Welcome to

Machine Learning Mondays

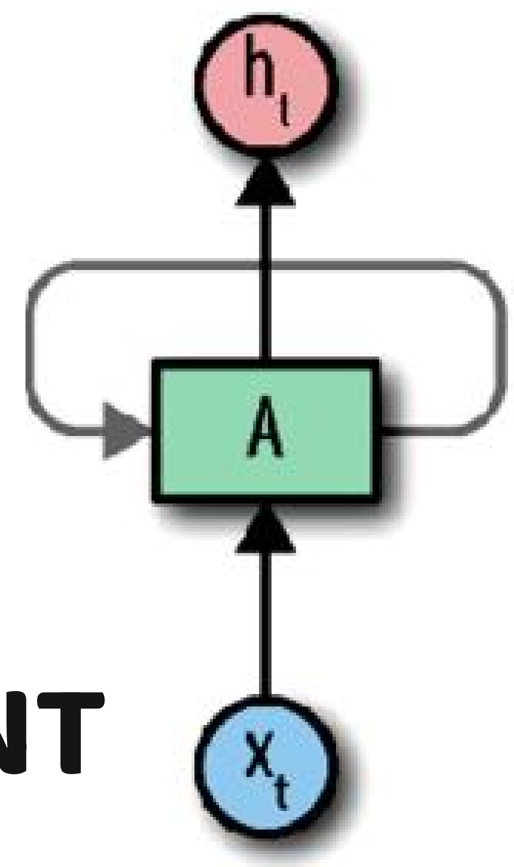
RNN

How Learning Happens

$$\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}(f(x))\right)\right)\right)\right)\right)$$


Arun Roy
@aktheroy

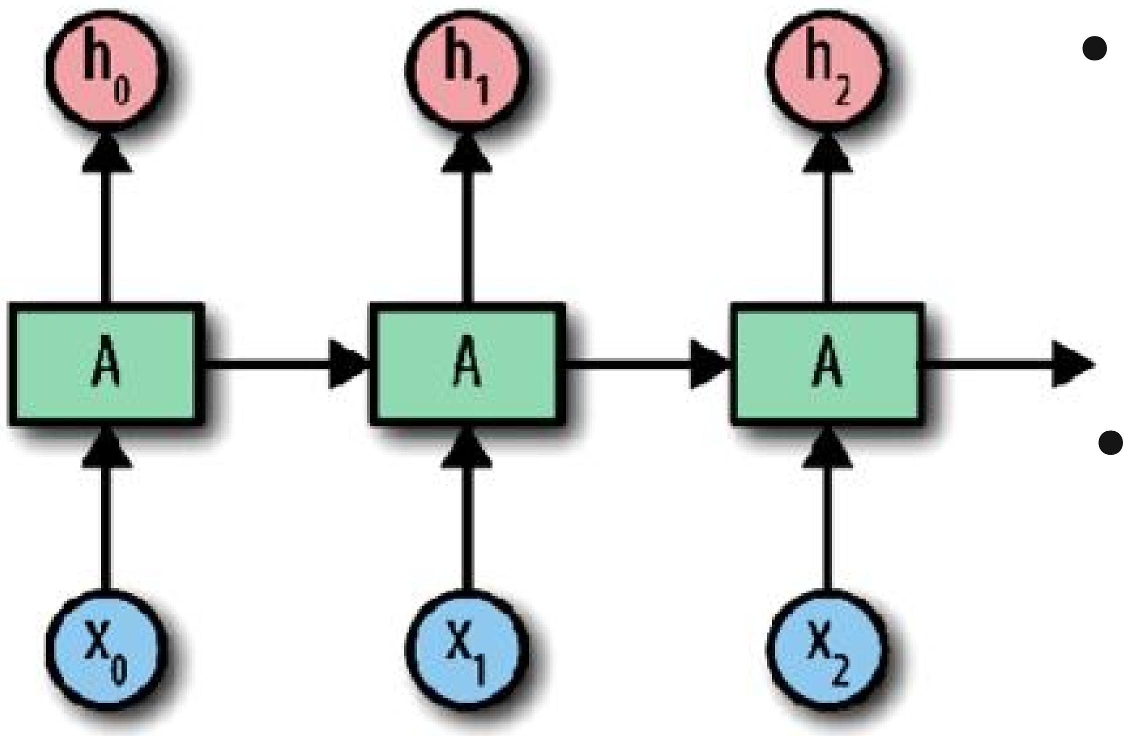
The Core Idea
Persistence of State



A compact, looped view.

