DS-501

Deep Learning: Lecture 4

Lecturer: Dr. Giant

BITTIGER

The Lifelong Learning Platform of Silicon Valley

Copyright Policy

All content included on the Site or third-party platforms as part of the class, such as text, graphics, logos, button icons, images, audio clips, video clips, live streams, digital downloads, data compilations, and software, is the property of BitTiger or its content suppliers and protected by copyright laws.

Any attempt to redistribute or resell will result in the appropriate legal action being taken.

We thank you in advance for respecting our copyrighted content.

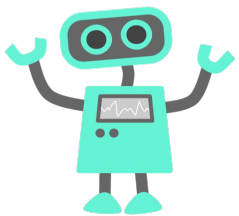
Agenda

- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- RNN Codelab
- Linear Regression in TensorFlow
- TensorBoard

Recurrent Neural Network (RNN)

Let’s Book a Ticket

Joseph: I would like to reach **Hsinchu** on **November 24th**



ticket booking robot

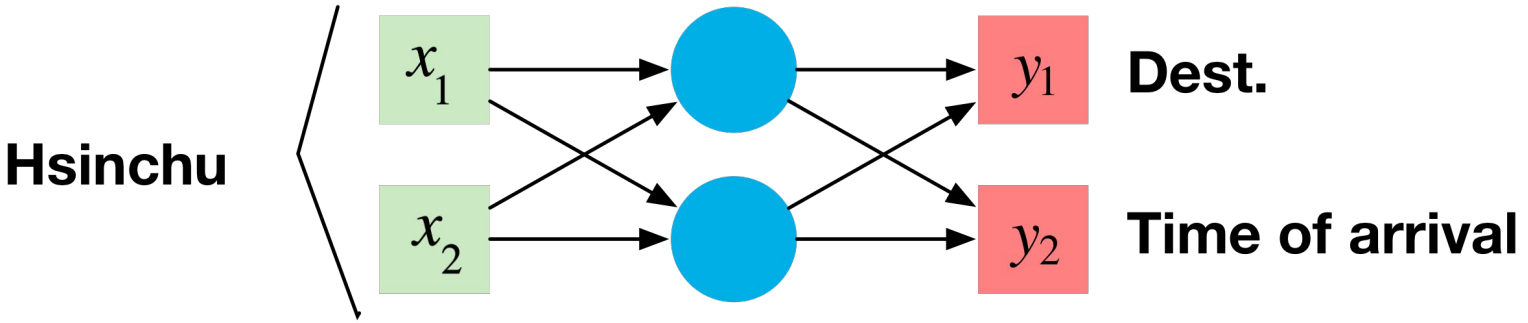
Destination: **Hsinchu**
Time of arrival: **November 24th**

Let's Book a Ticket (cont'd)

Solving slot filling by a feedforward neural network (FFNN)

Input: a word represented as a vector

Output: probability distribution that the input word belonging to the slots



Let's Book a Ticket: Problem

Dest.

