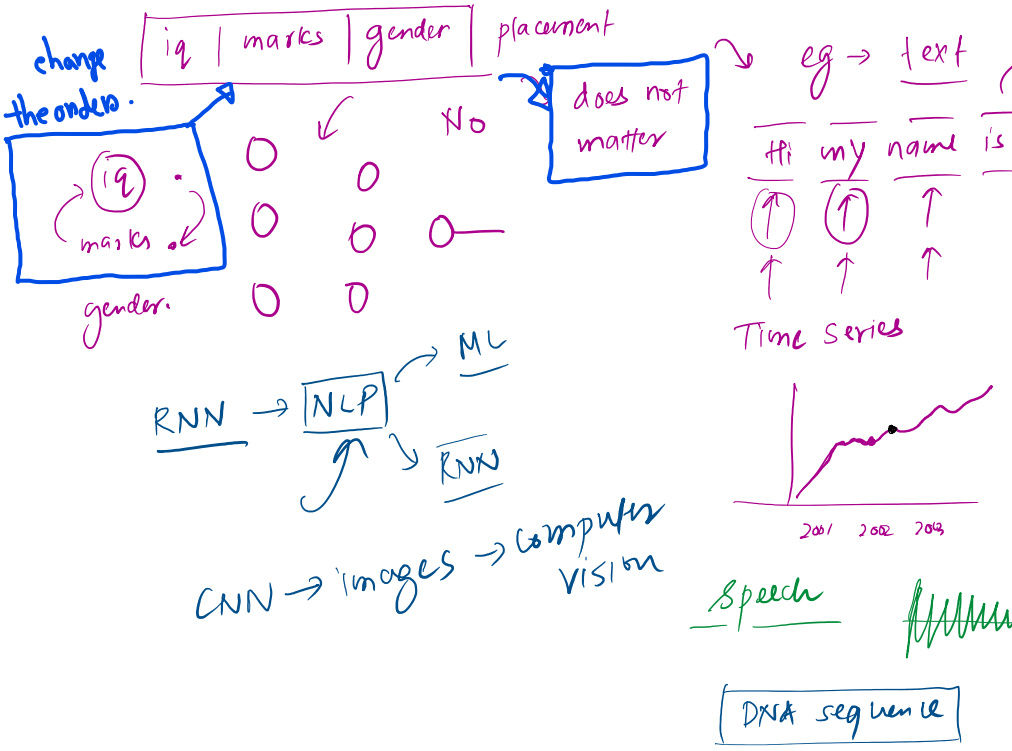
Lecture: 55

Sequential Data
22 October 2022 13:09

ANN → tabular data
CNN → images

RNN → Recurrent NN
↳ is type of sequential model to work on sequential data



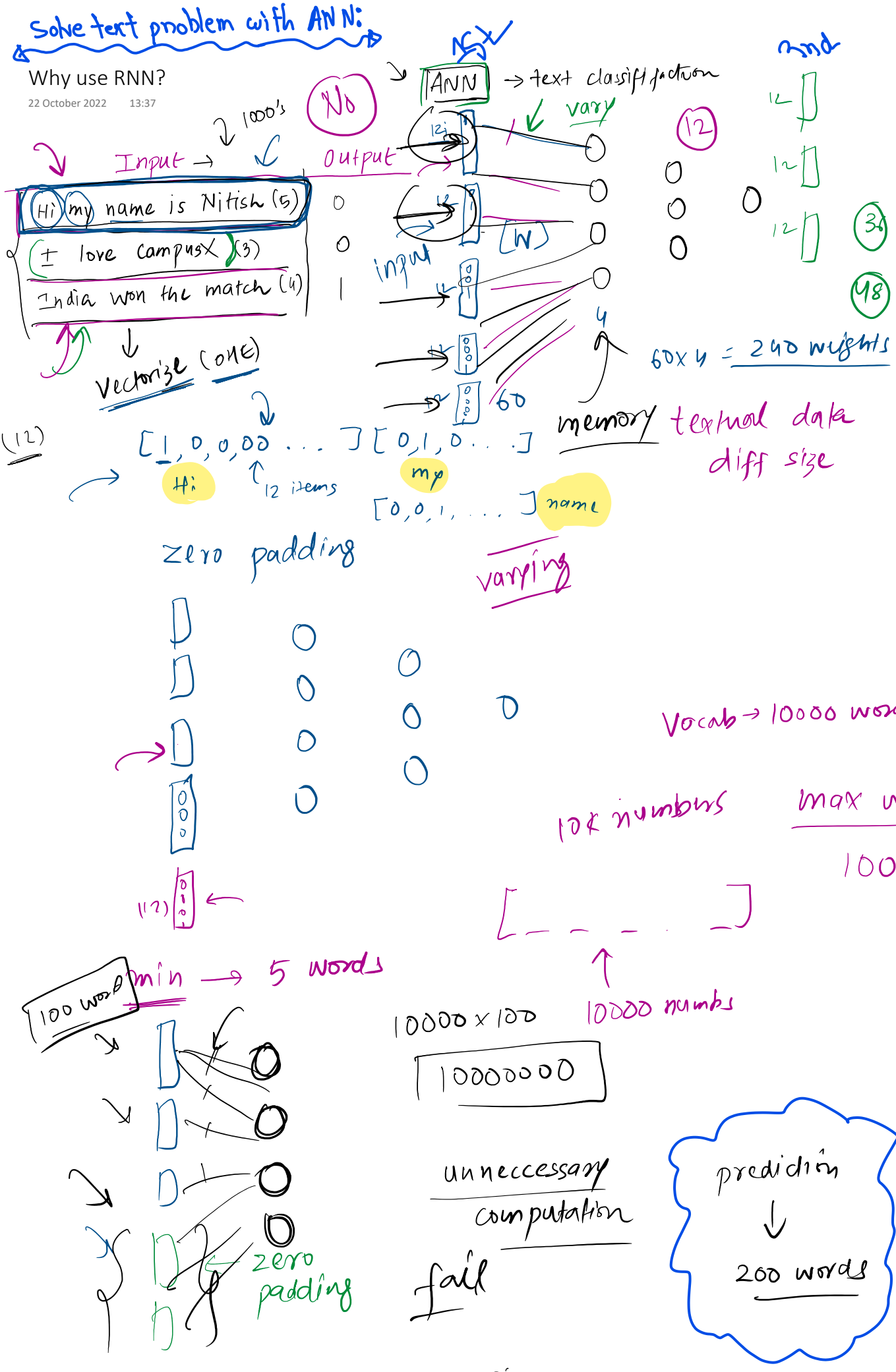
- ANN tabular data এর উপর কাজ করে।
- CNN image data এর উপর কাজ করে।
- RNN sequential data এর উপর কাজ করে।

sequential data:→ text ("my name is yasin") sequence matter করে। না হলে, sentence এর meaning change হয়ে যায়। time series data, speech, DNA sequence etc.

Solve text problem with ANN:

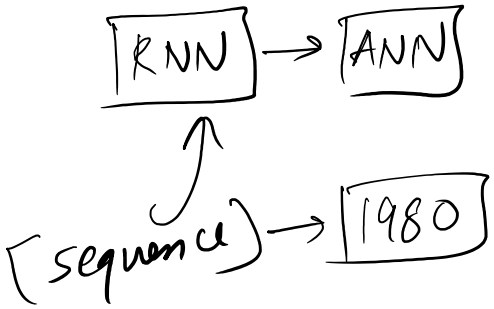
Why use RNN?

22 October 2022 13:37



✱ Sequential data ANN দিয়ে solve করলে, আমরা এই 4 টা problem face করি।

1) text input → varying size
2) zero padding → unnecessary computation
3) Prediction problem
4) Totally disregarding the sequence info



i) Model 200 টা character এর জন্য 200 টা neuron ব্যবহার করলাম কিন্তু, testing এর সময় 200 টার বেশি character এর input আসতে পারে।

ii) model এ 200 টা char এর জন্য 200 neuron রাখলে, যদি input এ 5 টা char এর input আসে তাহলে বাকি 195 টা char zero দিয়ে padding হয়। unnecessary computation.

iii) ANN এ সব data একসাথে যায়তে ফলে sequence maintain করে না। তাই, কোন text এর actual meaning হারিয়ে ফেলবে ANN।

