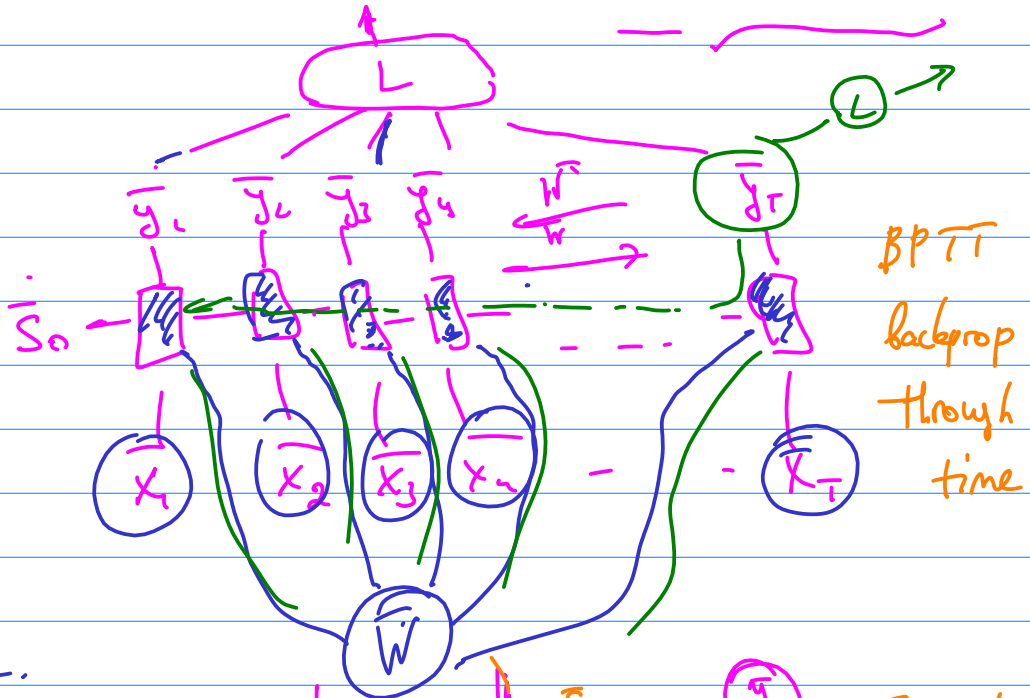
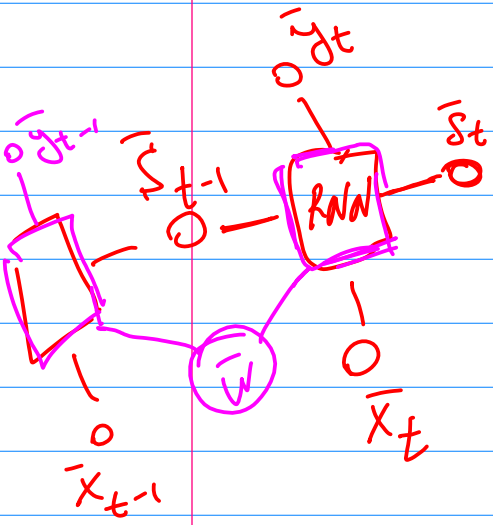
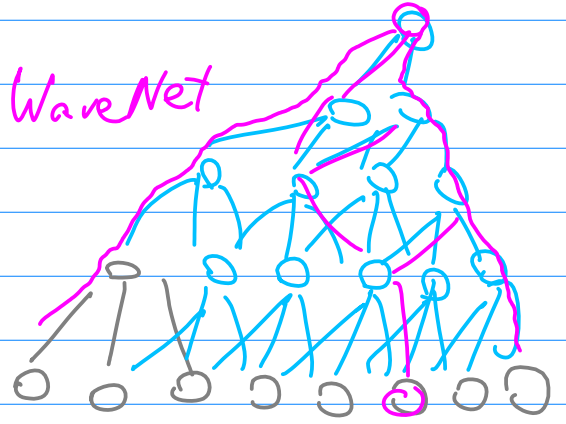
