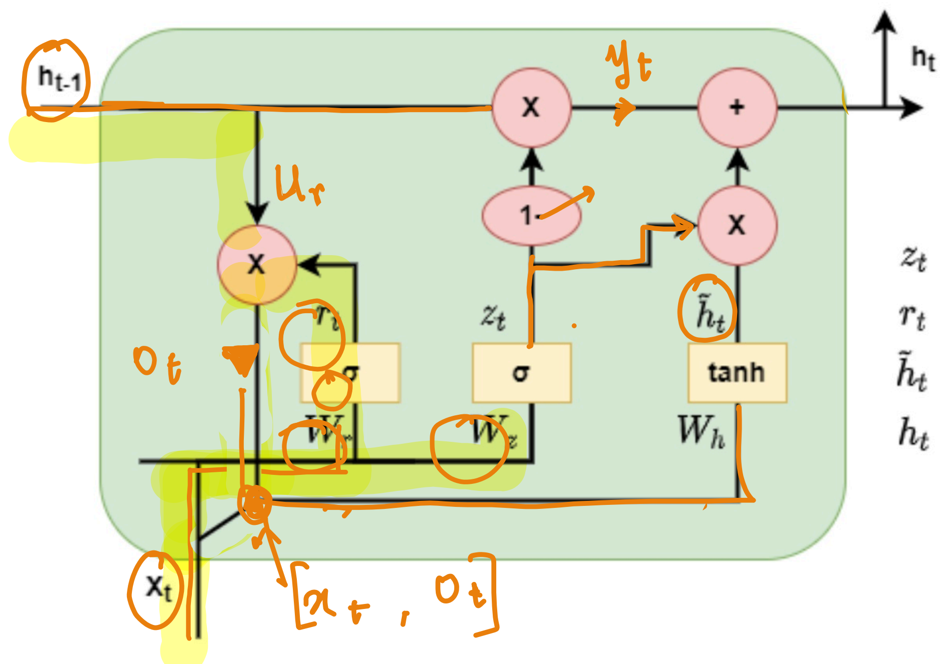


Agenda for today:

* Gated Recurrent Unit: LSTM variant with just one hidden state.
* Stacked LSTMs: Increasing the "depth" of LSTMs.
* LSTM code implementation.
* Named Entity Recognition
  - Rule-based
  - Conditional Random fields
  - Bi-directional LSTMs.
  - Hybrid approaches.
* NER implementation code.

Gated Recurrent Units.

⊛ → Elementwise multiplication.



Replace all x_t with $[x_t, h_{t-1}]$ ← peephole connection.

$$z_t = \sigma(W_z * [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r * [h_{t-1}, x_t])$$
$$\tilde{h}_t = tanh(W_h * [h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

These are wrong.

$$r_t = \sigma(W_r x_t + b_r)$$

$$o_t = h_{t-1} \circledast r_t \quad \text{OR} \quad (U_{rr} \cdot h_{t-1} + b_{rr}) \circledast r_t.$$