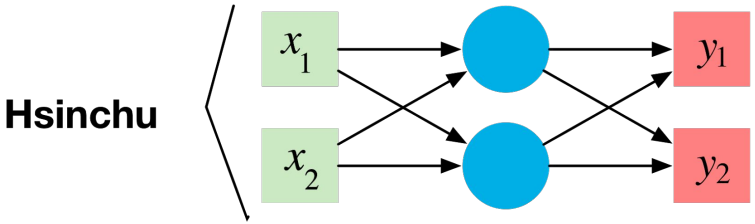
Joseph: reach Hsinchu on November 24th

or

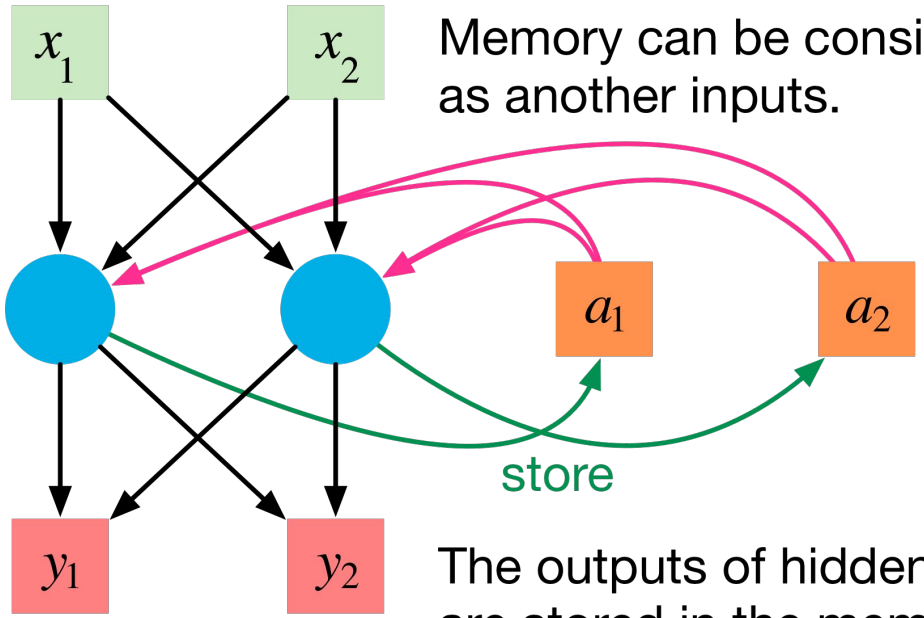
leave Hsinchu on November 24th

place of departure

Neural network needs memory!



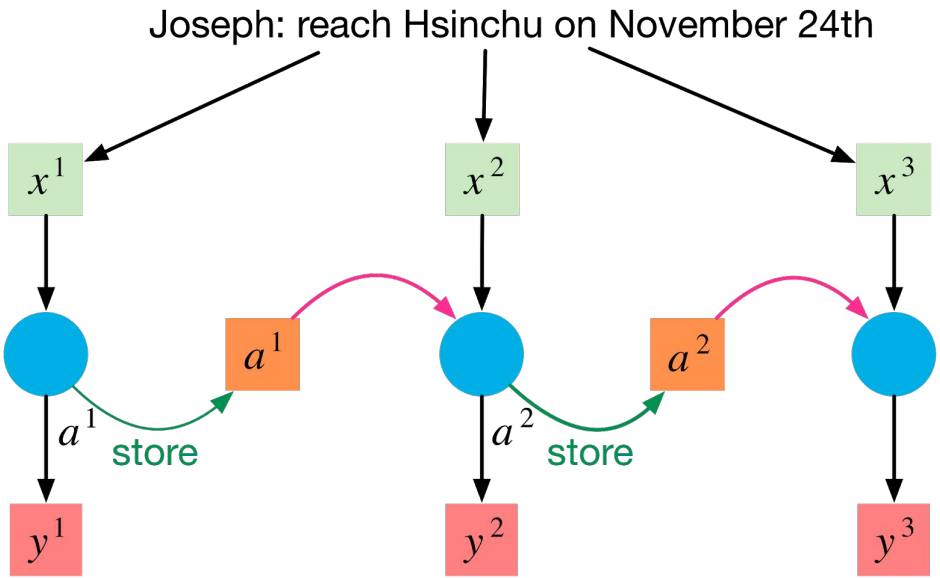
Network with Memory



Memory can be considered as another inputs.

The outputs of hidden layer are stored in the memory.