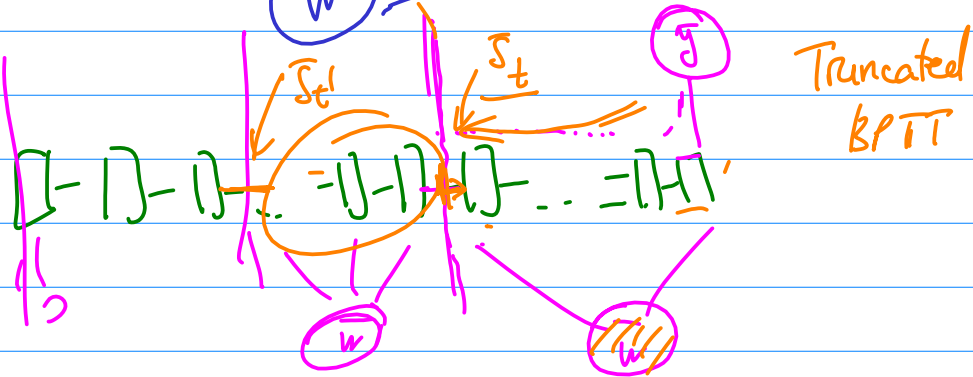
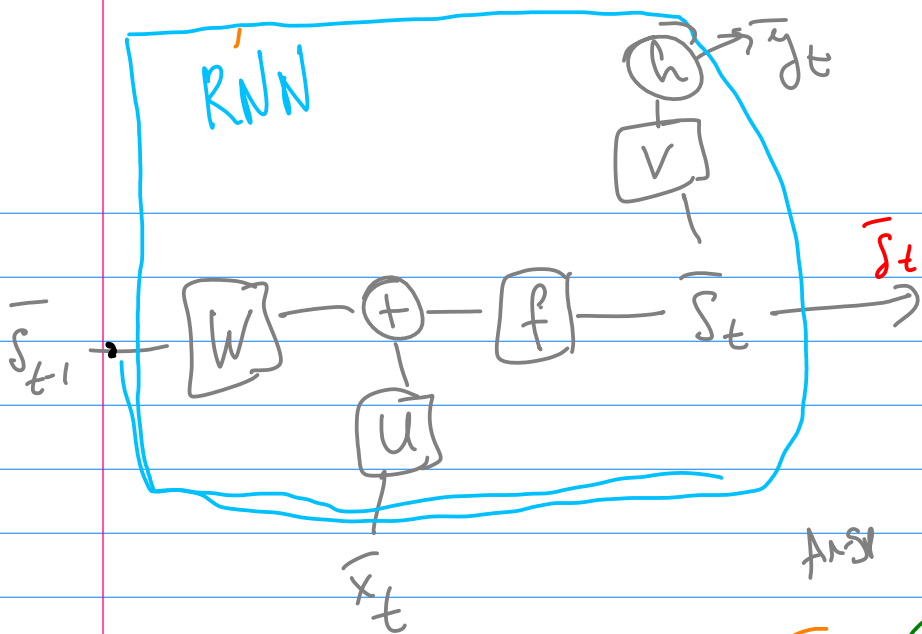


