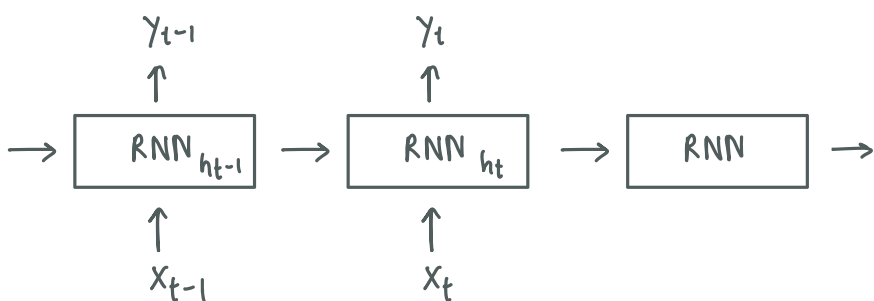
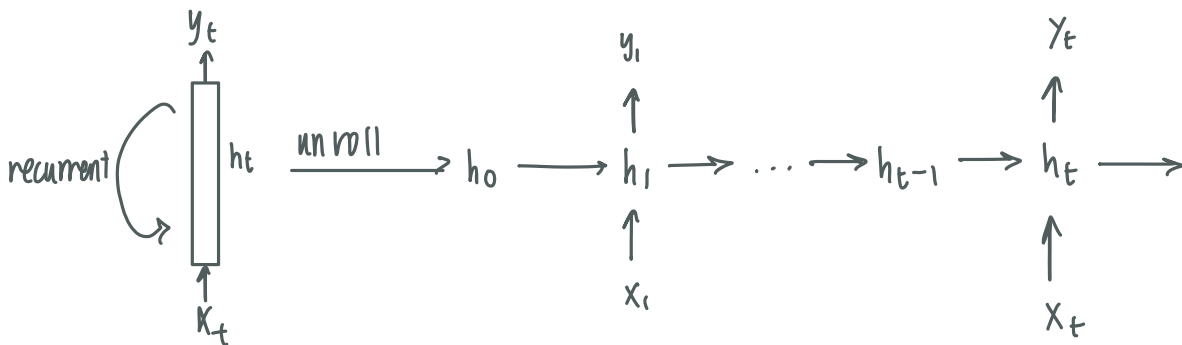


next HW assigned today for PyTorch.

