View this deck on Google Slides

Seriesly, Neural Networks

An Intro to RNNs, then LSTMs

Agenda

1. Administrative Meanderings (AM)
    a. RSO Update
    b. Coordinator Elections
    c. Future of SIGAI
2. Let's Review: Artificial & Convolutional Neural Nets
3. What is a Recurrent Neural Network?
4. Backpropagation Through Time
5. Long Short-Term Memory Networks

AM: RSO Update

- Constitution is written and submitted.
  - Currently deciding on where to publish it, but you'll have access to it by the end of the semester. :D
- We're on track to being an RSO in the fall semester.
  - Nothing should change for you, as a member; however, we'll let you know if anything does.

AM: Coordinator Elections

- Two stages:
  - **(Apr 05)** Pitches – RL P1 lecture
  - **(Apr 19)** Elections – MDP workshop
- Pitches:
  - 2-3 minute spiel about…
    - Your background
    - What part(s) of AI are you interested in?
    - What are some of the lecture/workshop content you'd like to bring?
- Coordinator meeting after finals **(May 02)**.
- We're looking to have 8-9 coordinators.

AM: Coordinator Roles (1-year commitment)

Coordinator:

- Must lecture at least 1 entire unit
- Expect to 10-15 hr commitment/week – may change if we get a graphic designer
- Preferable that you either do AI-related research, or work in AI
    - Overall, it's paramount that you're passionate about education and the field

Director:

- Same responsibilities as Coordinator, **plus:**
- You've previously been a Coordinator, *and* not been Director in the past year
- You're able to mentor Coordinators (e.g. you can pick up a lecture/workshop and be comfortable lecturing)
- You're simply a Coordinator who can occasionally `sudo`, and the external point of contact.

AM: Future of SIGAI

- We've secured access to Stokes! (The GPU cluster.)
  - This means more realistic workshops, among other things
- In the fall, we're looking to have two foci:
  - Industry (topics relating to Kaggle competitions, and generally data science)
  - Research (more advanced topics that haven't quite made it out labs)
- This is still somewhat open-ended, and will be a conversation that happens with new coordinators.

Agenda

1. Administrative Meanderings (AM)
   a. RSO Update
   b. Coordinator Elections
   c. Future of SIGAI
2. Let's Review: Artificial & Convolutional Neural Nets
3. What is a Recurrent Neural Network?
4. Backpropagation Through Time
5. Long Short-Term Memory Networks

Let's Review: ANNs and CNNs

- Great for independent data. *What does this mean?*
- **ANNs:**
  - Great for classification
  - Massage inputs to extract data from signal at each node
- **CNNs:**
  - Great feature extractors
  - Leverage convolution to enhance feature extraction
  - Attempt to keep only important values by pooling

Let's Review: Some Drawbacks

- We *assume* ***fixed*** input
- We *assume* ***independence*** of inputs and outputs
  - There is **no** sequential aspect to the data

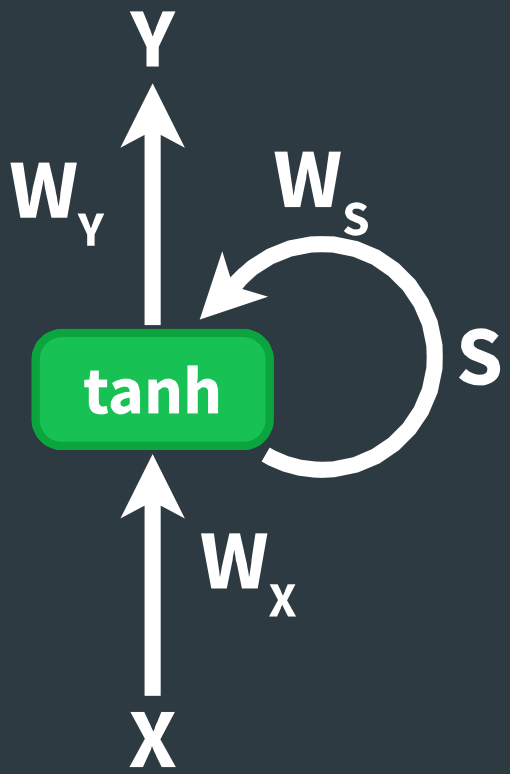
Let's Review: Some Drawbacks

- We *assume* ***fixed*** input
- We *assume* ***independence*** of inputs and outputs
  - There is **no** sequential aspect to the data

Agenda

1. Administrative Meanderings (AM)
   a. RSO Update
   b. Coordinator Elections
   c. Future of SIGAI
2. Let's Review: Artificial & Convolutional Neural Nets
3. What is a Recurrent Neural Network?
4. Backpropagation Through Time
5. Long Short-Term Memory Networks