RNN Applications

22 October 2022 13:37

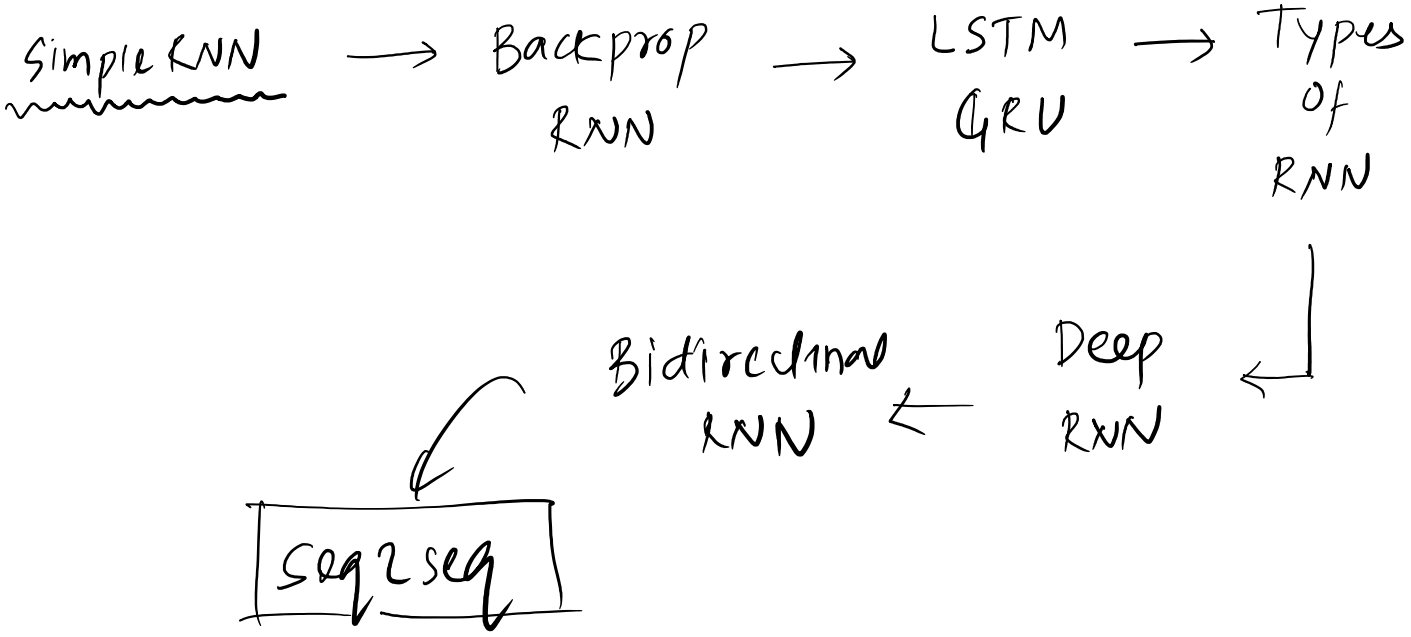
1. Natural Language Processing (NLP)
2. Time Series Prediction.
3. Speech and Audio Processing
4. Image and Video Processing
5. Healthcare
6. Robotics
7. Recommended Systems
8. Finance
9. Gaming
10. Others

Handwriting Recognition: Recognizing and digitizing handwritten text.
Autonomous Driving: Processing sequential sensor data to make driving decisions.

Roadmap

22 October 2022 13:37

(For Learning RNN)



Lecture:56 (Recap 55 & কি কি পড়লাম)

Why RNNs?

29 October 2022 13:30

1) sequence → any length → ANN → fixed input size

zero padding → cost of computation ↑

movie reviews

2) sequence contains some meaning

same

RNNs

NNs

memory feature

ANN → input

RNN architecture

RNN forward prop → prediction

↳ input → output

codes → solidify

Data for RNN

29 October 2022 13:30

[(timesteps, input_features)]

t=1, t=2, t=3

Review	Sentiment
movie was good	1
movie was bad	0
movie was not good	0

(4)

review 1

[[10000], [01000], [00100]]

vocab → 5 word (unique)

movie, was, good

[10000] [01000] [00100]

[00010] [00001]

bad, not

(3,5) ↑ no of timesteps → # of features

(3,5) (4,5)

Keras → SimpleRNN → batch size, timesteps, input_features

(3, 4, 5) → 3D tensor → RNN

↓

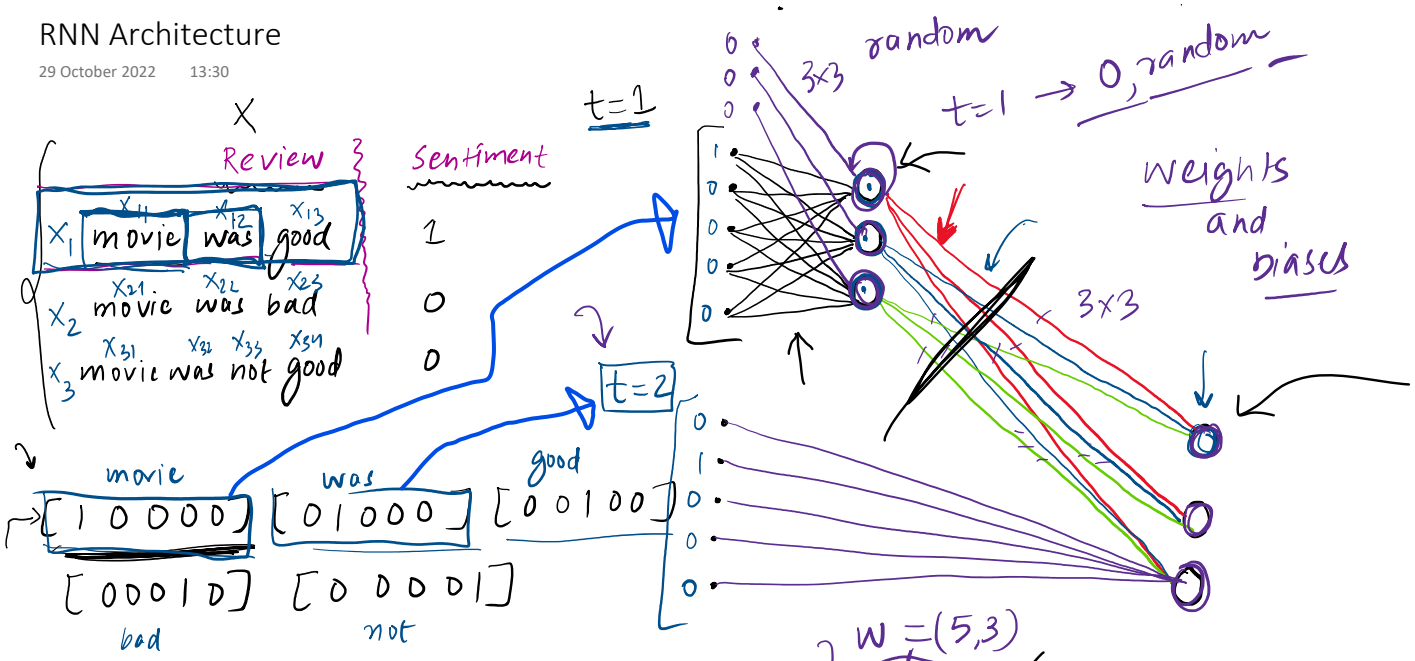
উপরের sentence গুলোর মধ্যে 3 নং টার timestep সবচেয়ে বেশি।

time step: Number of word in a sentence (train data set এর মধ্যে সবচেয়ে longest sentence এর word সংখ্যা খুঁজে বের করতে হবে।)

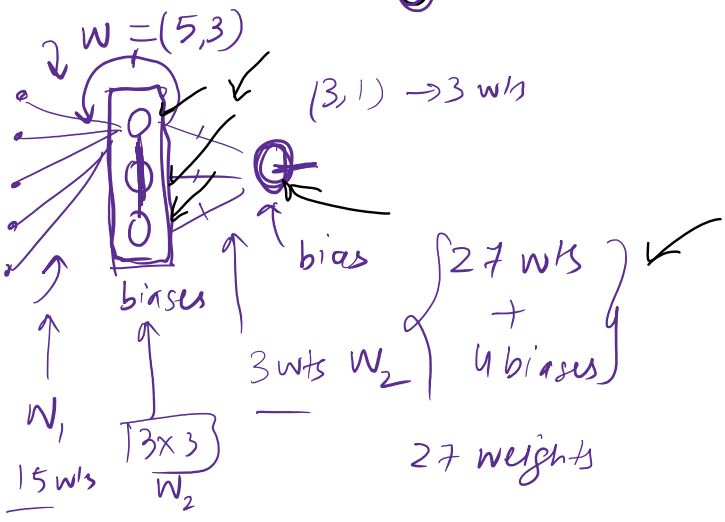
input feature: Total number of unique word in the train data set.

