# Recurrent Neural Networks

Geena Kim

University of Colorado **Boulder**

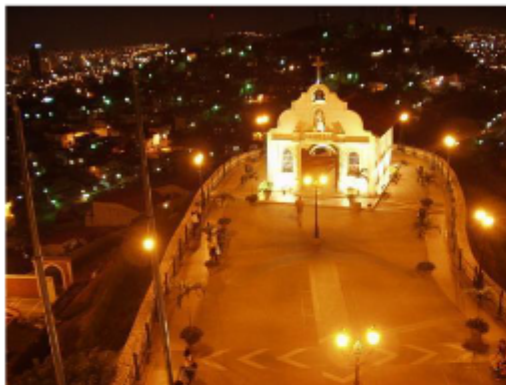# Dealing with sequential data

What is sequential data

- Sequence of numbers in time:
    - Stock price
    - Earthquake sensor data
    - EEG sensor data

- Sequence of words: texts
- Sound: speech, sound
- Image: videos

What tasks can you do with sequential data?

# Example Tasks

Text generation (Image captioning)



Retr.

1. Top view of the lights of a city at night, with a well-illuminated square in front of a church in the foreground;
2. People on the stairs in front of an illuminated cathedral with two towers at night;

1. Tourists are sitting at a long table with beer bottles on it in a rather dark restaurant and are raising their bierglaeser;
2. Tourists are sitting at a long table with a white table-cloth in a somewhat dark restaurant;

1. A dry landscape with light brown grass and green shrubs and trees in the foreground and large reddish-brown rocks and a blue sky in the background;
2. A few bushes at the bottom and a clear sky in the background;
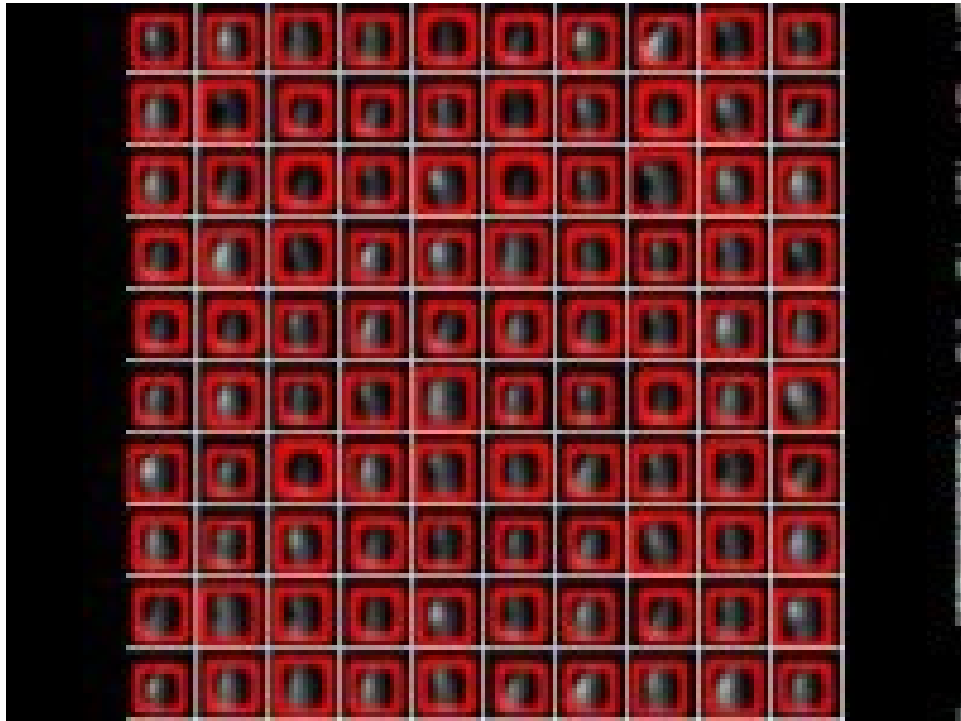
1. Group picture of nine tourists and one local on a grey rock with a lake in the background;
2. Five people are standing and four are squatting on a brown rock in the foreground;

Gen.

A square with burning street lamps and a street in the foreground;

Tourists are sitting at a long table with a white table cloth and are eating;

A dry landscape with green trees and bushes and light brown grass in the foreground and reddish-brown round rock domes and a blue sky in the background;

A blue sky in the background;

https://arxiv.org/abs/1412.6632

# Examples Tasks

How about hand-writing or drawing?



sketch-rnn mosquito predictor.

clear drawing | mosquito | random | predict

https://youtu.be/Zt-7MI9eKEo
http://arxiv.org/abs/1502.04623

https://magenta.tensorflow.org/sketch-rnn-demo

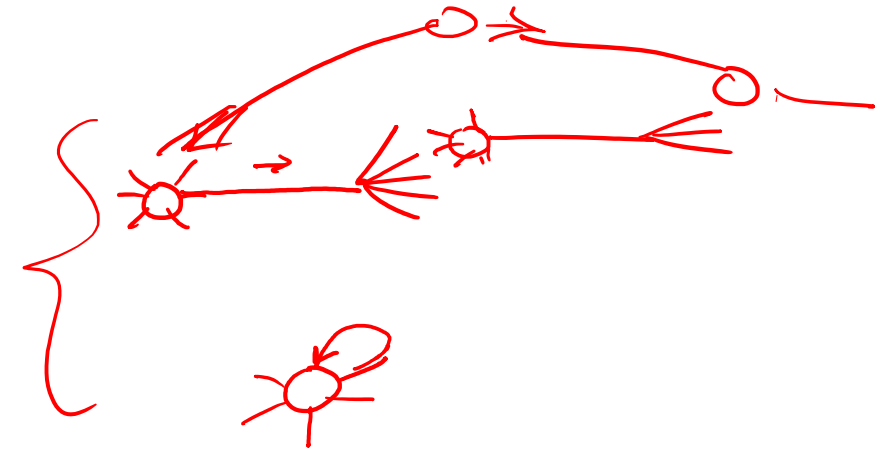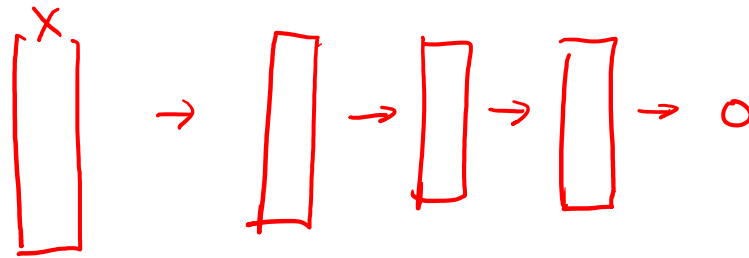https://ai.googleblog.com/2017/04/teaching-machines-to-draw.html
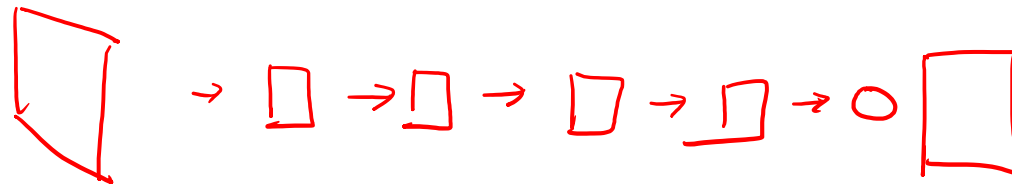
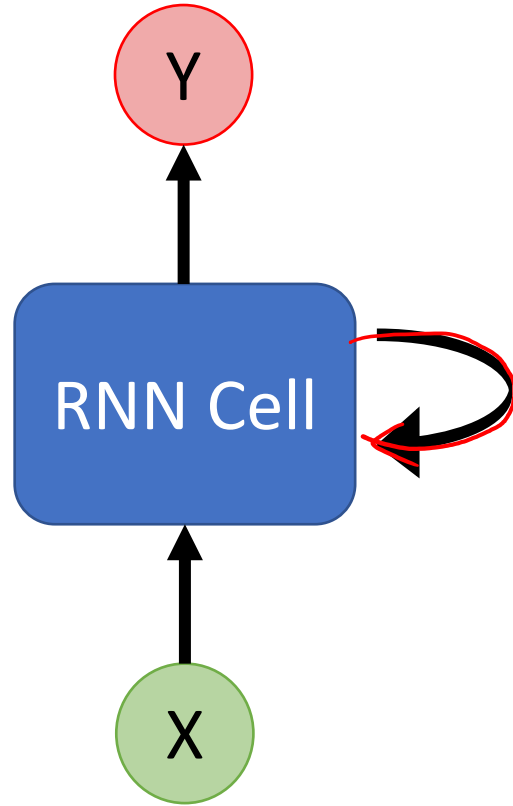# Example Tasks