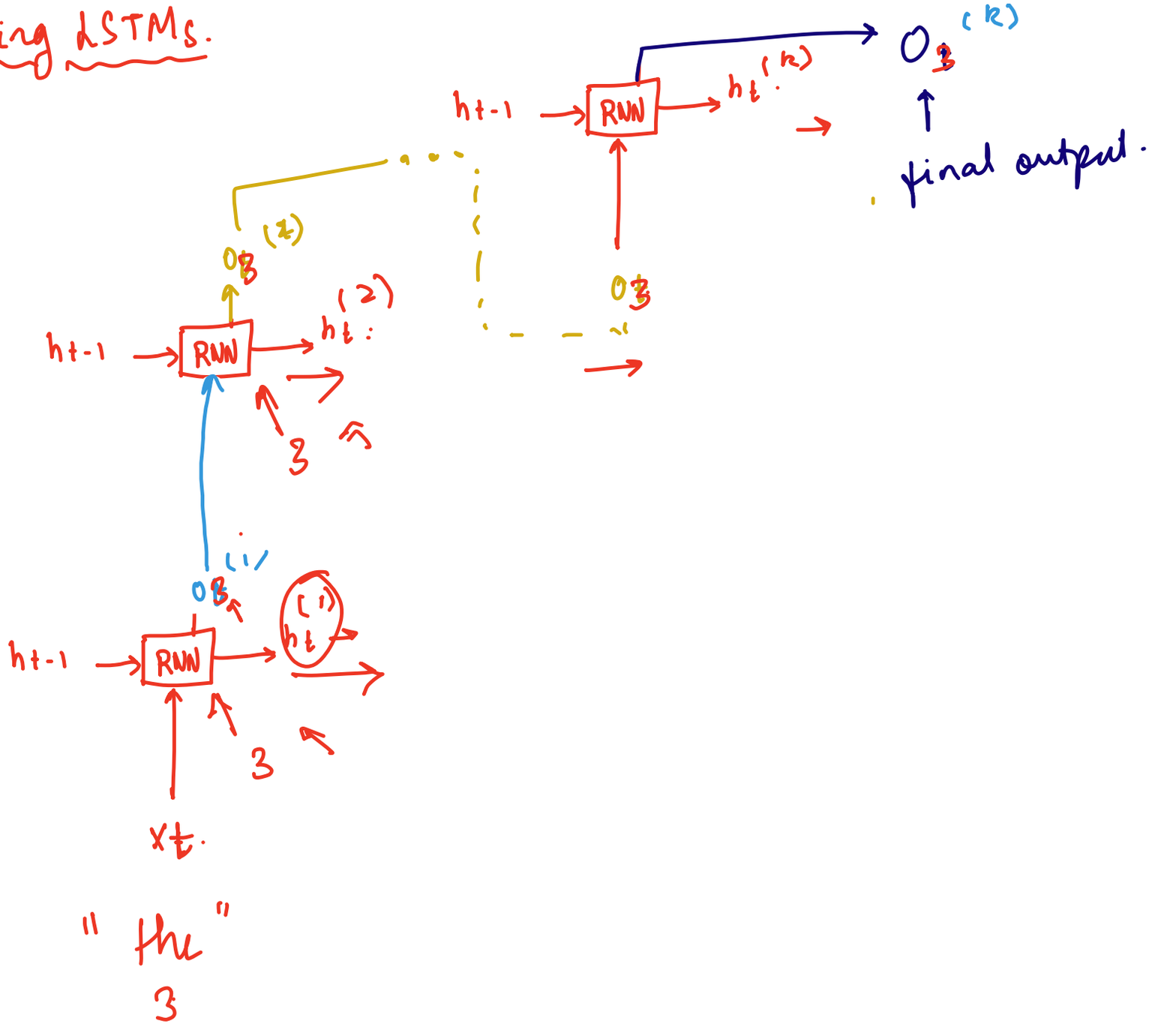
$$z_t = \sigma(W_z \cdot x_t + b_z)$$

$$y_t = h_{t-1} \circledast (1 - z_t)$$

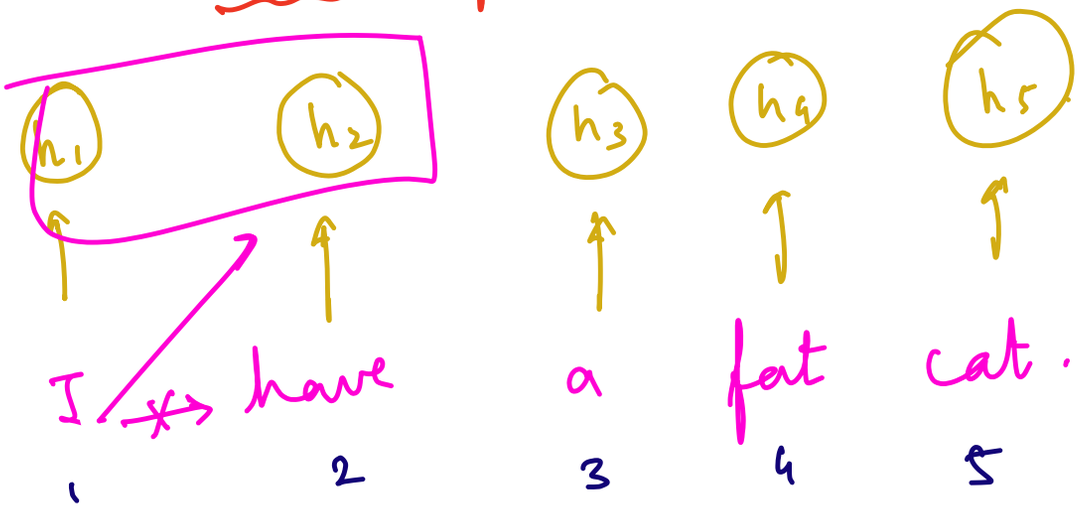
$$\tilde{h}_t = tanh(W_h \cdot [x_t, o_t] + b_h)$$

$$h_t = y_t + \tilde{h}_t \circledast z_t.$$
$$= h_{t-1} \circledast (1 - z_t) + \tilde{h}_t \circledast z_t.$$

Stacking LSTMs.



Conditional Random fields



→ p("have" | "I")

word word.

weight term feature.