RNN Architecture

29 October 2022 13:30



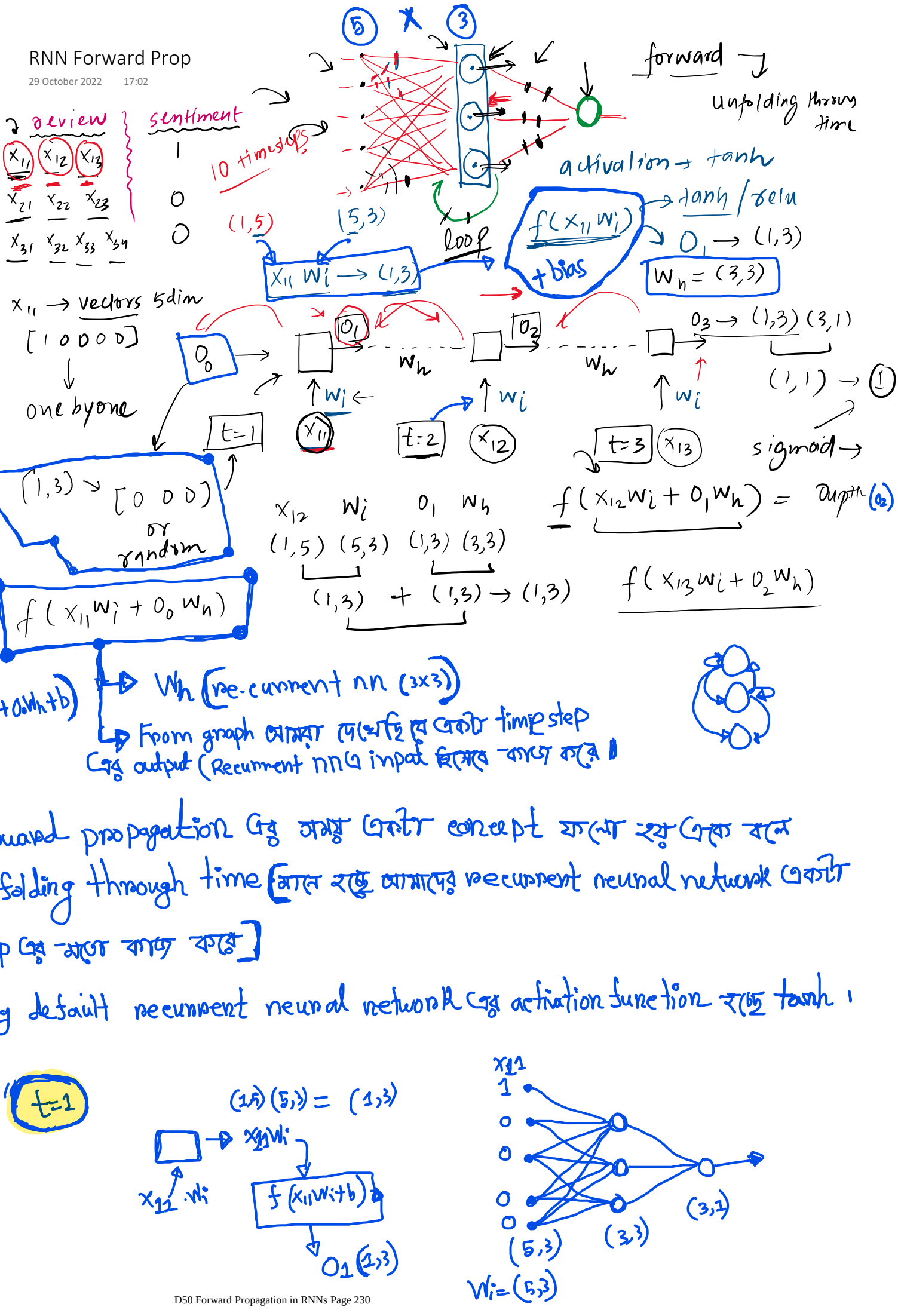
t=2, এ আমাদের আগের যেই output ছিল t=1 এর যেইটা neuron গুলোর সাথে input হিসেবে যাবে। কিন্তু, t=1 এর সময়ে t=0 এর কোন output ছিল না। তাই t=1এ আমরা zero বা random number input হিসেবে দেই।



- total trainable parameter 31
- see code example and find all the trainable parameter.

RNN Forward Prop

29 October 2022 17:02

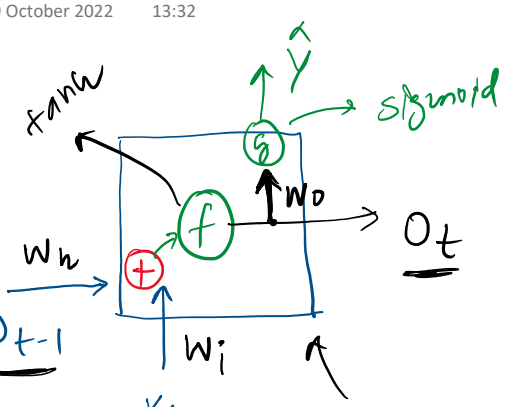


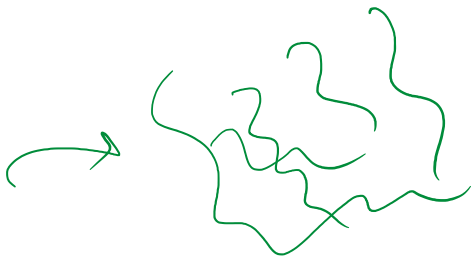
Forward propagation এর সময় একটা concept যেটা হয় সেটা বলে unfolding through time [মানে হচ্ছে আমাদের recurrent neural network একটা loop এর মতো কাজ করে]

⊕ By default recurrent neural network এর activation function হচ্ছে tanh।

Simplified Representation

29 October 2022 13:32





$\hat{y} = g(O_t W_o)$

$f(X_{it} W_i + O_{t-1} W_h) \rightarrow$ vector $\rightarrow O_t$

i = row #

t = time step

$X_{11} \rightarrow$

Hello how are you?

in general,

$f(X_{it} W_i + O_{t-1} W_h + b)$

- X_{it}: (t) current timestep org input
- W_i: input node
- O_{t-1}: (t-1) previous timestep org input
- W_h: re-current neural network

Code

29 October 2022 13:32

State and Memory

29 October 2022 13:33

RNN Sentiment Analysis

Keras Code Example

06 November 2022 09:14

Lecture: 57

আমরা দুই ভাবে text কে number এ convert করি।

i) Integer Encoding

ii) Embeddings

Integer Encoding :

Unique word: Hi, there, how, are, you

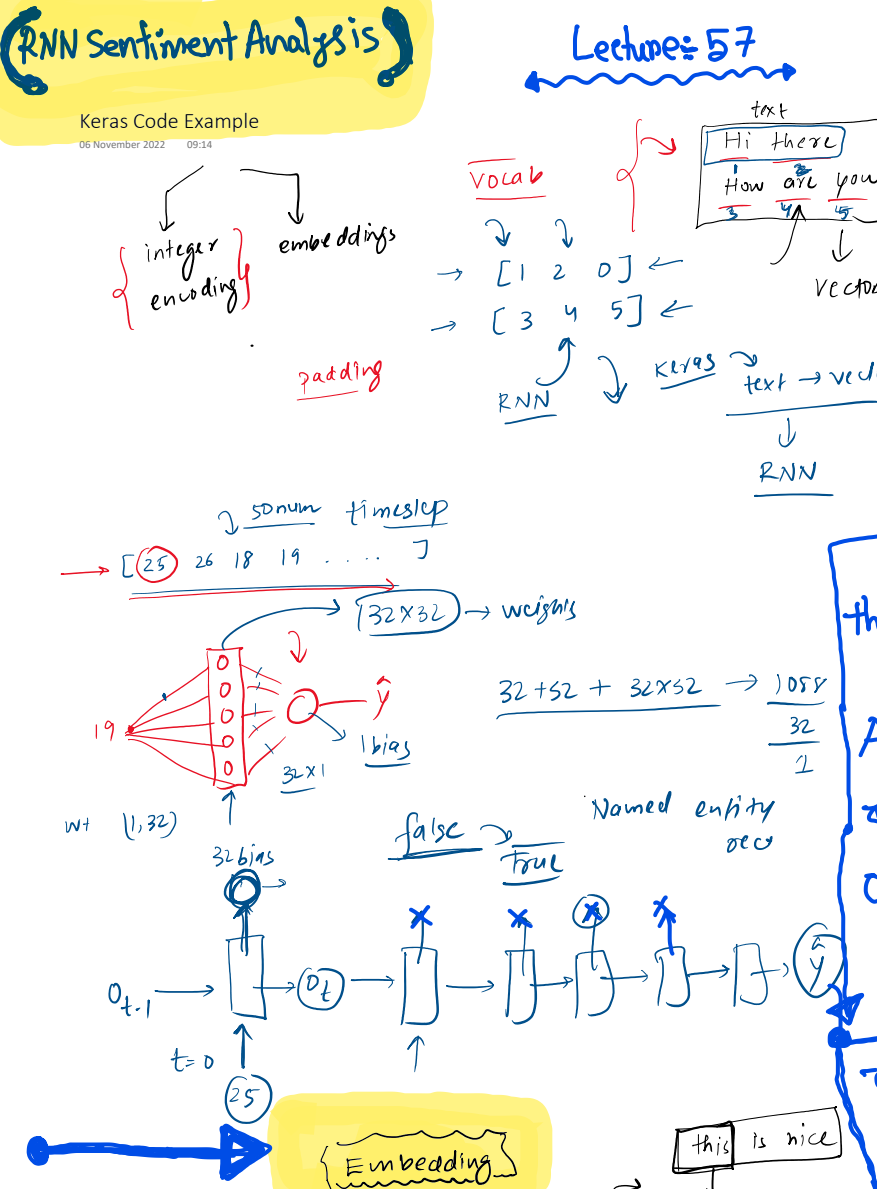
1 2 3 4 5

then, represent the sentence [1 2]

[3 4 5]

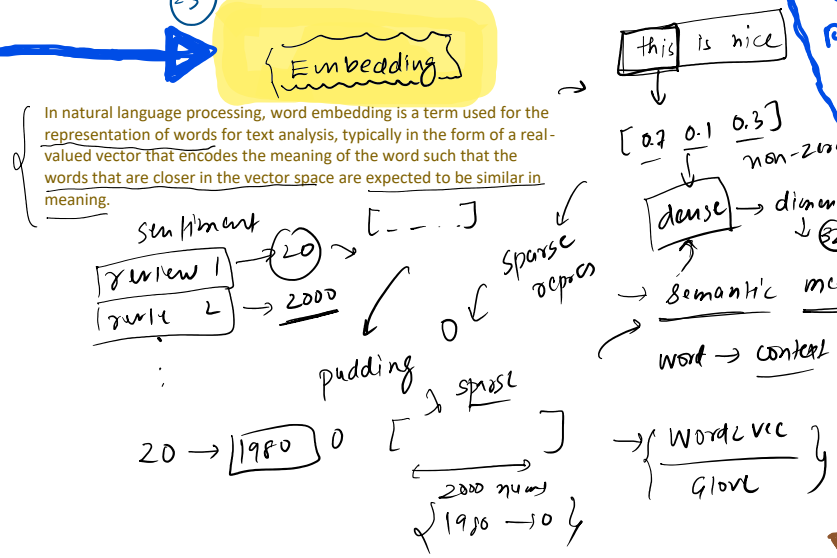
Array আলাদা হওয়ার যার array size সবচেয়ে বড় হবে তার সমান বাকী array গুলো বানাতে 0 padding করে।