Recurrent Neural Network



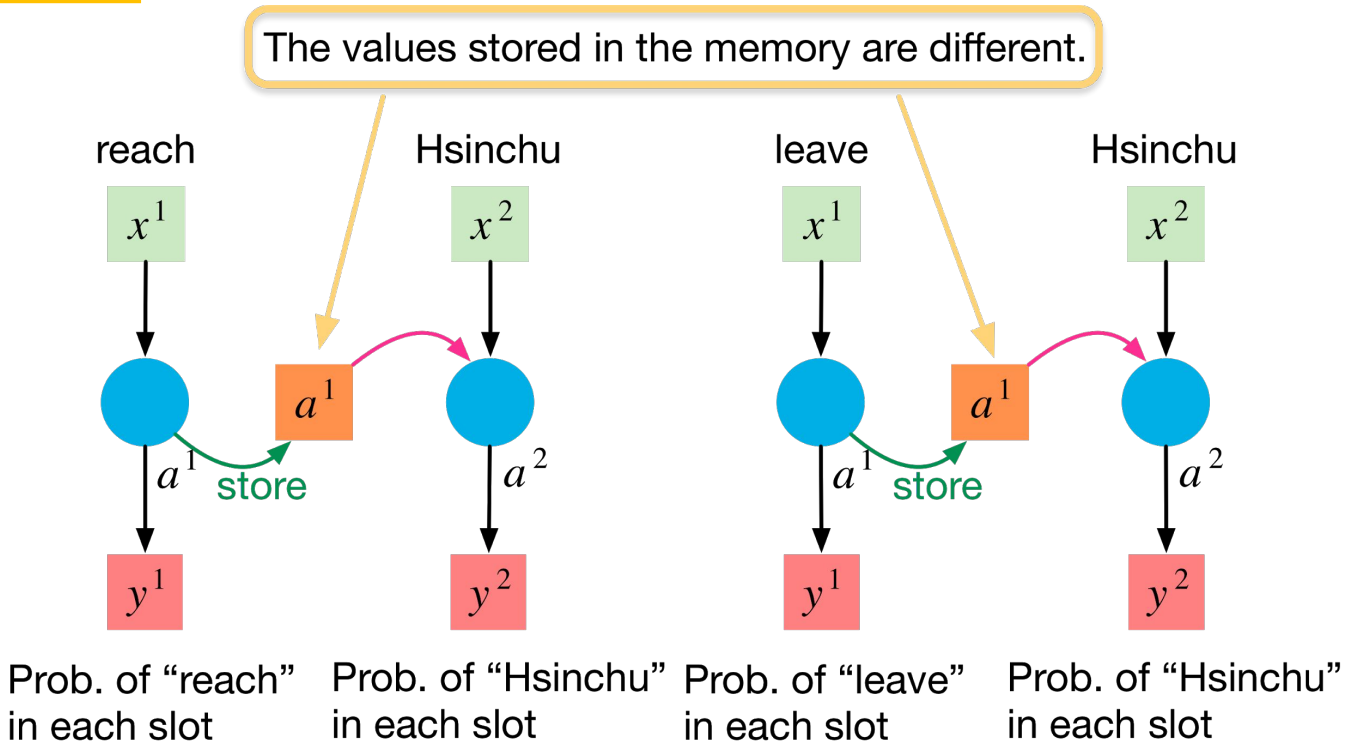
use same network again and again

Probability of "reach" in each slot

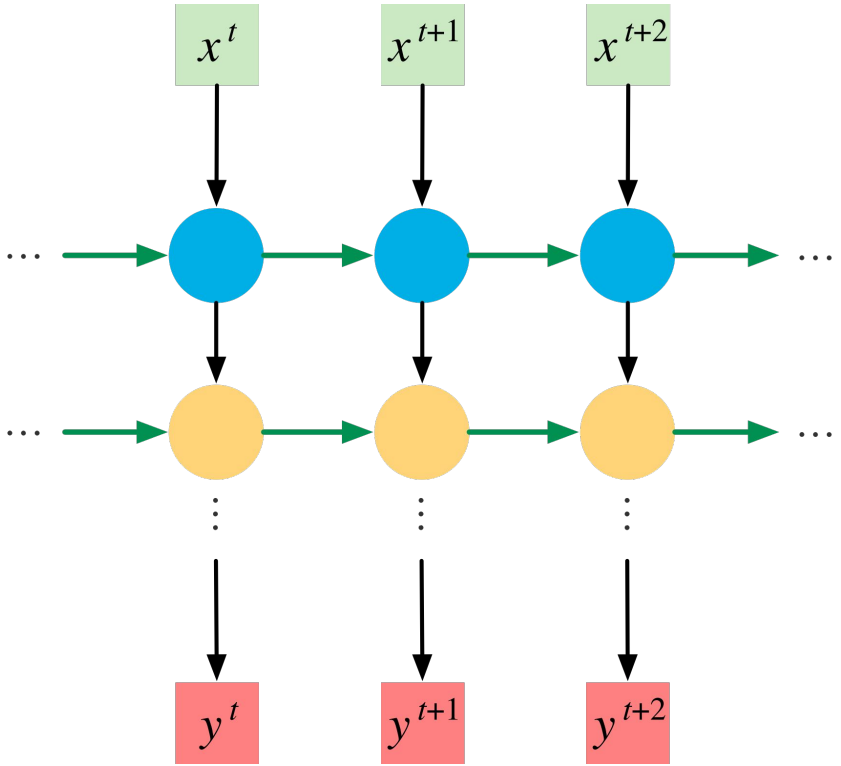
Probability of "Hsinchu" in each slot

Probability of "on" in each slot

Recurrent Neural Network (cont'd)