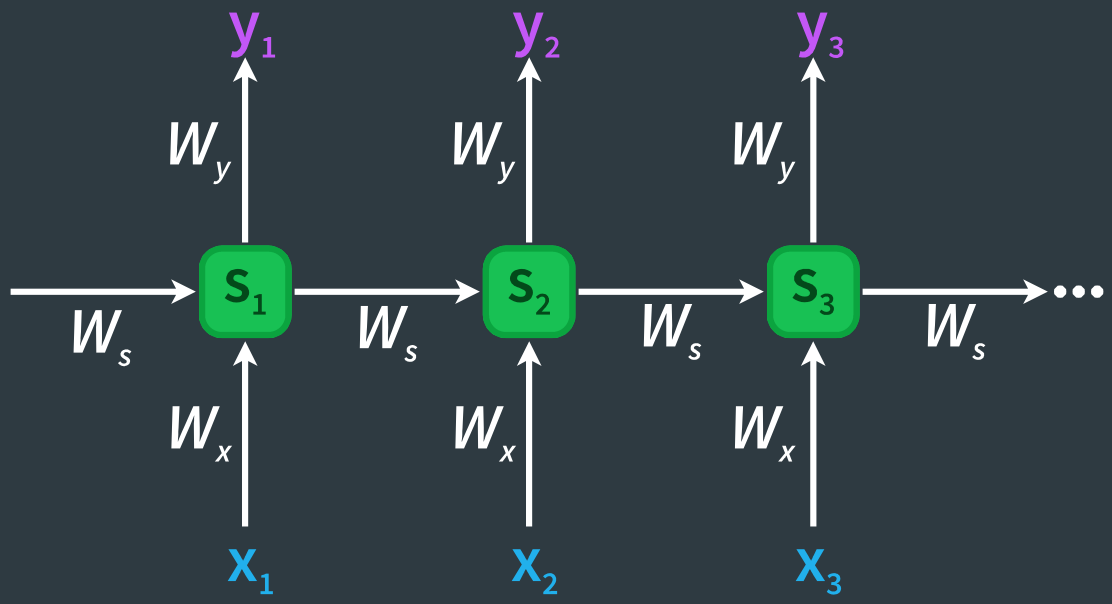
What is a Recurrent Neural Network? (RNN)



- X is our input
- Y is our output
- S is a hidden layer called “State”
- W_{...} is the weight matrix for each input/output

- We’ll expand this on the next slide

Expanding the RNN

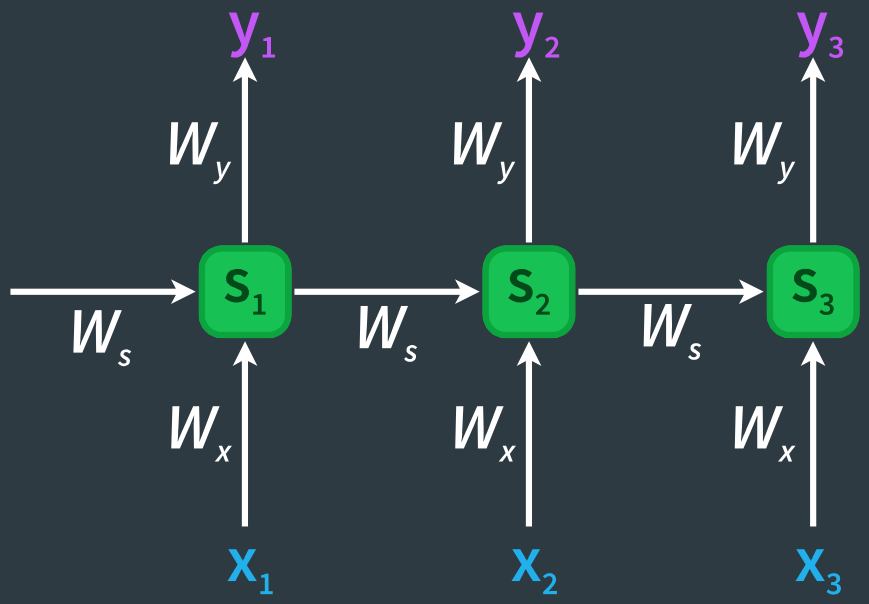


Observations?