Now see the code Example



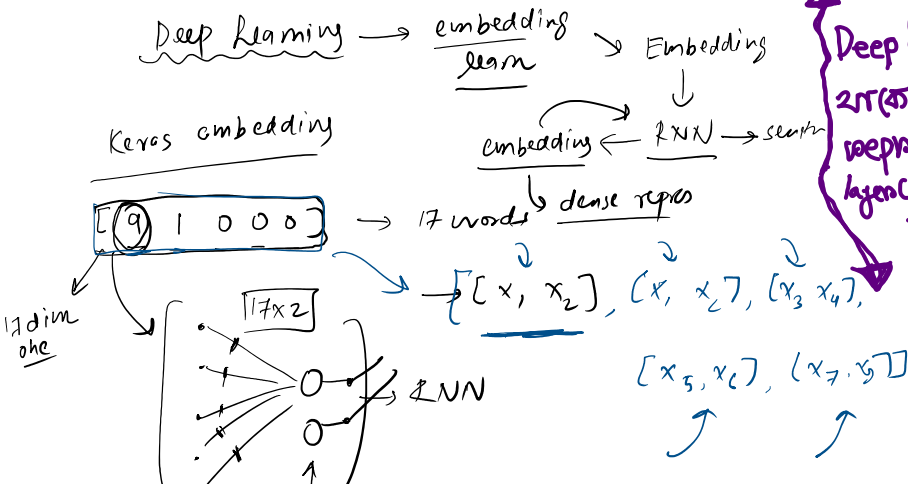
আমাদের প্রত্যেক time step এ output দরকার নেই শুধু final শুধু টা হলেই চলবে। তাই, return_sequences = False. এটা true হওয়াতে যখন আমার দরকার হবে ও টা word — Noun, Verb এর something.

In natural language processing, word embedding is a term used for the representation of words for text analysis, typically in the form of a real-valued vector that encodes the meaning of the word such that the words that are closer in the vector space are expected to be similar in meaning.

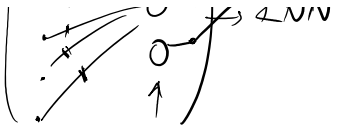


Embedding System

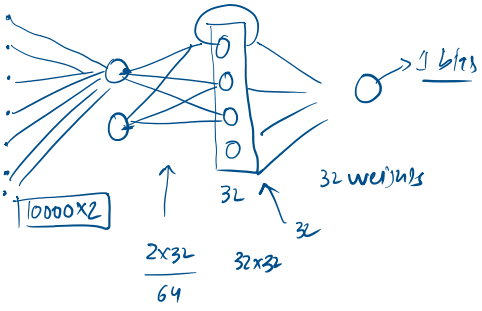
এখানে, word গুলো vector এ convert হয়। এই vector কে আমরা Dense বলি। কারণ, এতে non-zero element বেশি থাকে। অন্যদিকে, Integer encoding কে sparse representation বলে। কারণ padding করার পর এতে অনেক শূন্য চলে আসে। এজন্য dense representation semantic meaning capture করতে পারে। যদি দুটো word এ একই মানে হয় তাহলে সেইটা বুঝতে পারে।



Deep learning-এ আমাদের Embedding নামে layer থাকে। এই layer আমাদের জন্য dense representation তৈরি করে। Embedding layer এ data পাঠানোর আগে থেকে আমাদের কে Integer Encoding করতে হয়।



17 nodes



population in 9th year = x

$x + 10\%\text{ of } x = 10000$

$x + 0.1x = 10000$

$1.1x = 10000$

$x = \frac{10000}{1.1}$

$\frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{2}\left(\frac{x-1}{x}\right)^3 + \frac{1}{2}\left(\frac{x-1}{x}\right)^4 + \cdots$

t(n-1)

+ 2 +

+ (n-1)

50 → 500

A

A[35] → t sec

A[35] →

35

1 x 4 x35

x

1 25 50 100

12

O(n)

$O(n^2)$ → nested loops

input → 10 (100x100)

time → $(x)^2$ 4

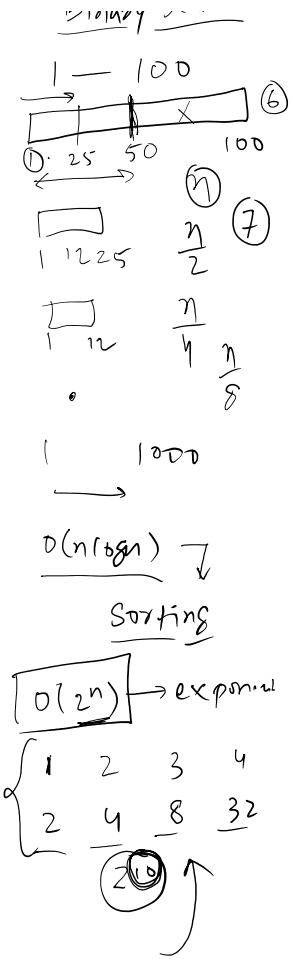
O(log(n))

Binary search

1 — 100

→ (6)

$\sqrt{(x_1}$


Binary
1 — 100
100
25 50
n/2
n/4
n/8
1000
O(nlogn)
Sorting
O(2^n) → exponential
1 2 3 4
2 4 8 32
for i in scan
O(n)
for j in (n^2)
O(n)
O(n+n) →
O(2n)
→ O(n)
O(n+n^2) O(n^2)
for i in arr:
for j in
25 → '25'
str()
n = 345%10
digits[5]
5 + ''
= '5'
345//10 → 34 != 0
34//10 → 3

34 // 10 → 4

4 + '5' →
45

34 // 10 → 3

3 != 0

3 // 10 => 3

digits[3]

3 + '45'
= '345'

Loop n → 10 times

345 = n ⇒ 345 0

3 → 4

inp → 10
time → 1 → log

O(n) → O(n+n)
O(n) O(2n) → O(n)

O(n) → O(n*n)
O(n) O(n²)

O(n)
O(n)
O(10000000)

n^2 10000000
×
$O(n^2)$

log(2)/2
1 → $\frac{n}{2}$

4 $\frac{n}{2}$ 2n

O(n) n=100

50, 100 $\frac{n}{2}$

2 → 100=n $\frac{n}{2}$

j=1 → 2
j=2 → 4 2-100
j=3 → 8
j=4 → 32

$\frac{n}{2} \times \log n$

n log n

O(1) → constant

n = 345

3+4+5 → 12

5

3 4 4

3 3

0

3450 → 4

log

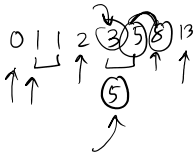
inp → 10 100

out → 1 2 3

fibonacci

↳ function

↳ recursion



n=1 n=0 → 1

fib(n)

↓

function calls

input → n → 10

#fcalls →

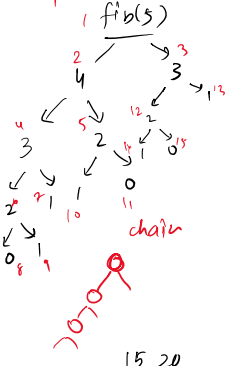
fib(3) n=3 1

fib(2) 2 fib(1) 3 → 1

fib(1) fib(0)

1 1

exponential → bad



O(2^n) 15 20 50

input 1 2 3 4

1 2 4 8 16

$O(2^n)$ (50)

input 1 2 3 4

t1 2 4 8 16

$n = 50, 100, 500$

↑ weeks

exponential

↳ days/weeks

subset → O(?)

power set

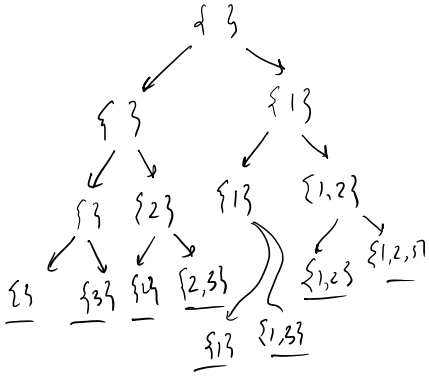
{1,2} → {{},{1},{2},{1,2}}

{1,2,3} →

{},{1},{2},{3},{1,2},{2,3}

{3,1} {1,2,3}

{1,2,3}



reduce → divide → log

Increase → multi → exp

→ exponential

{1,2} → 4 $2^2 = 4$

{1,2,3} → 8 $2^3 = 8$

$2^4 = 16$

$O(2^n)$

O(?)

$$T(n) = \begin{cases} 3T(n-1) & \text{if } n>0 \\ 1, & \text{otherwise} \end{cases}$$

$n > 0$

$T(n) = 3T(n-1)$

$= 3[3T(n-2)]$

$= 3^2 T(n-2)$

$= 3^2[3T(n-3)]$

$= 3^3 T(n-3)$

$= 3^n T(n-n)$

$= 3^n T(0)$

$T(n) = \boxed{3^n} \rightarrow O(3^n)$

$$T(n) = \begin{cases} 2T(n-1)-1 & \text{if } n>0 \\ 1, & \text{otherwise} \rightarrow \text{constant} \end{cases}$$