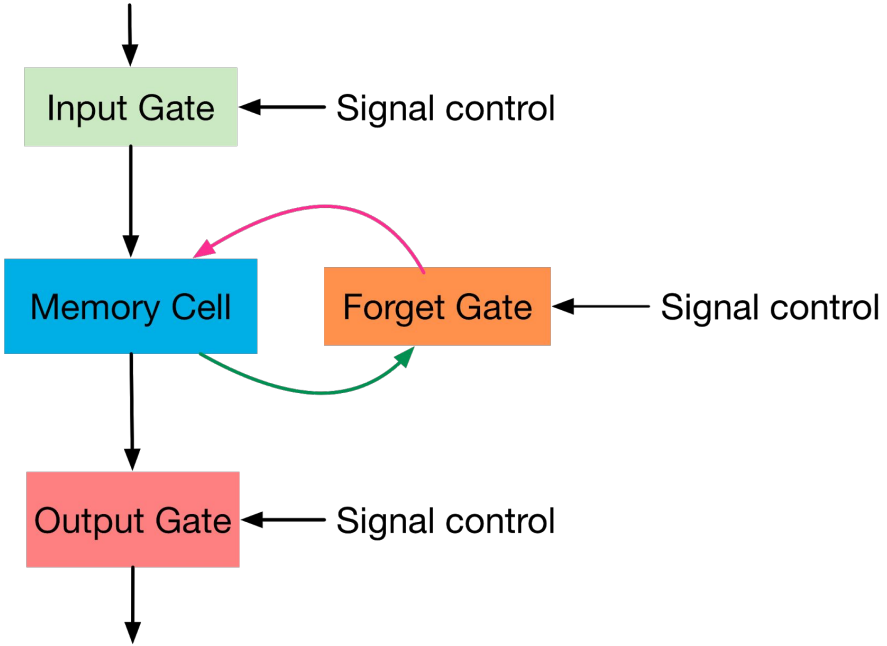
Recurrent Neural Network: Deeper!



Long Short-Term Memory (LSTM)

