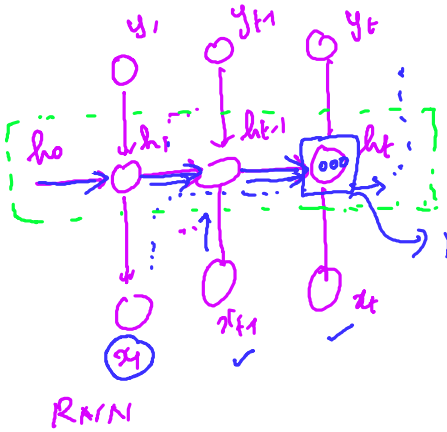
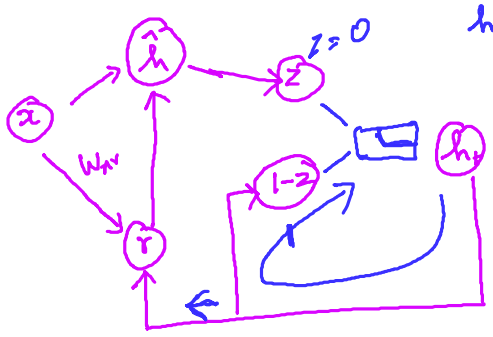
Topics

1. LSTM vs GRU
2. Encoders & Decoders – Seq2Seq modelling
   - English → Spanish → Machine Translation
   - Text → SQL Queries
   - Text → Programming language
3. Attention with Encoders & Decoders
4. Machine Translation problem
5. Text classification – LSTM – Glove Embeddings

LSTM → Long Short Term Memory



Ram ... He is ...



$h_t = f(z_t)$

$z \neq 0$

h_t — $f(\hat{h}) \xrightarrow{z} f(x, r)$

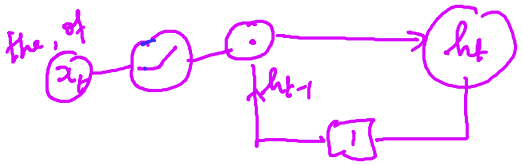
$1-z$ — $f(h_{t-1})$

$$r_t = \sigma(x_t W_{xr} + h_{t-1})$$

$$r_t = f[x_t, h_{t-1}]$$

$$\hat{h} = f(r_t, x_t)$$

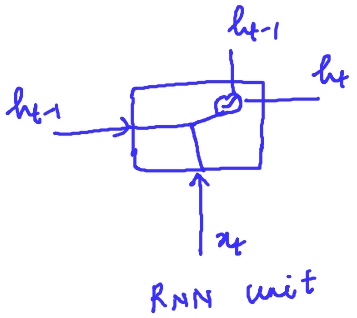
$$z_t = f(\hat{h}_t)$$



$$h_t = f(x_t, h_{t-1})$$

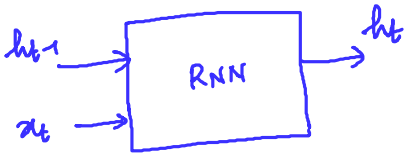
LSTM

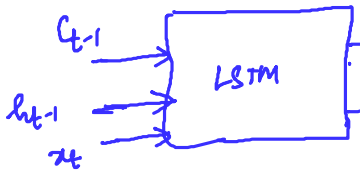
① forget gate
② Input gate/add gate
③ Output gate