Agenda

1. Administrative Meanderings (AM)
   a. RSO Update
   b. Coordinator Elections
   c. Future of SIGAI
2. Let's Review: Artificial & Convolutional Neural Nets
3. What is a Recurrent Neural Network?
4. Backpropagation Through Time
5. Long Short-Term Memory Networks

Let's Try Backprop



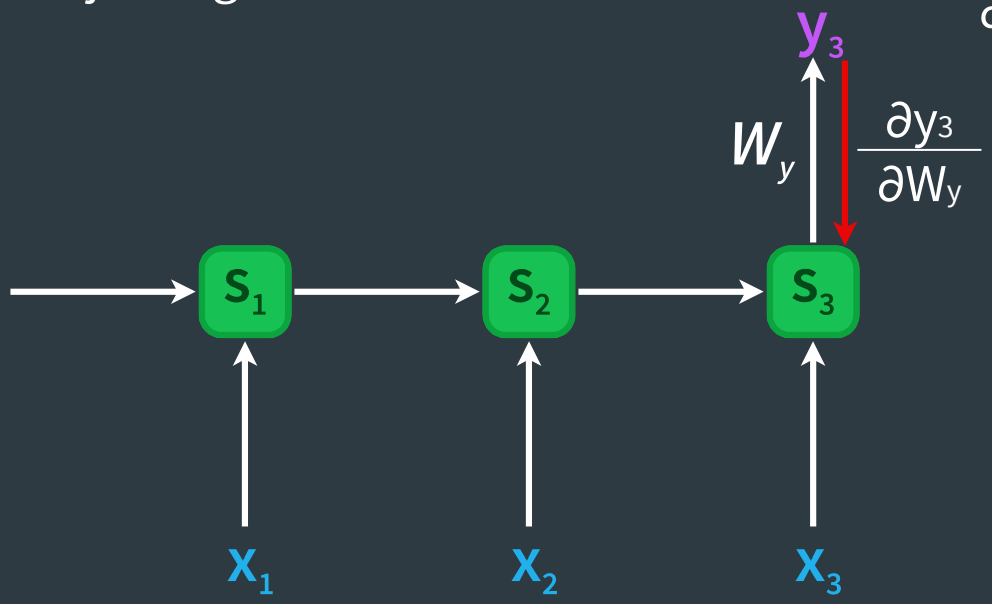
- We have 3 matrices to update.
- Taking the simplest case…
  - Let **t=3**

Let's Try Backprop

At time, **t=3**

Adjusting: W_y

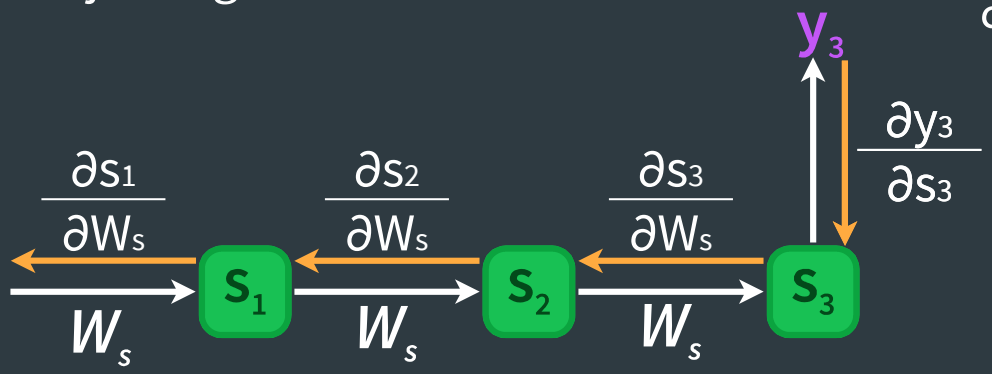
$$\frac{\partial E_3}{\partial W_y} = \frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial W_y}$$



