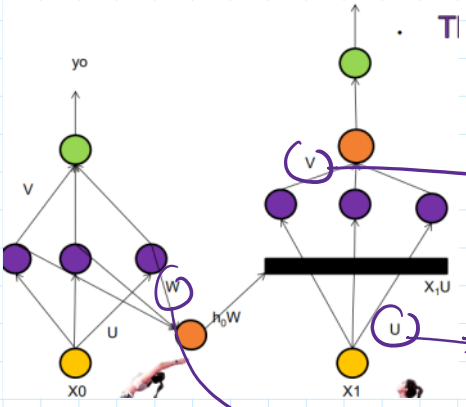
RNN Speedrun

24 September 2024 06:39



Input: n-dimensions, hidden layer: d neurons, Output: k classes.

Weight $V = d \times k$ (d: neurons in layer; k: no. of class labels) } VDK (vashaswini DK) [no. of neurons × no. of klasses]

Weight $U = n \times d$ (n: dimension of input-x; d: no. of neurons in layer) } Undries [dimension of x × no. of neurons]

Weight $W = d \times d$ (no. of neurons) } Wednesday for weights [no. of neurons × no. of neurons]

Eg: input: $n \to n$
hidden layer: $n/2$ neurons. $\to d$
output layer: $2 \to k$
total no. of weights: 161

$V + U + W = 161$

$(d \times k) + (n \times d) + (d \times d) = 161$

$\frac{n}{2}(2) + (n)\left(\frac{n}{2}\right) + \left(\frac{n}{2}\right)\left(\frac{n}{2}\right) = 161$

$\frac{2n}{2} + \frac{n^2}{2} + \frac{n^2}{4} = 161$

$\frac{4n}{4} + \frac{3n^2}{4} = 161$

$3n^2 - 4(161) + 4n = 0 \longrightarrow$ solve.

Problems RNN faces

① Vanishing grad
gradients of loss fn become very small
→ can't learn long term dependencies
→ cuz of repeated mult of small gradient values

② Exploding grad
gradients of loss fn become v. big.
→ instability during training
→ repeated mul of large grad.

LSTM over 1:1 vanilla RNN
→ no vanish/explode → use gated recurrent unit
→ memory for long term memory req.
→ var. length input + output.

LSTM — 3 gates
GRU — 2 gates.

cell state changes slow (memory)
hidden changes fast. (current state)

LSTM 3gate 2states more params

LSTM	3 gate	2 states	more param
GRU	2 gate	1 state	↓ param

No. of params for LSTM:

lstm layer = 100 units (m = no. of units in LSTM layer)

embedding vector length = 32. (n = embedding vector len)

LSTM has 4 functional units — 3 sigmoid, 1 tanh.

① For one function — $n+m+1$ params

② for 4 functions — $4(n+m+1)$

→ m units: $4m(n+m+1)$

↓

for GRU, 3

$3m(n+m+1)$ (m: no. of units)

(n: embedding length)

For simple RNN:

$nm+m^2+1$