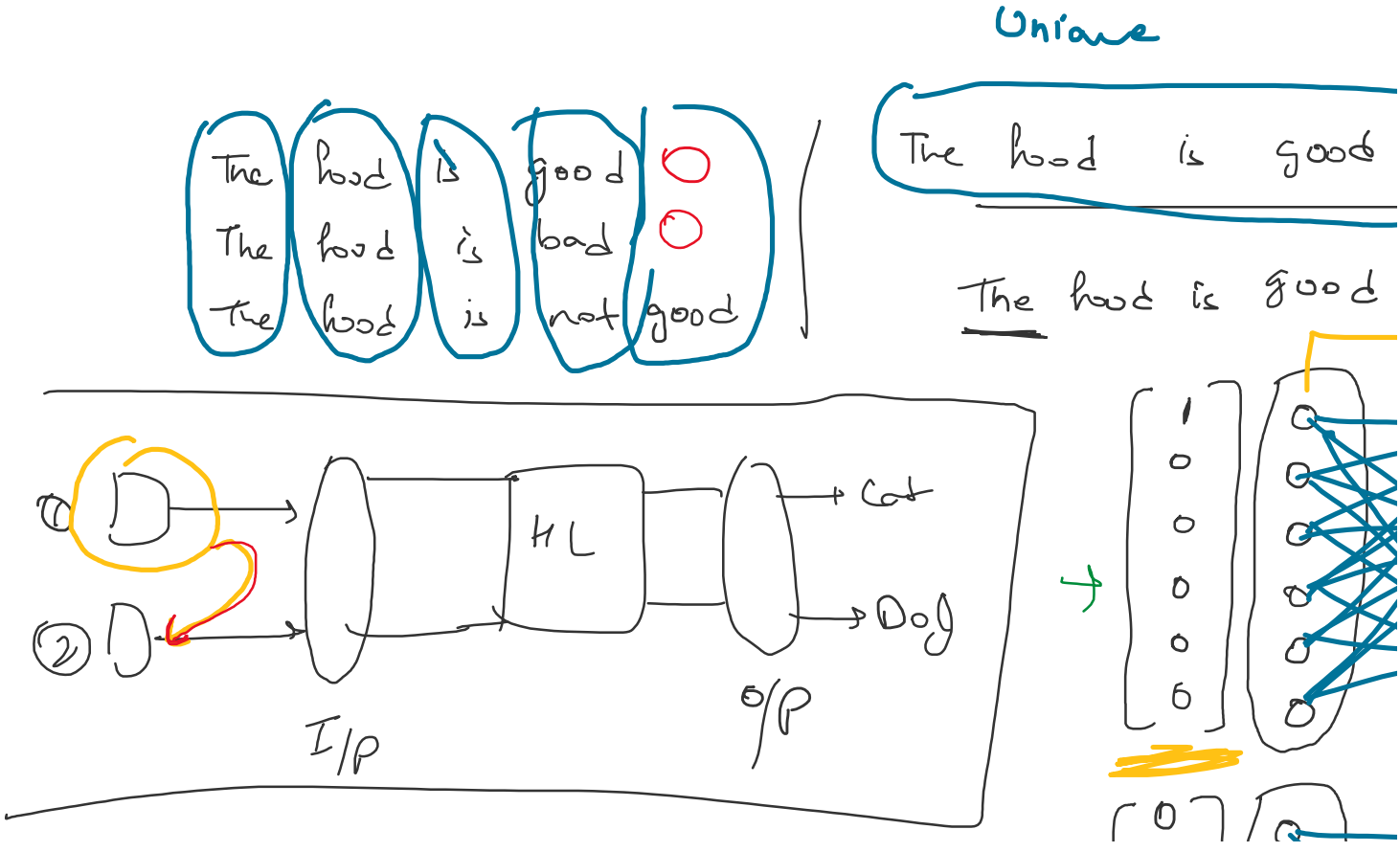
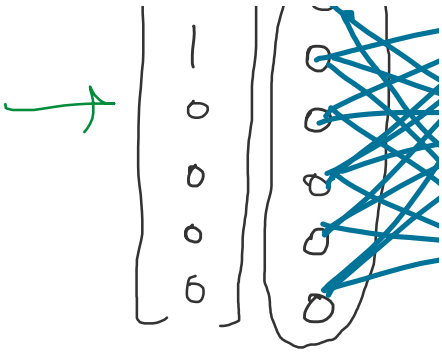
RNN