# Stories so far

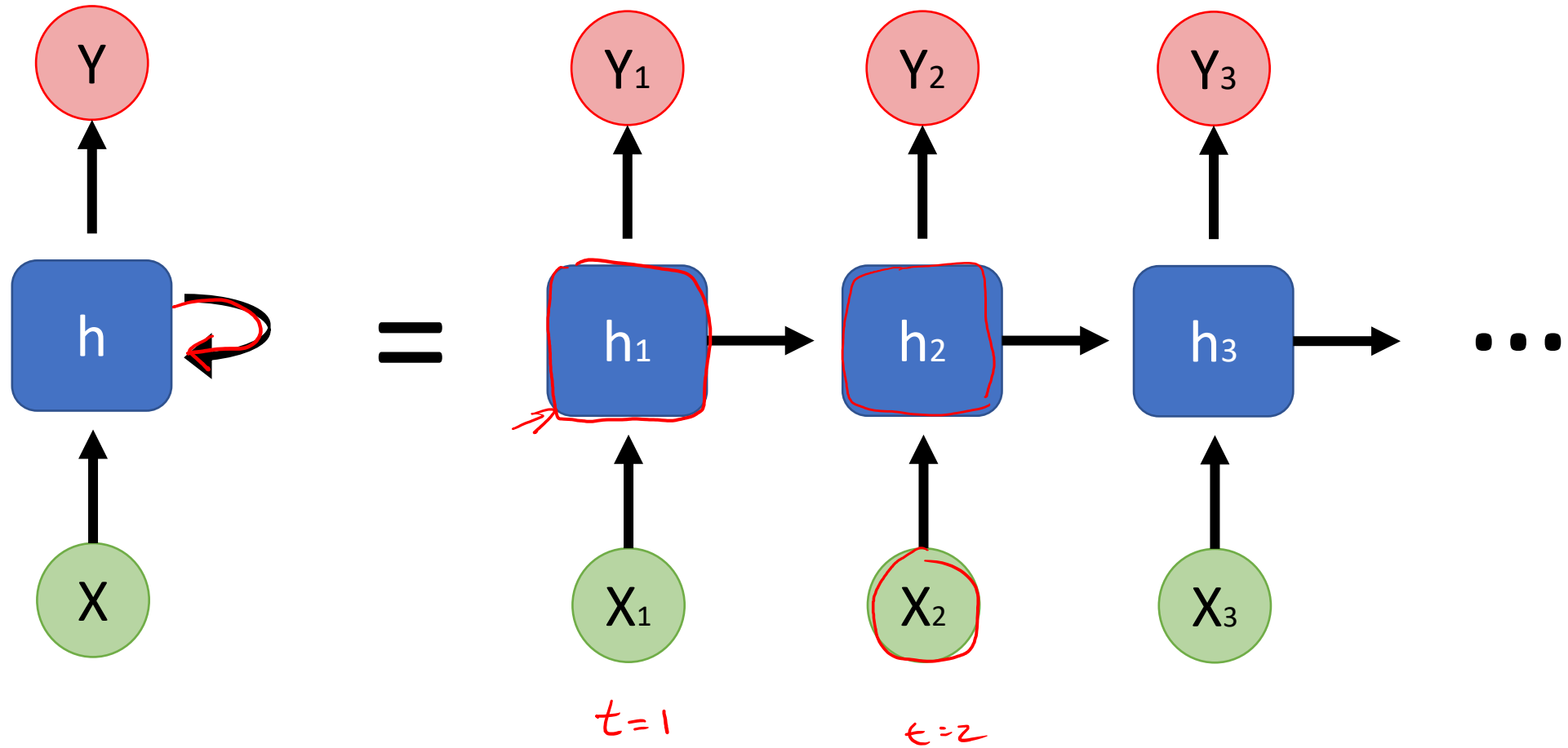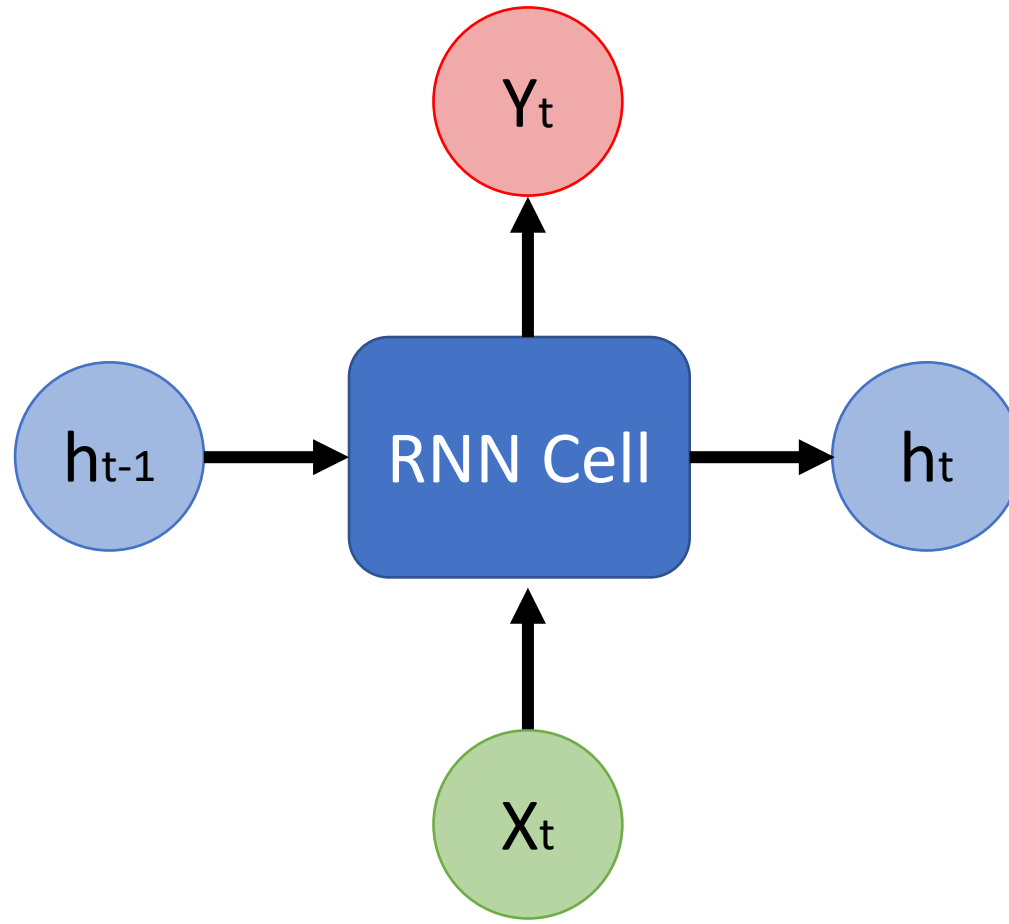Feed forward neural networks

