SNS COLLEGE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTION)

Approved by AICTE & Affiliated to Anna University
Accredited by NBA & Accrediated by NAAC with 'A+' Grade,
Recognized by UGC saravanampatti (post), Coimbatore-641035.

Department of Biomedical Engineering

Course Name: 19ECT303 & Artificial Intelligence and machine learning

III Year : V Semester

Unit V–DEEP LEARNING

Topic : Deep Recurrent Neural Networks

Computation in RNNs: parameter blocks

The computation in most recurrent neural networks can be decomposed into three blocks of parameters and associated transformations:
1. From the input to the hidden state
2. From the previous hidden state to the next hidden state
3. From the hidden state to the output

19ECT303/Artificial Intelligence and Machine learning/Unit 5/ Mr. Kathik G.L./AP/BME

Blocks of parameters as a shallow transformation

- With the RNN architecture shown each of these three blocks is associated with a single weight matrix, i.e.,

  - When the network is unfolded, each of these corresponds to a shallow transformation.

  - By a shallow Transformation we mean a transformation that would be represented a single layer within a deep MLP.

- Typically this is a transformation represented by a learned affine transformation followed by a fixed nonlinearity
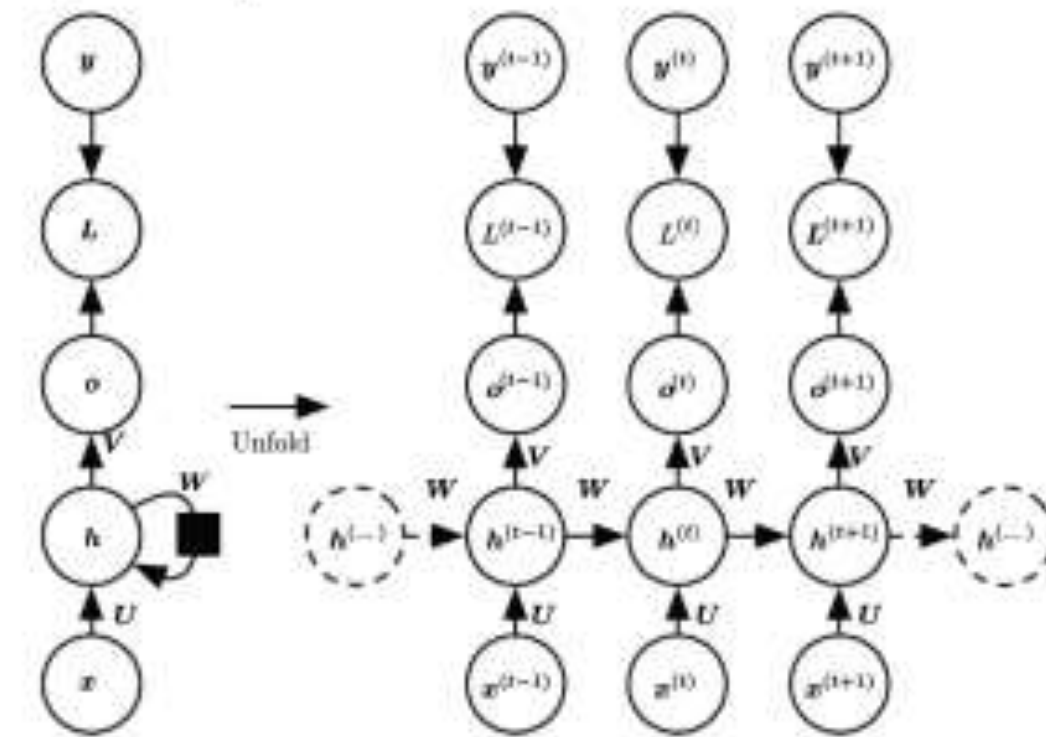
Would it be advantageous to introduce depth into each of these operations?
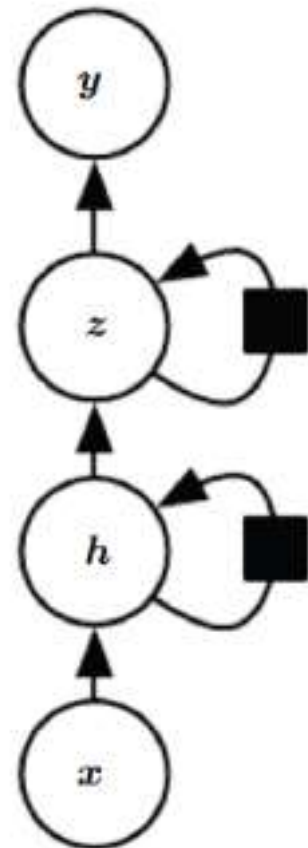
Experimental evidence strongly suggests so.
• That we need enough depth in order to perform the required transformations
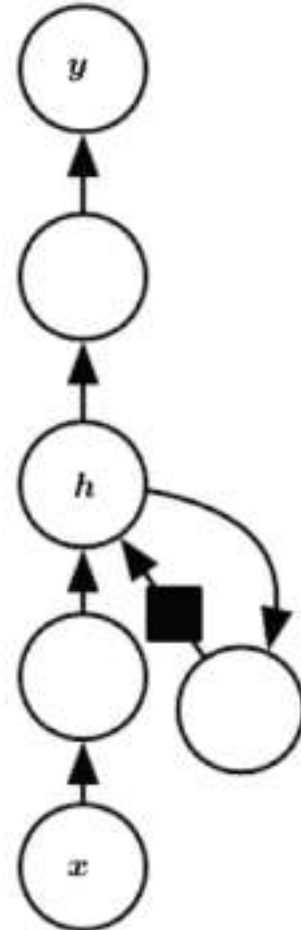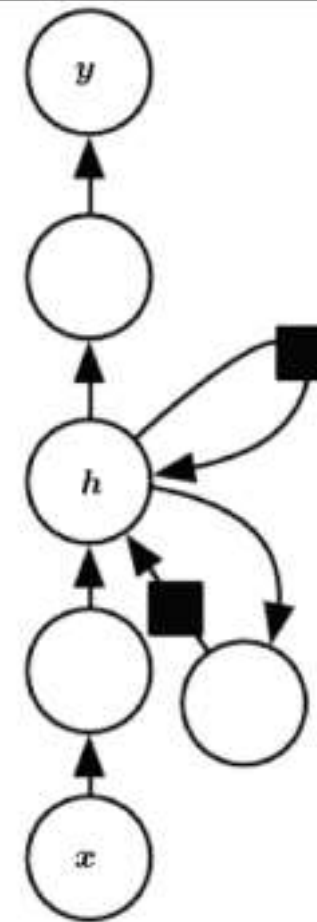
Ways of making an RNN deep

