Recurrent Neural Networks

Danna Gurari
University of Colorado Boulder
Fall 2022

https://home.cs.colorado.edu/~DrG/Courses/NeuralNetworksAndDeepLearning/AboutCourse.html

Review

- Last week:
  - Problems
  - Applications
  - PASCAL VOC detection challenge: region proposals with CNN features
  - PASCAL VOC semantic segmentation challenge: fully convolutional networks
- Assignments (Canvas):
  - Lab assignment 2 due earlier today
  - Lab assignment 3 due in 1.5 weeks
- Questions?

Today’s Topics

- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial

Today’s Topics

- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial

Sequence Definition: Data of Arbitrary Length

e.g., Document

THE SEATTLE PRESS-TIMES, TUESDAY, JUNE 16, 1891.

SEATTLE SHIP CANAL

Survey Authorized by Congress Completed.

Canal Should be Built and Controlled by Government.

Statement to Secretary of War by Chamber of Commerce.

ADVANTAGES TO ASIATIC POINTS.

Commercial Interests of Northwest Make it Imperative.

The board of trustees of the chamber of commerce, at their meeting this morning, listened to the reading of the report of the Lake Washington and Lake Union canal committee. The memorial prepared by the committee to the honorable secretary of war was adopted, and the paper will be forwarded at once.

Following is the paper:

The survey authorized by congress of the proposed ship canal to connect Lake Washington and Lake Union with Puget Sound at Seattle, having been completed, the attention of the honorable the secretary of war is respectfully invited to the following statement presenting some of the reasons why this canal should be promptly constructed:

The city of Seattle is situated almost in the center of Puget Sound region and has a frontage on salt water of more than six miles. Lake Washington lies abreast of the city to the eastward, parallel with the Sound, and stretches north and south 25 miles. Deeply embayed by a cordon of mountains, its surface is always smooth and safe; its shores are bold but not rocky, its waters are ample depth, soundings showing in places 100 fathoms. Its width near its central section is four miles and the average nearly three, affording ample room for handling vessels of the largest size. The absence of tides and currents renders its waters almost tranquil and of even stage, and preserves its banks without washing. Its outlet by an easy channel enters the Sound almost within the present city limits.

Bordered with fertile lands, backed with inexhaustible forests of finest timber and immense fields of coal, iron and the precious minerals, the situation of this lake has not failed to attract the attention of every one versed in the ways of business or commerce who has visited the region.

Another such body of fresh water, lying within two miles of a noble ocean roadstead with which it may easily be joined, capable of being converted into anchorage ground for the fleets of the world, does not exist elsewhere; it would seem to have been provided by nature as a complement to our harbor system, which it will make perfect in its every feature.

A Focus of Great Enterprises.

Puget Sound having been for many years marked as the seat of coming great commercial expansion has become, within the last decade, the focus of enterprise looking to the realization of its future. Four great transcontinental railway lines have selected its waters for termini of their systems. Already two of them are establishing ocean steamship lines to Asia and the Oriental islands.

From all parts of the world come fleets of vessels to this island sea to compete for the cargoes which lie close at hand or arrive as freights of railways spanning the continent. The tide of progress in trade and commerce is so rapid as to run ahead of the facilities provided, and such is likely to be the case for years to come. The eyes of the industrial world are set this way.

A cursory glance at the resources of the state show good reasons for faith in its future:

The wheat yield of 1890 was about 20,000,000 bushels; that for the current year is estimated at over 25,000,000 bushels. The capacity of the natural wheat fields of the state is conservatively placed at 200,000,000 bushels per year. Oats, barley, hay, fruits of all kinds, every farm crop, reward the husbandman with prodigious returns.

The lumber cut of the state for 1890 was 1,222,830,042 feet, an increase of 538,647,042 feet over the preceding year; nearly 100 per cent. This does not include the lumber cut into laths, pickets, etc., estimated at 164,185,800 feet.

The coal output of the state for 1890 was 1,349,773 tons; an increase of 438,246 tons, or about 50 per cent for the year.

The coal measures are practically limitless in extent and the business of shipping coal is in its infancy. The amount of coal mined hereafter will be limited only by the demand for it.

Iron ores of the best quality—fit for steel making—cover an immense field. Works for its manufacture are in course of construction at several points.

The precious metals are distributed over a wide area and promise enormous development of wealth. Gold, silver, lead, copper, tin, asbestos, graphite, limestone, marble, granite, sandstone, are to be added to iron and coal. All these combined furnish an opulence of mineral resources rarely encountered in a single field.

A state possessing unrivaled agricultural advantages, and having superadded to these such incomparable wealth in timber and minerals, cannot fail to develop manufacturing interests on a gigantic scale. Such development is already entered upon.

With the completion of the railway systems now speeding to their terminals at this point, the state will have an enjoyment of transportation lines greater and more perfect than any one state has hitherto equipped. She will be prepared to handle with utmost dispatch the traffic which throngs at her gates.

Scope of Foreign Commerce.

Great as the outlook for domestic trade may appear, it does not exceed the just scope for foreign commerce.

The trade of China alone is conservatively estimated at $150,000,000 annually. That country has just begun to make use of American products in large amounts. Every year she adds to her demands for timber and flour and manufactured goods. But the special field for expansion of trade is in manufactures of iron and steel. A country without railways, her manifest destiny is to develop, on a scale demanding enormous consumption of iron and steel, these modern means of improving civilization. Asiatic Russia has already entered upon this work in her possessions north of China.

Australia is opening wide doors to American trade and cargoes to her ports multiply yearly in astonishing ratio.

Japan and the coast countries are adding steadily to the volume of commercial business with our people. In South American countries, several of them possessing immense timberless savannas, the lumber of Puget Sound finds a profitable and growing market. This trade is apparently only in its incipiency, and is being augmented constantly.

Machinery, domestic utensils, manufactures of iron and wood, every product of our workshops, whether made on the Pacific or coming by rail from Eastern centers, are now to be added to the list of articles for trade with the South American states—a people rich in resource but possessing meagre manufactures. Return cargoes of sugars, coffee, spices, precious woods, dyestuffs, gums, tropical fruits and nitrates will seek these waters for distribution by the railways to inland points.

The Nicaraguan canal, a necessity of the times and certain now of speedy construction, will open the door to a vast tide of commerce to the Atlantic seaboard, from ports both from the United States and South America. It will reduce by one-half the distance to European ports and bring a mighty current of traffic from that country.

Thus from Asia, the greater islands of the southern ocean, the rich East India ports, the immense stretch of the Atlantic seaboard, the fruitful lands of Western Europe, a rising volume of commercial business is sweeping this way, and must be received and cared for in Puget Sound, or else surrendered to our active and determined competitors to the north of us.

Commercial Expansion.

It is not too much to assert that a commercial expansion equal to that now existing at the port of New York is impending for these waters. A far greater shore line than that port has, will be essential for the proper accommodation of the docks and wharves required here for the reason that nearly the entire shipping of the Pacific coast must find harbor facilities at a few ports, embracing Puget Sound, San Francisco and the mouth of the Columbia river; whereas, many good harbors exist on the Atlantic seaboard within easy radius of New York.

Australia and all of Oceanica lie nearer to Puget Sound than to England, on the Pacific side; the Malay peninsula makes the half-way point. Japan and all Northern Asia are thousands of miles nearer to us than to London. Our commerce is now within less than three weeks of the coast of China. Statistics show that New York by way of Puget Sound is 1400 miles nearer to Canton than is Liverpool, and about 4000 miles nearer Shanghai.

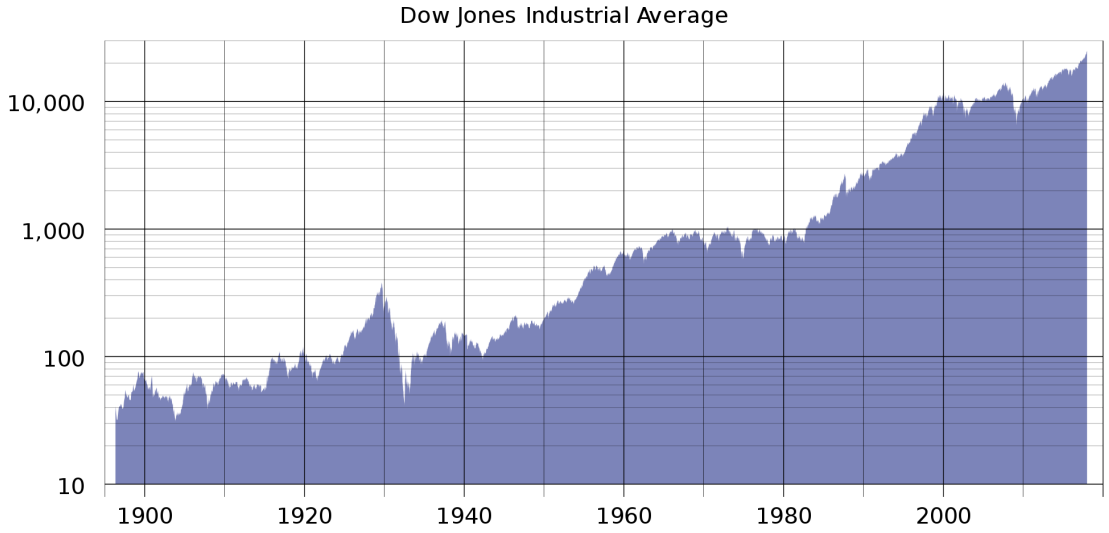
Advantages in Distances.

Puget Sound has also the advantage in distance to Asiatic points over San Francisco. The following table of distances, culled from Prof. Ruffer's interesting report upon the state of Washington, pre-

Concluded on next page.