Let's Try Backprop

At time, **t=3**

Adjusting: W_s



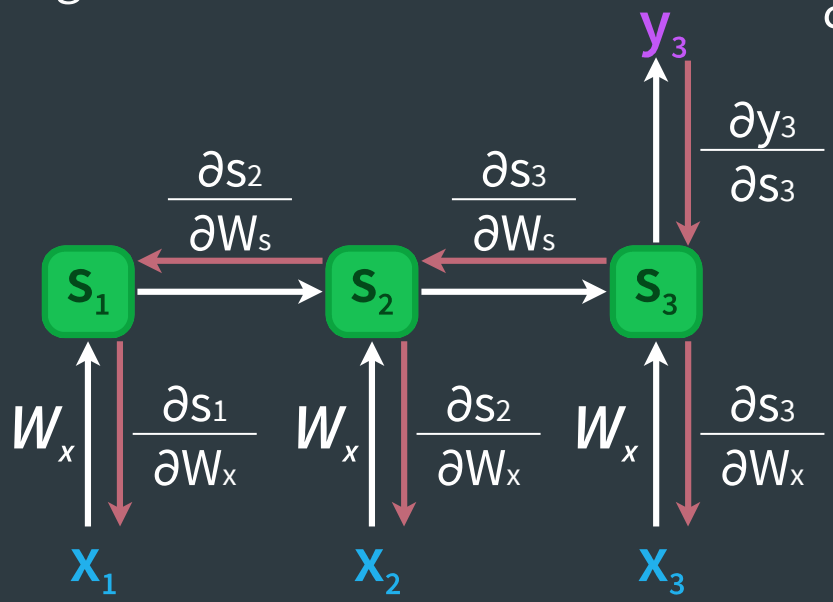
$$\frac{\partial E_3}{\partial W_s} = \frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_s} \cdot \frac{\partial s_3}{\partial W_s} +$$
$$\frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_s} \cdot \frac{\partial s_3}{\partial s_2} \cdot \frac{\partial s_2}{\partial W_s} +$$
$$\frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_s} \cdot \frac{\partial s_3}{\partial s_2} \cdot \frac{\partial s_2}{\partial s_1} \cdot \frac{\partial s_1}{\partial W_s}$$

This equation generalizes to a summation, we won't derive it though.[1]

Let's Try Backprop

At time, **t=3**

Adjusting: W_x



$$\frac{\partial E_3}{\partial W_x} = \frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial W_x} +$$
$$\frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial s_2} \cdot \frac{\partial s_2}{\partial W_x} +$$
$$\frac{\partial E_3}{\partial y_3} \cdot \frac{\partial y_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial s_2} \cdot \frac{\partial s_2}{\partial s_1} \cdot \frac{\partial s_1}{\partial W_x}$$

This equation generalizes to a summation, we won't derive it though.[1]

Some Notes on BPTT

- BPTT becomes extremely complex when you layer RNNs
  - You have to take all possible paths to the states and inputs
- We find two problems with BPTT
  - The Vanishing Gradient Problem[2]
  - The Exploding Gradient Problem[3]

- Information recall is difficult with only a single state, or even multiple in parallel (this ought to become clearer once we understand how LSTMs work)