WaveNet



$$\frac{\partial L}{\partial w_i} = \sum \dots$$





$$\bar{s}_t = f(W\bar{s}_{t-1} + U\bar{x}_t)$$

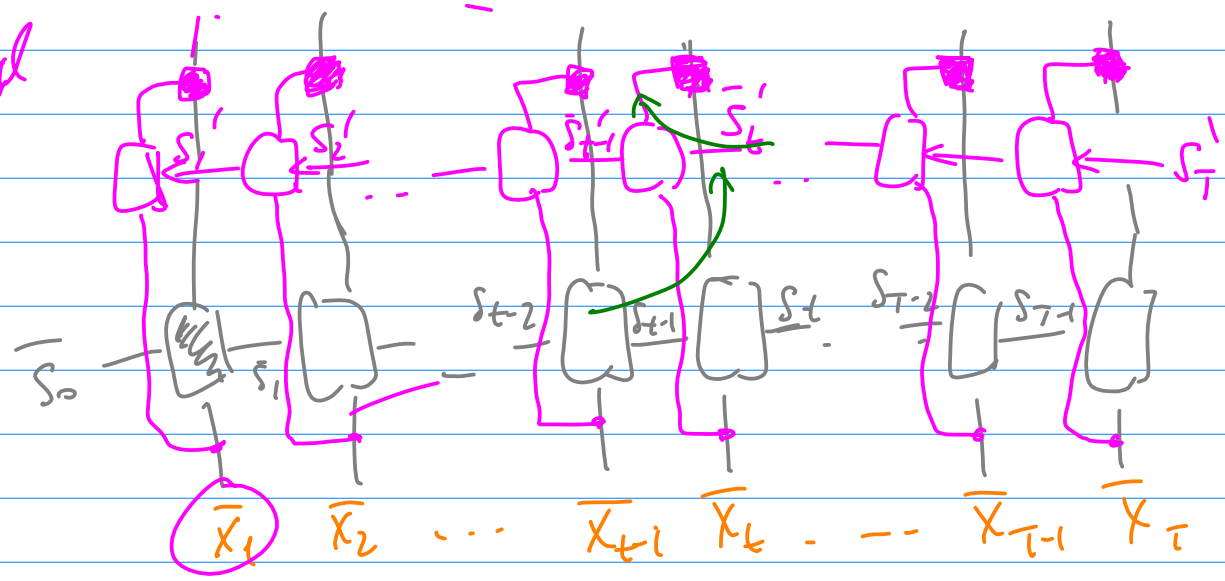
$$\bar{y}_t = h(V\bar{s}_t)$$

$W - d_s \times d_s$
 $U - d_s \times d_{in}$
 $V - d_{out} \times d_s$

ANSI Cap. Pycharm

$\bar{y}_1 \bar{y}_2 \dots \bar{y}_t \bar{y}_{t+1} \dots \bar{y}_{T-1} \bar{y}_T$

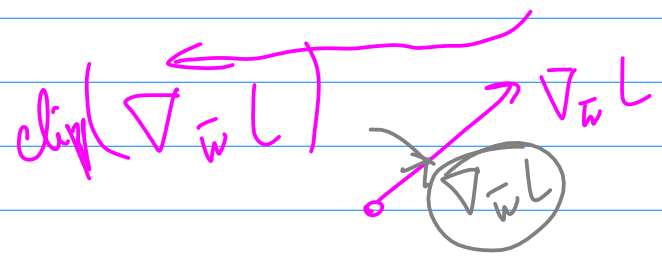
Bidirectional RNN



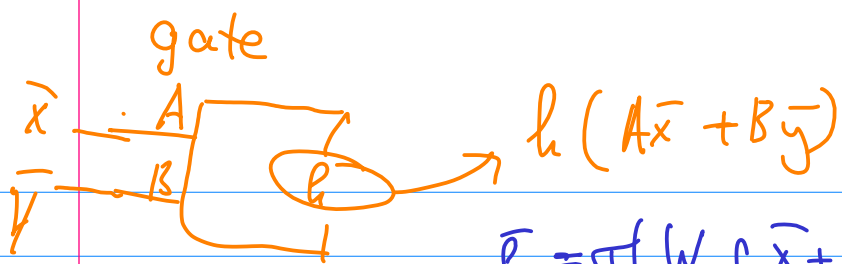
$$\bar{s}_t \approx W^{t-1} \bar{s}_0$$

$$\frac{\partial L}{\partial w} = 2^t$$

$\|W\| > 1 \Rightarrow$ exploding gradients
 $\|W\| < 1 \Rightarrow$ vanishing gradients