e.g., Images

e.g., Time-Series Data

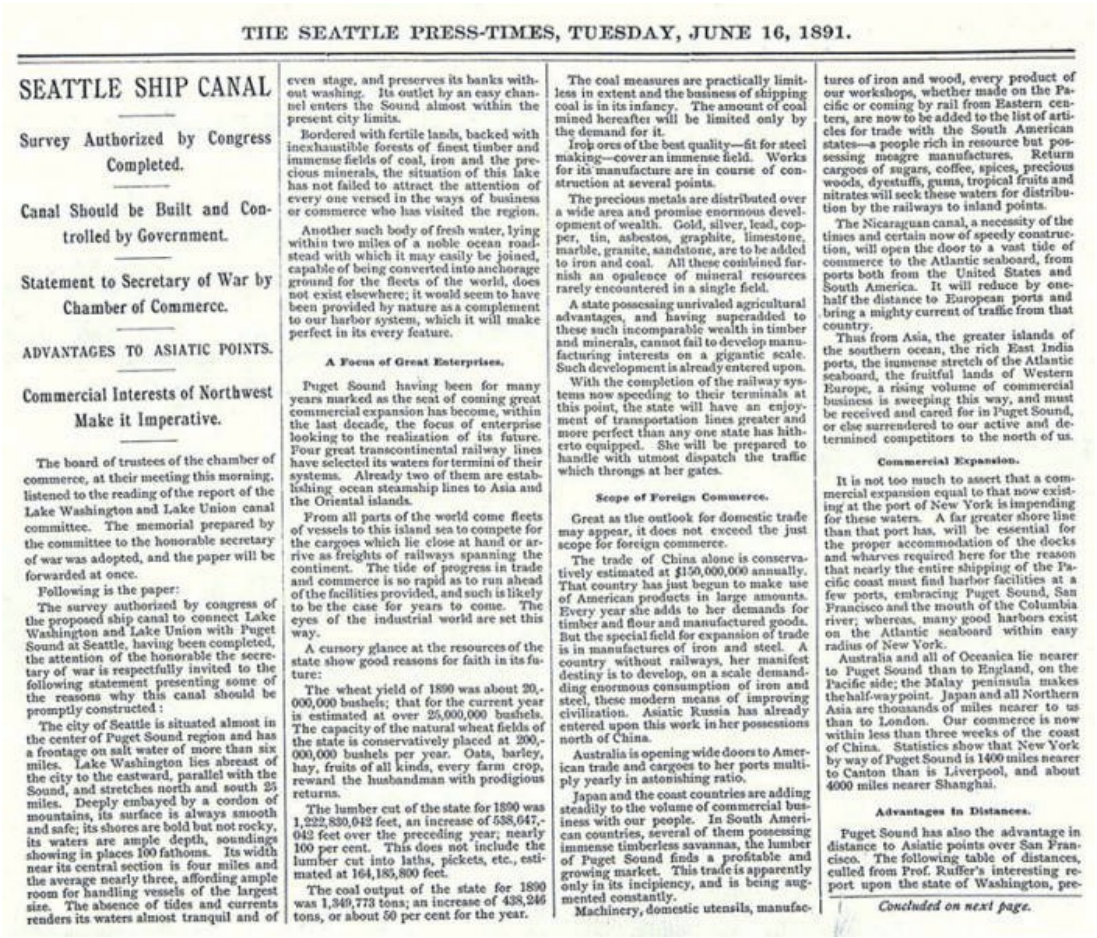


e.g., sentences, audio samples, brain waves, radio waves, air temperature

Properties of Sequences?

* Elements of a sequence occur in a certain order
* Elements depend on each other

e.g., Document



THE SEATTLE PRESS-TIMES, TUESDAY, JUNE 16, 1891.

SEATTLE SHIP CANAL

Survey Authorized by Congress Completed.

Canal Should be Built and Controlled by Government.

Statement to Secretary of War by Chamber of Commerce.

ADVANTAGES TO ASIATIC POINTS.

Commercial Interests of Northwest Make it Imperative.

The board of trustees of the chamber of commerce, at their meeting this morning, listened to the reading of the report of the Lake Washington and Lake Union canal committee. The memorial prepared by the committee to the honorable secretary of war was adopted, and the paper will be forwarded at once.

Following is the paper:

The survey authorized by congress of the proposed ship canal to connect Lake Washington and Lake Union with Puget Sound at Seattle, having been completed, the attention of the honorable the secretary of war is respectfully invited to the following statement presenting some of the reasons why this canal should be promptly constructed:

The city of Seattle is situated almost in the center of Puget Sound region and has a frontage on salt water of more than six miles. Lake Washington lies abreast of the city to the eastward, parallel with the Sound, and stretches north and south 25 miles. Deeply embayed by a cordon of mountains, its surface is always smooth and safe; its shores are bold but not rocky, its waters are ample depth, soundings showing in places 100 fathoms. Its width near its central section is four miles and the average nearly three, affording ample room for handling vessels of the largest size. The absence of tides and currents renders its waters almost tranquil and of even stage, and preserves its banks without washing. Its outlet by an easy channel enters the Sound almost within the present city limits.

Bordered with fertile lands, backed with inexhaustible forests of finest timber and immense fields of coal, iron and the precious minerals, the situation of this lake has not failed to attract the attention of every one versed in the ways of business or commerce who has visited the region.

Another such body of fresh water, lying within two miles of a noble ocean roadstead with which it may easily be joined, capable of being converted into anchorage ground for the fleets of the world, does not exist elsewhere; it would seem to have been provided by nature as a complement to our harbor system, which it will make perfect in its every feature.

A Focus of Great Enterprises.

Puget Sound having been for many years marked as the seat of coming great commercial expansion has become, within the last decade, the focus of enterprise looking to the realization of its future. Four great transcontinental railway lines have selected its waters for termini of their systems. Already two of them are establishing ocean steamship lines to Asia and the Oriental islands.

From all parts of the world come fleets of vessels to this island sea to compete for the cargoes which lie close at hand or arrive as freights of railways spanning the continent. The tide of progress in trade and commerce is so rapid as to run ahead of the facilities provided, and such is likely to be the case for years to come. The eyes of the industrial world are set this way.

A cursory glance at the resources of the state show good reasons for faith in its future:

The wheat yield of 1890 was about 20,000,000 bushels; that for the current year is estimated at over 25,000,000 bushels. The capacity of the natural wheat fields of the state is conservatively placed at 200,000,000 bushels per year. Oats, barley, hay, fruits of all kinds, every farm crop, reward the husbandman with prodigious returns.

The lumber cut of the state for 1890 was 1,222,830,042 feet, an increase of 538,647,042 feet over the preceding year; nearly 100 per cent. This does not include the lumber cut into laths, pickets, etc., estimated at 164,185,800 feet.

The coal output of the state for 1890 was 1,349,773 tons; an increase of 438,246 tons, or about 50 per cent for the year.

The coal measures are practically limitless in extent and the business of shipping coal is in its infancy. The amount of coal mined hereafter will be limited only by the demand for it.

Iron ores of the best quality—fit for steel making—cover an immense field. Works for its manufacture are in course of construction at several points.

The precious metals are distributed over a wide area and promise enormous development of wealth. Gold, silver, lead, copper, tin, asbestos, graphite, limestone, marble, granite, sandstone, are to be added to iron and coal. All these combined furnish an opulence of mineral resources rarely encountered in a single field.

A state possessing unrivaled agricultural advantages, and having superadded to these such incomparable wealth in timber and minerals, cannot fail to develop manufacturing interests on a gigantic scale. Such development is already entered upon.

With the completion of the railway systems now speeding to their terminals at this point, the state will have an enjoyment of transportation lines greater and more perfect than any one state has hitherto equipped. She will be prepared to handle with utmost dispatch the traffic which throngs at her gates.

Scope of Foreign Commerce.

Great as the outlook for domestic trade may appear, it does not exceed the just scope for foreign commerce.

The trade of China alone is conservatively estimated at $150,000,000 annually. That country has just begun to make use of American products in large amounts. Every year she adds to her demands for timber and flour and manufactured goods. But the special field for expansion of trade is in manufactures of iron and steel. A country without railways, her manifest destiny is to develop, on a scale demanding enormous consumption of iron and steel, these modern means of improving civilization. Asiatic Russia has already entered upon this work in her possessions north of China.

Australia is opening wide doors to American trade and cargoes to her ports multiply yearly in astonishing ratio.

Japan and the coast countries are adding steadily to the volume of commercial business with our people. In South American countries, several of them possessing immense timberless savannas, the lumber of Puget Sound finds a profitable and growing market. This trade is apparently only in its incipiency, and is being augmented constantly.

Machinery, domestic utensils, manufactures of iron and wood, every product of our workshops, whether made on the Pacific or coming by rail from Eastern centers, are now to be added to the list of articles for trade with the South American states—a people rich in resource but possessing meagre manufactures. Return cargoes of sugars, coffee, spices, precious woods, dyestuffs, gums, tropical fruits and nitrates will seek these waters for distribution by the railways to inland points.

The Nicaraguan canal, a necessity of the times and certain now of speedy construction, will open the door to a vast tide of commerce to the Atlantic seaboard, from ports both from the United States and South America. It will reduce by one-half the distance to European ports and bring a mighty current of traffic from that country.

Thus from Asia, the greater islands of the southern ocean, the rich East India ports, the immense stretch of the Atlantic seaboard, the fruitful lands of Western Europe, a rising volume of commercial business is sweeping this way, and must be received and cared for in Puget Sound, or else surrendered to our active and determined competitors to the north of us.

Commercial Expansion.

It is not too much to assert that a commercial expansion equal to that now existing at the port of New York is impending for these waters. A far greater shore line than that port has, will be essential for the proper accommodation of the docks and wharves required here for the reason that nearly the entire shipping of the Pacific coast must find harbor facilities at a few ports, embracing Puget Sound, San Francisco and the mouth of the Columbia river; whereas, many good harbors exist on the Atlantic seaboard within easy radius of New York.

Australia and all of Oceanica lie nearer to Puget Sound than to England, on the Pacific side; the Malay peninsula makes the half-way point. Japan and all Northern Asia are thousands of miles nearer to us than to London. Our commerce is now within less than three weeks of the coast of China. Statistics show that New York by way of Puget Sound is 1400 miles nearer to Canton than is Liverpool, and about 4000 miles nearer Shanghai.

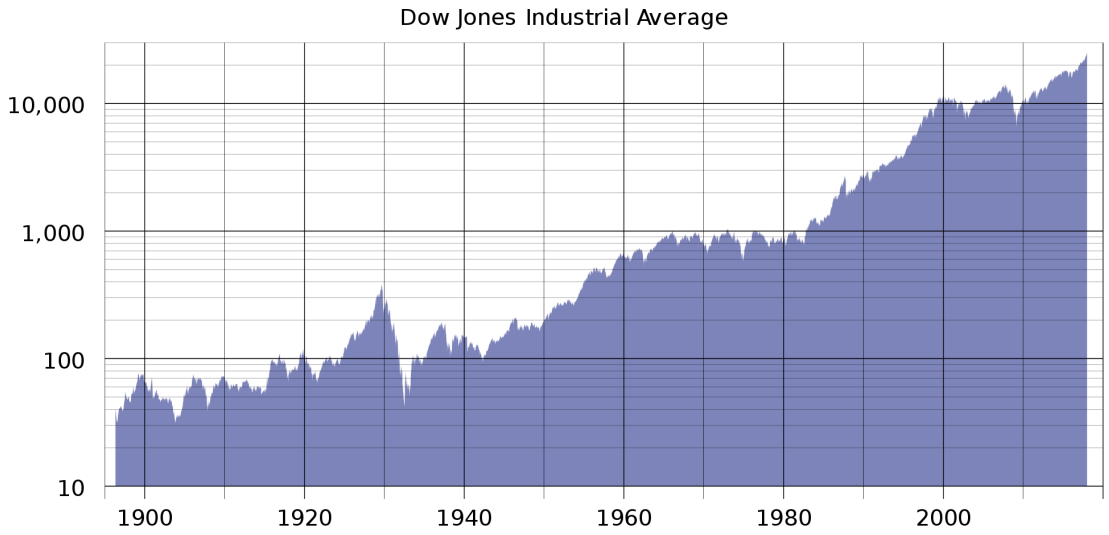
Advantages in Distances.