| 1. Hidden recurrent state can be broken down into groups organized hierarchically	2. Deeper computation can be introduced in the input-hidden, hidden-hidden and hidden-output parts. This may lengthen the shortest path linking different time steps	3. The path-lengthening effect can be mitigated by introducing skip connections.

1. Recurrent states broken down into groups

.

We can think of lower levels of the hierarchy play a role of transforming the raw input into a representation that is more appropriate at the higher levels of the hidden state
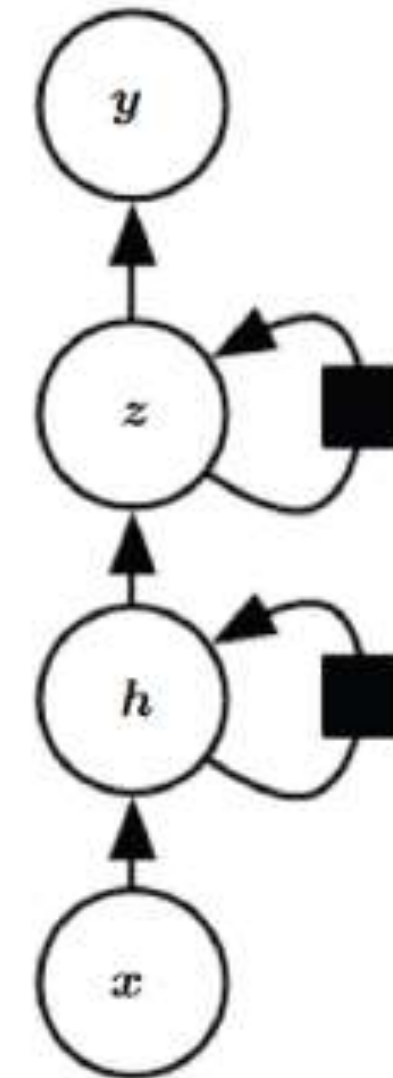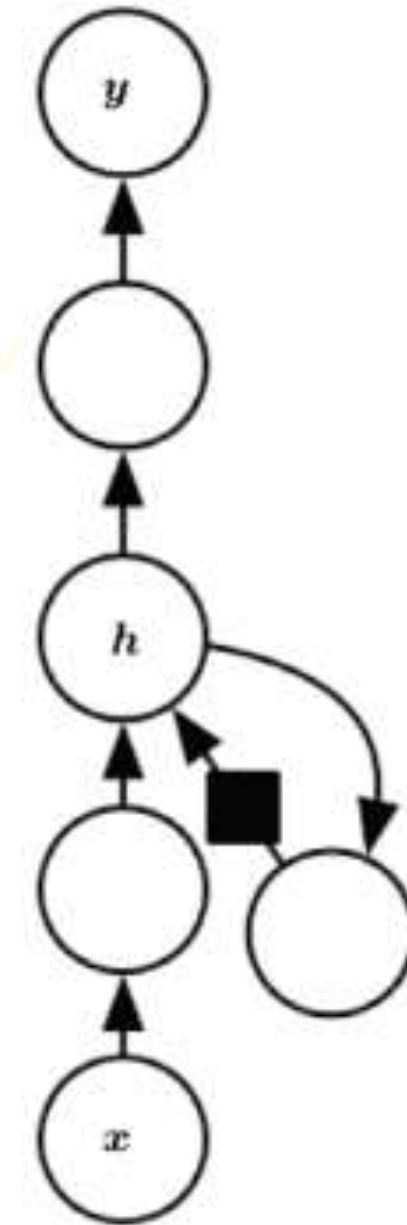
2. Deeper computation in hidden-to-hidden

-

- Go a step further and propose to have a separate MLP (possibly deep) for each of the three blocks:
    1. From the input to the hidden state
    2. From the previous hidden state to the next hidden state
   3. From the hidden state to the output

- Considerations of representational capacity suggest that to allocate enough capacity in each of these three steps

  - But doing so by adding depth may hurt learning by making optimization difficult

  - In general it is easier to optimize shallower architectures

  - Adding the extra depth makes the shortest time of a variable from time step t to a variable in time step $t+1$ beome longer

3. Introducing skip connections

- For example, if an MLP with a single hidden layer is used for the state-to-state transition, we have doubled the length of the shortest path between variables in any two different time steps compared with the ordinary RNN.
- This can be mitigated by introducing skip connections in the hidden-to-hidden path as illustrated here
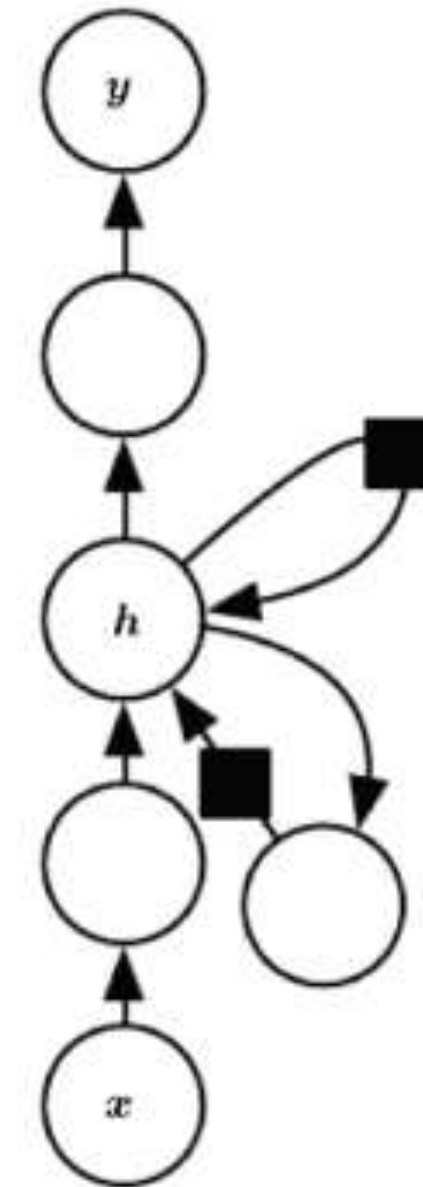
THANK YOU !!!
HAPPINESS ISN'T OUTSIDE, ITS WITHIN