CSE 465/L-21/06.04.2025/

* RNN:
  - Recurrent Neural Networks
  - used for sequential data, where order matters.
    - time seriese data.
    - text data
    - speech or audio
    - videos

* Sequential vs Non-sequential comparison:

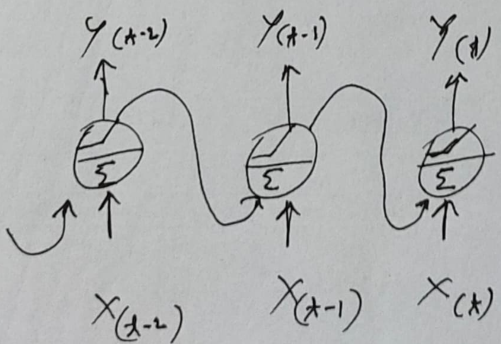
  $\boxed{\text{slide - 4}}$

* Reasons, why ANN and CNN are not suitable for sequential data.

  $\boxed{\text{Slide - 7}}$.

* Some applications of RNNs - $\boxed{\text{Slide - 8-9}}$

*

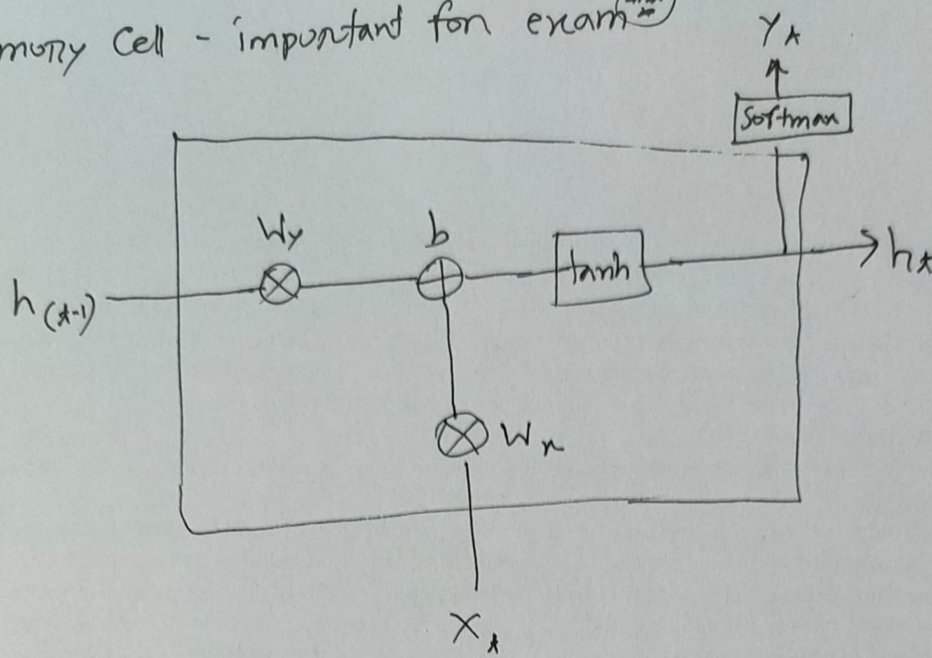


form a layer,

activation function

$$Y_{(t)} = \phi\left(W_x{}^T X_{(t)} + W_y{}^T Y_{(t-1)} + b\right)$$

Memory Cell - important for exam **



Types of RNNs:

(i.) Sequence to sequence

- one to one
- time seriese data, stock price

(ii) Sequence to vector

- many to one
- text data

(iii) Vector to sequence

- one to many
- image to caption

(iv) Encoder to Decoder

- sequence to vector and vector to sequence
- language translation.

L-22 / 08.04.2025 /

Shortcomings of RNNs:

Unstable gradient problem:

- solved ~~caused~~ by . parameter initialization, faster optimizers, dropouts.
- use tanh
- Layer normalization.

Batch normalization:

- not good for RNNs but good for CNN.

$$\hat{x}^{(i)} = \frac{x^{(i)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

avoid division by zero.

Here,

$$\mu_B = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu_B)^2$$

Slide - 23-ৰ

Then, scale and shift,

$$y^{(i)} = \gamma \hat{x}^{(i)} + \beta$$

Learnable parameter

Layer normalization:

- good for RNNs

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

Here,

$$\mu = \frac{1}{d} \sum_{i=1}^{d} x_i \Big\} \text{ mean}$$

$$\sigma^2 = \frac{1}{d} \sum_{i=1}^{d} (x_i - \mu)^2 \Big\} \text{ variance}$$

Then, scale and shift,

$$y_i = \gamma_i \hat{x}_i + \beta_i$$

Shortcoming of RNNs.

- short time memory loss
- have no trace of the first inputs.
- Solutions!

  LSTM, GRU

LSTM cell diagram : must in final

$\boxed{\text{Slide-29}}$

- equation of parameters: $\boxed{\text{Slide-32}}$

~~GRU~~

Diagram of GRU cell and equation of parameters

$\boxed{\text{Slide - 35}}$ (***)

L-23 / 13.04.2025

NLP:

- Natural Language Processing
- RNN

Control diversity:

temperature ↑ ⇒ more diverse

Previous RRNs were stateless

in NLP, RRNs are stateful. $\boxed{\text{Slide - 12,13}}$

Tokenizations:

- splits text into smaller units, which are often words, subwords, or even characters.
- makes raw text manageable for processing
- needs to be done before applying model.

⊛ Word embedding:

- maps each token to a dense numerical vector in a high-dimensional space.
- capture the semantic meaning of words.
- some pre-trained word embedding:

  slide - 17

⊛ NMT:

- Neural Machine Translation
- How to translate English sentence?

⊛ Attention Mechanism diagram

slide - 25

L-24 / 15. 04. 2025

⊛ Basic information about XAI, from lecture slide 12.

Final Exam

24. 04. 2025

CNN, RNN, Optimizer, NLP, XAI