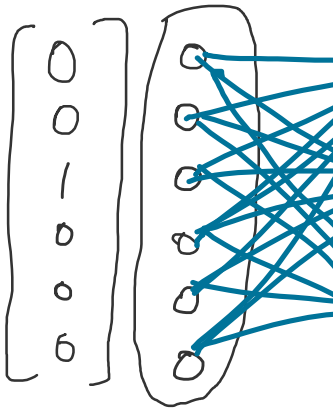
09 July 2025 14:37

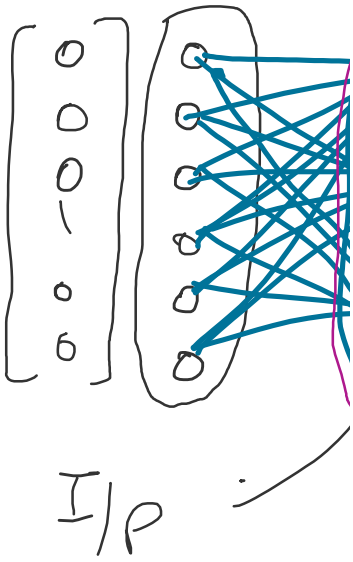
Recurrent Neural Network

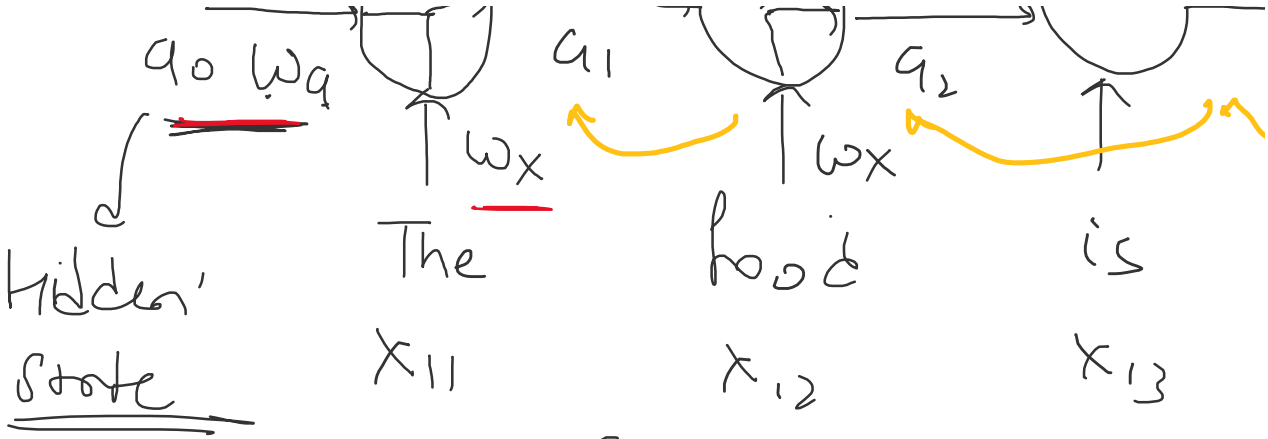
1. RNN → LSTM / GRU → Bidirectional RNN → Encoder Decoder
2. Self-Attention
3. Transformers
4. LLMs . . . . .








I/p



Forward

$$a_1 = f(n_{11} w_x + c$$

$$a_2 = f(n_{12} w_x + c$$

$$a_3 = f(n_{13} w_x + c$$

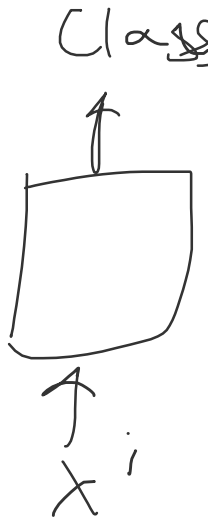
$$\vdots$$

$$a_t = f(n_{1t} w_x + a$$

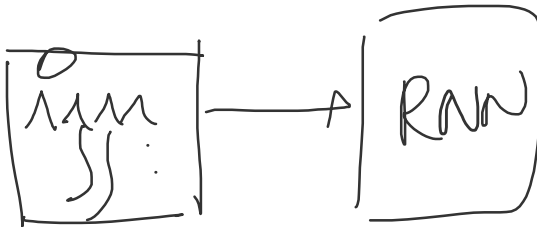
Context vector of entire Sentence

Different Applications:

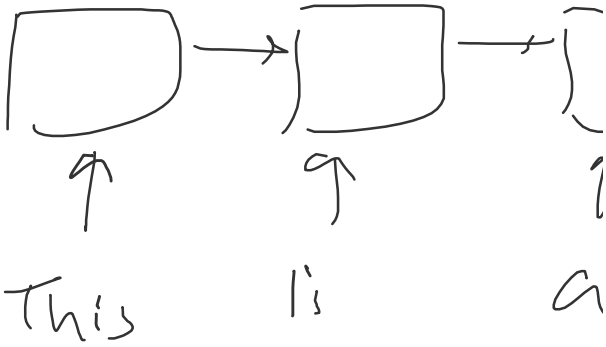
(1) One to one : FCN



(2) One to many :



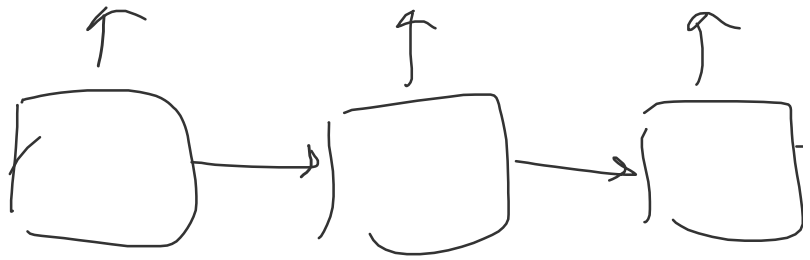
(3) many to one :