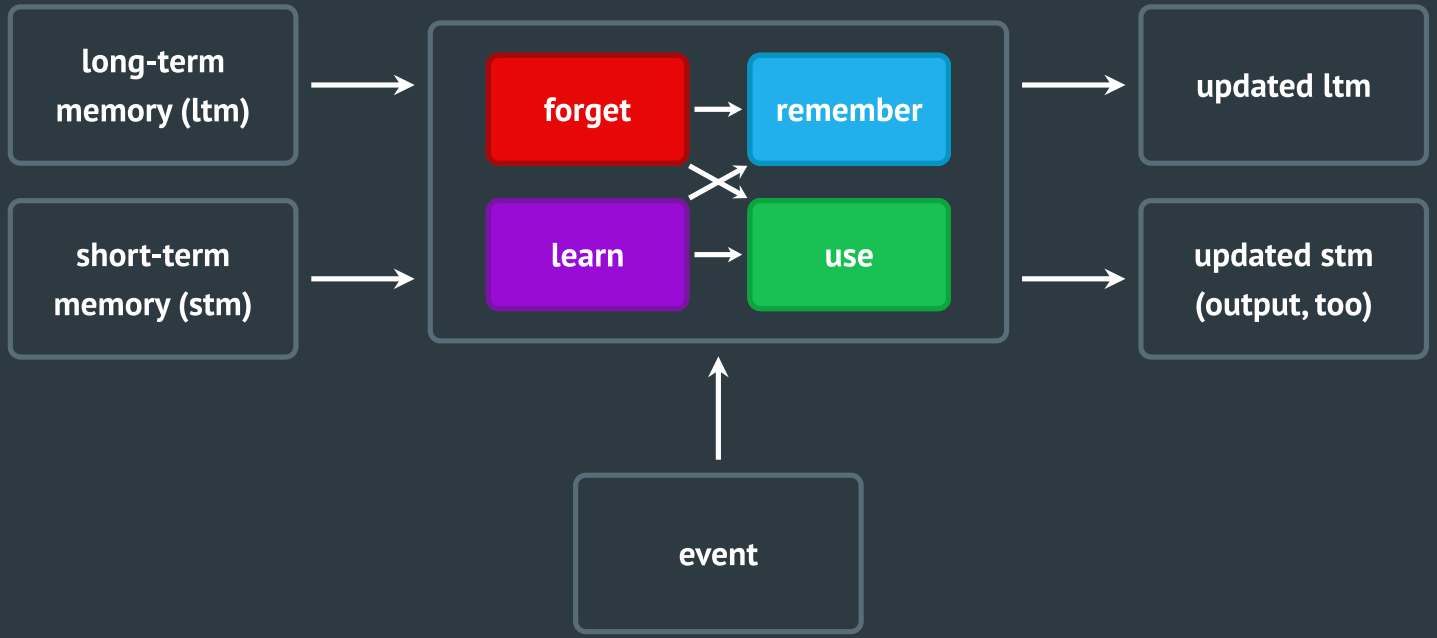
Agenda

1. Administrative Meanderings (AM)
   a. RSO Update
   b. Coordinator Elections
   c. Future of SIGAI
2. Let's Review: Artificial & Convolutional Neural Nets
3. What is a Recurrent Neural Network?
4. Backpropagation Through Time
5. Long Short-Term Memory Networks

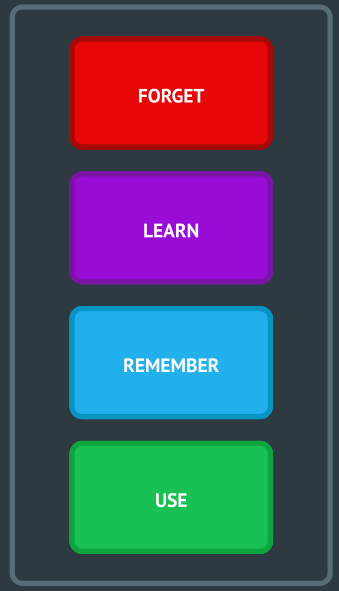