Puget Sound has also the advantage in distance to Asiatic points over San Francisco. The following table of distances, culled from Prof. Ruffer's interesting report upon the state of Washington, pre-

Concluded on next page.

e.g., Images

e.g., Time-Series Data

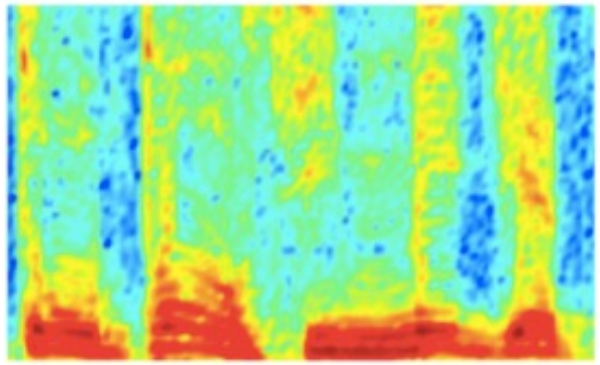


e.g., sentences, audio samples, brain waves, radio waves, air temperature

Sequence Sources

* Elements of a sequence occur in a certain order
* Elements depend on each other

AUDIO



Audio Spectrogram

IMAGES

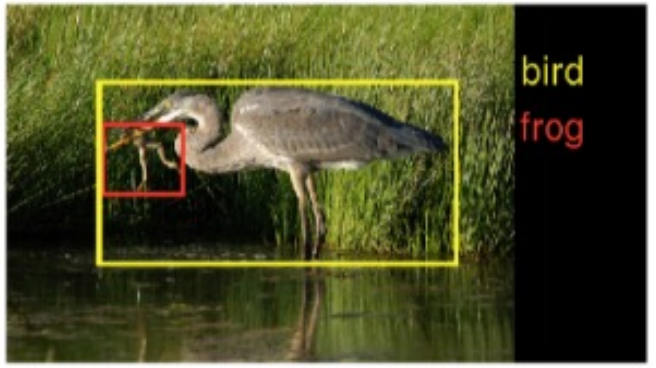


Image pixels

TEXT

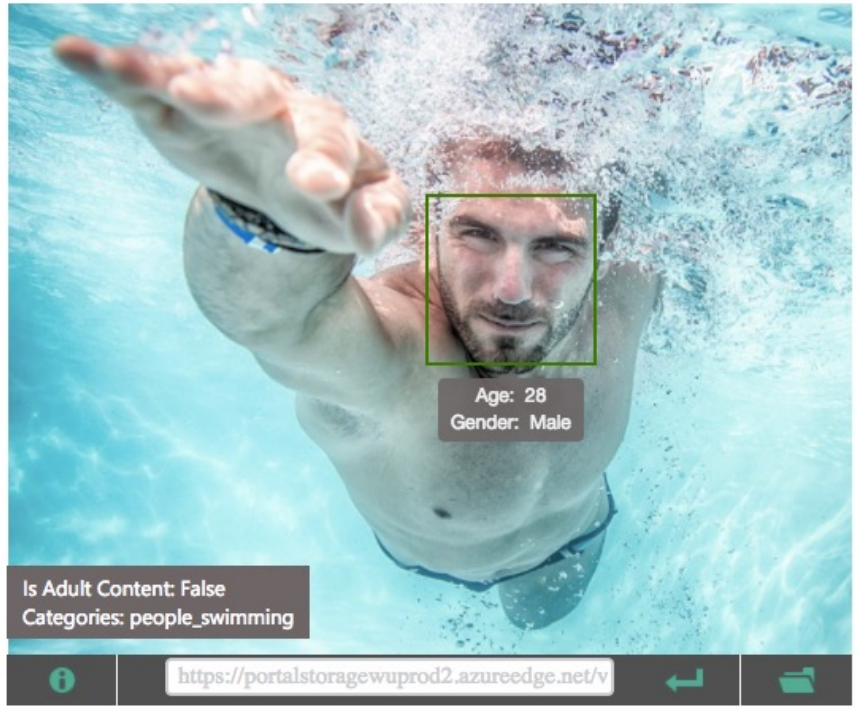
0	0	0	0.2	0	0.7	0	0	0	...	...

Word, context, or document vectors

https://www.tensorflow.org/tutorials/representation/word2vec

Sequence Applications: One-to-Many

- **Input**: fixed-size
- **Output**: sequence
- e.g., image captioning



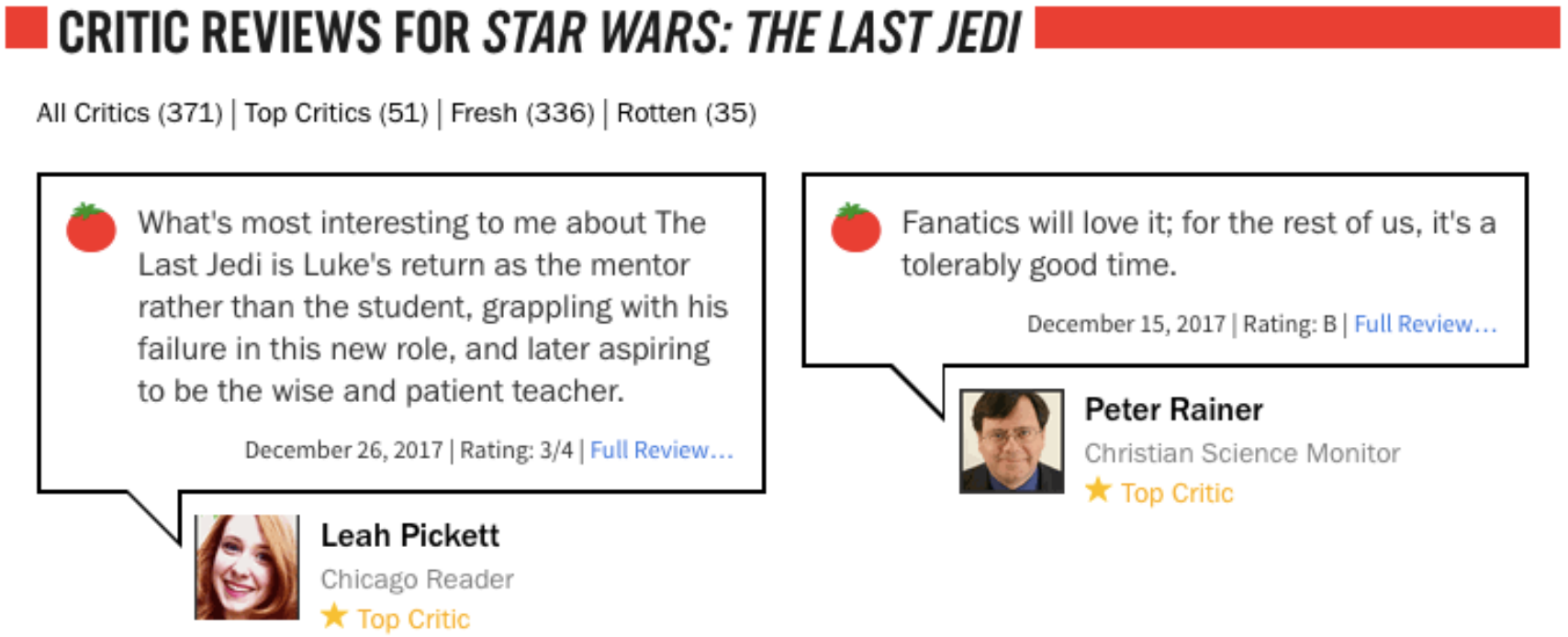
Features:

Feature Name	Value
Description	{ "type": 0, "captions": [{ "text": "a man swimming in a pool of water", "confidence": 0.7850108693093019 }] }
Tags	[{ "name": "water", "confidence": 0.9996442794799805 }, { "name": "sport", "confidence": 0.9504992365837097 }, { "name": "swimming", "confidence": 0.9062818288803101, "hint": "sport" }, { "name": "pool", "confidence": 0.8787588477134705 }, { "name": "water sport", "confidence": 0.631849467754364, "hint": "sport" }]
Image Format	jpeg
Image Dimensions	1500 x 1155
Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-LineDrawing
Black & White Image	False

Captions: https://www.microsoft.com/cognitive-services/en-us/computer-vision-api

Sequence Applications: Many-to-One

- **Input**: sequence
- **Output**: fixed-size
- e.g., sentiment analysis (hate? love?, etc)