(Multi-Layer) Perceptrons

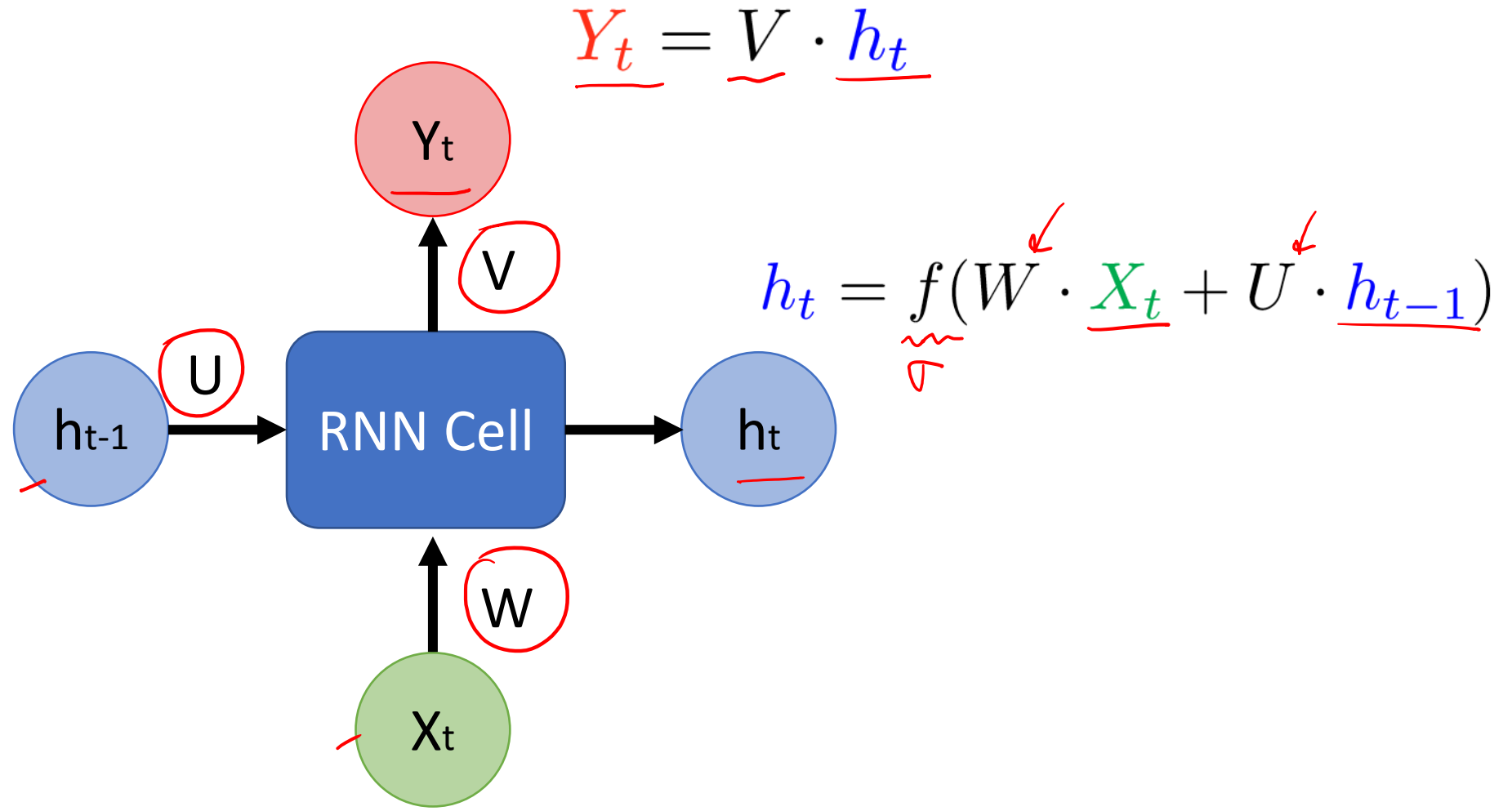Convolutional Neural Nets

# What is Recurrent Neural Network?

# What is Recurrent Neural Network?

# What's inside the RNN Cell?
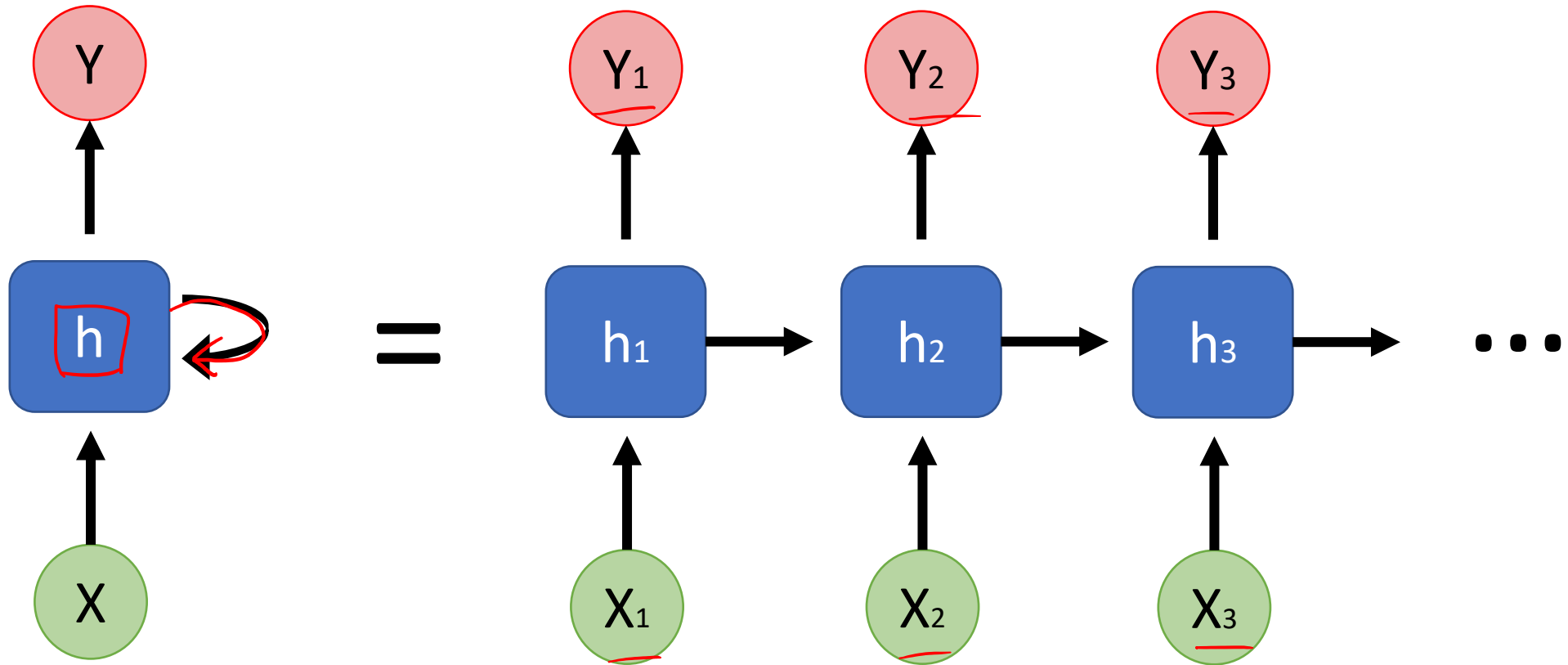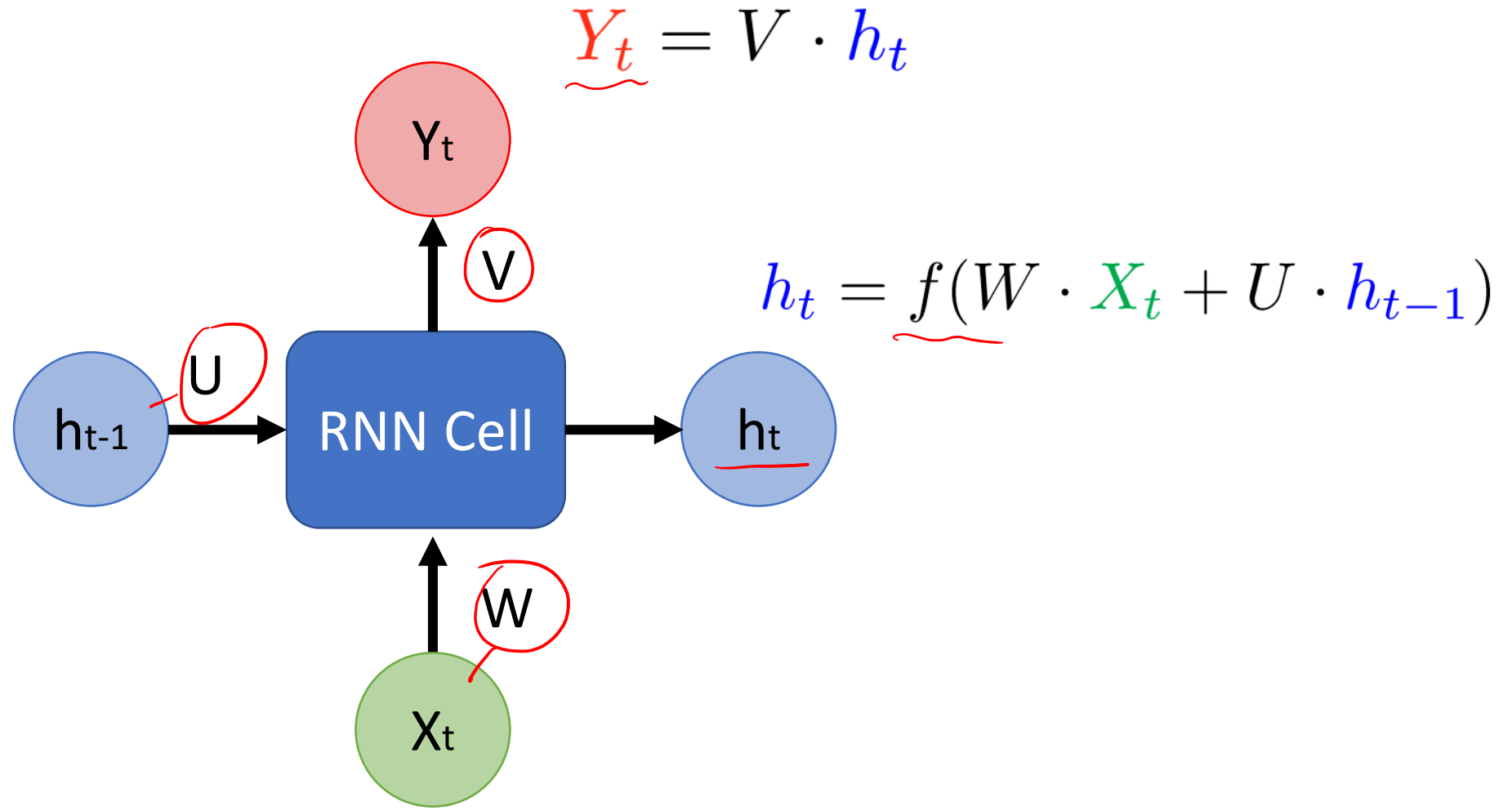
# What's inside the RNN Cell?



