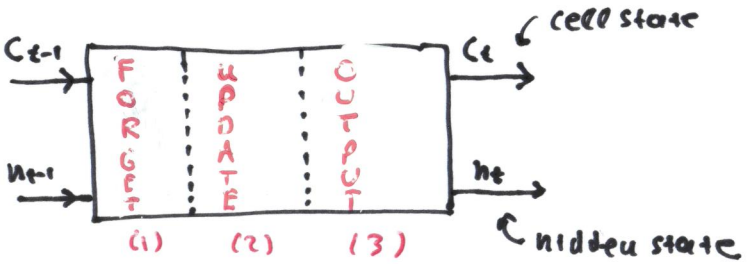
LSTMs: [Long-Short Term Memory Networks]

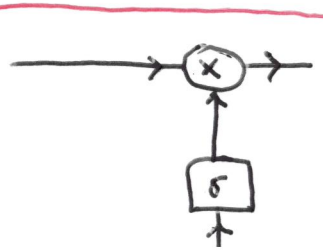
I: High Level Overview:

LSTMs' operation is based on a 3 steps process:



(1): Forget irrelevant parts of the previous state: For example, if we are modelling a sentence and we see a new subject, we might want to forget information concerning the old subject because we anticipate that future words will most probably relate to the new subject.

- Information is added or removed to the cell state through structures called gates



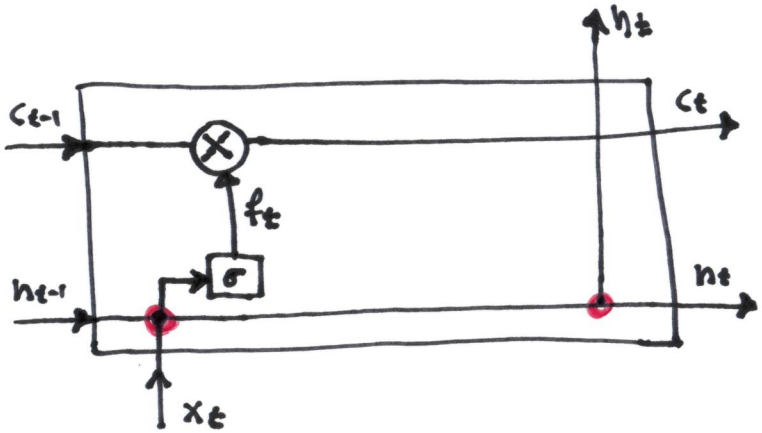
- Gates optionally let information through, via a sigmoid neural net layer and point-wise multiplication

(2): Update the cell state to reflect information according to the new input: In the context of the previous example, we update the cell state according to the new subject, i.e. its gender or whether it is in plural or singular form.

(3): Output certain parts of the cell state: For example, if we have just seen a new subject, we should predict information about the corresponding new verb as its tense.

- A cell state (C_t) is maintained making it easy for information to flow within the cell.
- An output gate (h_t) is utilized to output parts of the cell state.

LSTMs : Forget Irrelevant Information

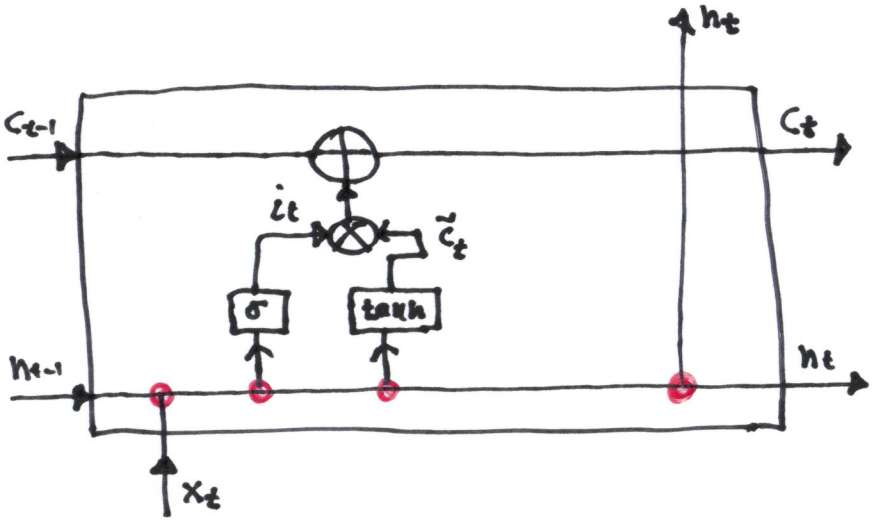


$$f_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_f\right)$$

- Use previous cell output and input.
- Sigmoid value within $[0, 1]$ range:
  - 0 indicates to completely forget
  - 1 indicates to completely remember

- For example, the gender pronoun of the previous subject in the sentence.