$=$

$T(n) = 2T(n-1)-1$

$= 2[2T(n-2)-1]-1$

$= 2^2 T(n-2) - 2 - 1$

$= 2^2[2T(n-3)-1] - 2 - 1$

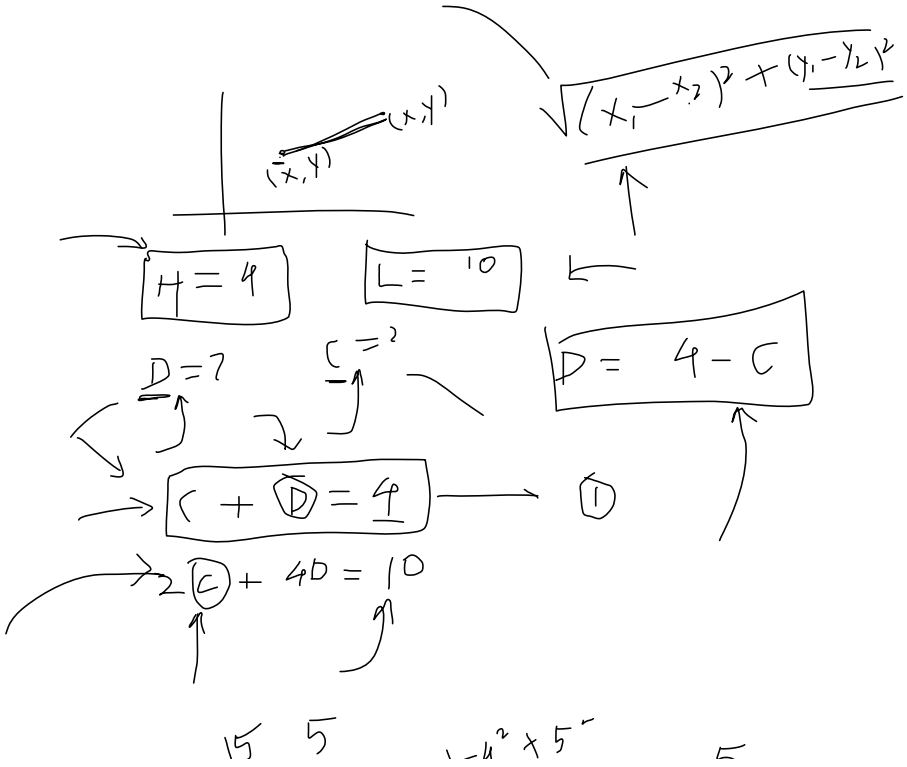
$= 2^3 T(n-3) - 2^2 - 2^1 - 2^0$

$= 2^n T(n-n) - 2^{n-1} - 2^{n-2} \ldots - 2^1 - 2^0$

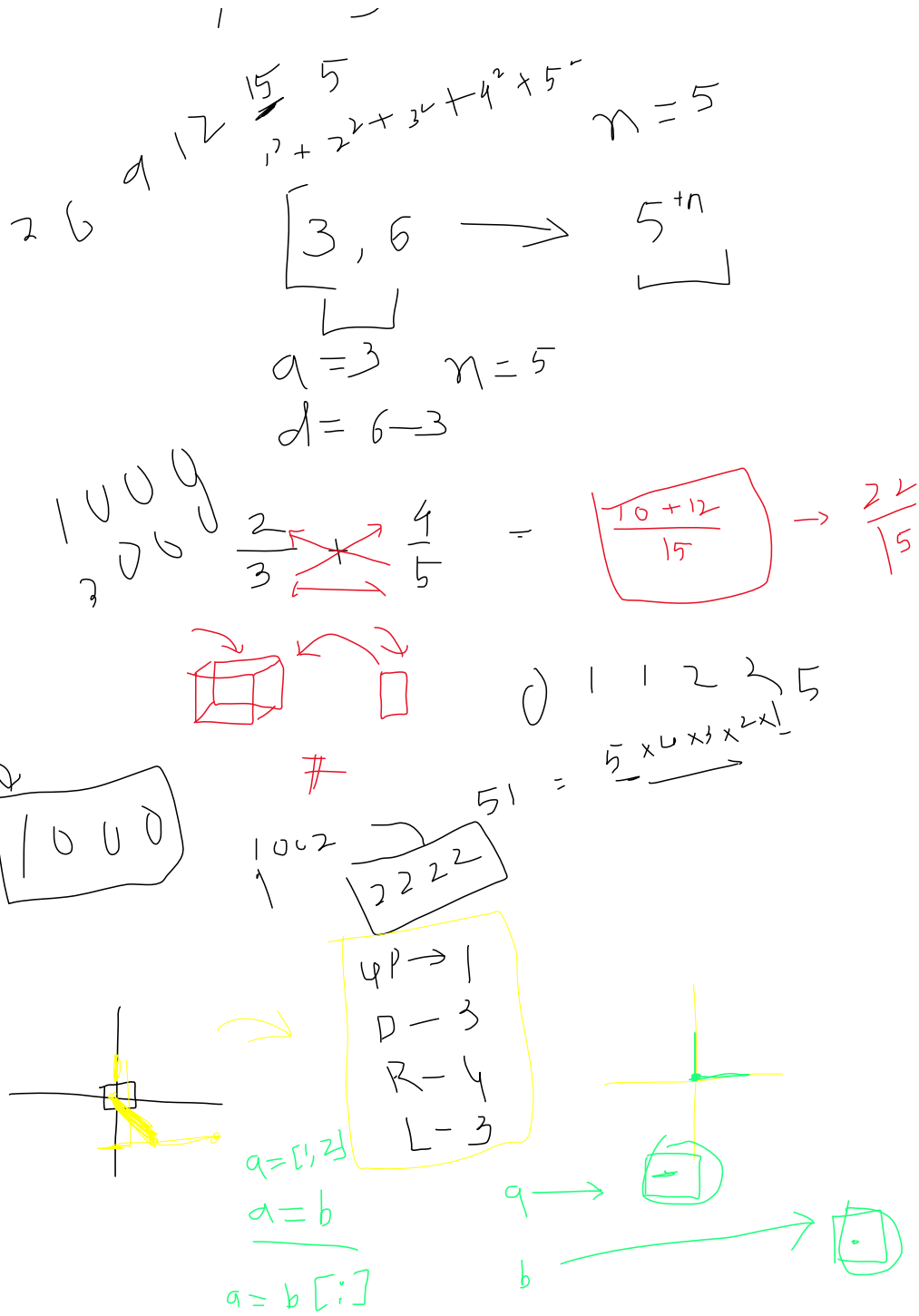
$= 2^n - [2^{n-1} + 2^{n-2} + \ldots + 2^1 + 2^0]$

$= 2^n - [2^n - 1] = 2^n - 2^n + 1$

$O(1) \rightarrow$ constant



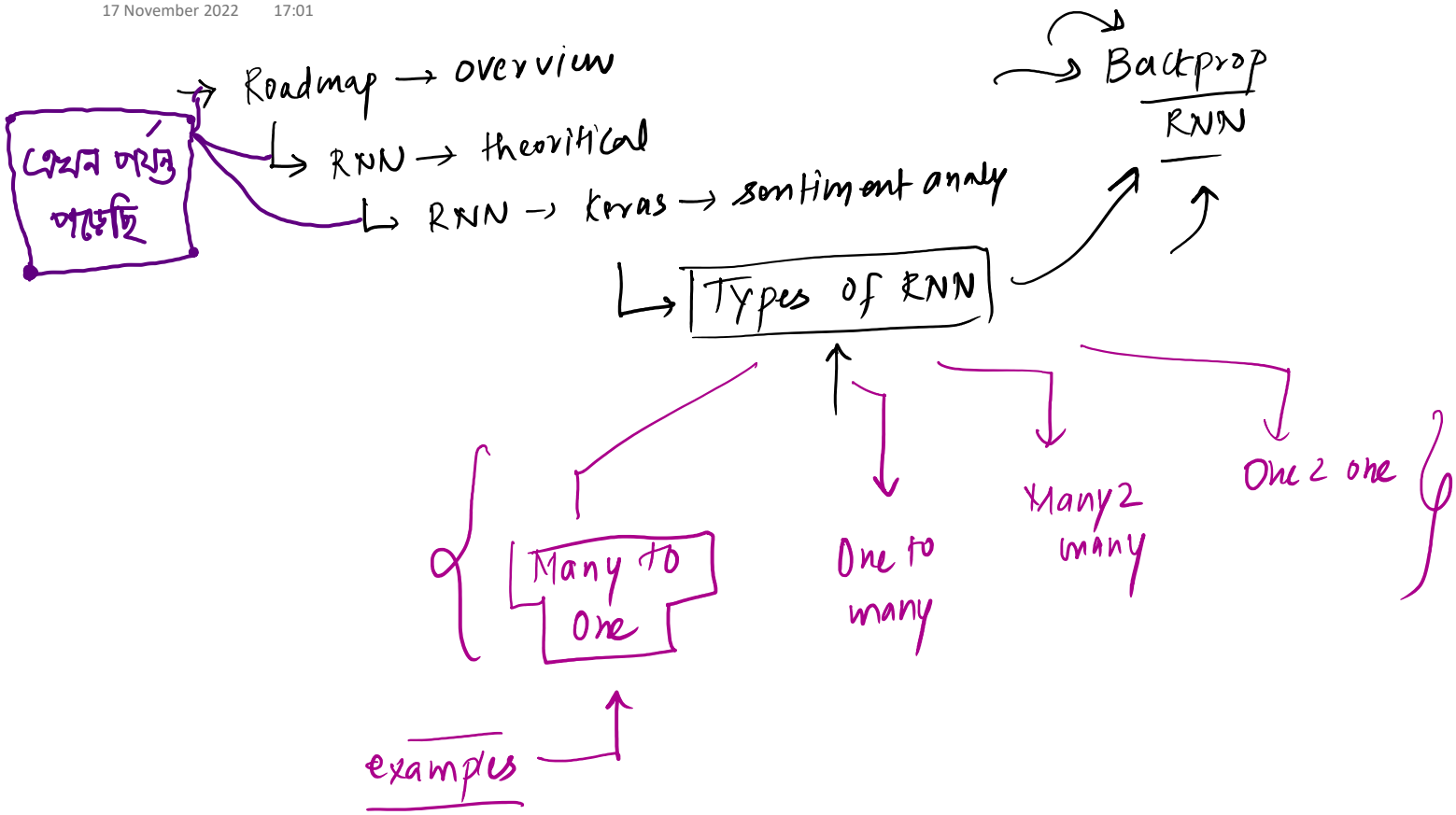
15 5 $\quad -4^2 \times 5^{-}$ $\quad -$


n = 5
[3, 6 → 5^n
q = 3 n = 5
d = 6 − 3
10 + 12 / 15 → 22/5
0 1 1 2 3 5
5 x 4 x 3 x 2 x 1
1000
1002
2222
51 =
up → 1
D − 3
R − 4
L − 3
q = [1, 2]
a = b
a = b [:]
q
b