$$Y_t = V \cdot h_t$$

$$h_t = f(W \cdot X_t + U \cdot h_{t-1})$$

# Recurrent Neural Networks (2)

Geena Kim

University of Colorado **Boulder**

# What is Recurrent Neural Network?

# What's inside the RNN Cell?
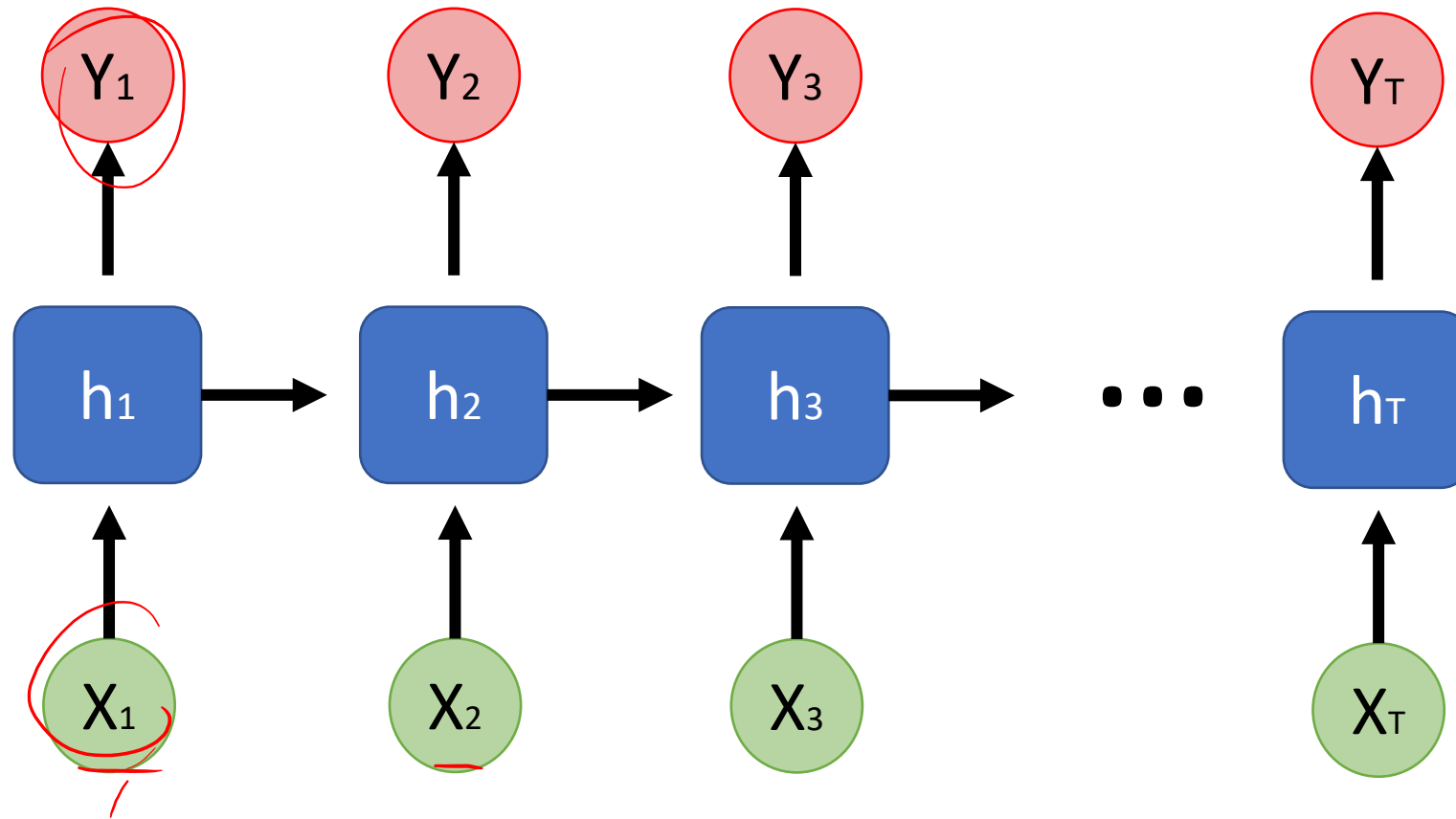


$$Y_t = V \cdot h_t$$

$$h_t = f(W \cdot X_t + U \cdot h_{t-1})$$

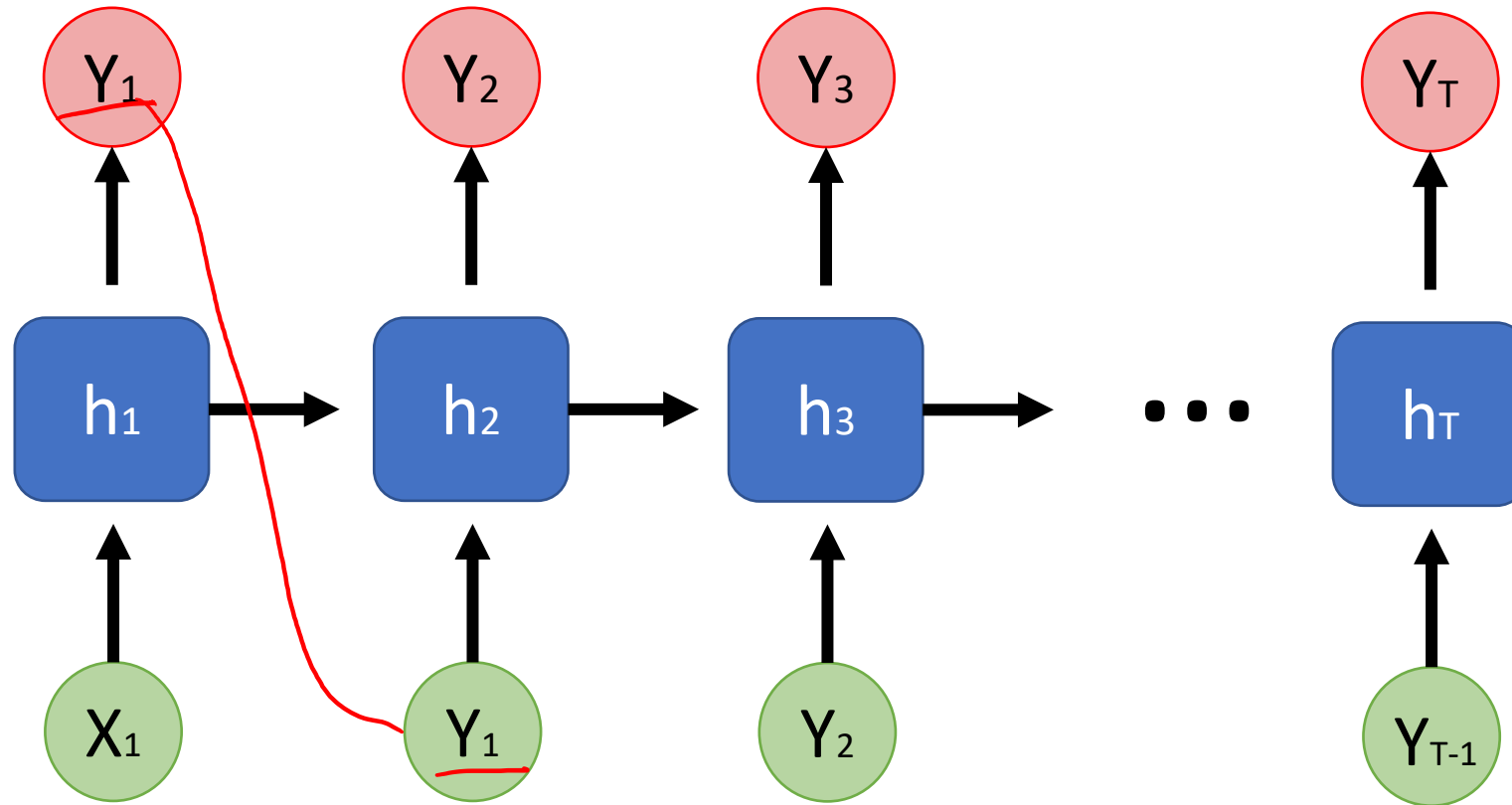# Vanilla RNN unrolled in time
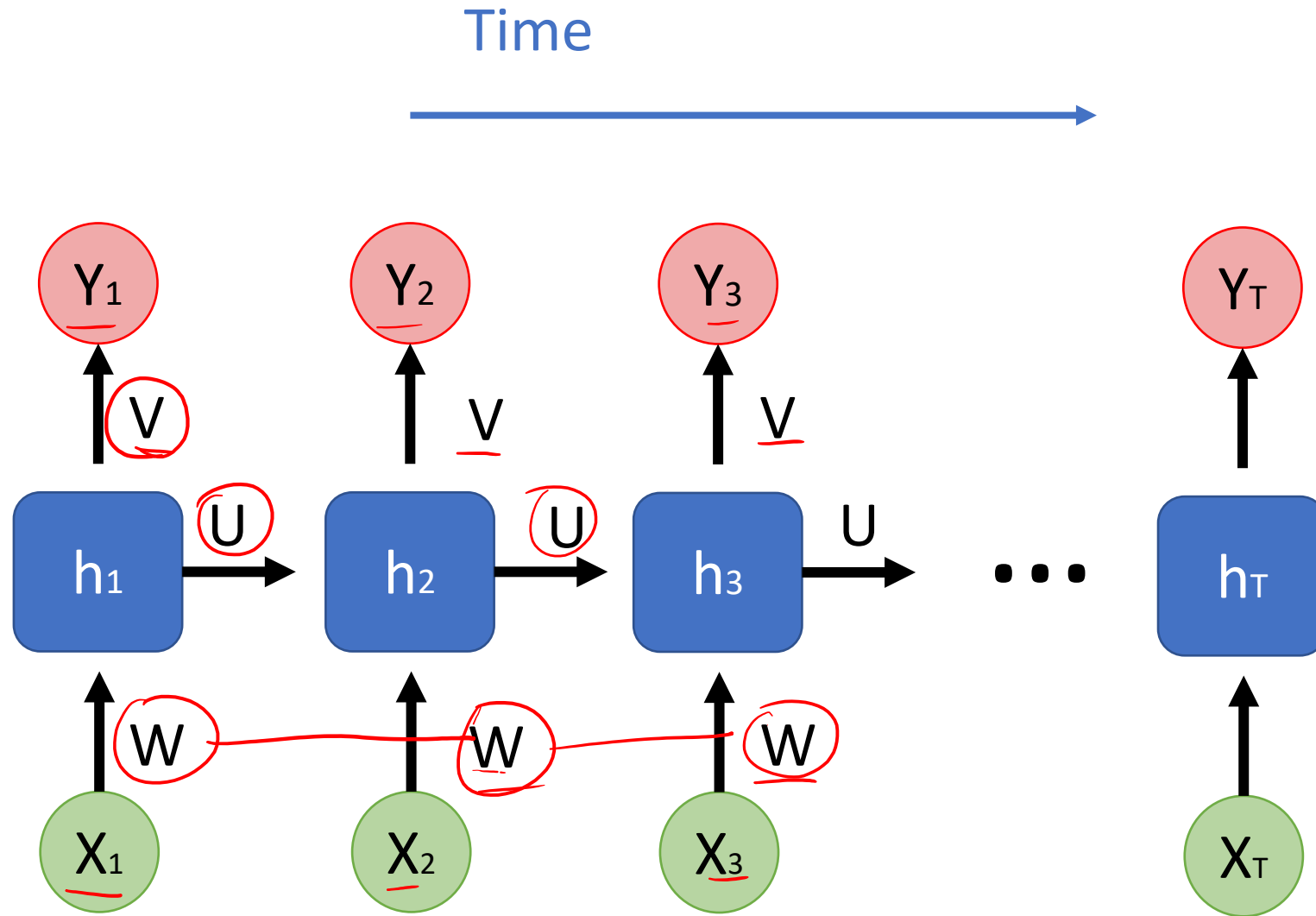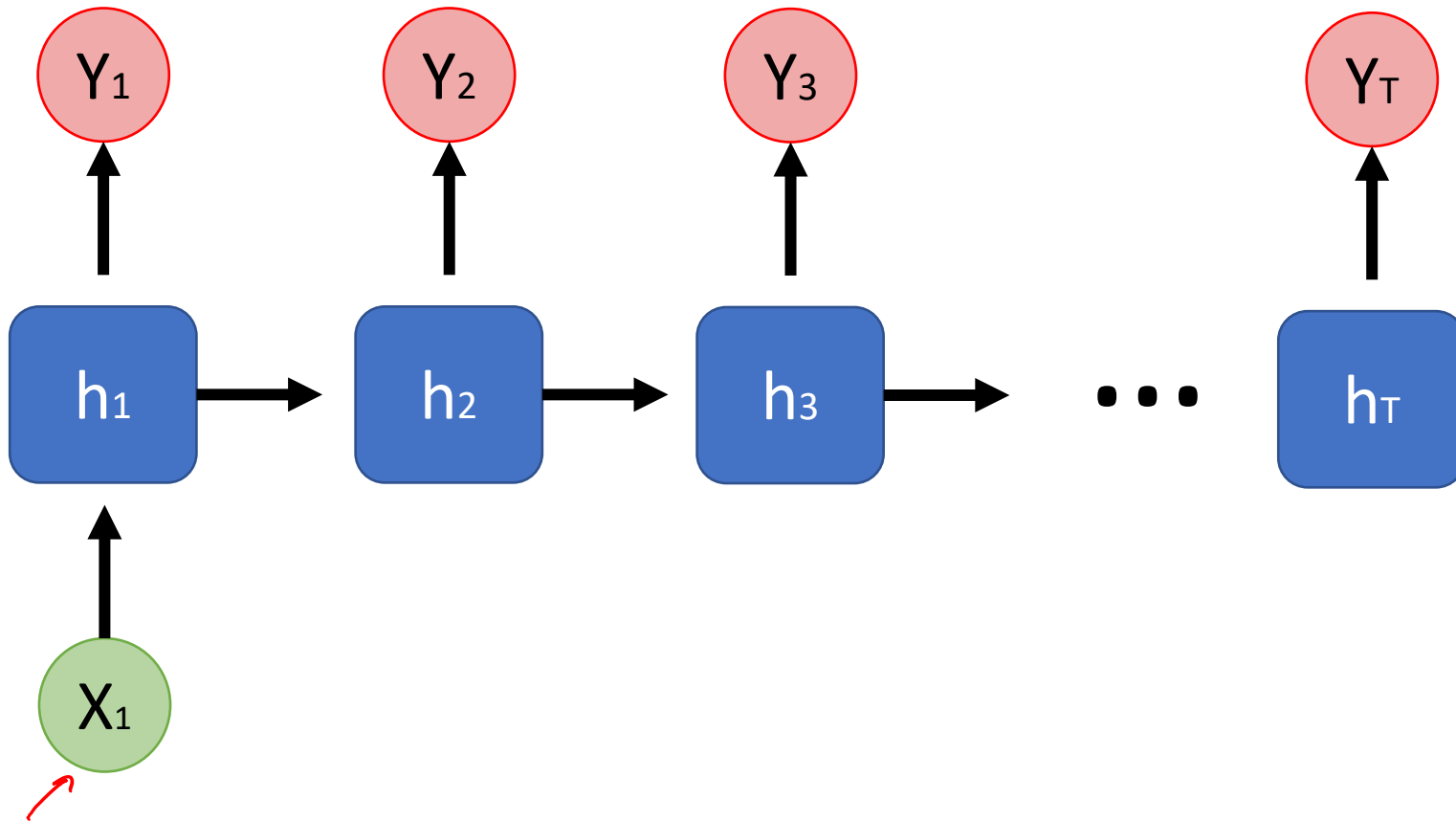
# RNN output types