https://www.rottentomatoes.com/m/star_wars_the_last_jedi

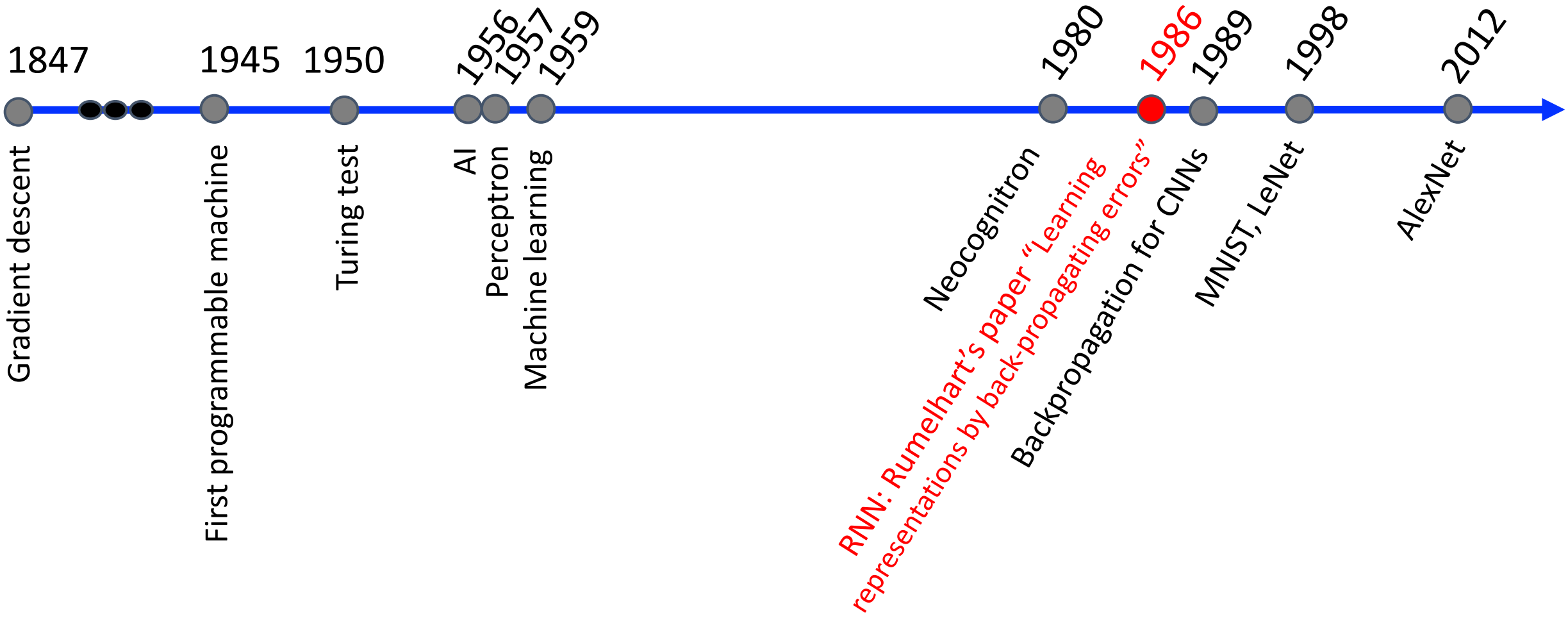
Sequence Applications: Many-to-Many

- **Input**: sequence
- **Output**: sequence
- e.g., language translation

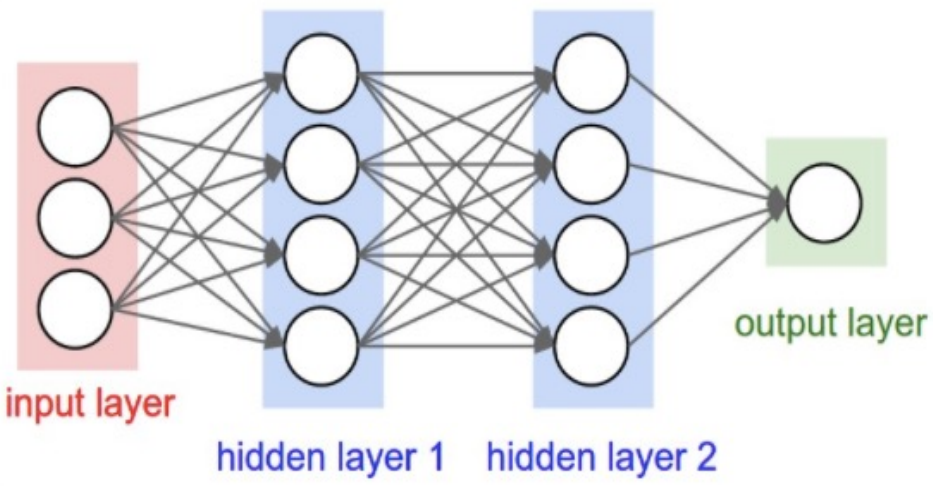
Today's Topics

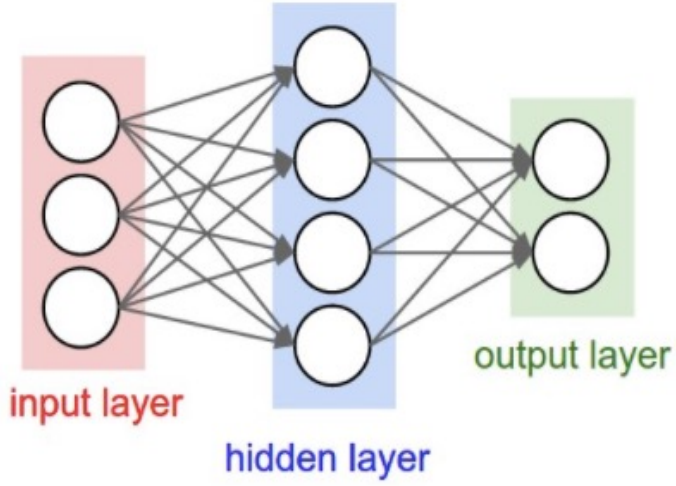
- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial

Historical Context



Recall: Feedforward Neural Networks





Problem: many model parameters

Problem: inputs/output sizes are fixed

Problem: no memory of past since weights learned independently

Each layer serves as input to the next layer with no loops

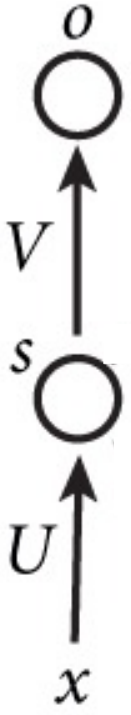
Figure Source: http://cs231n.github.io/neural-networks-1/

Recurrent Neural Networks (RNNs)

- Main idea: use hidden state to capture information about the past

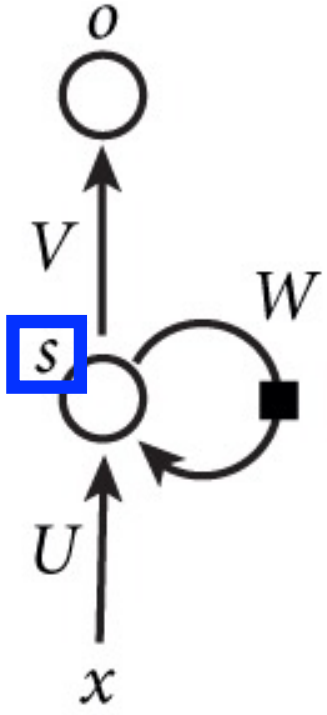
Feedforward Network

Each layer receives input from the previous layer with no loops



Recurrent Network

Each layer receives input from the previous layer and the output from the previous time step (i.e., time-delayed connections)



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Recurrent Neural Networks (RNNs)

- Main idea: use hidden state to capture information about the past

Recurrent: same function is applied to previous result

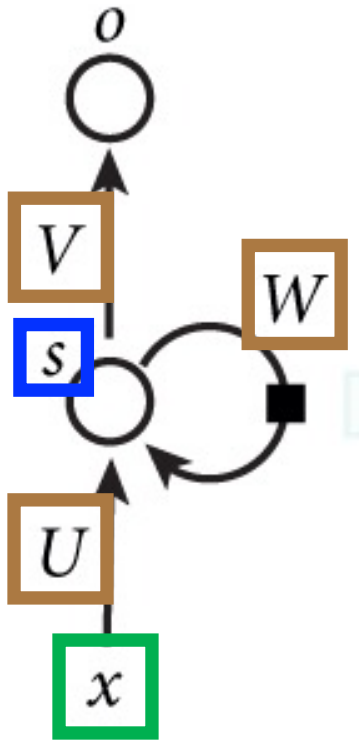
Model parameters

$$s_t = f_m(s_{t-1}, x_t)$$

New state — Old state — Input at time step

Recurrent Network

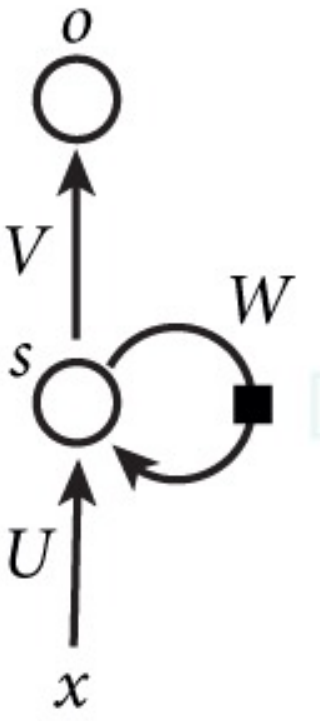
Each layer receives input from the previous layer and the output from the previous time step (i.e., time-delayed connections)



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Time Step 1

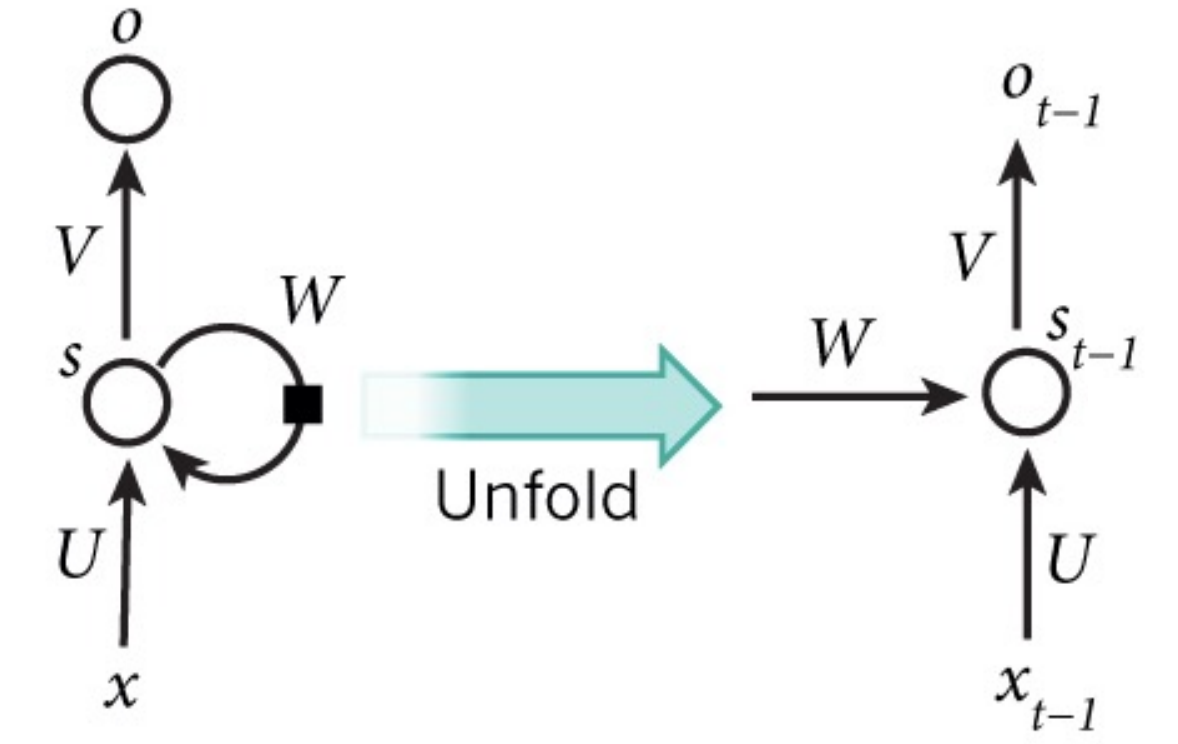
- Main idea: use hidden state to capture information about the past



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Time Step 1

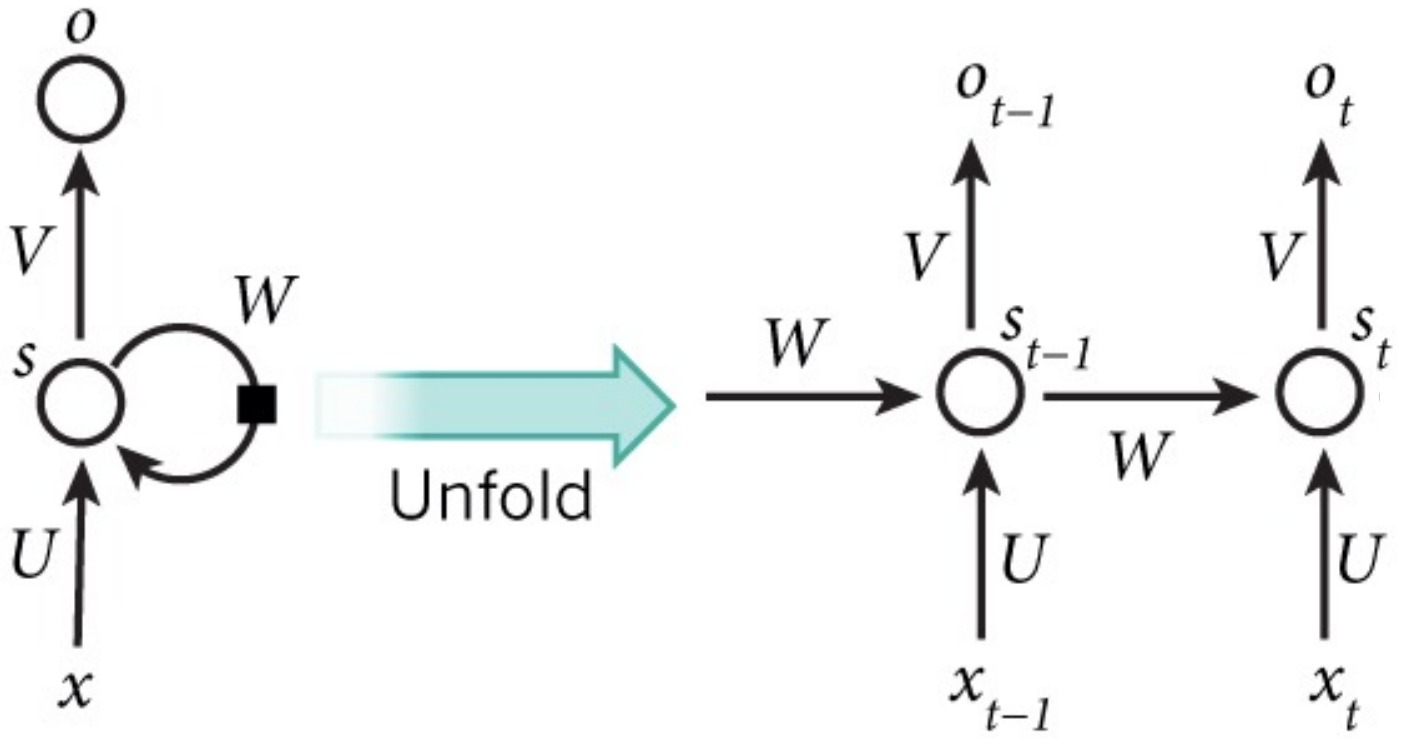
- Main idea: use hidden state to capture information about the past