RNN unit

$$h_t = \tanh(W_{xh} x_t + W_{hh} h_{t-1})$$





---

forget gate = $f_t = \sigma(W_f h_{t-1} + W_f x_t)$

$$K_t = C_{t-1} \odot f_t$$

Retained information

---

input/add gate $g_t = \tanh(W_3 h_{t-1} + W_4 x_t)$ RNN

$i_t = \sigma(W_i h_{t-1} + W_i x_t)$ add gate

$J_t = g_t \odot i_t$ information that needs to be updated

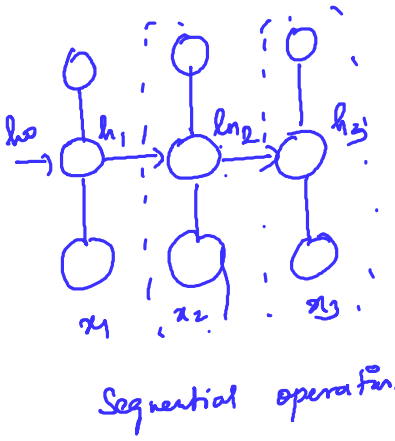
New information

$C_t = J_t + K_t$

New Cell state = Retained information + New information

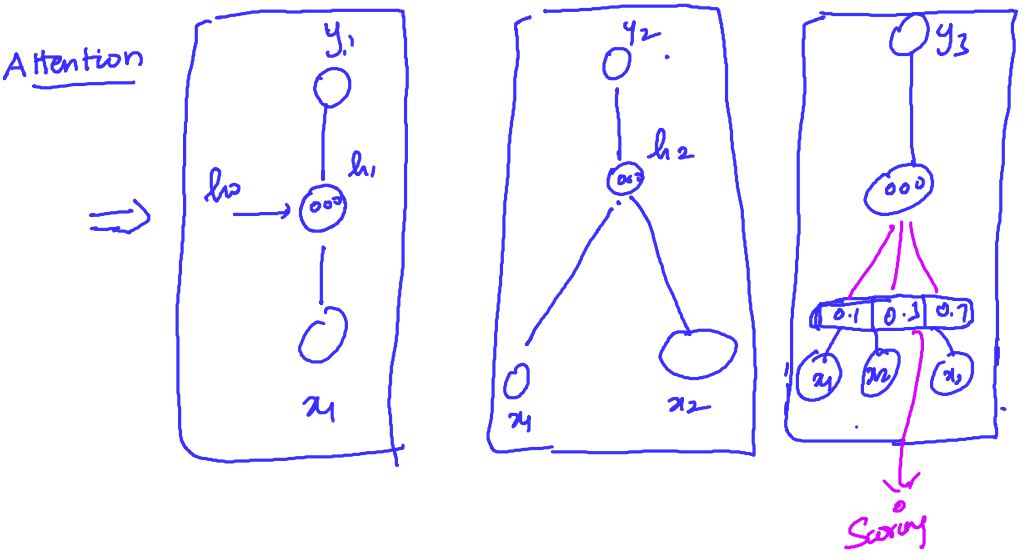
$O_t = \sigma(W_6 h_{t-1} + W_7 x_t)$ output gate

$h_t = O_t \odot \tanh(C_t)$



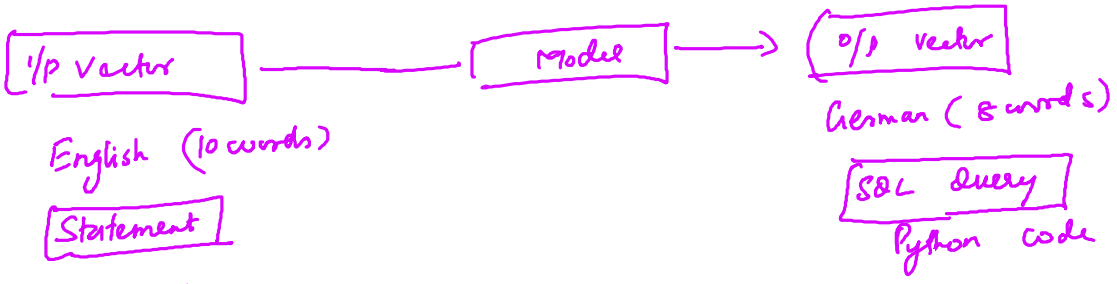
Sequential operation

Attention



Encoder - Decoder (seq 2 seq modelling)