Lecture No: 58

Till Now

17 November 2022 17:01



RNN ৪ ধরনের হয়ে থাকে। এই ৪ ধরনের RNN এর উপর ভিত্তি করে RNN এ Backpropagation আলাদা আলাদা ভাবে implement হয়।

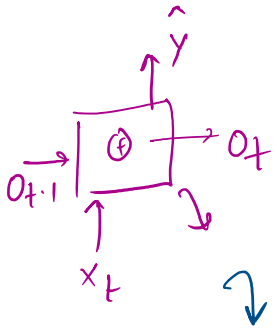
Many to One

17 November 2022 17:02

input → sequence → sent, chars, time series

out → non-sequnl → int/num → scalars → (1,0)

↳ sentiment → Anany



input ↳ sequential data i) sentence ii) character iii) time series

output ↳ non-sequential data i) int ii) number iii) scalar থাকে।

example: i) Sentement Analysis

ii) movie review to star rating (1 to 5)

input output

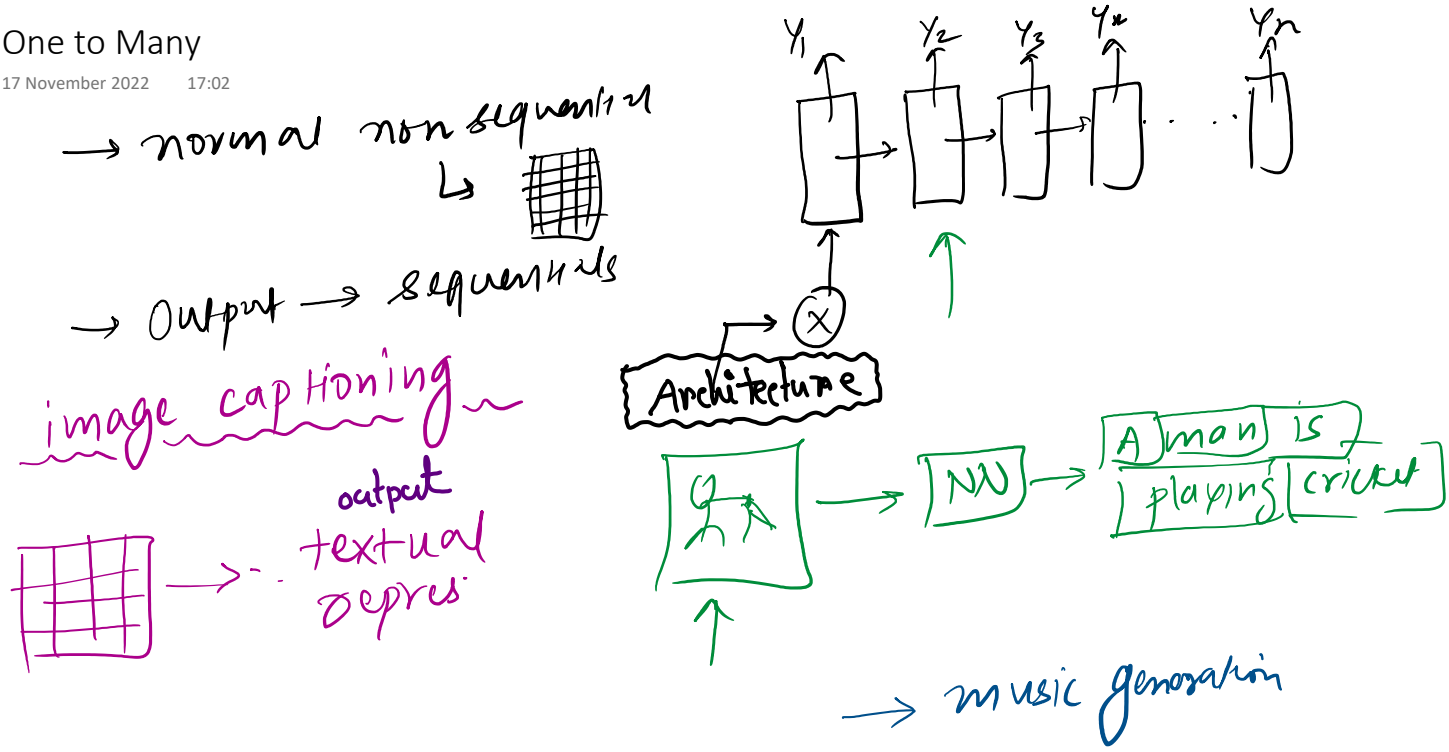
Multi-class classification problem

→ RNN- Architecture ↳ blocks হলো হচ্ছে recurrent blocks.

→ M to O বলতেই বোঝায়, input হচ্ছে multiple কিন্তু output হচ্ছে একটা।

One to Many

17 November 2022 17:02



Input: non-sequential data (image is also a non-sequential data)

output: Sequential data

example: i) input এ image দিলে তার caption লিখে দিবে। (image captioning)

ii) input এ non-sequential data দেই যেখান থেকে music পাই। (music Generation)

iii) Architecture Analysis করলে আমরা পাই-যে, একটা non sequential data (image) এর জন্য আমরা sequential data (text) পাচ্ছি।

Many to Many

17 November 2022 17:02

input → sequent → seq2 seq
out → sequen

input: sequential data
output: " "
এই থেকে seq2 seq ও বলে।
seq to seq দুই প্রকার।

Same length

Variable length

input seq == output seq

POS: Part of speech, where input seq is equal to output seq.

pos tagging → NLP

My(1) name(2) is(3) Nitish(4)

pronoun(1) noun(2) verb(3) Noun(4)

Current output

আগের output input হিসেবে

Current input

same length many2many RNN

NER → Name entity recognition {Another example} {same length}

Lets meet at 7pm at the airport

Variable length

machine trans

1 lang → 2 lang

google translate

Hi my name is Nitish

মেরা নাম নীতিশ হৈ

encoder decoder

encoder

অনুবাদে ভিন্ন ভিন্ন timestep input নেই

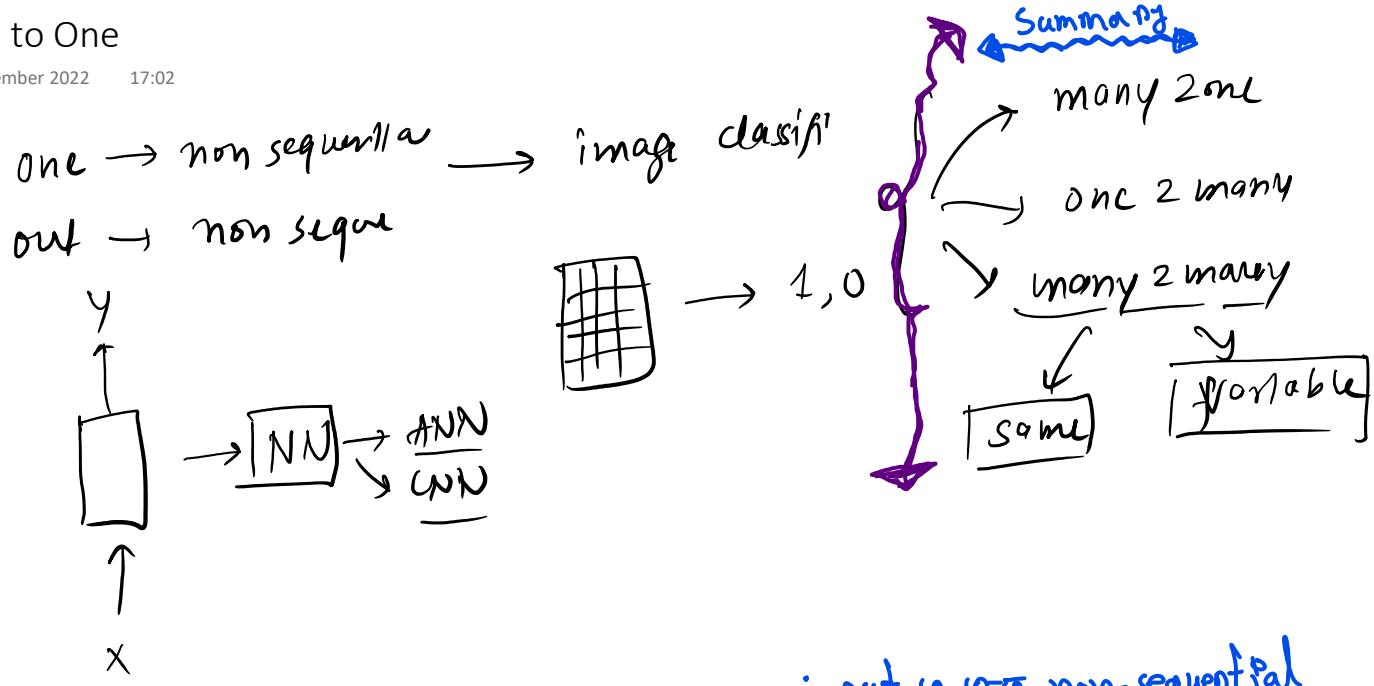
decoder

একবারে last input দেওয়ার পর সবগুলোর output পাই

Variable length এর উদাহরণ হলো google translate. English এ কোন sentence এ 15 টি word থাকলে বাংলায় সেইটা 15 এর চেয়ে কম বা বেশি বা সমান হতে পারে। MtoM Architecture এ আমরা -অনুবাদে input নেই তারপর output দেওয়া শুরু করি। একে encoder decoder বলে। input টুকু encoder আর output টুকু decoder বলে।