e.g. video frame classification
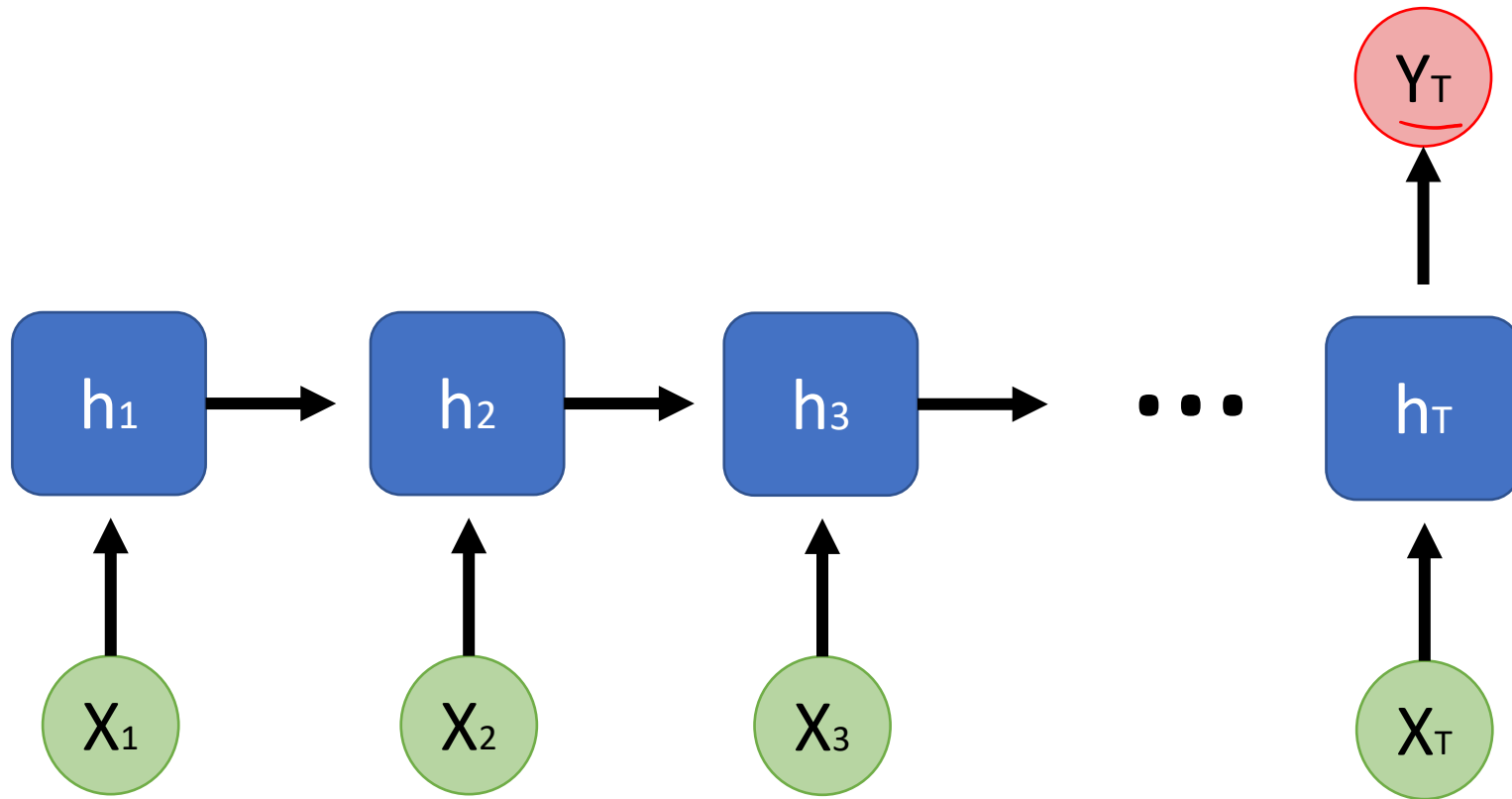
# RNN output types