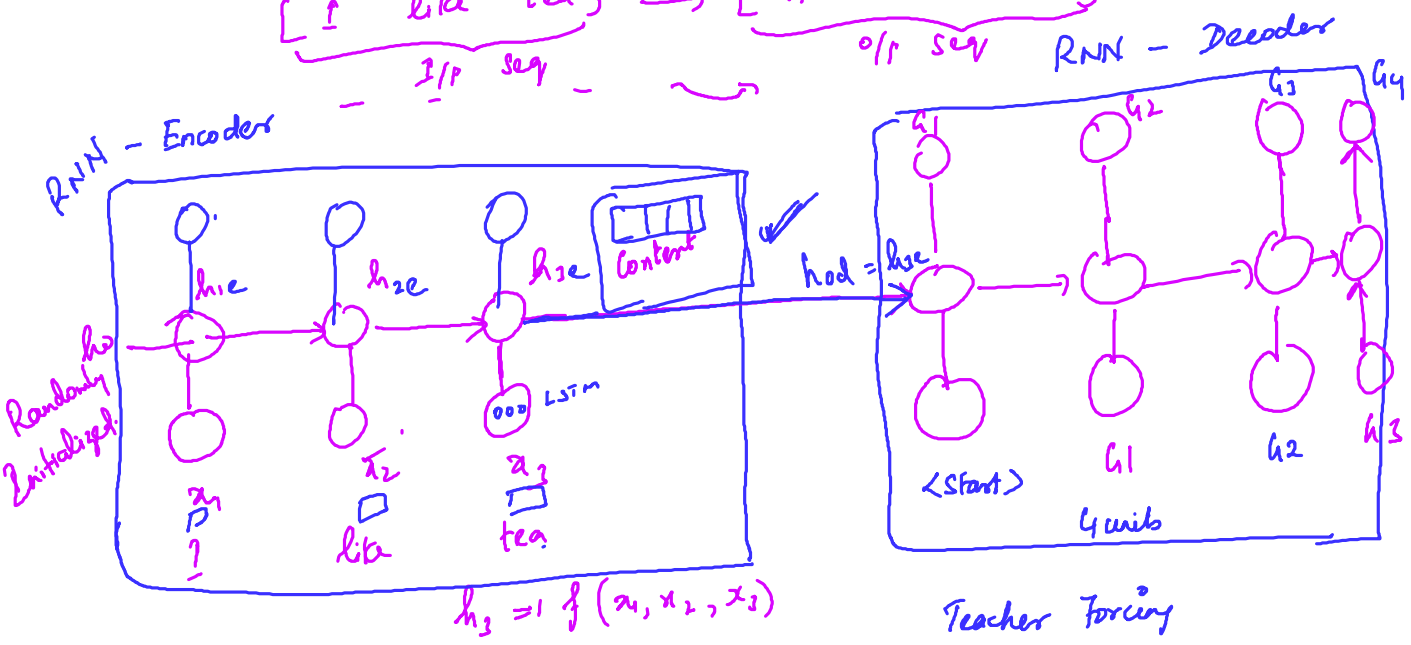
- Machine Translation
- Natural query → SQL
- Natural Language → Program



[I like tea] → [G1 G2 G3 G4]

I/p seq — O/p seq

RNN - Encoder — RNN - Decoder



Input language	Target language
I like tea	G_1 G_2 G_3 G_4
– – –	– – – –

} I/p data

"I like tea → M_1 M_2 M_3 M_4"

["I like tea"] ["M_1, M_2, M_3, M_4"]

Target I/p

Gated Recurrent Units

$$r_t = \sigma(U_r h_{t-1} + W_r x_t)$$

$$z_t = \sigma(U_z h_{t-1} + W_z x_t)$$

$$\bar{h} = \tanh(U(r_t \odot h_{t-1}) + W x_t)$$

$$h_t = (1 - z_t) h_{t-1} + z_t \bar{h}$$