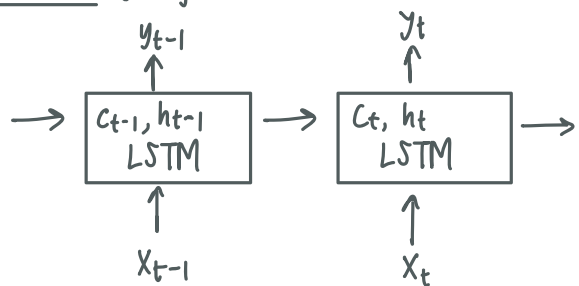
- code for CNN

- compute on Google Colab for GPU

RNN

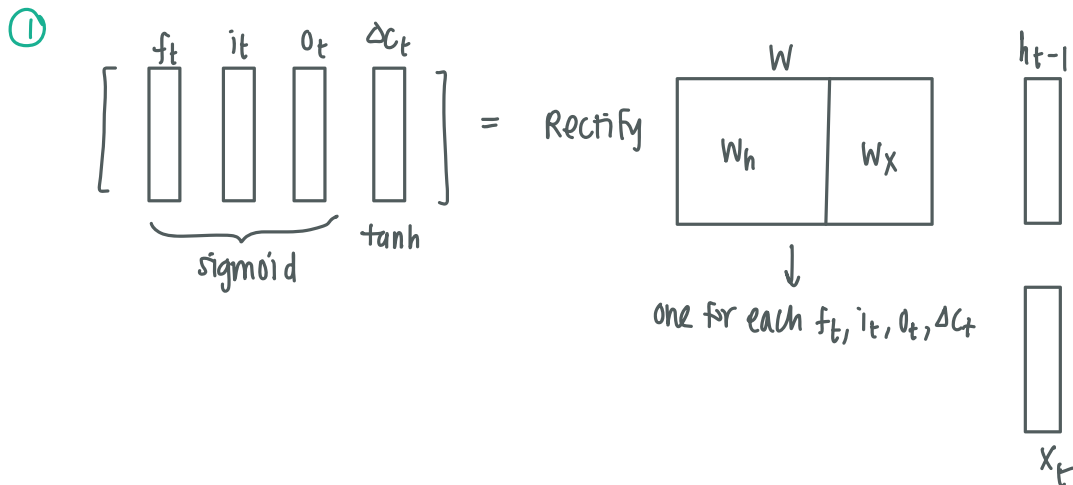
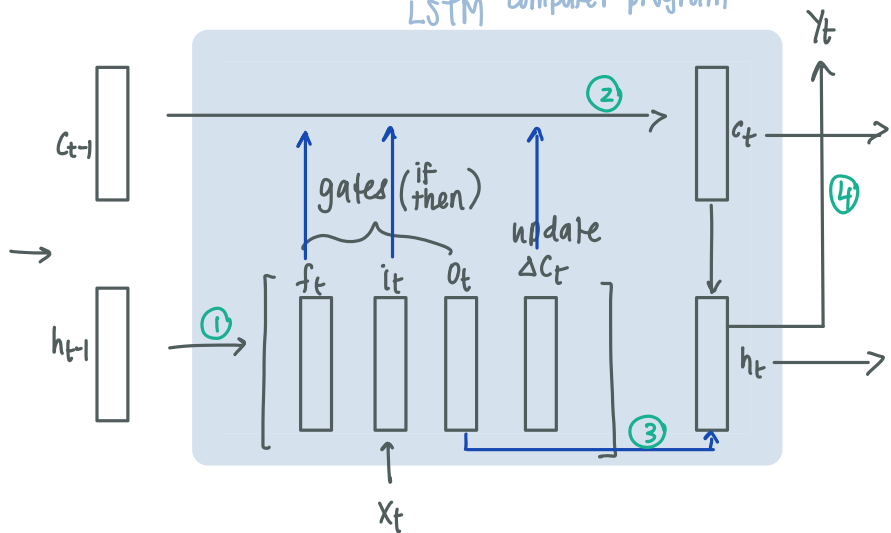


LSTM (long short term memory)



c stands for (memory) cells

LSTM computer program



②
$$c_t = f_t \odot c_{t-1} + i_t \odot \Delta c_t$$

③
$$h_t = \tanh \left(o_t \odot c_t \right)$$

④
$$y_t = \text{Rectify} \left(W h_t \right)$$

Gated Recurrent Unit (GRU)

①
$$\begin{bmatrix} r_t \\ z_t \end{bmatrix} = \text{sigmoid} \left(W \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \right)$$

\downarrow
 one for r_t
 one for z_t

②
$$\tilde{h}_t = \tanh \left(W \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \right)$$

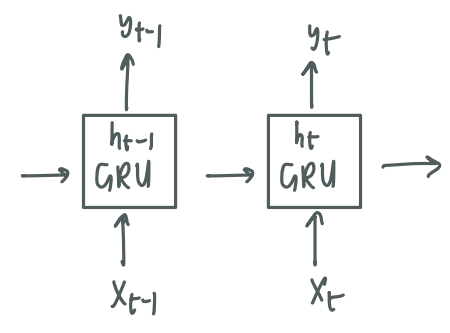
remember or not?

③
$$h_t = z_t \odot \tilde{h}_t + (1 - z_t) \odot h_{t-1}$$

\downarrow
replace or not?

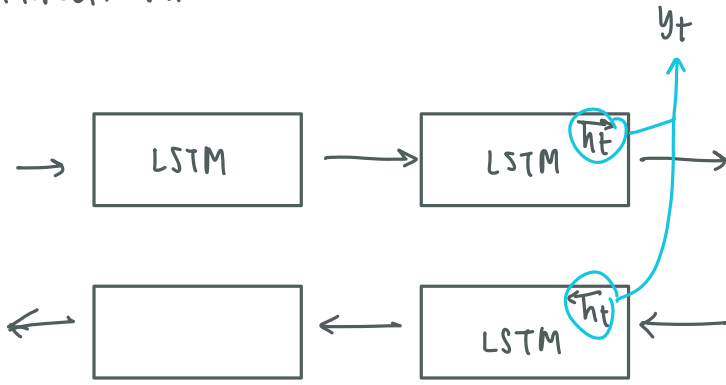
④
$$y_t = \text{Rectify} (W h_t)$$

overall global picture:



h_{t-1} controls initial generation of operations
 care most abt h_t

Bidirectional



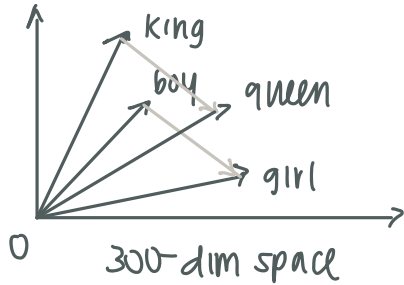
Natural Language Processing (NLP)

words \in dictionary 6000 words

i.e. king queen
boy girl

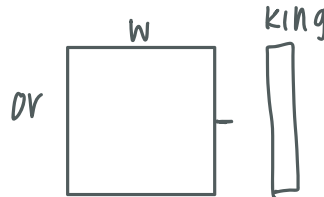
tokens \in 50,000

word \rightarrow vec / embedding

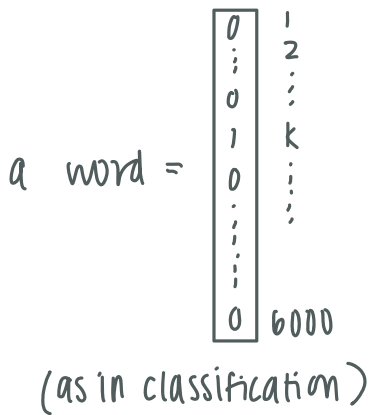


arithmetic

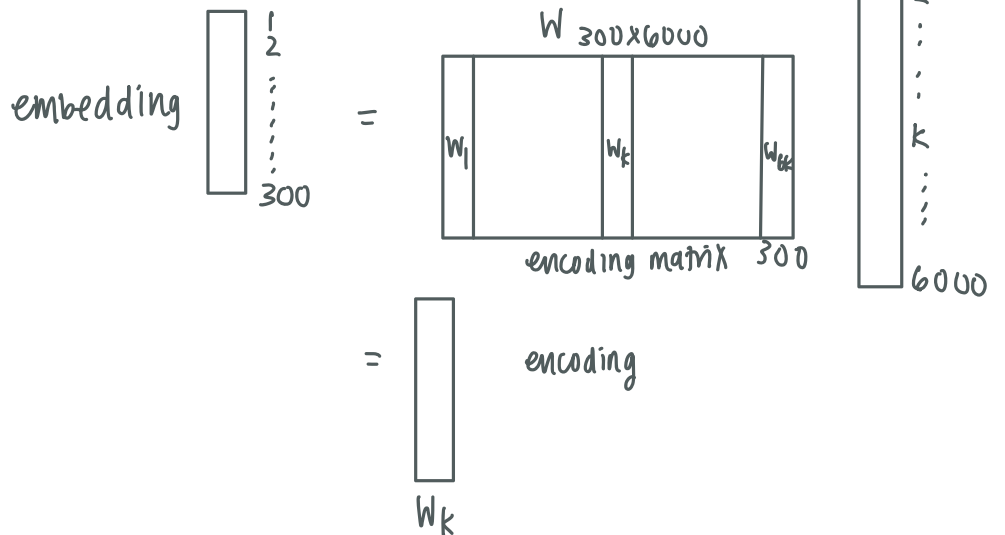
$$\text{queen} = \text{king} + \text{girl} - \text{boy}$$