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Time Step 2

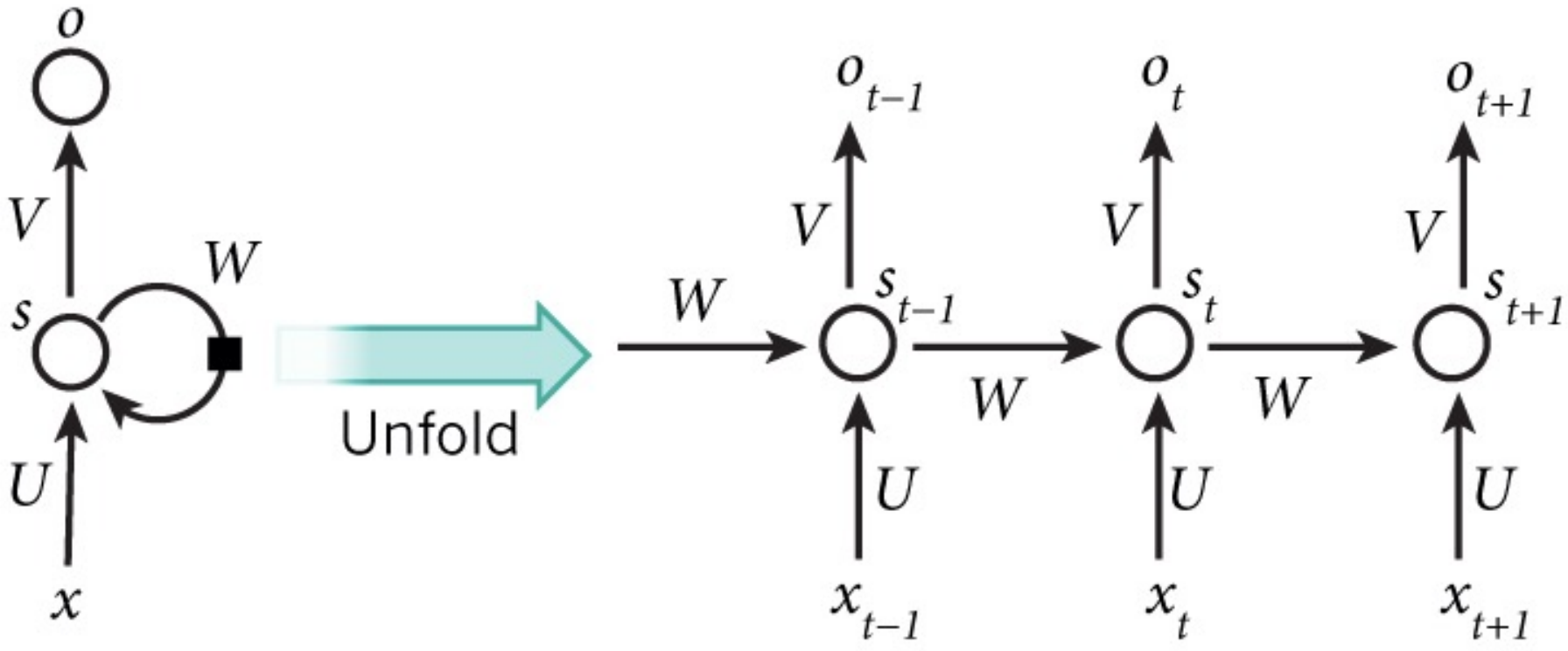
- Main idea: use hidden state to capture information about the past



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Time Step 3

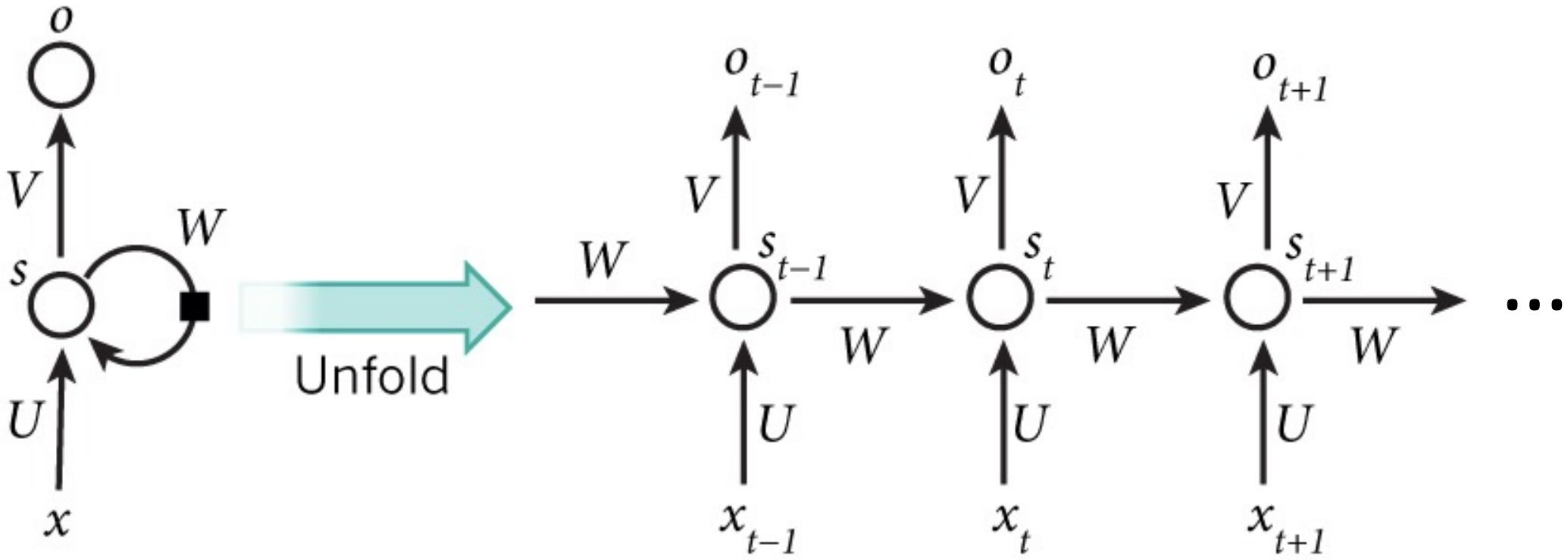
- Main idea: use hidden state to capture information about the past



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: And So On…

- Main idea: use hidden state to capture information about the past



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Model Parameters and Inputs

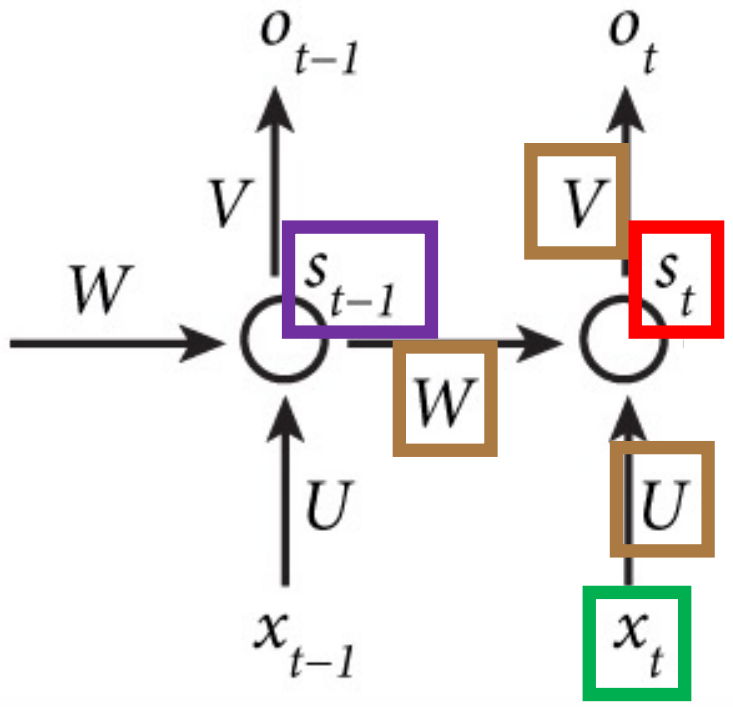
- Main idea: use hidden state to capture information about the past

Recurrence formula applied at every time step:

Model parameters

$$s_t = f_m(s_{t-1}, x_t)$$

New state | Old state | Input at time step



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Model Parameters and Inputs

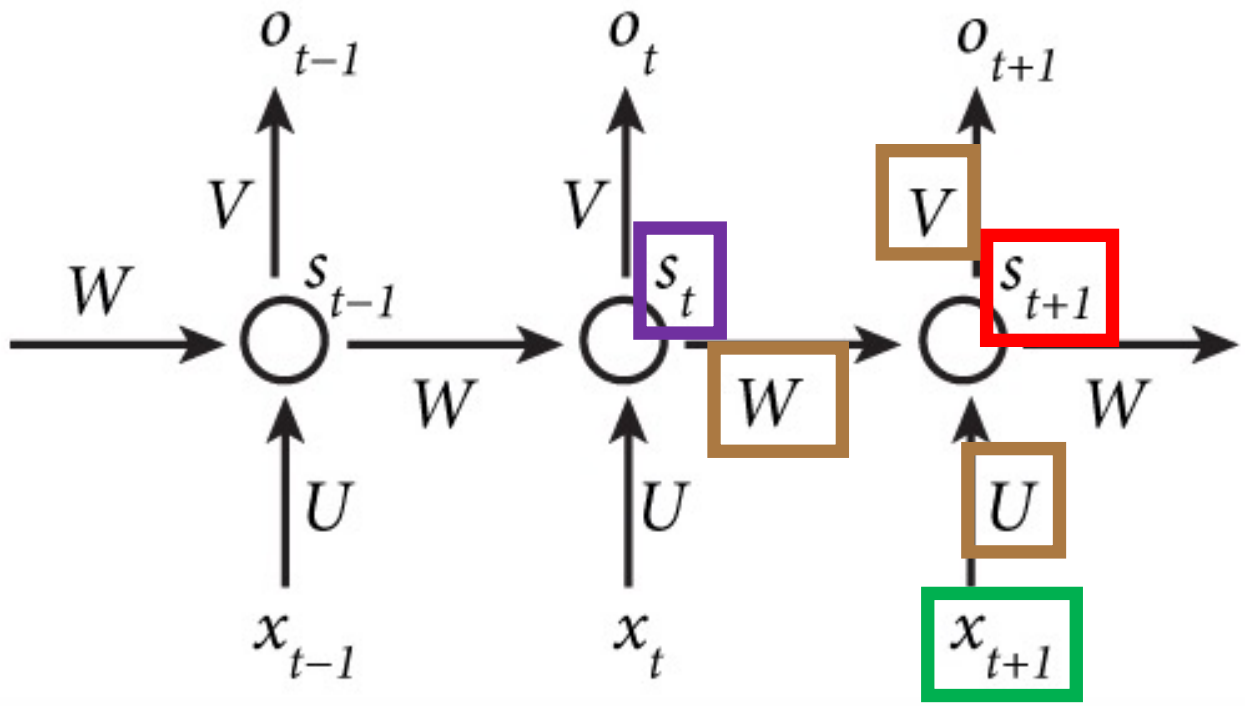
- Main idea: use hidden state to capture information about the past