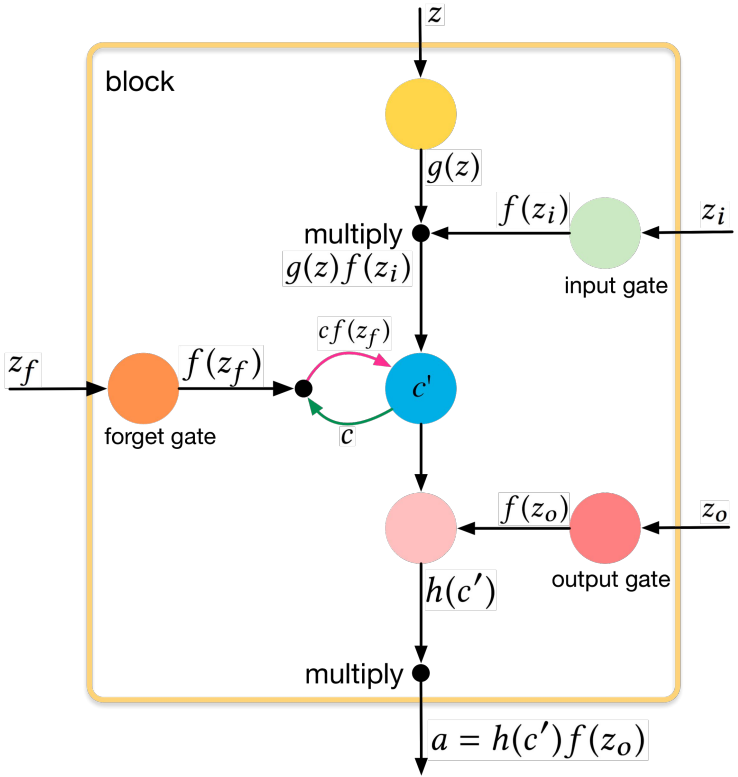
Long Short-Term Memory: Concept

special neuron:
4 inputs, 1 output

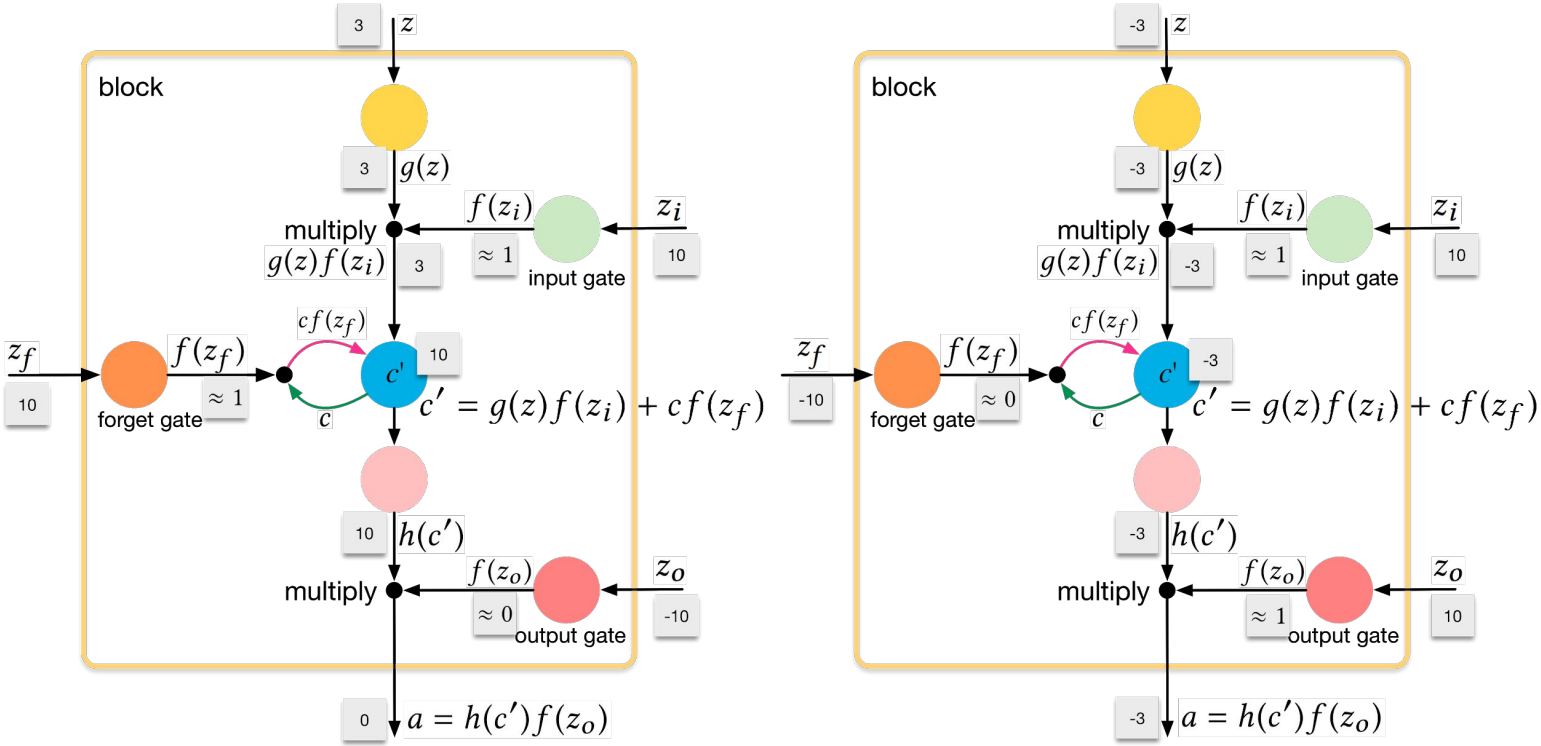


Long Short-Term Memory: Details

- Activation function f of gates is usually a sigmoid function
  - Between 0 and 1
- Mimic open and closed gates