CRF:

$$p_\theta(y|x) = \frac{exp(\sum_j w_j F_j(x,y))}{\sum_{y'} exp(\sum_j w_j F_j(x,y'))}$$

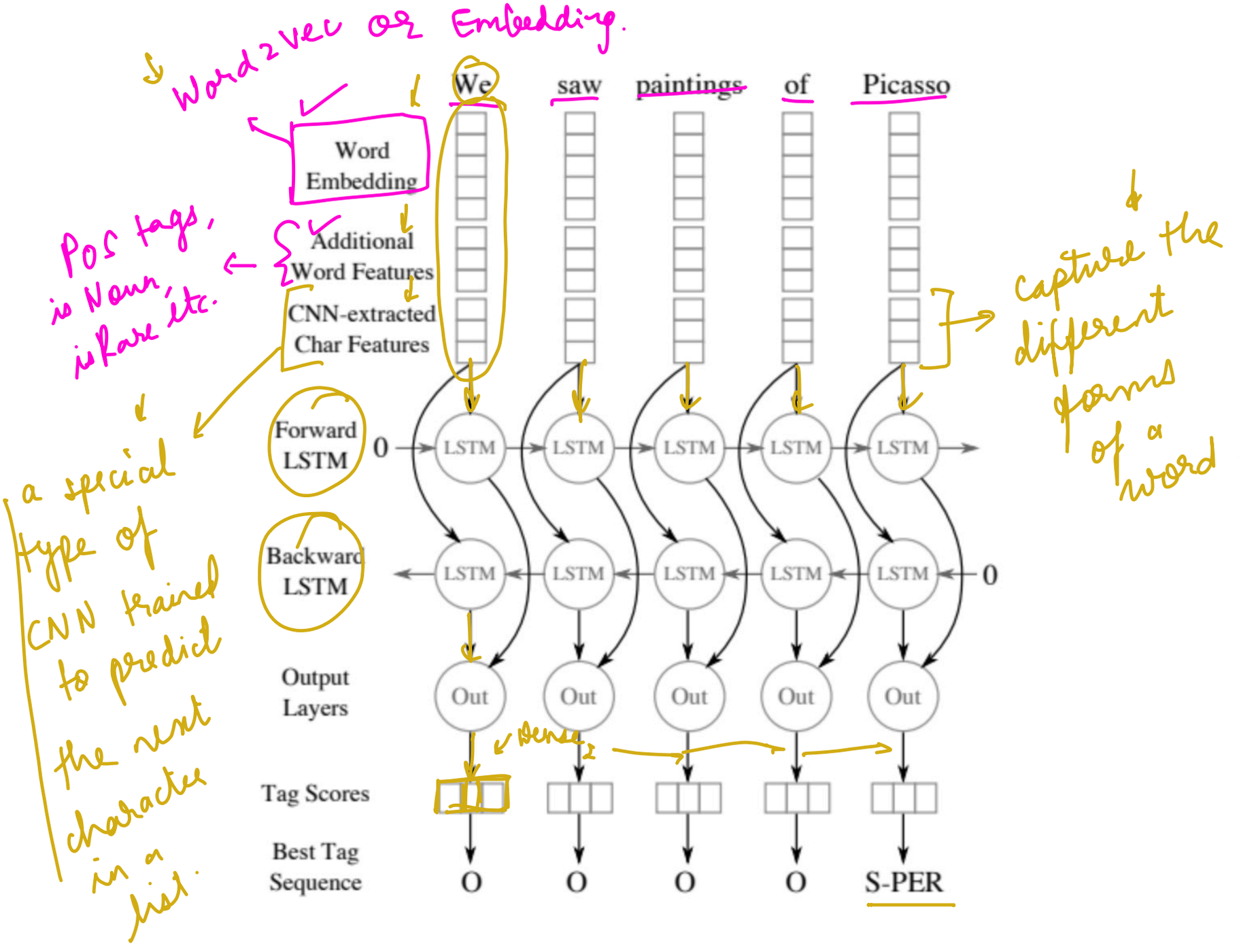
, where $F_j(x,y) = \sum_{i=1}^{L} f_j(y_{i-1}, y_i, x, i)$

L → sentence length.

Bi-directional RNN:

→ We have 1 hidden state for start-to end, and

1 hidden state for end-to-start.


Word2vec or Embedding.
We
saw
paintings
of
Picasso
Word Embedding
Additional Word Features
CNN-extracted Char Features
POS tags, is Noun, is Rare etc.
a special type of CNN trained to predict the next character in a list.
capture the different forms of a word.
Forward LSTM
Backward LSTM
0
LSTM
Out
Output Layers
Dense
Tag Scores
Best Tag Sequence
O
O
O
O
S-PER