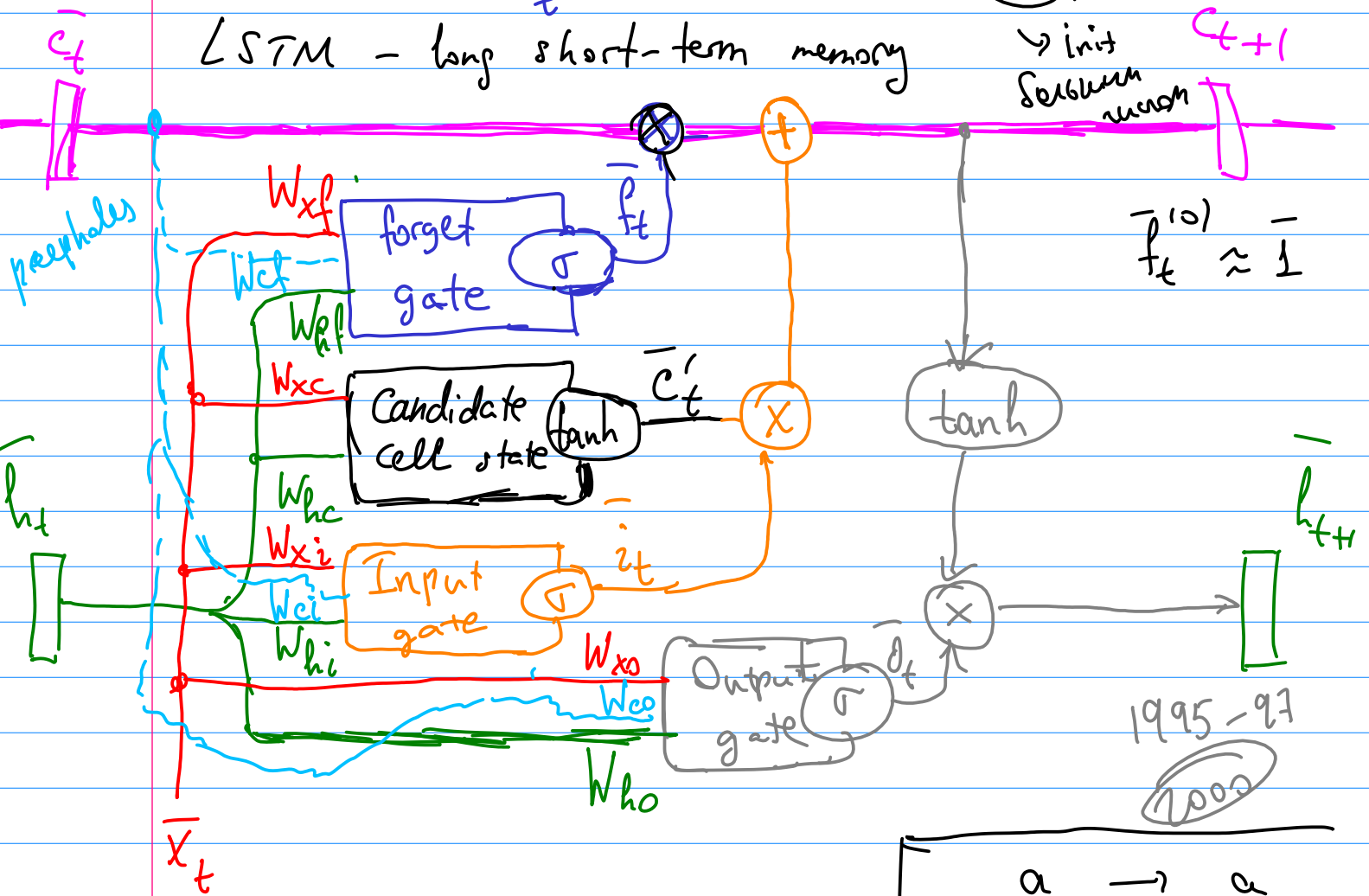


Memory networks



$$\bar{f}_t = \sigma(W_{xf} \bar{x} + W_{hf} \bar{h} + W_{of} \bar{c})$$

LSTM - long short-term memory



init
Serbuen
unoch

$$|\bar{f}_t^{(10)}| \approx 1$$

1995-97
2000

$$\bar{f}_t = \sigma(W_{df} \bar{x} + W_{hf} \bar{h} + W_{cf} \bar{c})$$

$$\bar{c}'_t = \tanh(W_{xc} \bar{x} + W_{hc} \bar{h})$$

$$\bar{i}_t = \sigma(W_{xi} \bar{x} + W_{hi} \bar{h} + W_{ci} \bar{c})$$

$$\bar{o}_t = \sigma(W_{xo} \bar{x} + W_{ho} \bar{h} + W_{co} \bar{c})$$

$d_x \times d_{in}$ $d_h \times d_h$ $d_x \times d_h$

a → a
ab → bc
LSTM → ntlc
memory → gromem

Seq2Seq LSTM