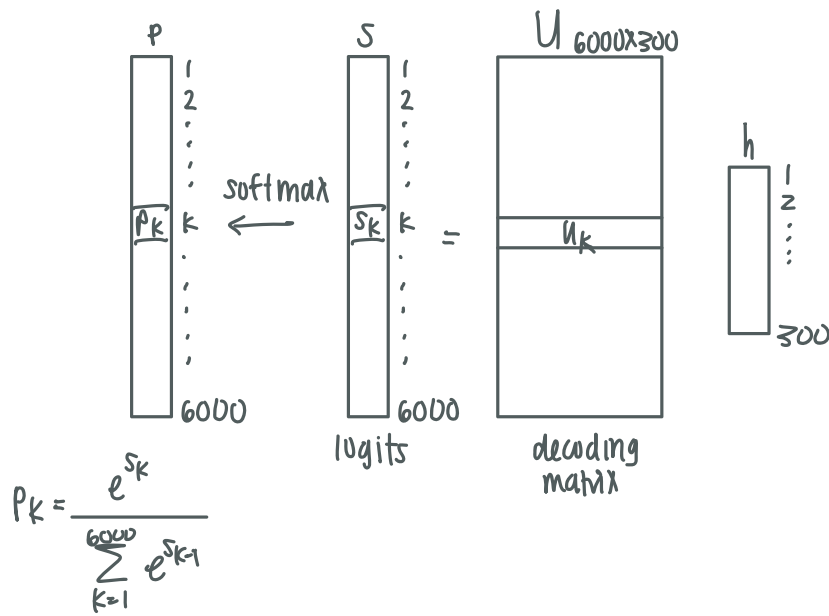
one-hot



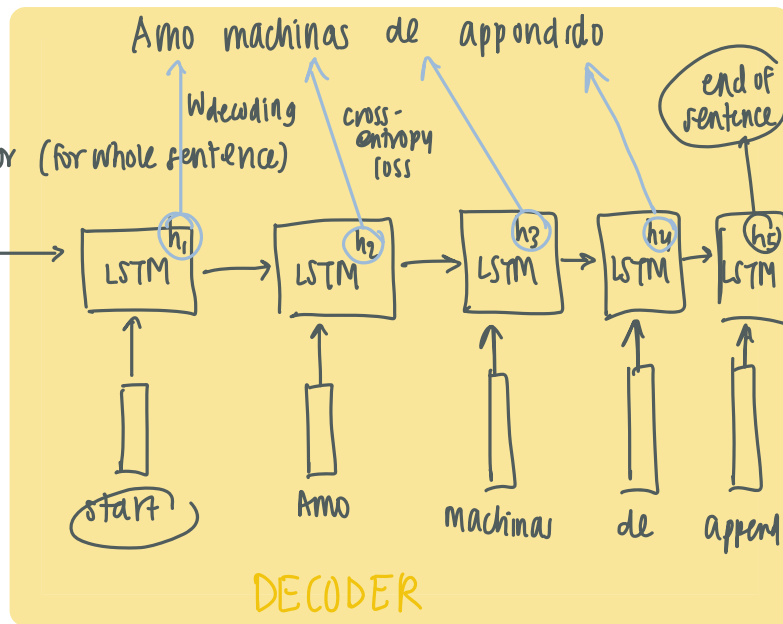
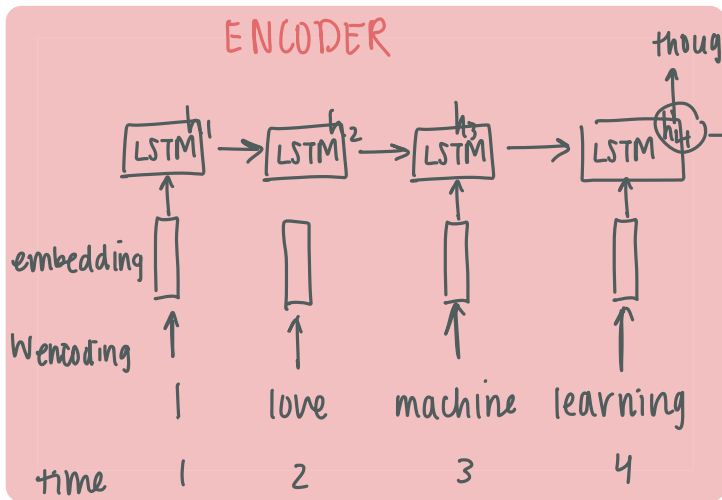
dense \leftarrow (any element can be nonzero)
distributed \leftarrow (distributed representation)



Decoding



Translation



Auto-regressive