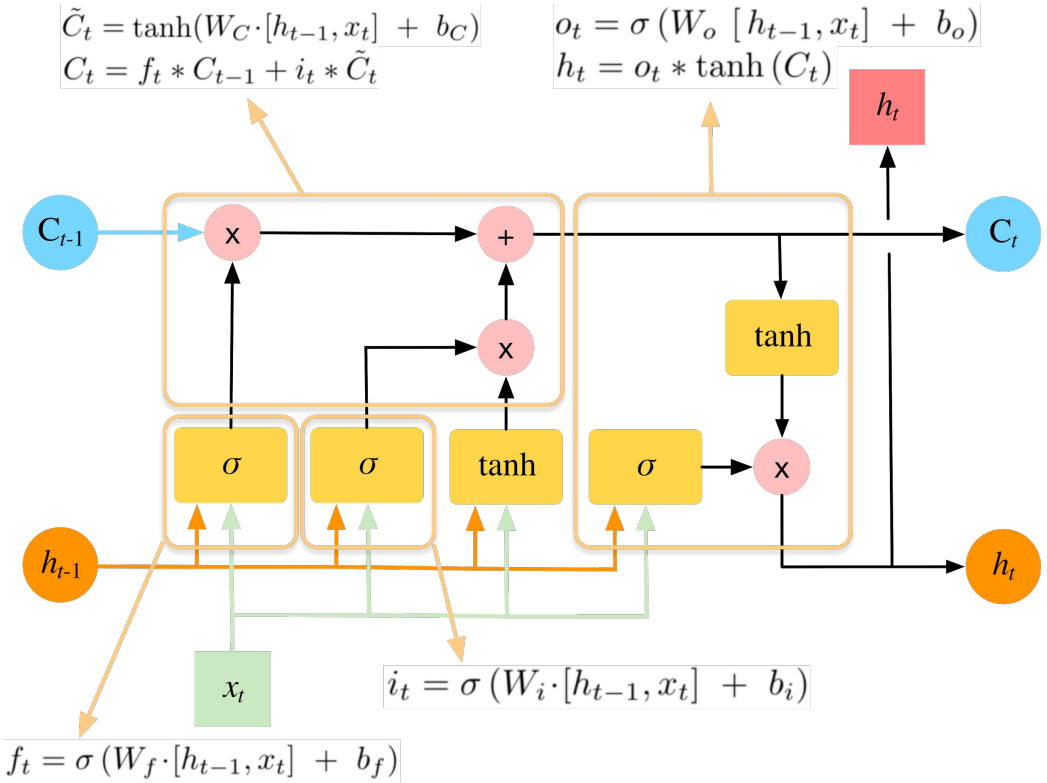


Long Short-Term Memory: Simple Example



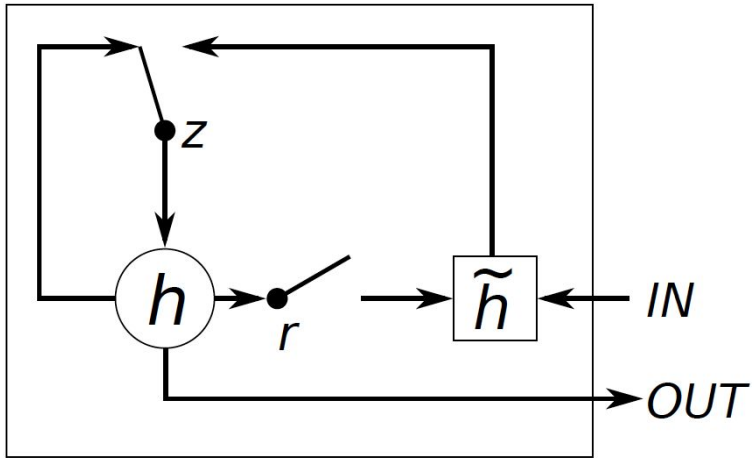
Long Short-Term Memory: Implementation

- An LSTM cell includes:
  - f (**f**orget)
  - i (**i**nput)
  - o (**o**utput)
  - a memory **C**
- Notice the σ on gates



Gated Recurrent Unit (GRU)