- The cell "A" reads an input (X_t) with previous state to produce a new hidden state (h_t)
- This loop represents the **RECURRENT**

it's the same network unrolled through time



- At time t = 0, it processes X_0 and creates hidden state h_1.
- At time t = 1, it processes X_1 and memory from the previous step h_1 to create h_2

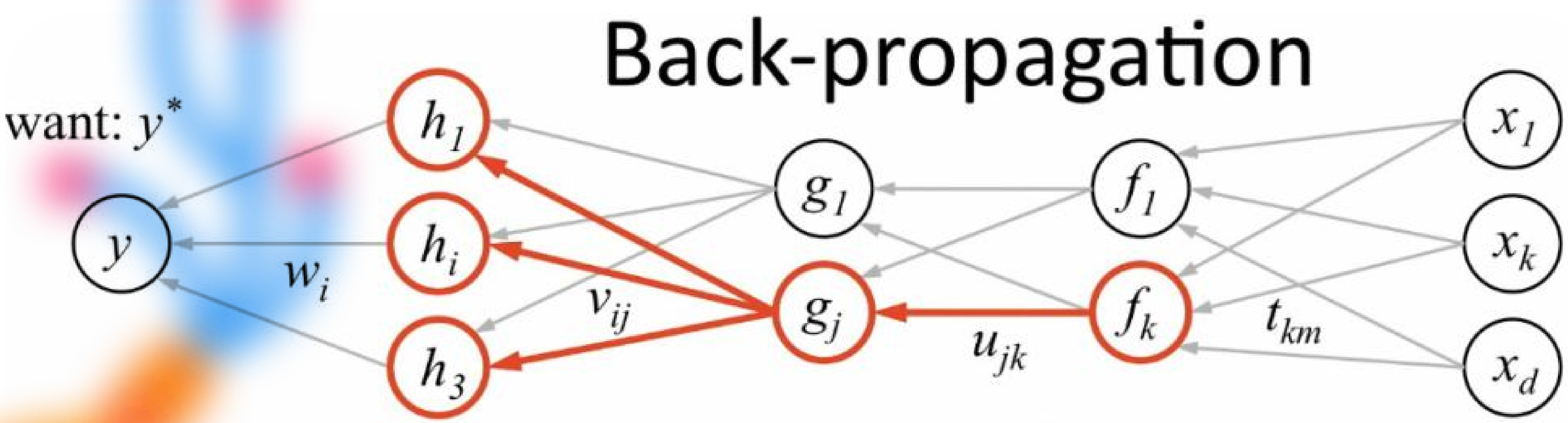
This process repeats, building a chain of memory.

Arun Roy
@aktheroy

How Learning Happens
Back Propagation Through Time (BPTT)

How does our unrolled network learn?

1. Calculate the error at the final output (h_4).
2. Travel backwards through the unrolled network, step-by-step.
3. At each step t, calculate the gradient (the "blame") for the error with respect to the weights in A.
4. Because the same weights W are used at every step, the total update is the sum of the gradients from all time steps.



1. receive new observation $\boldsymbol{x} = [x_1 \dots x_d]$ and target y^*
2. **feed forward:** for each unit g_j in each layer $1 \dots L$
   compute g_j based on units f_k from previous layer: $g_j = \sigma\left(u_{j0} + \sum_k u_{jk} f_k\right)$
3. get prediction y and error $(y - y^*)$
4. **back-propagate error:** for each unit g_j in each layer $L \dots 1$

Arun Roy
@aktheroy

The Fundamental Challenge
Vanishing Gradients

BPTT has a major flaw. To update the weights for the early steps (X_0, X_1), the error signal from h_4 must travel a long way back.

This signal is multiplied by the same weight matrix W over and over again.

If these weights are small ($|w| < 1$), the gradient shrinks exponentially and vanishes.

$$\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}\left(\frac{d}{dx}(f(x))\right)\right)\right)\right)\right)$$

Arun Roy
@aktheroy

Follow for more

Arun Roy
@aktheroy

https://aktheroy.com/

https://github.com/aktheroy

https://www.youtube.com/@aktheroy