e.g. next word/character generation
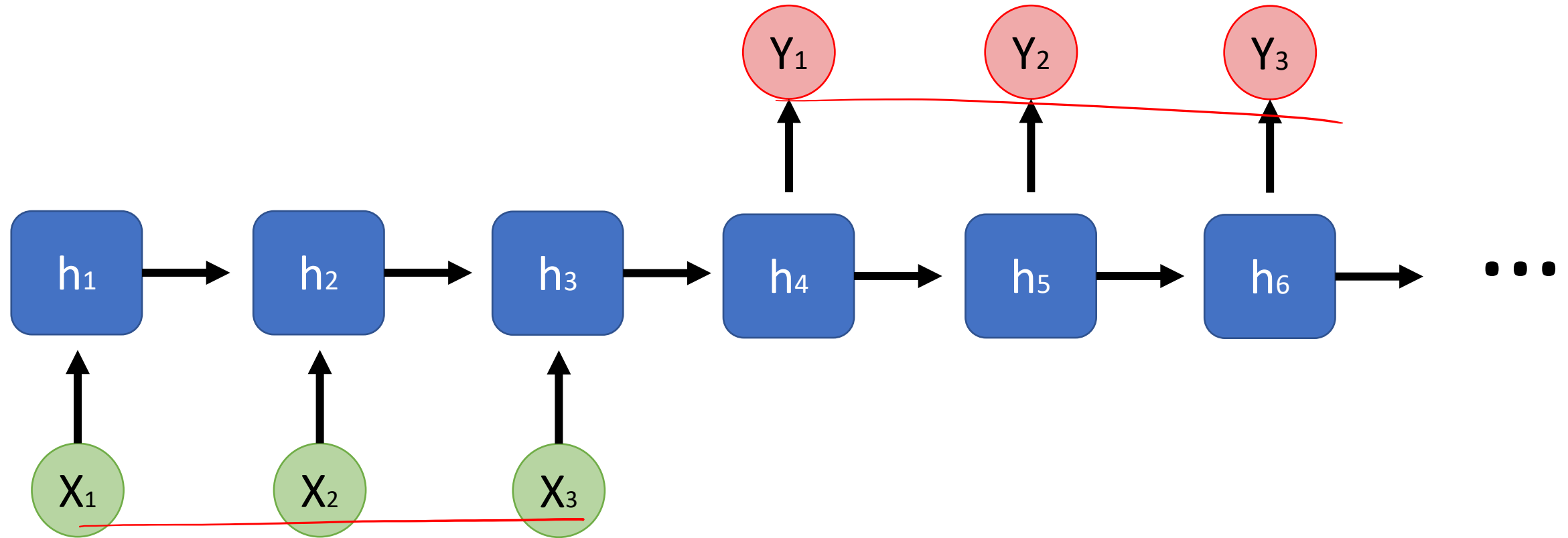
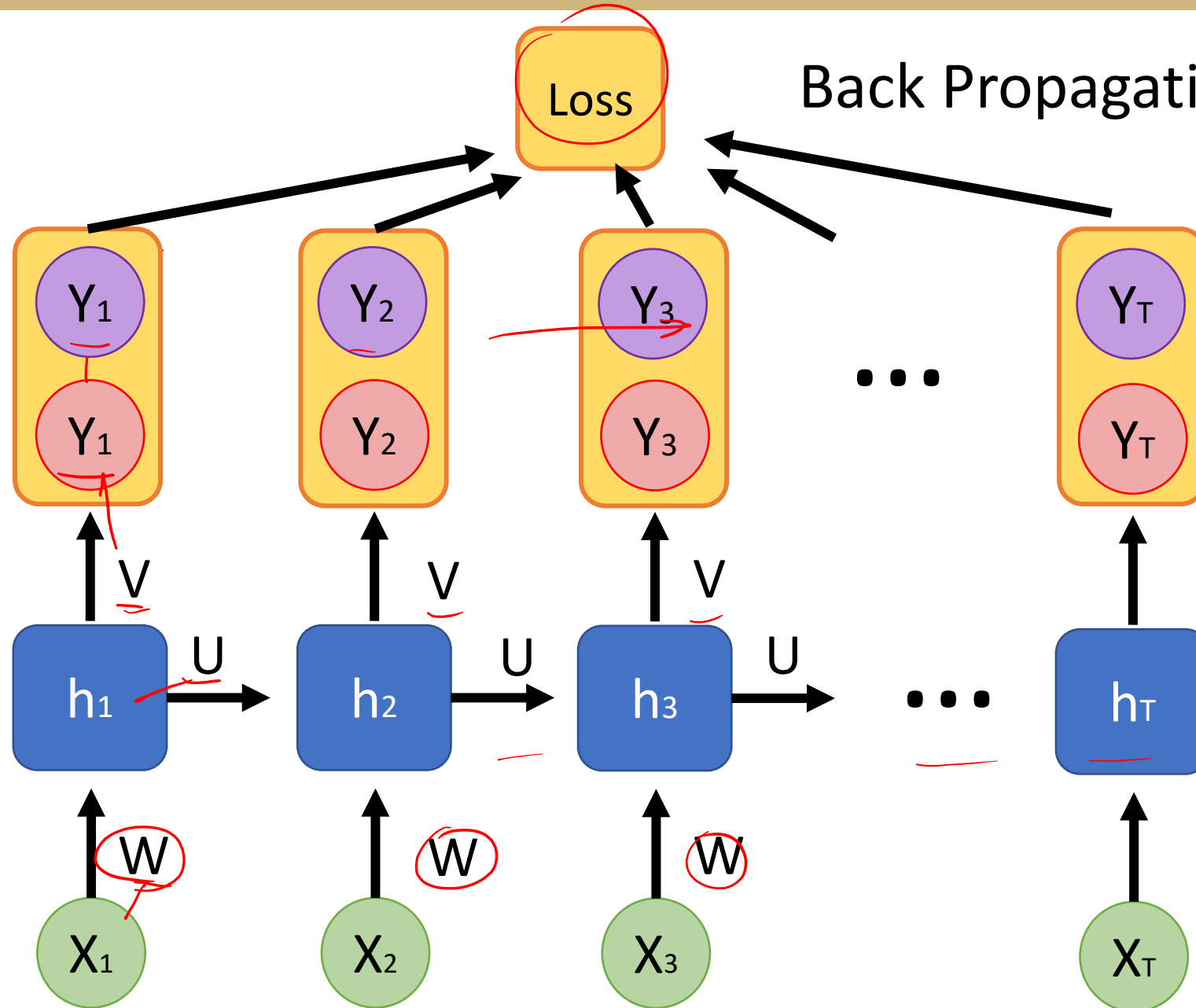# RNN output types



e.g. image captioning

# RNN output types



e.g. Sentiment analysis, document/video classification

# RNN output types



e.g. machine translation

# How a weight update work in an RNN?

**Snow Fall**

**The Avalanche at Tunnel Creek**

The snow burst through the trees with no warning but a last-second whoosh of sound, a two-story wall of white and Chris Rudolph's piercing cry: "Avalanche! Elyse!"

The very thing the 16 skiers and snowboarders had sought — fresh, soft snow — instantly became the enemy. Somewhere above, a pristine meadow cracked in the shape of a lightning bolt, slicing a slab nearly 200 feet across and 3 feet deep. Gravity did the rest.

Snow shattered and spilled down the slope. Within seconds, the avalanche was the size of more than a thousand cars barreling down the mountain and weighed millions of pounds. Moving about 7o miles per hour, it crashed through the sturdy old-growth trees, snapping their limbs and shredding bark from their trunks.

The avalanche, in Washington's Cascades in February, slid past some trees and rocks, like ocean swells around a ship's prow. Others it captured and added to its violent load.

Somewhere inside, it also carried people. How many, no one knew.

The slope of the terrain, shaped like a funnel, squeezed the growing swell of churning snow into a steep, twisting gorge. It moved in surges, like a roller coaster on a series of drops and high-banked turns. It accelerated as the slope steepened and the weight of the slide pushed from behind. It slithered through shallower pitches. The energy raised the temperature of the snow a couple of degrees, and the friction carved striations high in the icy sides of the canyon walls.

Elyse Saugstad, a professional skier, wore a backpack equipped with an air bag, a relatively new and expensive part of the arsenal that backcountry users increasingly carry to ease their minds and increase survival odds in case of an avalanche. About to be overtaken, she pulled a cord near her chest. She was knocked down before she knew if the canister of compressed