(4) Many to many : Named Entity

Amitabh is a

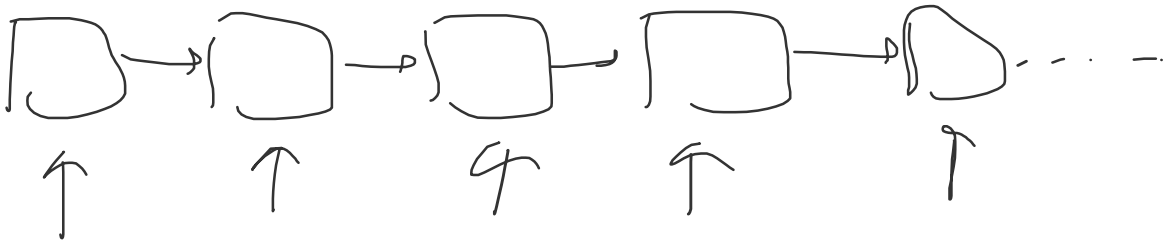
1 0 0



(5) Encoder - Decoder

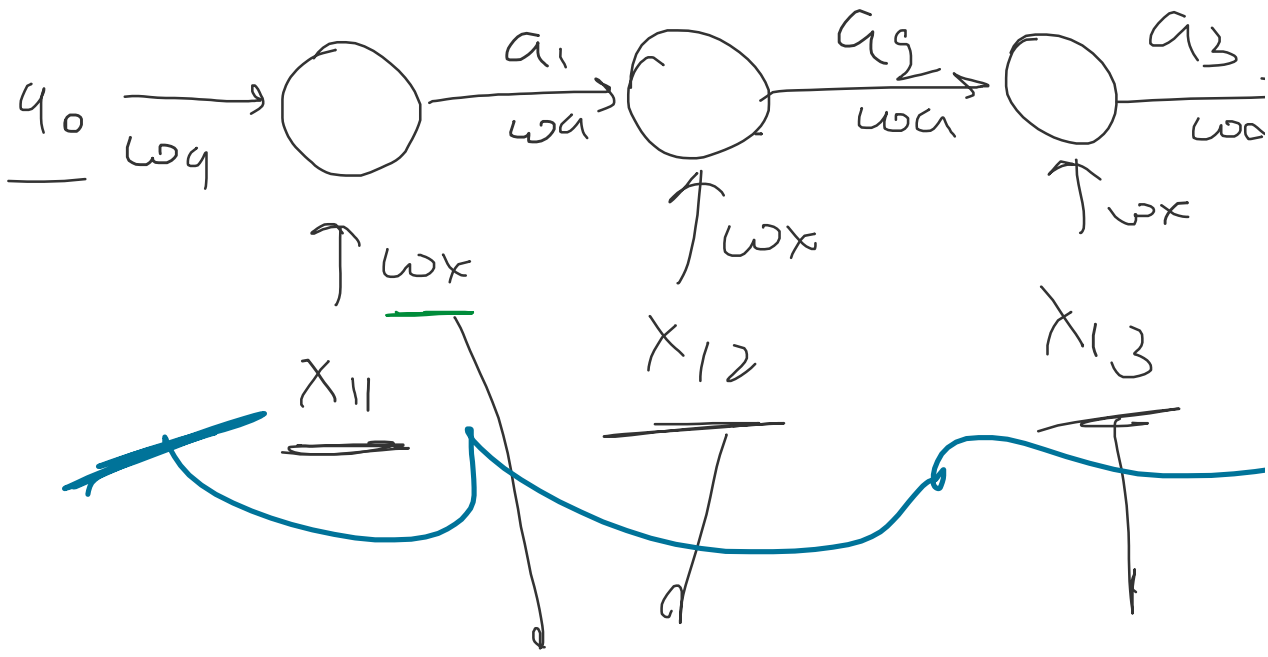
Speech to text

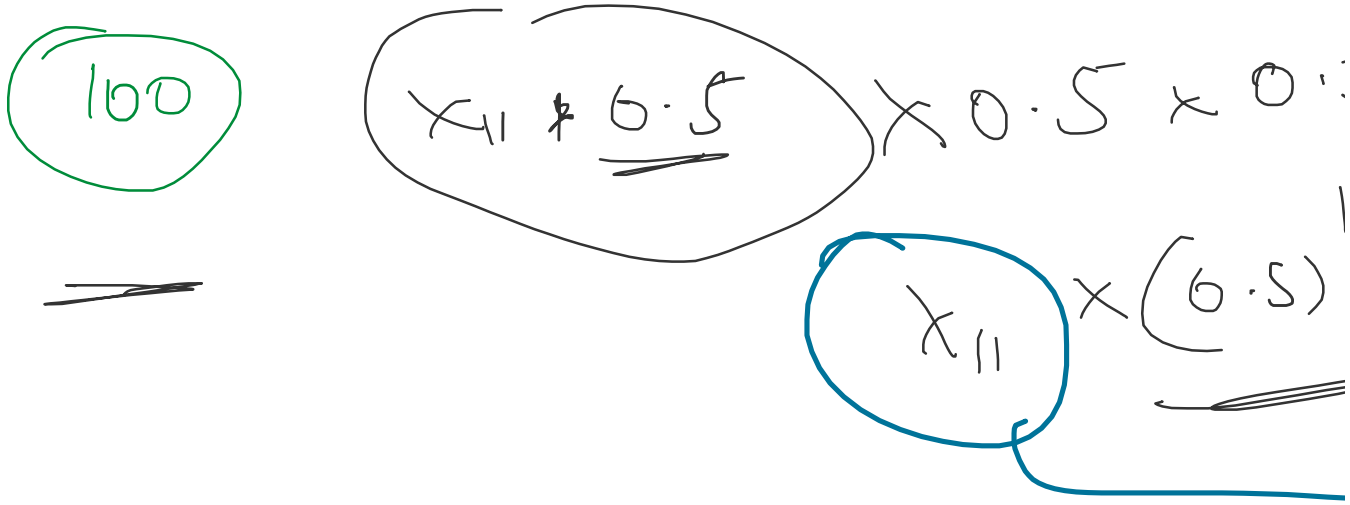
Translation



Word Embedding → Word2

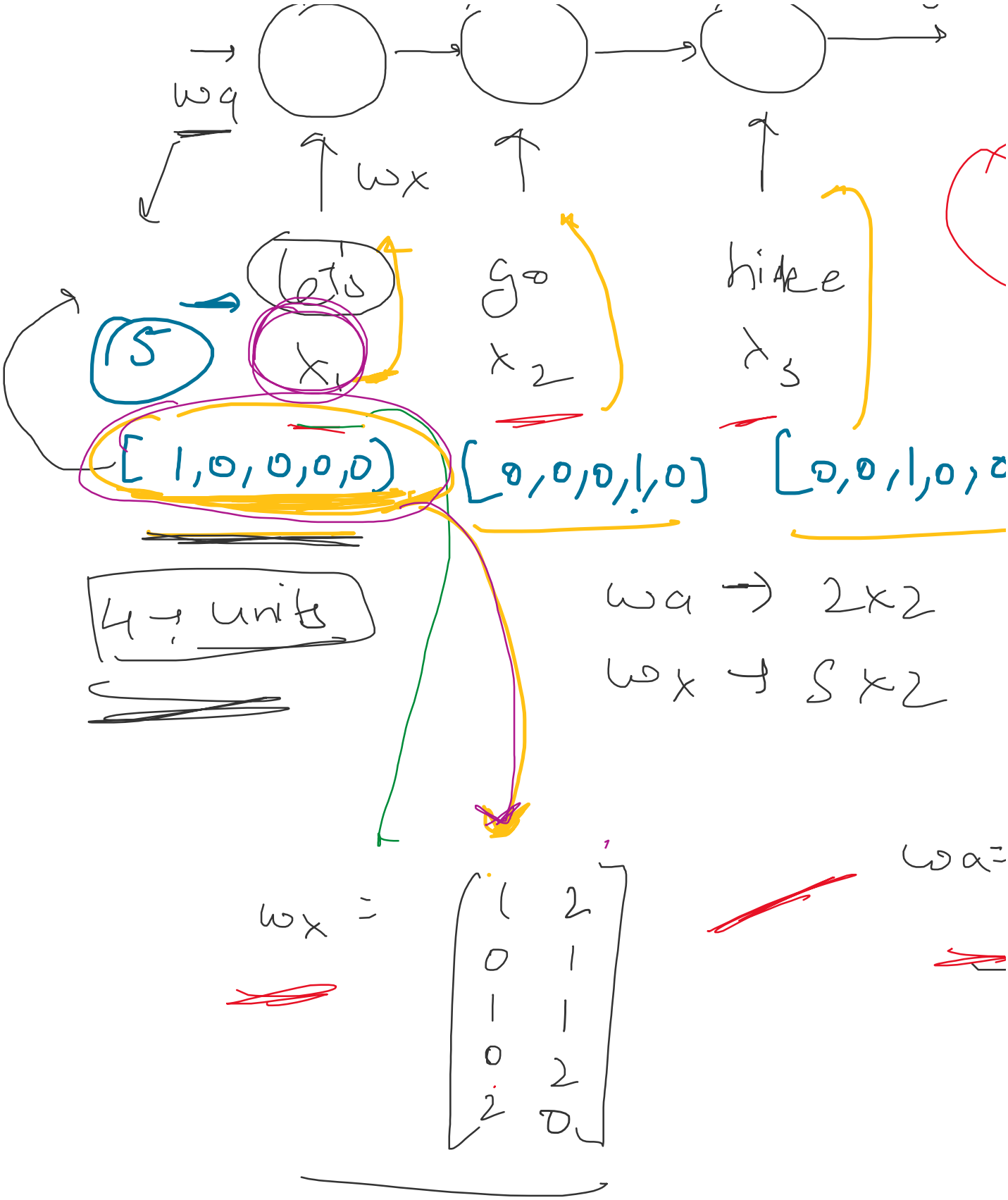
Disadvantages





(1) long term dependency

(2) Vanishing gradient


wq
wx
lets
go
hide
S
x1
x2
x3
[1,0,0,0,0]
[0,0,0,1,0]
[0,0,1,0,0
4 units
wq → 2x2
wx → 5x2
wx =
1 2
0 1
1 1
0 2
2 0
wq=

RNN

1. Vanishing gradient proble
   Exploding gradient Proble

2. don't remember Lo

↓

LSTM Lor