- Similar with LSTM
  - But simpler (only 2 gates, z and r)
- GRU is also designed to fight vanishing gradient



[Source: Chung et al., arXiv'14]

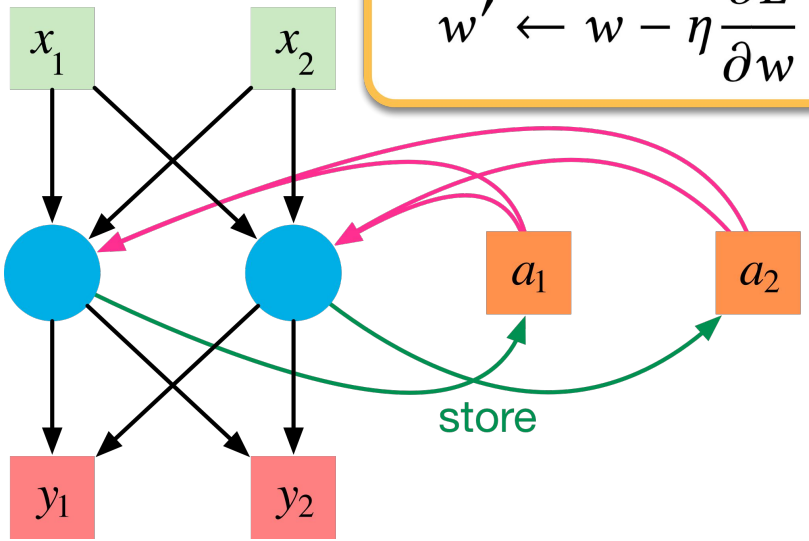
LSTM v.s. GRU: Which is Better?

- No clear winner
- Use sample/parameter ratio to choose
- If you have A LOT of data, use LSTM.

How to Learn RNN (LSTM, GRU)?

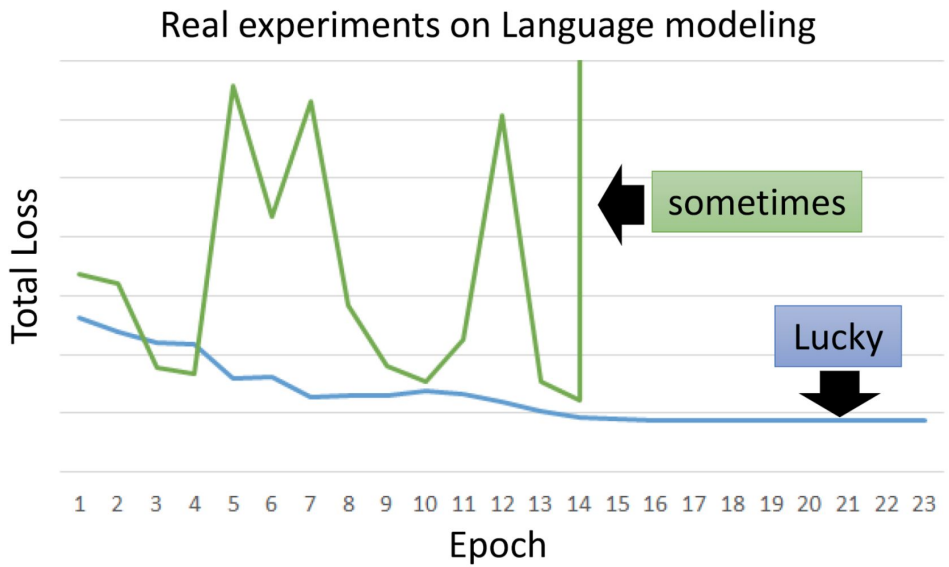
Backpropagation through time (BPTT)