# How do we run a backprop in RNN?

Truncated Back Propagation Though Time
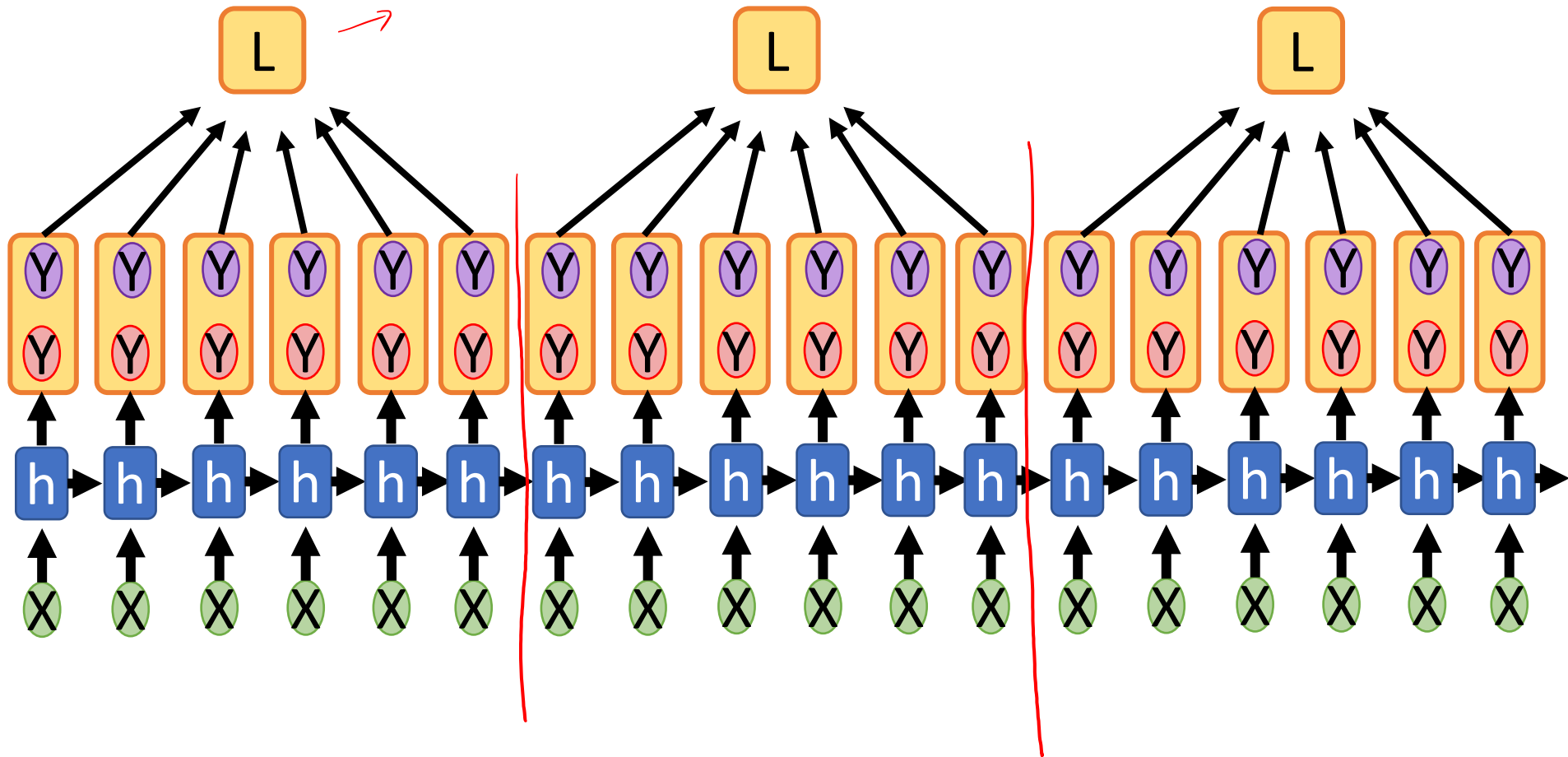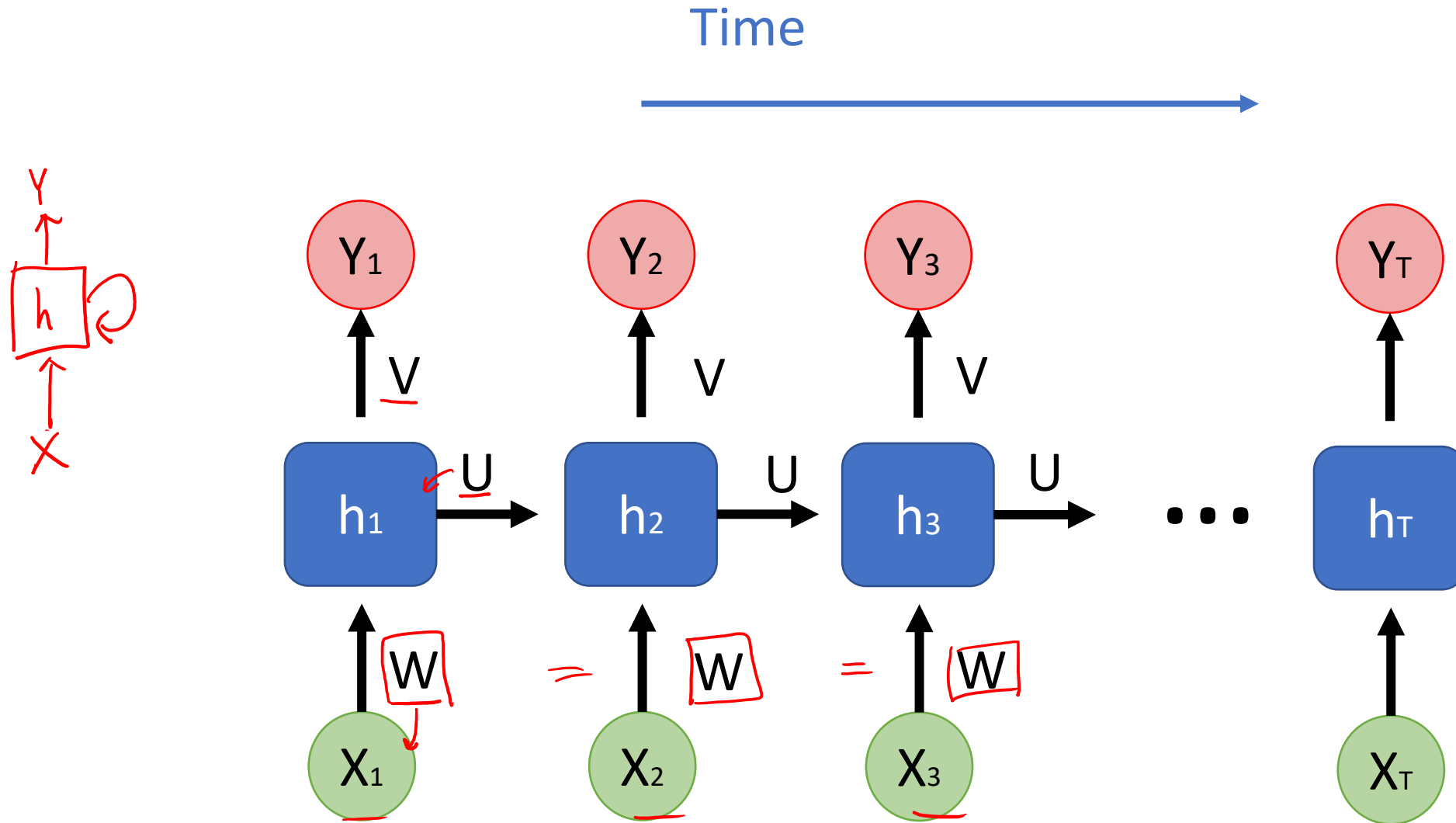
# Cool Examples and where to get data