Recurrence formula applied at every time step:

Model parameters

$$s_t = f_m(s_{t-1}, x_t)$$

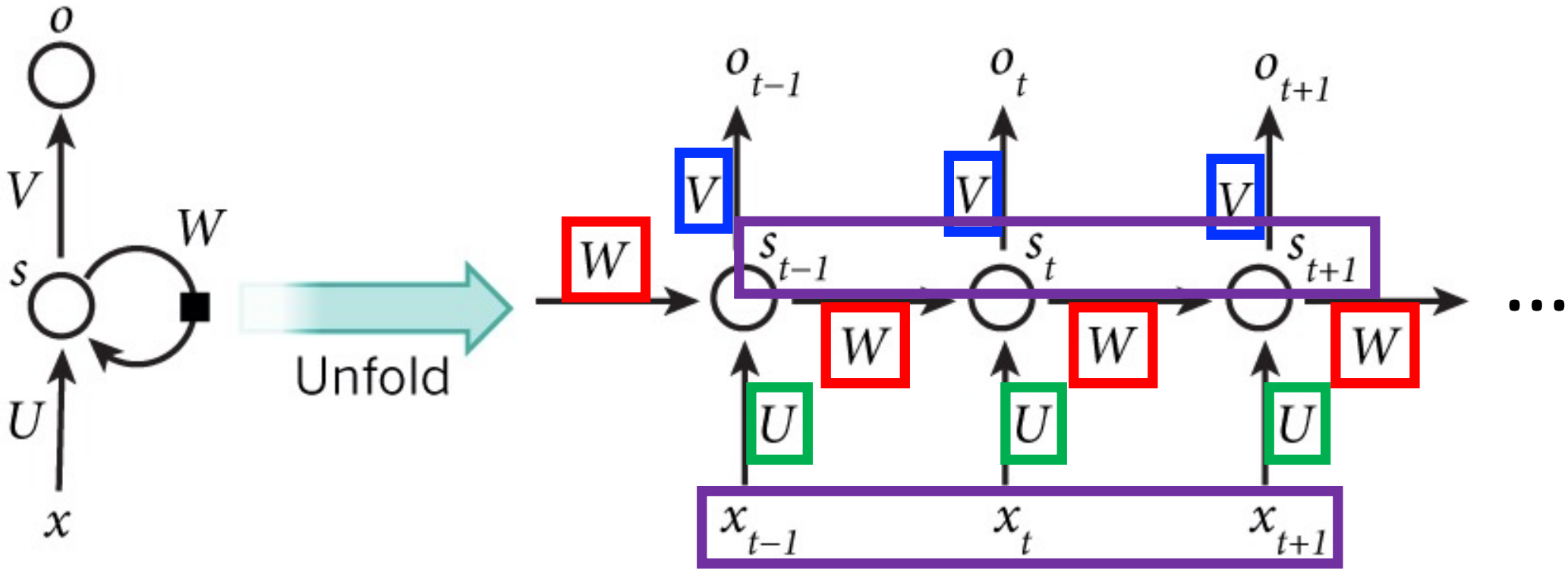
New state | Old state | Input at time step



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Model Parameters and Inputs

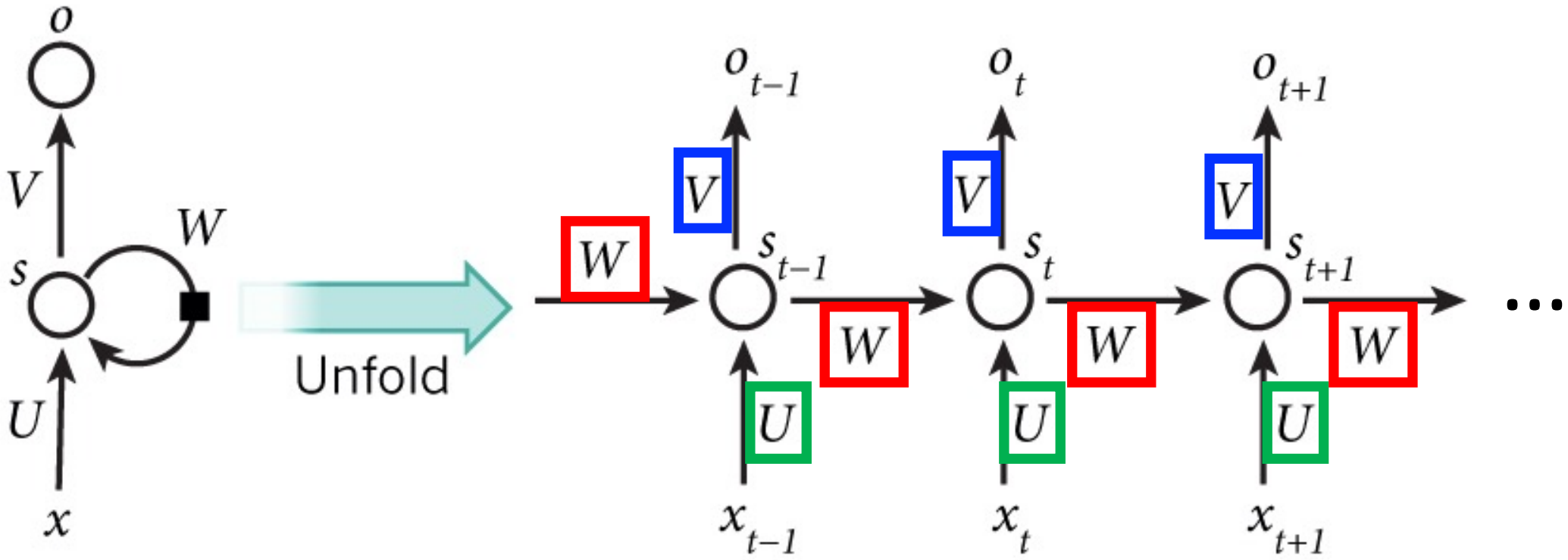
- All layers share the same model parameters (U, V, W)
  - What is different between the layers?



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Model Parameters and Inputs

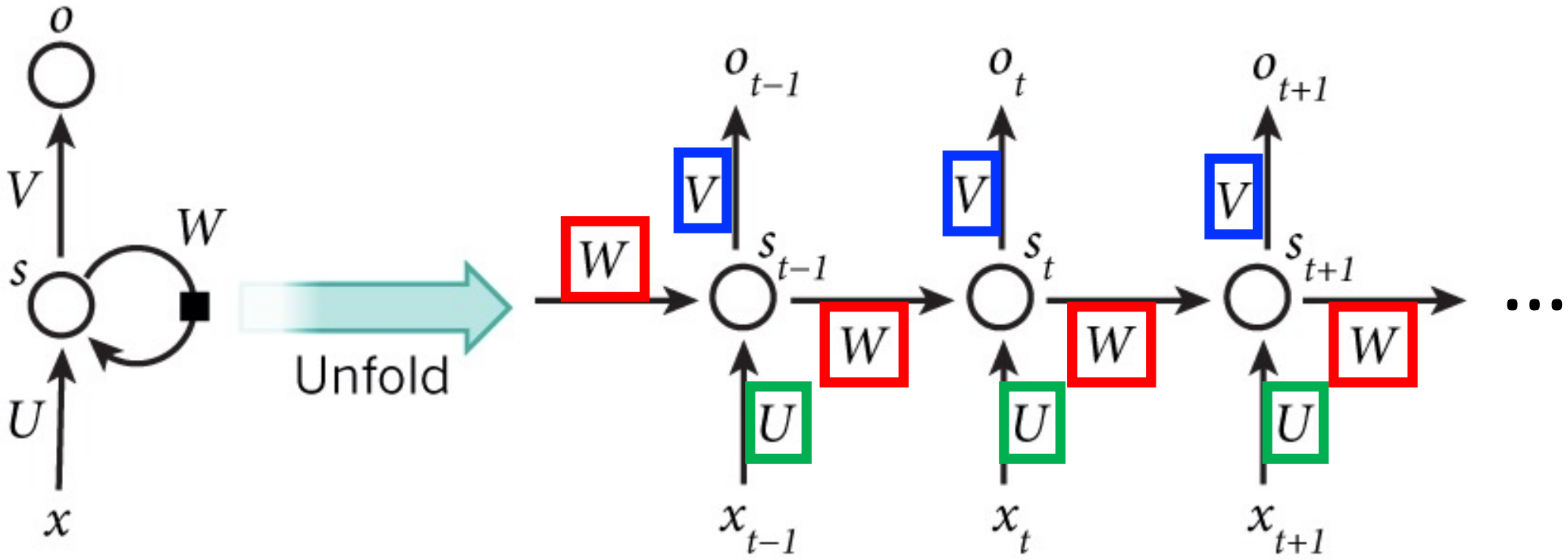
- When unfolded, a RNN is a deep feedforward network with shared weights!



