

# Introduction to Deep Learning

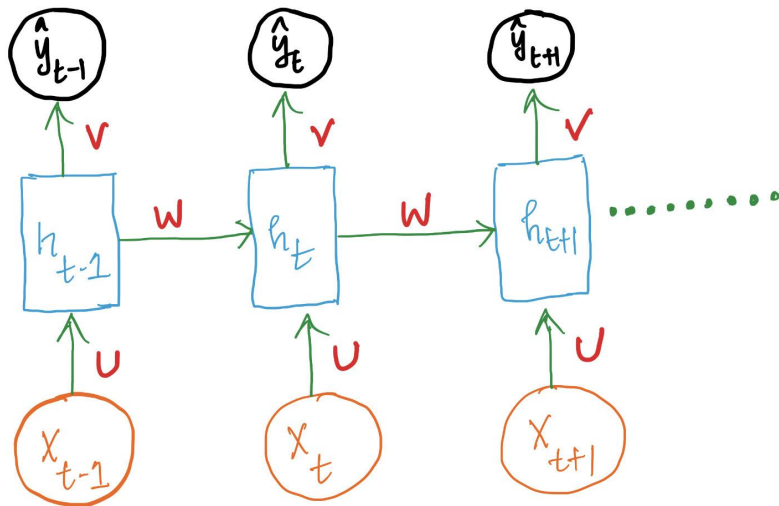
## Recurrent Neural Networks

# Recurrent Neural Networks

- Why not use Conv Networks?
- The order of input in sequential data is considered unlike treating it as independent inputs.
- Types of RNN
  - One to One
  - One to Many
  - Many to One
  - Many to Many

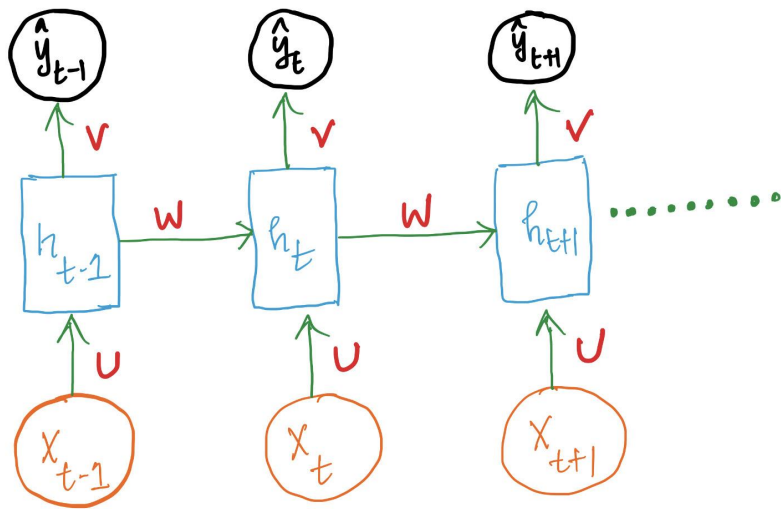
# Recurrent Neural Networks

- The order of input in sequential data is considered unlike treating it as independent inputs.
- Recurrent connection allows information flow from one time point to another



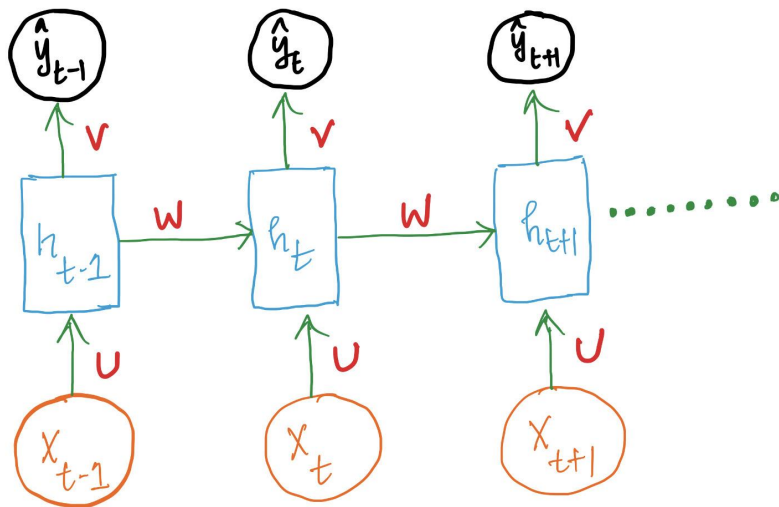
# Recurrent Neural Networks

- At each time point, the hidden state depends on current input and previous hidden state.



# Recurrent Neural Networks

- Parameters of the model are updated based on **Backpropagation through time (BPTT)**
- Multilayer RNNs



## Long Short Term Memory

- Improvement over RNNs: capture long term dependencies.
- Has an additional cell state, that stores long term information.
- Cell state is altered through three gates:- input gate, forget gate, output gate.
- Values of the states vary from 0 to 1.

## Gated Recurrent Unit

- Improvement over LSTMs with cell state and hidden state combined.
- Fewer parameters than LSTMs

# RNN

Tensorflow example

# Transformers

- Can capture long term dependencies
- Uses attention mechanism instead of recurrence
- Tokens: smallest unit of text
- Each token has a query, key and value vectors
- Query = Which other words in this sentence help understand the current word better?
- Key = represents the kind of information they hold
- If query and keys are similar, then more attention -> dot product
- Apply softmax to get attention scores
- Value = The actual content
- Weighted sum of attention scores and value.