The example works are from Andrej Karpathy's blog:
http://karpathy.github.io/2015/05/21/rnn-effectiveness/
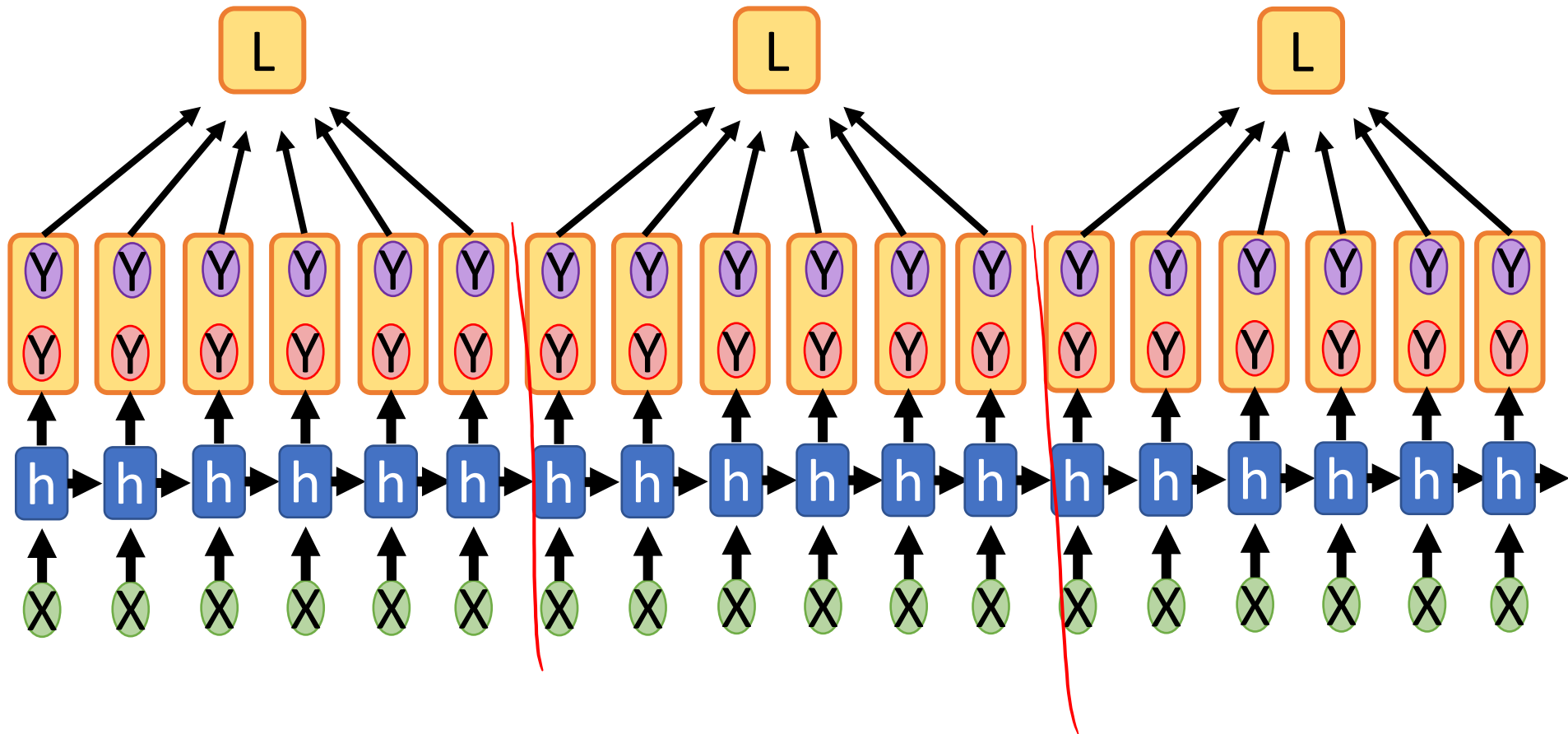
# Vanilla RNN Limitations
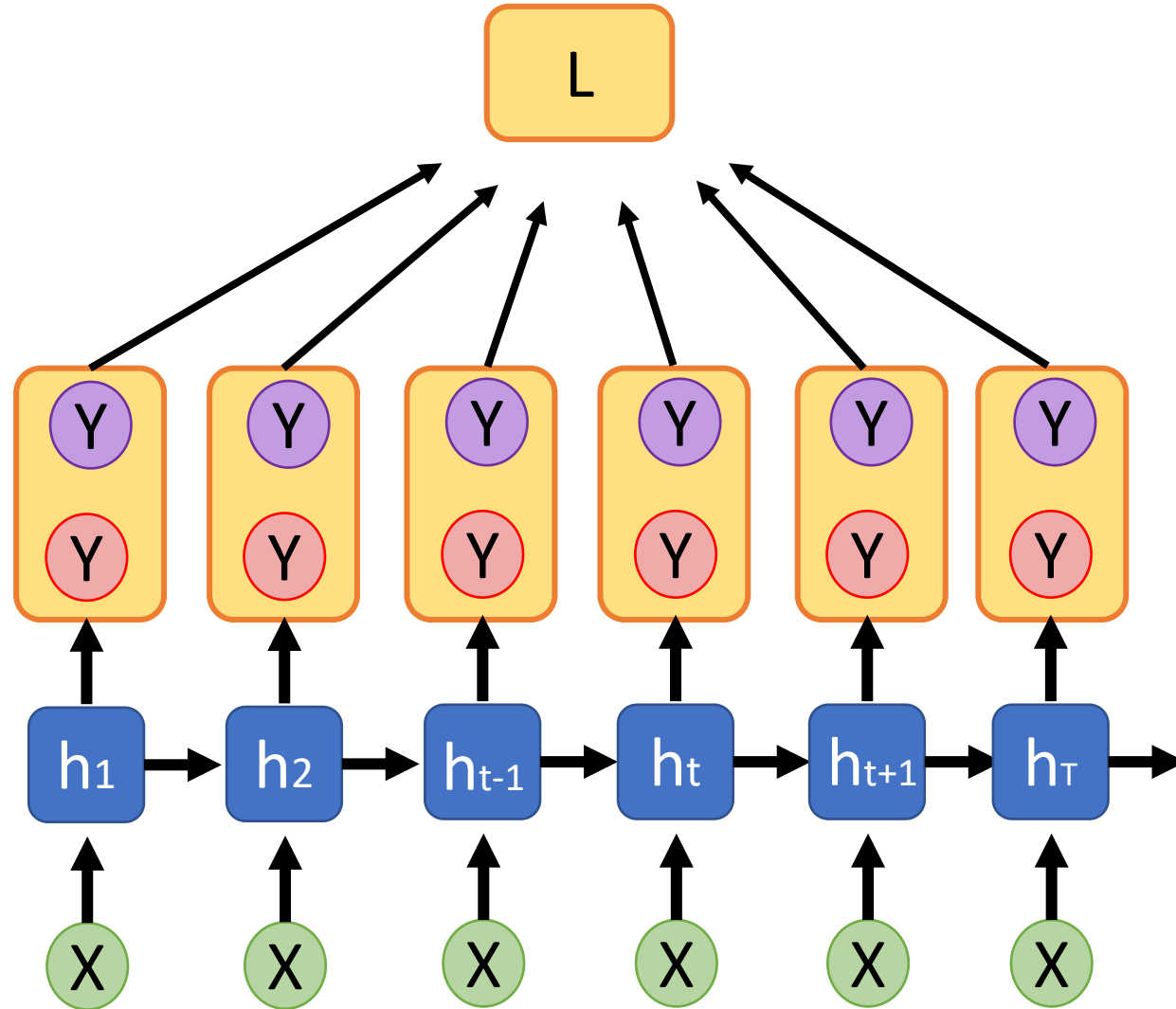
Geena Kim

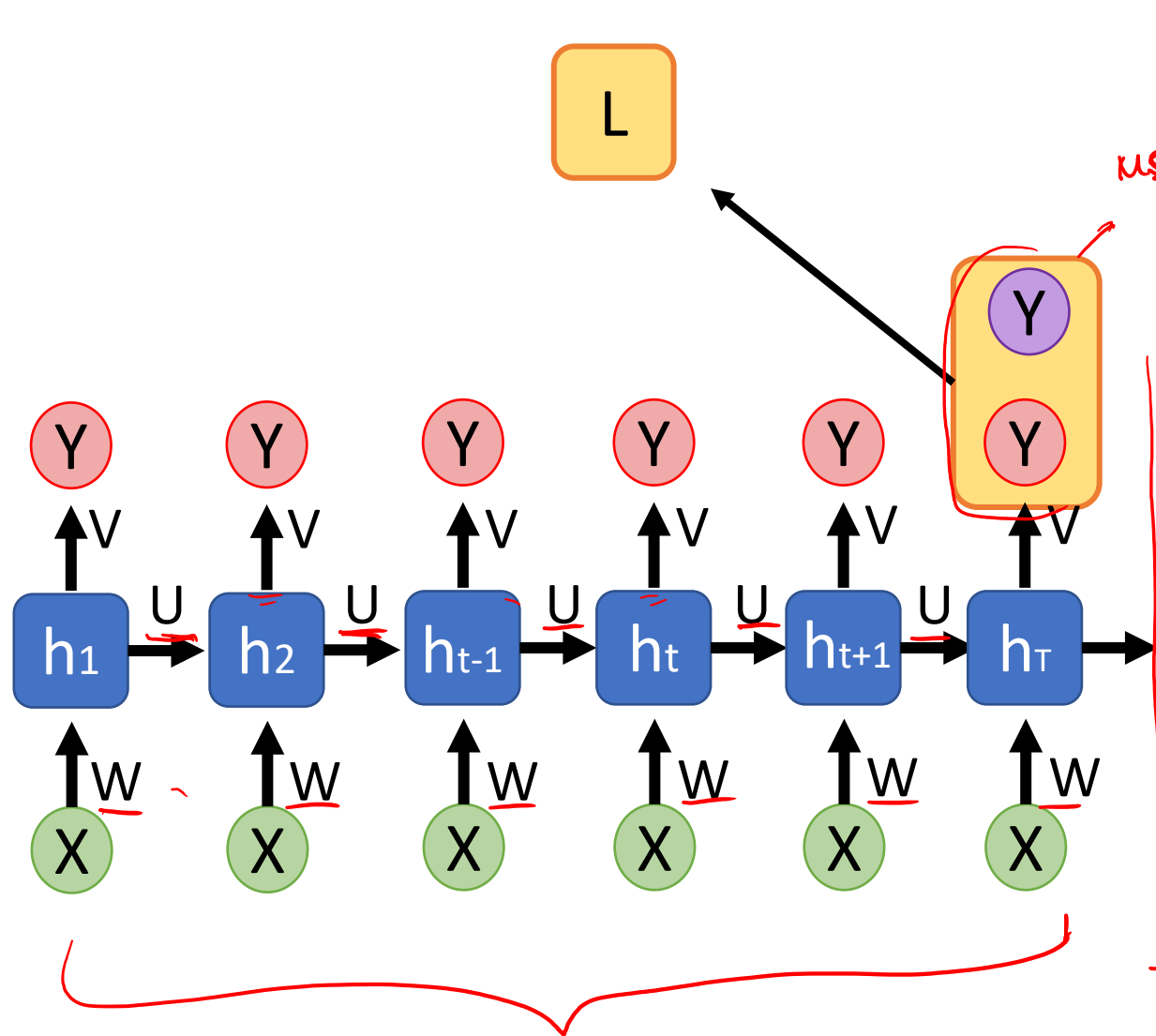# Vanilla RNN unrolled in time

# Vanilla RNN training with Truncated BPTT

# Problems with RNN training

# Vanishing/Exploding Gradient in RNN



$$\frac{\partial L}{\partial U} = (y - \hat{y}) \cdot \frac{\partial \hat{y}}{\partial u}$$

$$= (y - \hat{y}) \cdot V \cdot \left(\frac{\partial h_t}{\partial U}\right)$$

$$h_t = f(\underbrace{W \cdot X_t + U \cdot h_{t-1}}_{U})$$

MSE

$V \cdot h_t$

$$\frac{\partial h_t}{\partial u} = f' \cdot \left(f + u \cdot \frac{\partial f}{\partial u}\right)$$

$$\frac{\partial f}{\partial u} = f' \left(f + \boxed{u} \frac{\partial f}{\partial u}\right)$$

$f \cdot f'$

$\sigma \cdot (1 - \sigma) \sigma \rightarrow 0$

$\boxed{u^{\# N}}$

$U < 1 \rightarrow$ Vanishing

$> 1 \rightarrow$ Exploding

# Remedies

- ReLU activation function
- Truncated BPTT
- Clip gradient
- Use learning rate scheduling
- Add residual connection
- Change architectures- LSTM, GRU