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Advantages

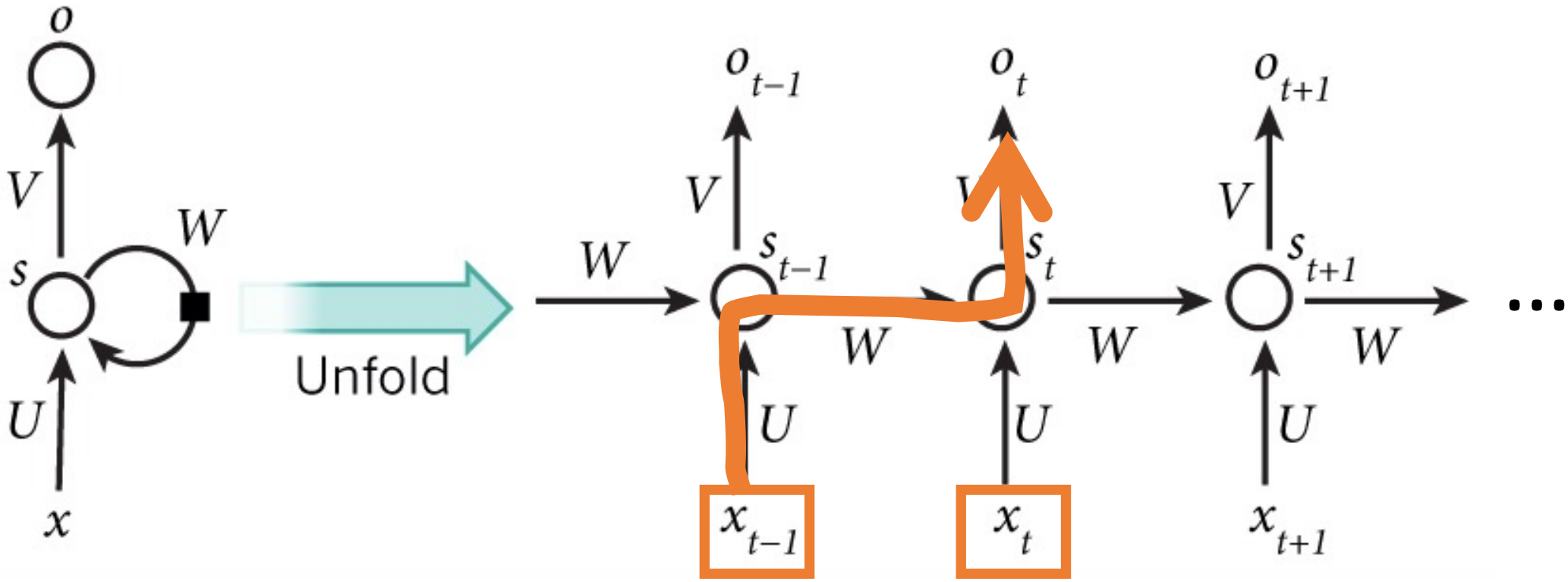
- Overcomes problem that weights of each layer are learned **independently** by using previous hidden state
- Overcomes problem that model has many parameters by sharing the same weights for all input sizes



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Advantages

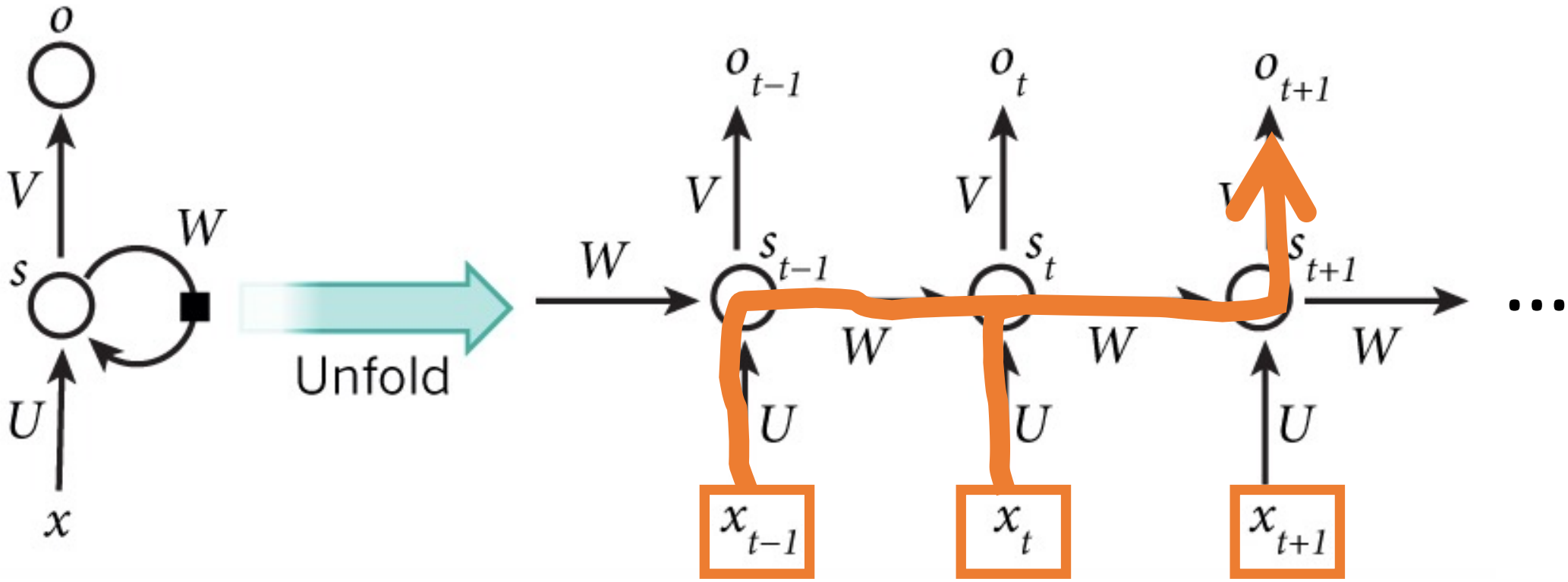
- Retains information about past inputs for an amount of time that depends on the model's weights and input data rather than a fixed duration selected a priori



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Advantages

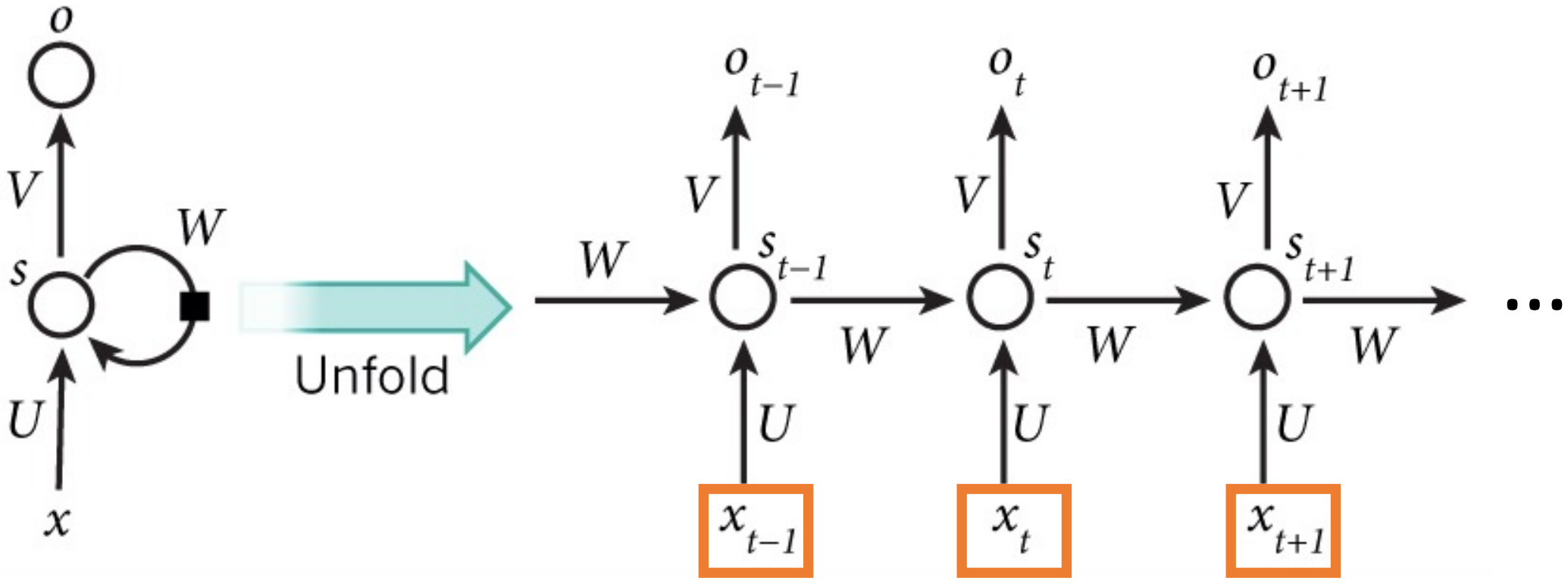
- Retains information about past inputs for an amount of time that depends on the model's weights and input data rather than a fixed duration selected a priori



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN: Advantages

- Can theoretically handle any input size by unfolding less/more time steps



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RNN Example: Predict Sequence of Characters

- Goal: predict next character in text (i.e., automatic text completion)
- Training Data: sequence of characters represented as one-hot vectors