Long Short-Term Memory Networks (LSTMs)

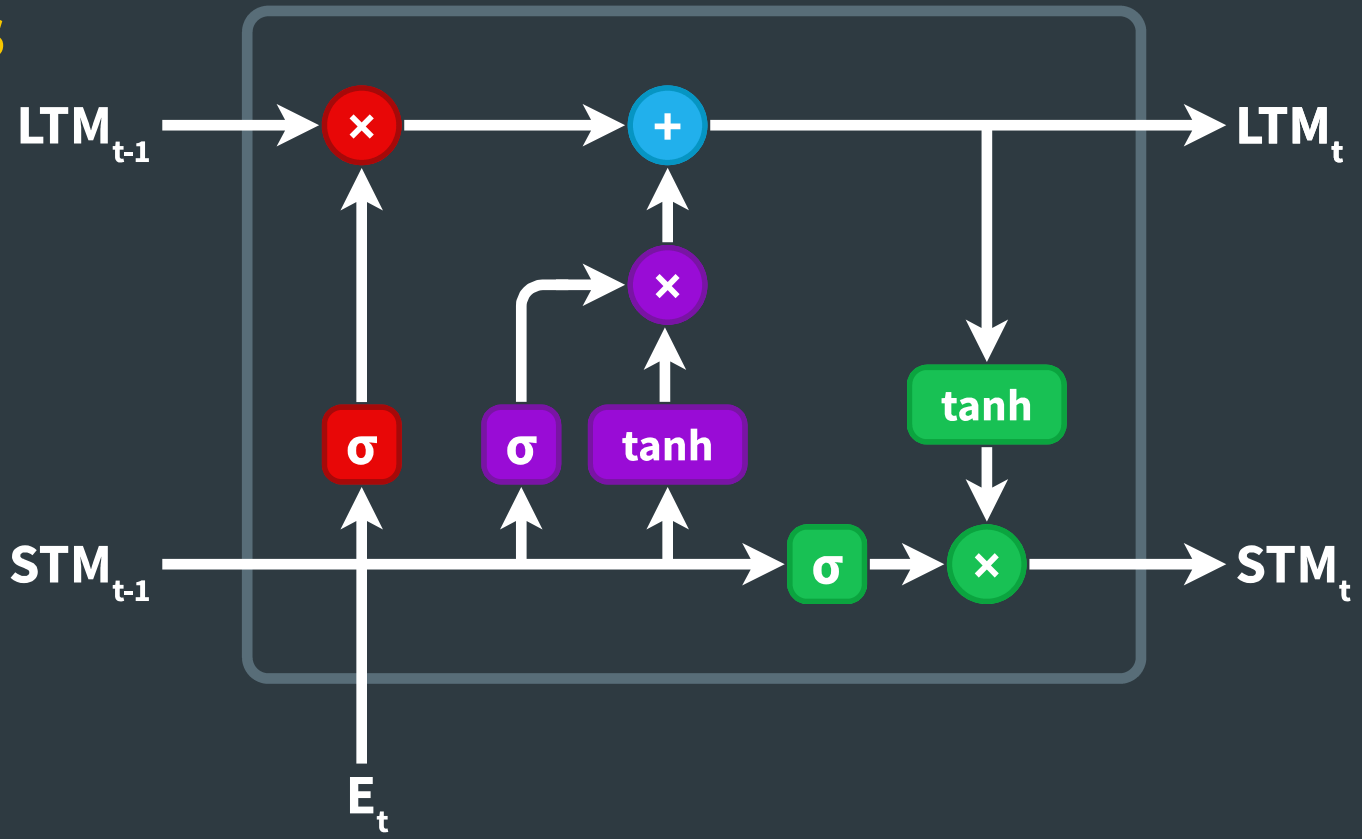
- Proposed in 1997[4]
  - Still heavily used today, as well as other gated architectures[5][6][7]
- Overcome the long-term dependency issue of RNNs
- Have an interesting architecture
  - Each “cell” is composed of 4 hidden layers
  - Layers act as “gates” to control the flow of information