One to One

17 November 2022 17:02



- One to One এইটার RNN এর মধ্যে পড়েনা। কারণ, input এ যেহেতু non-sequential data আর output এ আমরা non-sequential data পাচ্ছি।

Lecture: 59

Backpropagation in RNN

01 December 2022 16:43

RNN intro → RNN → practical → types of RNN

→ RNN → Backprop → BPTT

RNN এর backpropagation কে Backpropagation through time (BPTT) বলে।

Many to One RNN

Sentiment Analysis

(Lets see an example of sentiment Analysis) [Binary classification problem]

text → 1/0

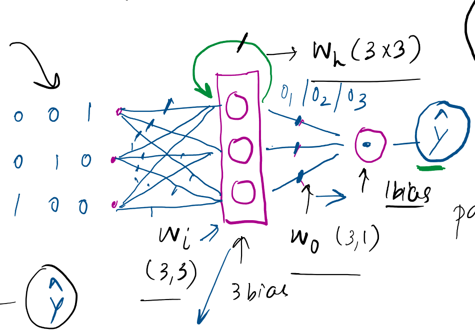
text			y
cat	mat	rat	1
rat	rat	mat	1
mat	mat	cat	0

unique words / vocab → cat [1 0 0], mat [0 1 0], rat [0 0 1]

	o_1	o_2	o_3	y
X_1	[1 0 0]	[0 1 0]	[0 0 1]	1
X_2	[0 0 1]	[0 0 1]	[0 1 0]	1
X_3	[0 1 0]	[0 1 0]	[1 0 0]	0

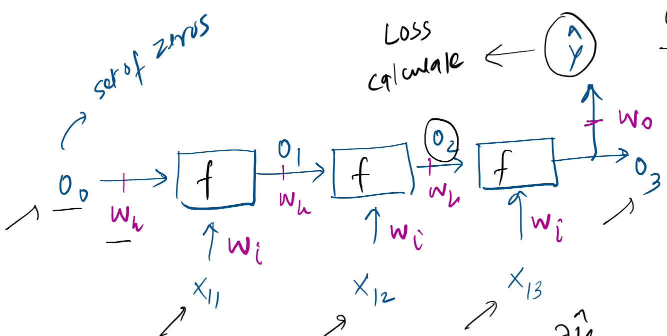
$\hat{y} \to L$

Parameters: W_i, W_h, W_o



প্রথমে x_1 input এ যাবে word by word. [1 0 0] যাওয়ার পর O_1 calculate হবে। O_1 & [0 1 0] এরপর input এ যাবে। এরপর, O_2 & [0 0 1] input এ যাবে। যেহেতু আর কোন word বাকী নেই অবশেষে আমরা final output পাবো। O_3

forward prop



$$O_1 = f(X_{11} W_i + O_0 W_h)$$

$$O_2 = f(X_{12} W_i + O_1 W_h)$$ (Put in activation fn)

$$O_3 = f(X_{13} W_i + O_2 W_h)$$

$$\hat{y} = \sigma(O_3 W_o)$$

Calculate loss & minimize with gradient descent:

loss calculate → minimize → L min → gradient descent

loss function Binary cross entrophy

$$L = -y_i \log \hat{y}_i - (1-y_i)\log(1-\hat{y}_i)$$

W_i W_h W_o

$\frac{\partial \hat{y}}{\partial W_o}$, $\frac{\partial L}{\partial \hat{y}}$

$$W_i = W_i - \eta \frac{\partial L}{\partial W_i}$$

$$W_h = W_h - \eta \frac{\partial L}{\partial W_h}$$

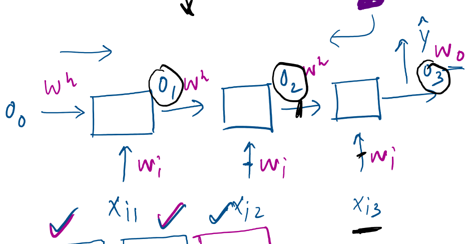
$$W_o = W_o - \eta \frac{\partial L}{\partial W_o}$$

Backprop

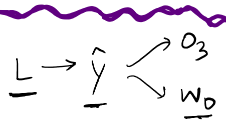
unfold → 3 times

$\frac{\partial L}{\partial W_o}$ $L \to \hat{y} \to O_3, W_o$

now, calculate the derivate $\frac{\partial L}{\partial w_i}$, $\frac{\partial L}{\partial w_h}$, $\frac{\partial}{\partial w_o}$



$\frac{\partial L}{\partial w_o}$



$$\frac{\partial L}{\partial w_o} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_o}$$

$\frac{\partial L}{\partial w_o}$, $\frac{\partial L}{\partial w_i}$, $\frac{\partial L}{\partial w_h}$

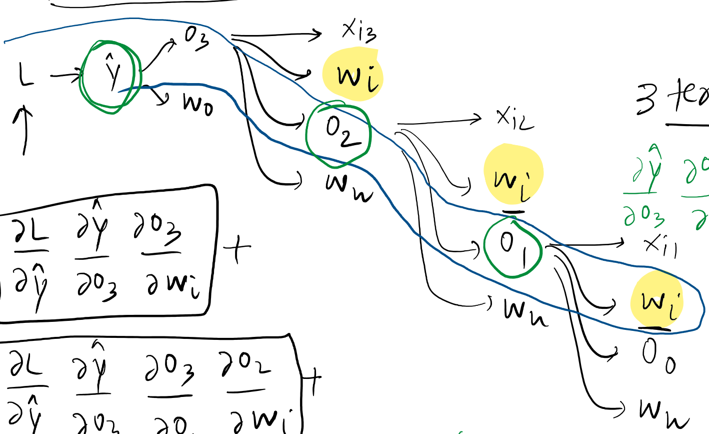
$\rightarrow O_1 = f(x_{i1} w_i + O_0 w_h)$

$\rightarrow O_2 = f(x_{i2} w_i + O_1 w_h)$

$\rightarrow O_3 = f(x_{i3} w_i + O_2 w_h)$

$\hat{y} = \sigma(O_3 w_o)$

$\frac{\partial L}{\partial w_i}$



3 terms

$\frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2}$

⊛ ⊛ তিন ঘরের যোগফলের সমান হবে।

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial w_i} + \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial w_i} + \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial w_i}$$

$$\frac{\partial \hat{y}}{\partial O_1} = \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial O_1}$$

10 words ধরলে 10টা

10 terms

t = 1

$$\frac{\partial L}{\partial w_i} = \sum_{j=1}^{3} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_j} \frac{\partial O_j}{\partial w_i}$$

j=3 → $\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial w_i}$

j=1 → $\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_1} \frac{\partial O_1}{\partial w_i} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial w_i}$

j=2 → $\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_2} \frac{\partial O_2}{\partial w_i} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial w_i}$

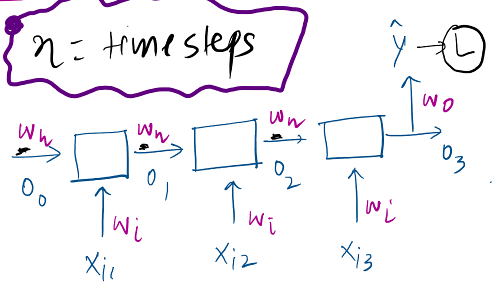
Therefore,

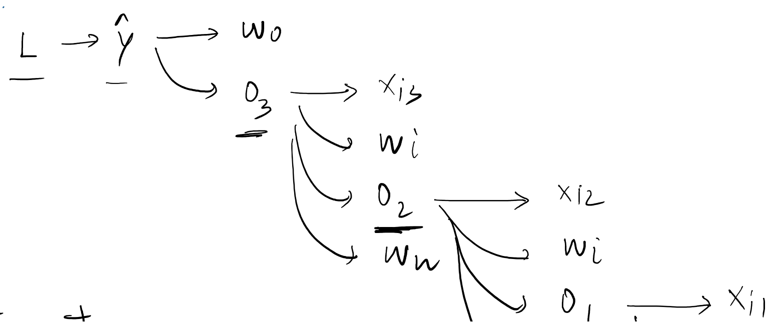
$$\frac{\partial L}{\partial w_i} = \sum_{j=1}^{n} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_j} \frac{\partial O_j}{\partial w_i}$$

n = timesteps

W_i বের করার সাথে মিল রয়েছে

Now → $\frac{\partial L}{\partial w_h}$





$$\frac{\partial L}{\partial w_h} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial w_h} +$$

$\partial L \quad \partial \hat{y} \quad \partial O_2 \quad \partial O_2 \quad +$

∂w_h ∂y ∂o_3 ...

