Solution 15_ rnn

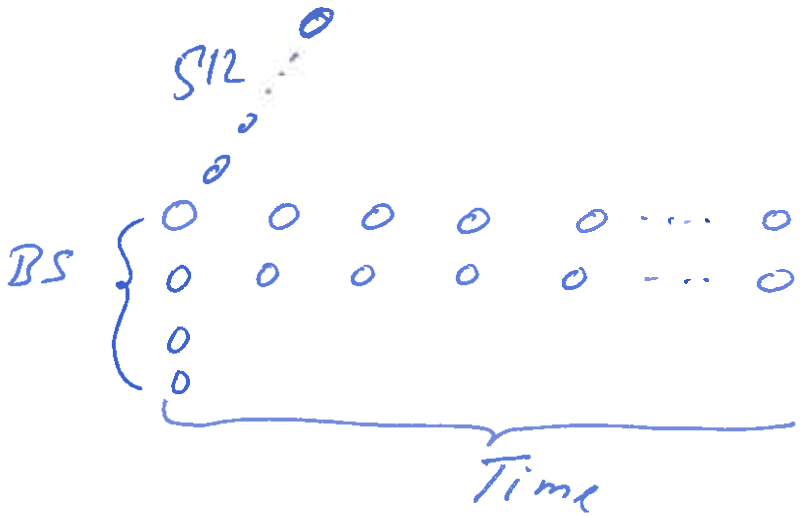
a) $[\underbrace{h_{t-1}, x_{-t}}_{512+106}] \cdot \{\overbrace{(W_f)}^{512} + \underbrace{(\cdot b_f)}_{512}$

⇒ Parameters for the forget gate

$(512 + 106) \cdot 512 + 512 = 316'928$

Times four for whole layer $1'267'712$

last layer



Foreach timepoint and batch, we transform a 512 dimensional hidden vector to a 106 dimensional pros. distribution.

Hence we have $512 \cdot 106 + 106 = 54'378$

↑ bias