Information Flow

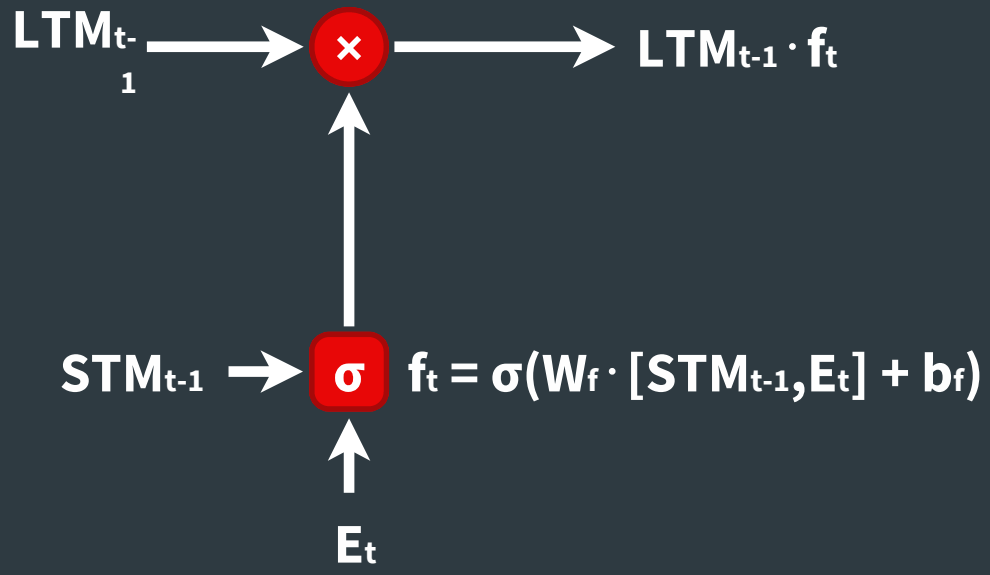


Cell Internals

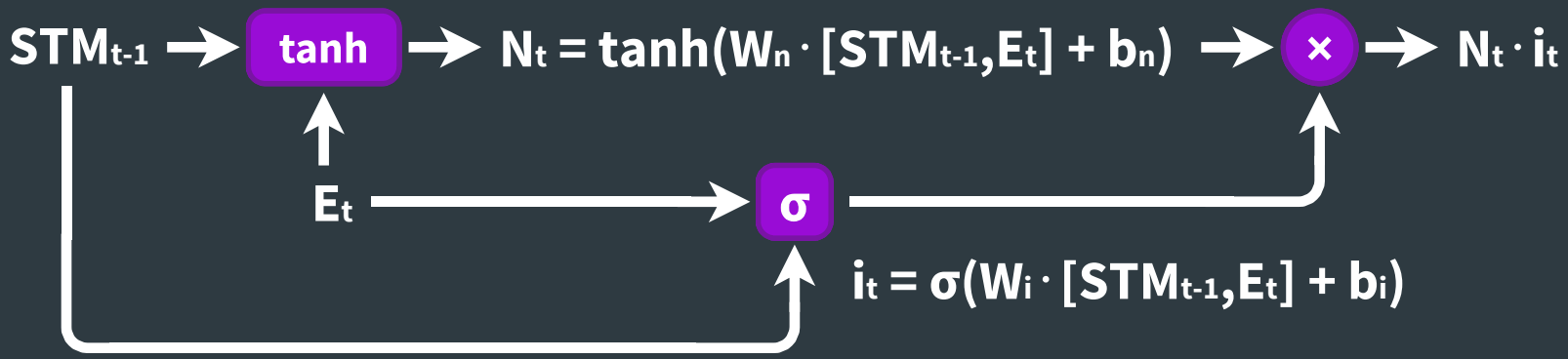




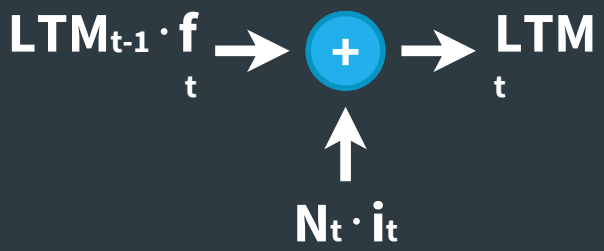
The Forget Gate



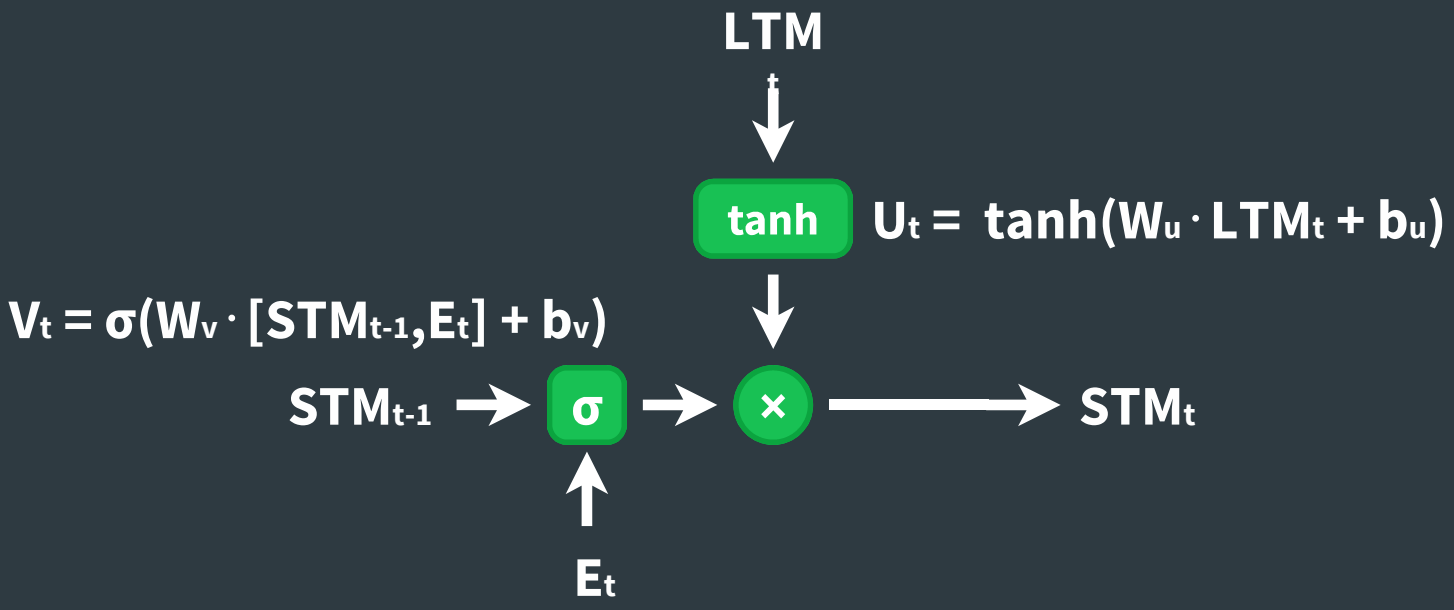
The Learn Gate



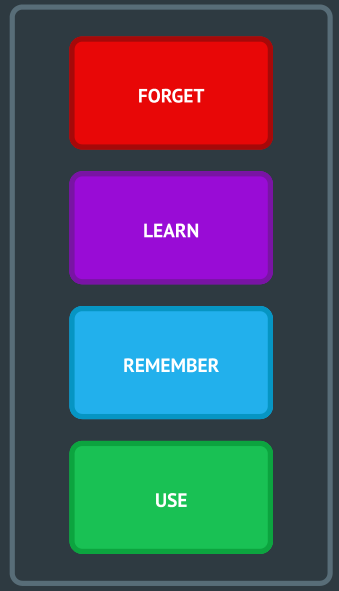
The Remember Gate

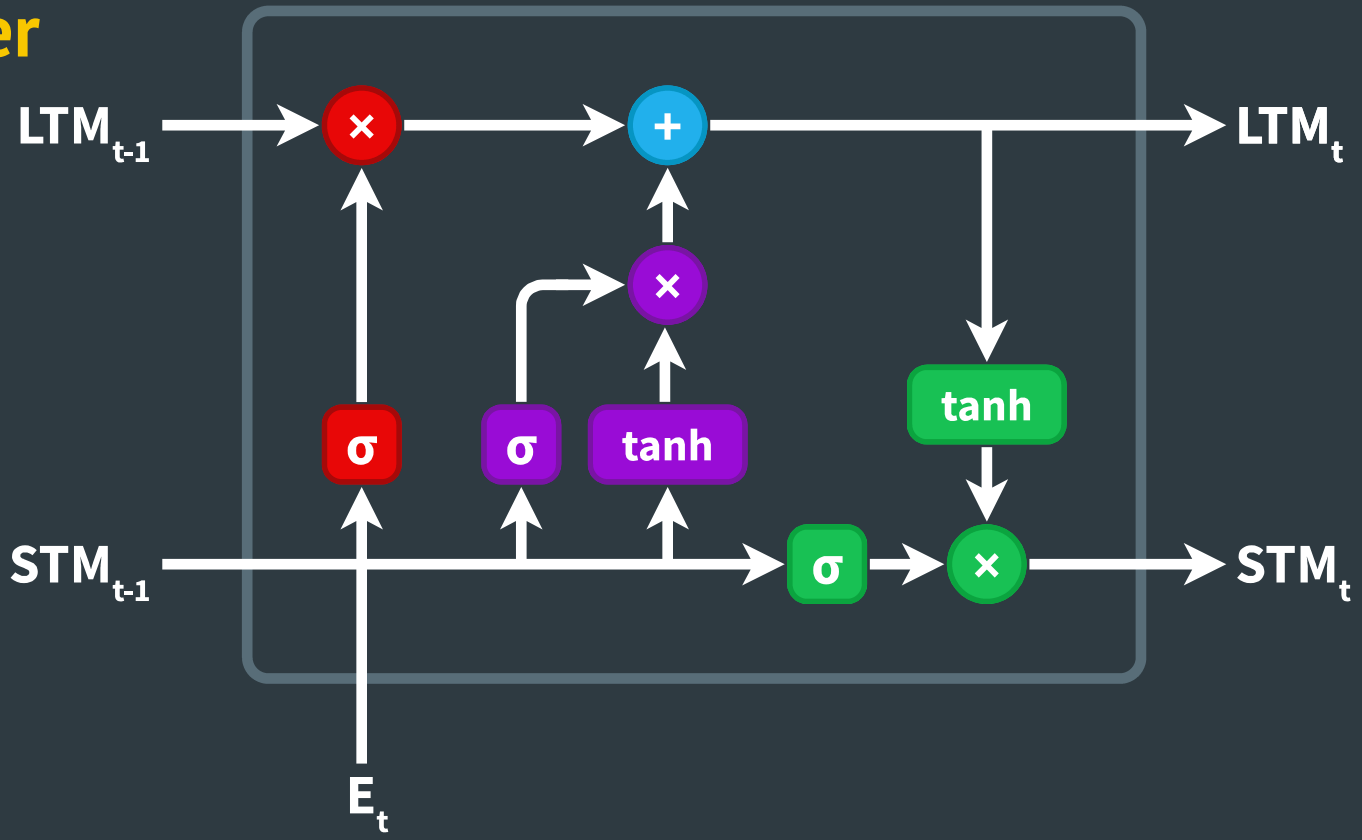


The Use Gate


LTM
t
tanh
$U_t = \tanh(W_u \cdot LTM_t + b_u)$
$V_t = \sigma(W_v \cdot [STM_{t-1}, E_t] + b_v)$
STM_{t-1}
σ
×
STM_t
E_t

Back Together





Extras & Looking Ahead

[1] Summations to calculate input and state weight matrix updates

[2] What is the vanishing gradient problem?

(Albeit Quora, this is a pretty good answer.)

[3] A Gentle Introduction to Exploding Gradients in Neural Networks

[4] Long Short-Term Memory

(original paper)

[5] Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation

(Intro'd Gated Recurrent Units, alternative to LSTMs, comparable in capability)

[6] LSTM: A Search Space Odyssey

(Comparison of various gated architectures)

[7] Grid Long Short-Term Memory

Questions?