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial w_h} +$$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial o_1} \frac{\partial o_1}{\partial w_h}$$

$$\frac{\partial L}{\partial w_h} = \sum_{j=1}^{n} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_j} \frac{\partial o_j}{\partial w_h}$$

n = timesteps

for $j=3$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial w_h} \rightarrow \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_1} \frac{\partial o_1}{\partial w_h}$$

for $j=10$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_{10}} \frac{\partial o_{10}}{\partial o_1} \frac{\partial o_1}{\partial w_h}$$

$$\prod_{t=2}^{j} \frac{\partial o_t}{\partial o_{t-1}}$$

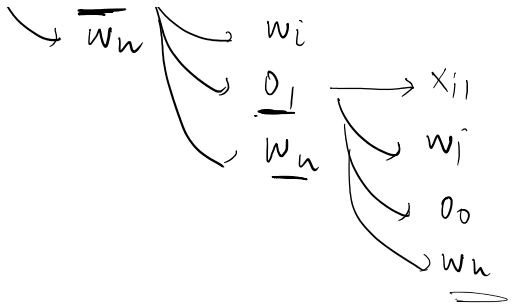
$$t=3 \quad \frac{\partial o_t}{\partial o_{t-1}} = \frac{\partial o_2}{\partial o_1} \frac{\partial o_3}{\partial o_2}$$

$t=2$

$$o_t = f(x_{it} w_{inp} + o_{t-1} w_h)$$

$$\frac{\partial o_t}{\partial o_{t-1}} = \prod_{t=2}^{j} f'(x_{it} w_{inp} + o_{t-1} w_h) w_h$$

$[0-1]$



Lecture : 60

Problem with RNN

19 December 2022 16:33

আমরা RNN পড়েছি sequential data এর জন্য। কিন্তু RNN ২টা major problem থাকার কারণে RNN ব্যবহার না করে আমরা LSTM Architecture build করেছি।

RNNs → sequential data → textual, time series

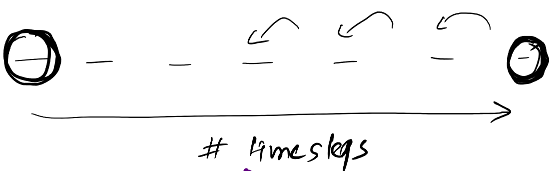
suffer (2) major problems → LTSMs (Why)

Prob:1 → Problem of long term dependency → Unstable gradients

Prob:2 → ~~Unstable gradients~~ Stagnated values

RNN

Application: next word predicts



→ RNN

\# timesteps

→ Marathi is spoken in Mahar... (short)

→ Maha is a beautiful place. I went there last year, but I could not enjoyed properly b/c I dont understand marathi

RNN → keypad

long term

timestep অনেক বেশি হলে RNN আগের জিনিসগুলো মনে রাখতে পারেনা। তখনই long term dependency prob বলে একে মূলে রয়েছে যার কারণে LTDP হয়।

RNN → Vanishing gradient prob

2nd : Exploding gradient problem → RNN → যার কারণে training ঠিকঠাক হয় না।

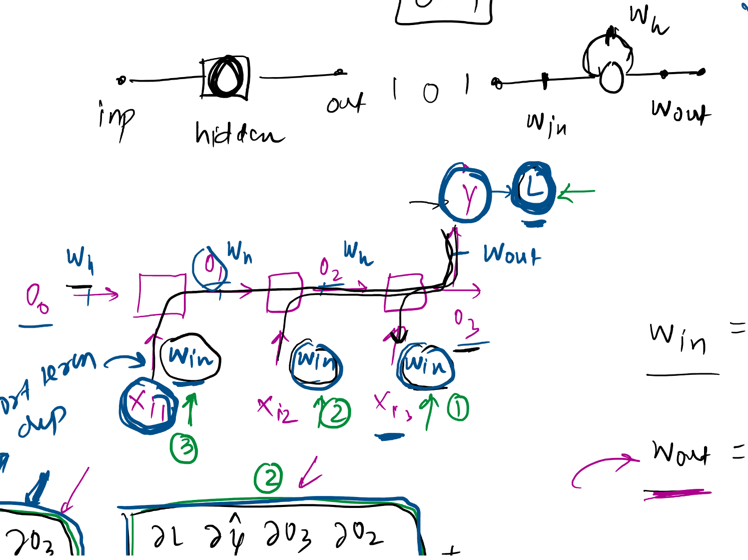
possible sol → remove

Problem # 1 → Problem of long term dependency → Vanishing

inp			out
1	0	1	1
0	0	1	0
0	0	0	0
1	1	1	1

3 time stamps

0-1



L → min

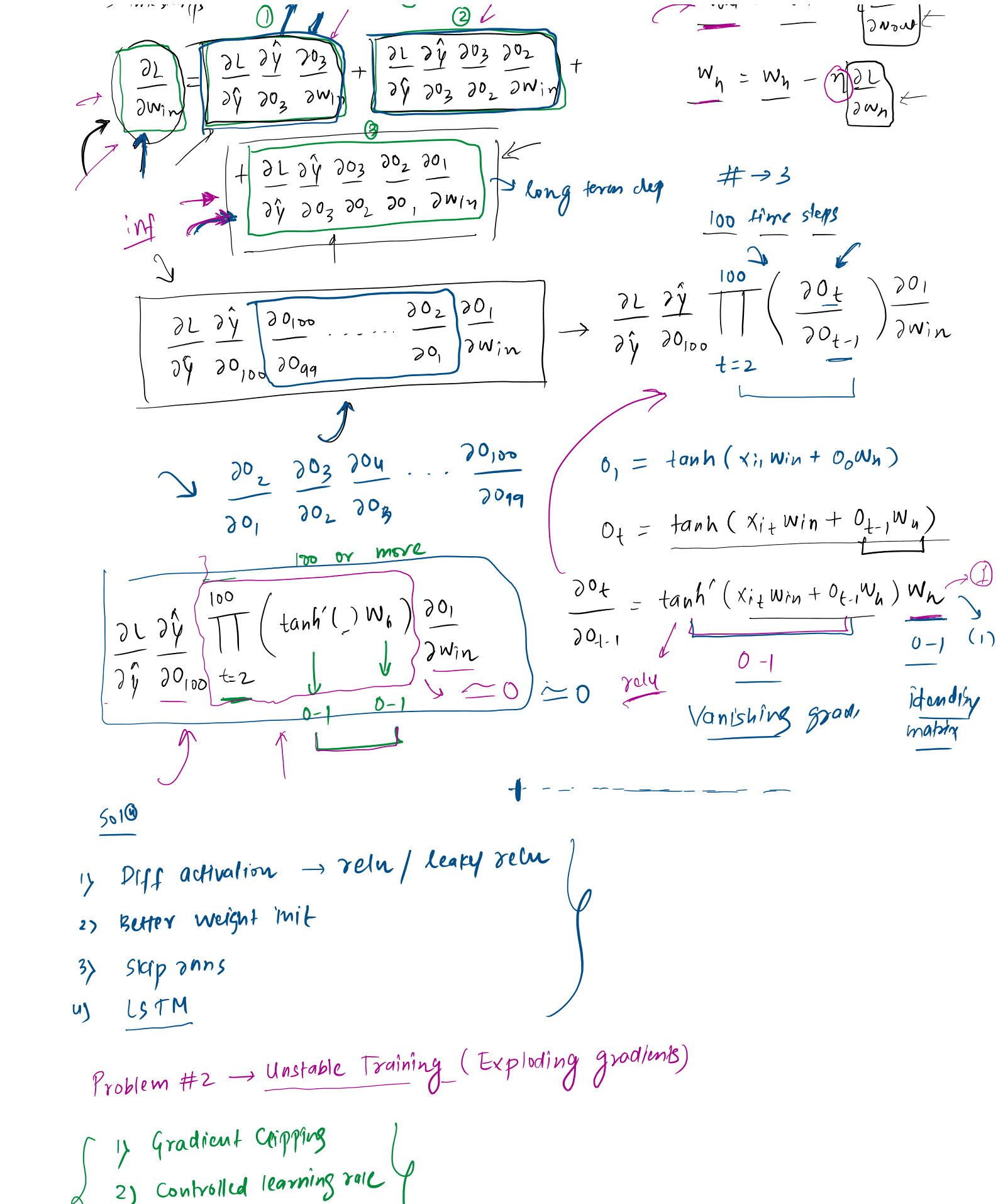
W_h, W_{in}, W_{out} → gradient des

$W_{in} = W_{in} - \eta \frac{\partial L}{\partial W_{in}}$

$W_{out} = W_{out} - \eta \frac{\partial L}{\partial W_{out}}$

$\partial L \quad \partial \hat{y} \quad \partial O_3$

$\partial L \quad \partial \hat{y} \quad \partial O_3 \quad \partial O_2$


Long term dep
100 time steps
100 or more
Vanishing grad
identity matrix
Sol(n)
1) Diff activation → relu / leaky relu
2) Better weight init
3) Skip anns
4) LSTM
Problem #2 → Unstable Training (Exploding gradients)
1) Gradient Clipping
2) Controlled learning rate
3) LSTM

Recap

21 August 2023 11:55