$$w' \leftarrow w - \eta \frac{\partial L}{\partial w}$$



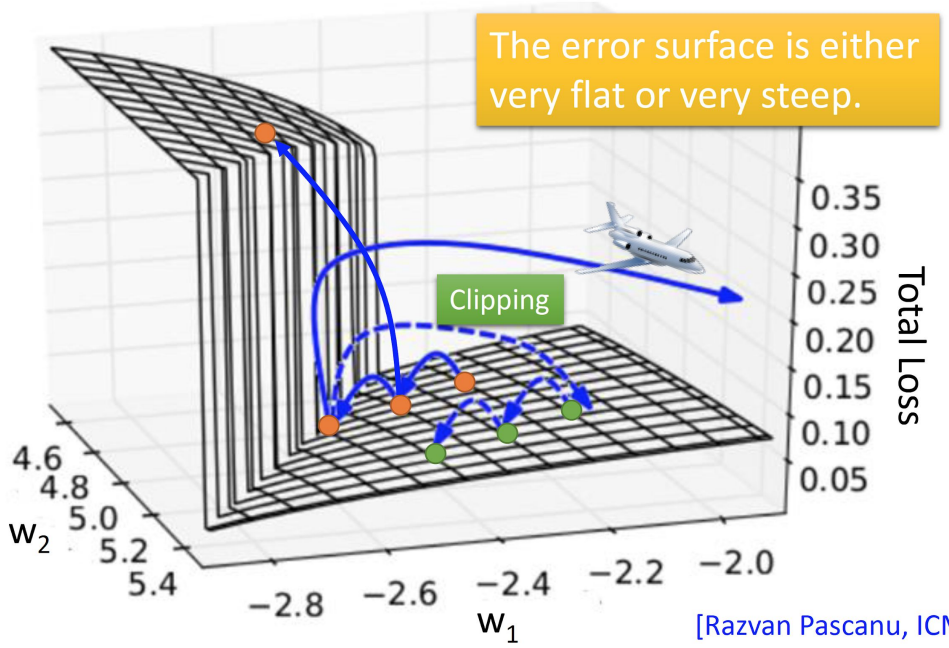
Learning RNN is Difficult

- RNN-based network is not always easy to learn



[Source: Hung-Yi Lee, http://www.slideshare.net/tw_dsconf/ss-62245351]

Learning RNN: Steep Loss Surface

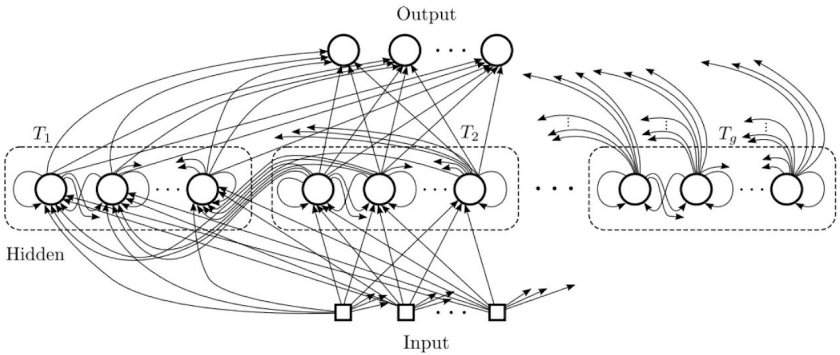


[Razvan Pascanu, ICML'13]

[Source: Hung-Yi Lee, http://www.slideshare.net/tw_dsconf/ss-62245351]

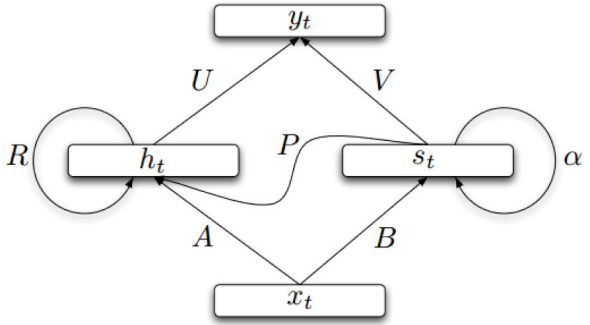
Learning RNN: Techniques

Clockwise RNN



[Jan Koutnik, JMLR'14]

Structurally Constrained Recurrent Network (SCRN)



[Tomas Mikolov, ICLR'15]

[Source: Hung-Yi Lee, http://www.slideshare.net/tw_dsconf/ss-62245351]

Linear Regression

Tensorboard

- How to invoke?
  - tensorboard --logdir="/YOUR/DIRECTORY/" --port=3101
- Add metrics of interests to “tf.summary”
- E.g., tf.summary.scalar('mean', mean)

Questions?

BITTIGER

The Lifelong Learning Platform of Silicon Valley