LSTMs: Identify New Information to be Stored



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, X_t] + b_i\right)$$

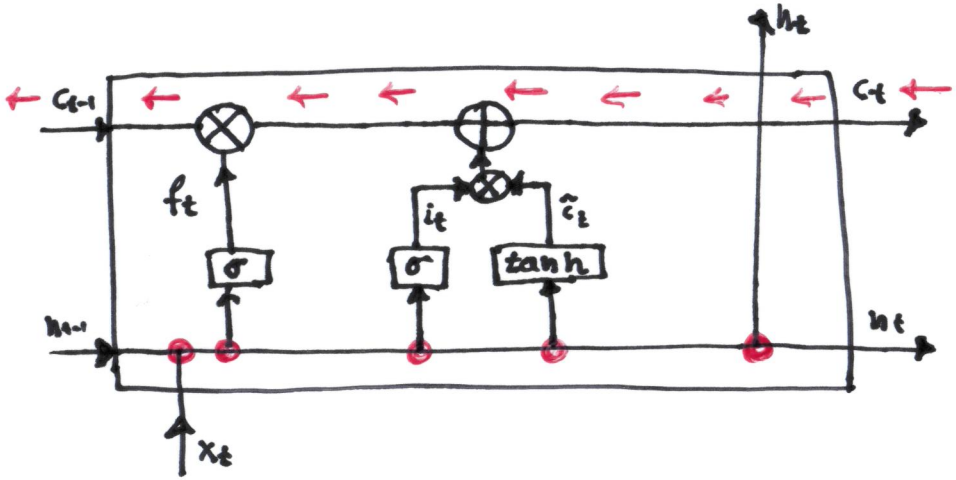
$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, X_t] + b_c\right)$$

① Sigmoid Layer decides which values to update.

② Tanh Layer generates a new vector of "candidate values" that could be added to the state.

For example, add gender of new subject to replace the old subject.

LSTMs: Update the Cell State



For example, drop all previous information and add new information about subject's gender.

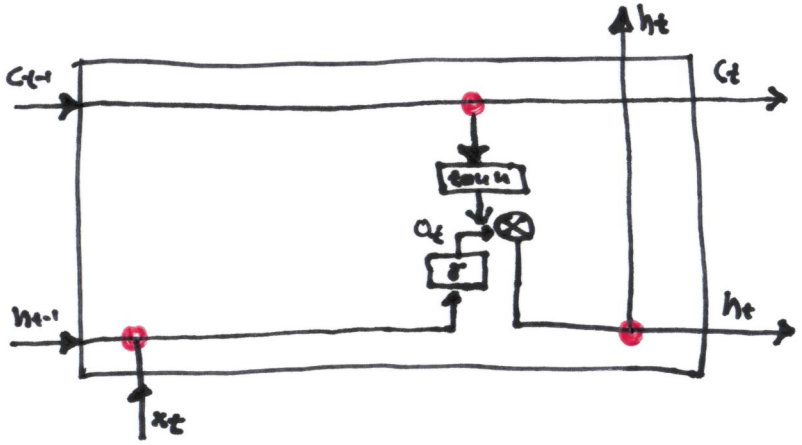
$$C_t = f_t \circledast C_{t-1} + i_t \circledast \tilde{C}_t$$

- Apply forget operation to previous internal cell state: $f_t \circledast C_{t-1}$
- Add new candidate values, scaled by how much we decide to update: $i_t \circledast \tilde{C}_t$

S.O.S:

- Backpropagation from C_t to C_{t-1} requires only elementwise multiplication!
- No matrix-multiplication is required → AVOID VANISHING GRADIENT PROBLEM

LSTMs : Output a Filtered Version of the Cell State



$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \circledast \tanh(C_t)$$

- Sigmoid Layer decides which parts of the cell state to output.
- Tanh Layer squases values within $[-1, +1]$ range.
- Output is a filtered version of the cell state:

$$o_t * \tanh(C_t)$$

⊛ For example, having seen the new subject, we may output information relating to a verb.