RNN Example: Predict Sequence of Characters; e.g., To Write a Wikipedia Page

Training Input

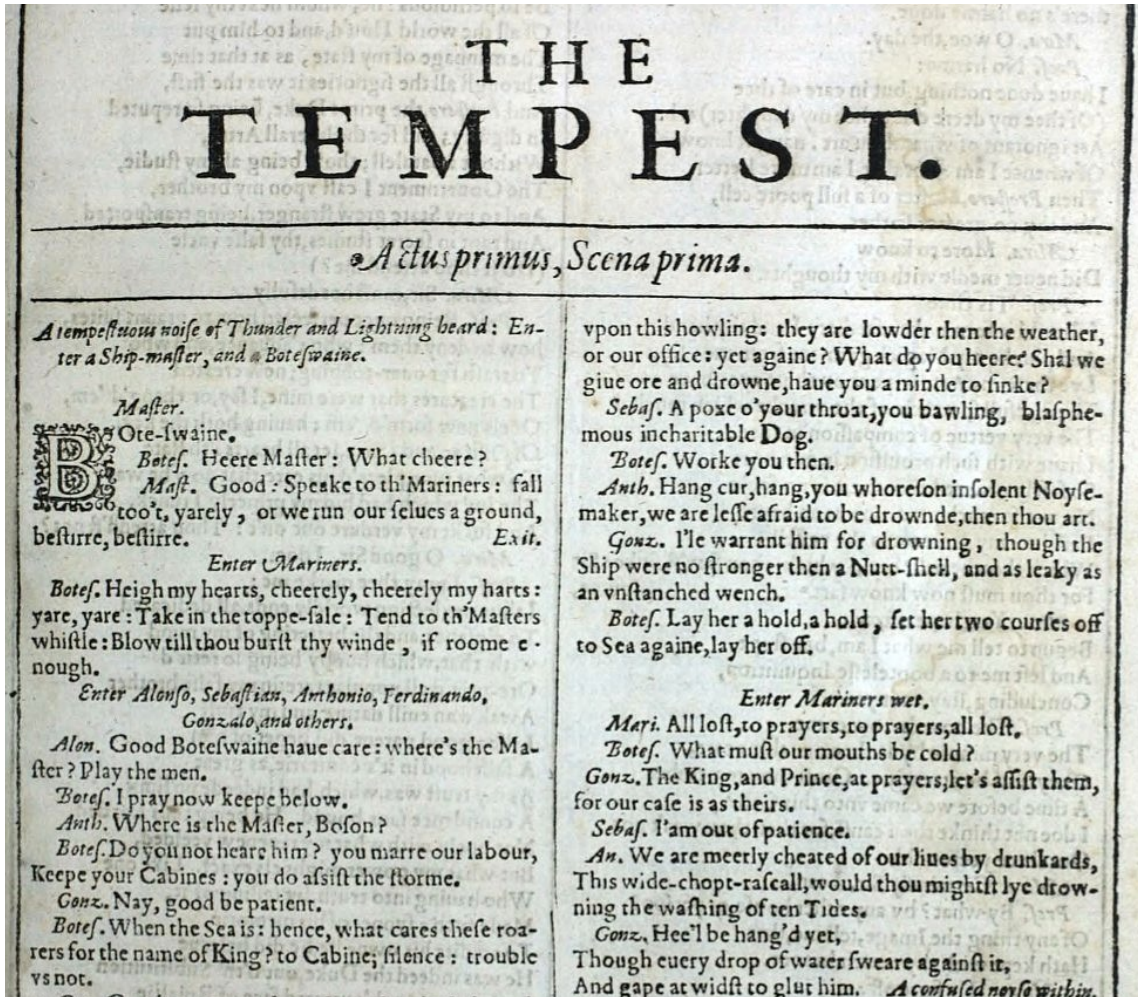
Predicted Output

```
Naturalism and decision for the majority of Arab countries' capitalide was grounded
by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated
with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal
in the [[Protestant Immineners]], which could be said to be directly in Cantonese
Communication, which followed a ceremony and set inspired prison, training. The
emperor travelled back to [[Antioch, Perth, October 25|21]] to note, the Kingdom
of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known
in western [[Scotland]], near Italy to the conquest of India with the conflict.
Copyright was the succession of independence in the slop of Syrian influence that
was a famous German movement based on a more popular servicious, non-doctrinal
and sexual power post. Many governments recognize the military housing of the
[[Civil Liberalization and Infantry Resolution 265 National Party in Hungary]],
that is sympathetic to be to the [[Punjab Resolution]]
(PJS)[http://www.humah.yahoo.com/guardian.
cfm/7754800786d17551963s89.htm Official economics Adjoint for the Nazism, Montgomery
was swear to advance to the resources for those Socialism's rule,
was starting to signing a major tripad of aid exile.]]
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNN Example: Predict Sequence of Characters; e.g., To Write Like Shakespeare

Training Input (All Works of Shakespeare)



THE
TEMPEST.

Actus primus, Scena prima.

A tempestuous noise of Thunder and Lightning heard: Enter a Ship-master, and a Boteswaine.

Master.

BOte-swaine.

Bote. Heere Master: What cheere?

Mast. Good: Speake to th'Mariners: fall too't, yarely, or we run our selues a ground, bestirre, bestirre. *Exit.*

Enter Mariners.

Bote. Heigh my hearts, cheerely, cheerely my harts: yare, yare: Take in the toppe-sale: Tend to th'Masters whistle: Blow till thou burst thy winde, if roome enough.

Enter Alonso, Sebastian, Anthonio, Ferdinando, Gonzalo, and others.

Alon. Good Boteswaine haue care: where's the Master? Play the men.

Bote. I pray now keepe below.

Anth. Where is the Master, Boson?

Bote. Do you not heare him? you marre our labour, Keepe your Cabines: you do assist the storme.

Gonz. Nay, good be patient.

Bote. When the Sea is: hence, what cares these roarers for the name of King? to Cabine; silence: trouble vs not.

Gonz. Good, yet remember whom thou hast aboord.

Bote. None that I more loue then my selfe. You are a Counsellor, if you can command these Elements to silence, and worke the peace of the present, we will not hand a rope more; vse your authority: if you cannot, giue thankes you haue liu'd so long, and make your selfe ready in your Cabine for the mischance of the houre, if it so hap. Cheerely good hearts: Out of our way I say.

Gonz. I haue great comfort from this fellow: Methinks he hath no drowning marke vpon him; his complexion is perfect Gallowes. Stand fast good Fate to his hanging: make the rope of his destiny our cable, for our owne doth little aduantage. If he be not borne to be hang'd, our case is miserable.

Exeunt.

Enter Boteswaine.

Bote. Downe with the Top-mast: yare, lower, lower, bring her to try with main-course. A plague vpon this howling: they are louder then the weather, or our office: yet againe? What do you heere? Shal we giue ore and drowne, haue you a minde to sinke?

Sebast. A poxe o'your throat, you bawling, blasphemous incharitable Dog.

Bote. Worke you then.

Anth. Hang cur, hang, you whoreson insolent Noyse-maker, we are leſſe afraid to be drownde, then thou art.

Gonz. I'le warrant him for drowning, though the Ship were no stronger then a Nutt-shell, and as leaky as an vnstanched wench.

Bote. Lay her a hold, a hold, set her two courses off to Sea againe, lay her off.

Enter Mariners wet.

Mari. All lost, to prayers, to prayers, all lost.

Bote. What must our mouths be cold?

Gonz. The King, and Prince, at prayers, let's assist them, for our case is as theirs.

Sebast. I'am out of patience.

An. We are merely cheated of our liues by drunkards, This wide-chopt-rascall, would thou mightst lye drowning the washing of ten Tides.

Gonz. Hee'l be hang'd yet, Though euery drop of water sweare against it, And gape at widst to glut him. *A confused noyse within.*

Predicted Output

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fe
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/ https://github.com/martin-gorner/tensorflow-rnn-shakespeare

RNN Example: Predict Sequence of Characters; e.g., To Write Code

Training Input (C code on GitHub)

```c
/*
 * Bad block management
 *
 * - Heavily based on MD badblocks code from Neil Brown
 *
 * Copyright (c) 2015, Intel Corporation.
 *
 * This program is free software; you can redistribute it and/or modify it
 * under the terms and conditions of the GNU General Public License,
 * version 2, as published by the Free Software Foundation.
 *
 * This program is distributed in the hope it will be useful, but WITHOUT
 * ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or
 * FITNESS FOR A PARTICULAR PURPOSE.  See the GNU General Public License for
 * more details.
 */

#include <linux/badblocks.h>
#include <linux/seqlock.h>
#include <linux/device.h>
#include <linux/kernel.h>
#include <linux/module.h>
#include <linux/stddef.h>
#include <linux/types.h>
#include <linux/slab.h>
```

Predicted Output

```c
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
  int error;
  if (fd == MARN_EPT) {
    /*
     * The kernel blank will coeld it to userspace.
     */
    if (ss->segment < mem_total)
      unblock_graph_and_set_blocked();
    else
      ret = 1;
    goto bail;
  }
  segaddr = in_SB(in.addr);
  selector = seg / 16;
  setup_works = true;
  for (i = 0; i < blocks; i++) {
    seq = buf[i++];
    bpf = bd->bd.next + i * search;
    if (fd) {
      current = blocked;
    }
  }
  rw->name = "Getjbbregs";
  bprm_self_clearl(&iv->version);
  regs->new = blocks[(BPF_STATS << info->historidac)] | PFMR_CLOBATHINC_SECONDS << 12;
  return segtable;
}
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/ https://github.com/martin-gorner/tensorflow-rnn-shakespeare

RNN Example: Predict Sequence of Characters; e.g., To Write Facebook Messages

Training Input

Facebook messages from…

Hendrik J. Weideman

Ph.D. candidate at Rensselaer Polytechnic Institute

Predicted Output

1. *The meaning of life is to* find them? Oh, I don't know if I would be able to publish a paper on that be climbing today, but it will definitely know what that makes sense. I'm sure they wanted to socialis that I am bringing or

2. *What a cruel twist of fate, that we should* be persuate that 😂 And cook :D I will think that's mean I think I need to go to the phoebe? That's awesome though Haha, sorry, I don't know if it was more time to clas for it's badass though I jus

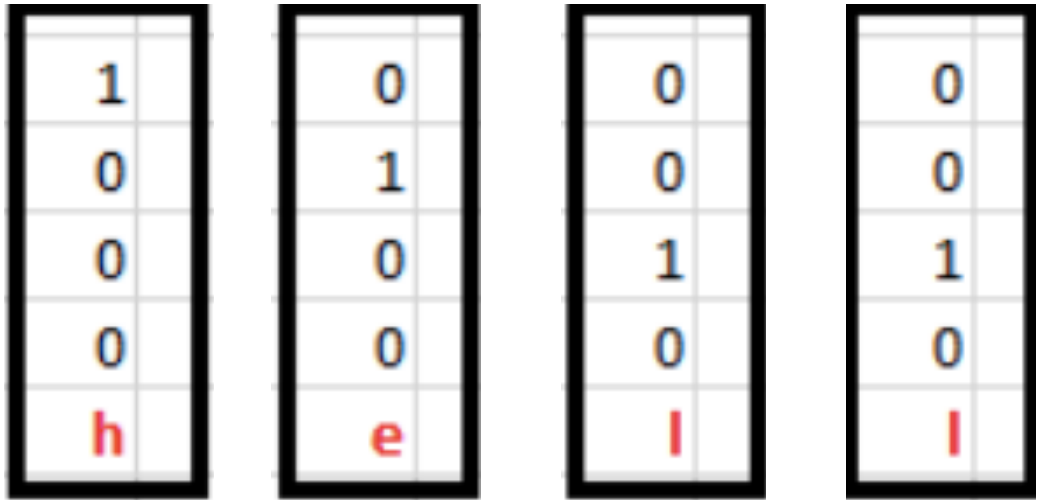
3. *The fact of the matter is* just the world to invite your stuff? I don't know how to right it wouldn't be as offriving for anything, so that would be awesome, thanks :) I have no idea... She would get to worry about it :P And I

4. *At the very least, you should remember that* as a house of a perfect problems 😂 Yeah :D I wonder how perfect for this trank though So it's probably foltower before the bathers will be fine and haven't want to make it worse Thanks for one of

https://hjweide.github.io/char-rnn

Example: Predict Sequence of Characters

- Goal: predict next character in text
- Prediction: feed training sequence of one-hot encoded characters; e.g., “hello”
  - For simplicity, assume the following vocabulary (i.e., character set): {h, e, l, o}
  - What is our input at time step 1?
  - What is our input at time step 2?
  - What is our input at time step 3?
  - What is our input at time step 4?
  - And so on…

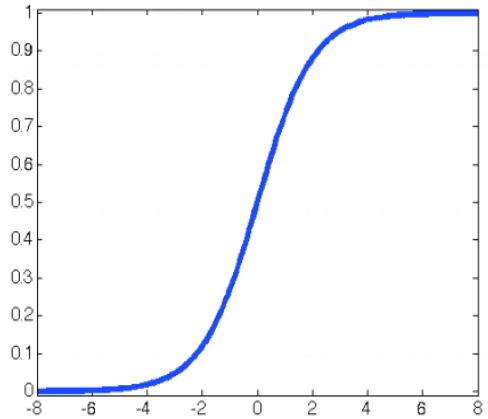


https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

Example: Predict Sequence of Characters

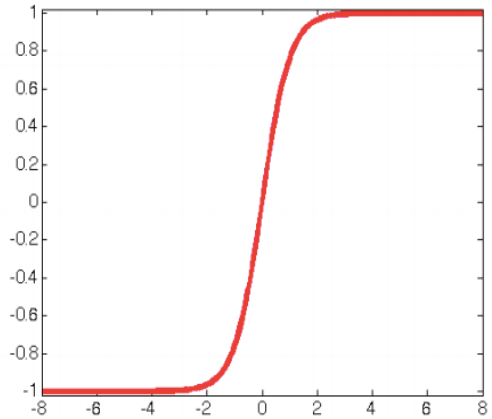
Recall activation functions: tanh is common choice for RNNs

Sigmoid



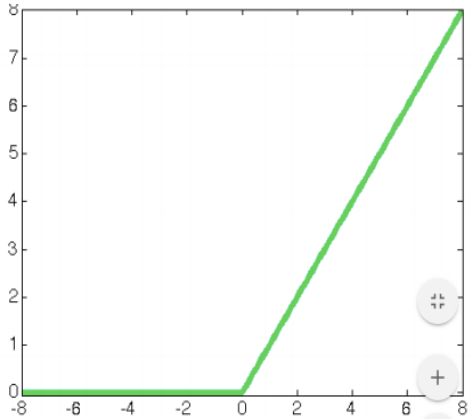
$$\sigma(z) = \frac{1}{1+\exp(-z)}$$

Tanh



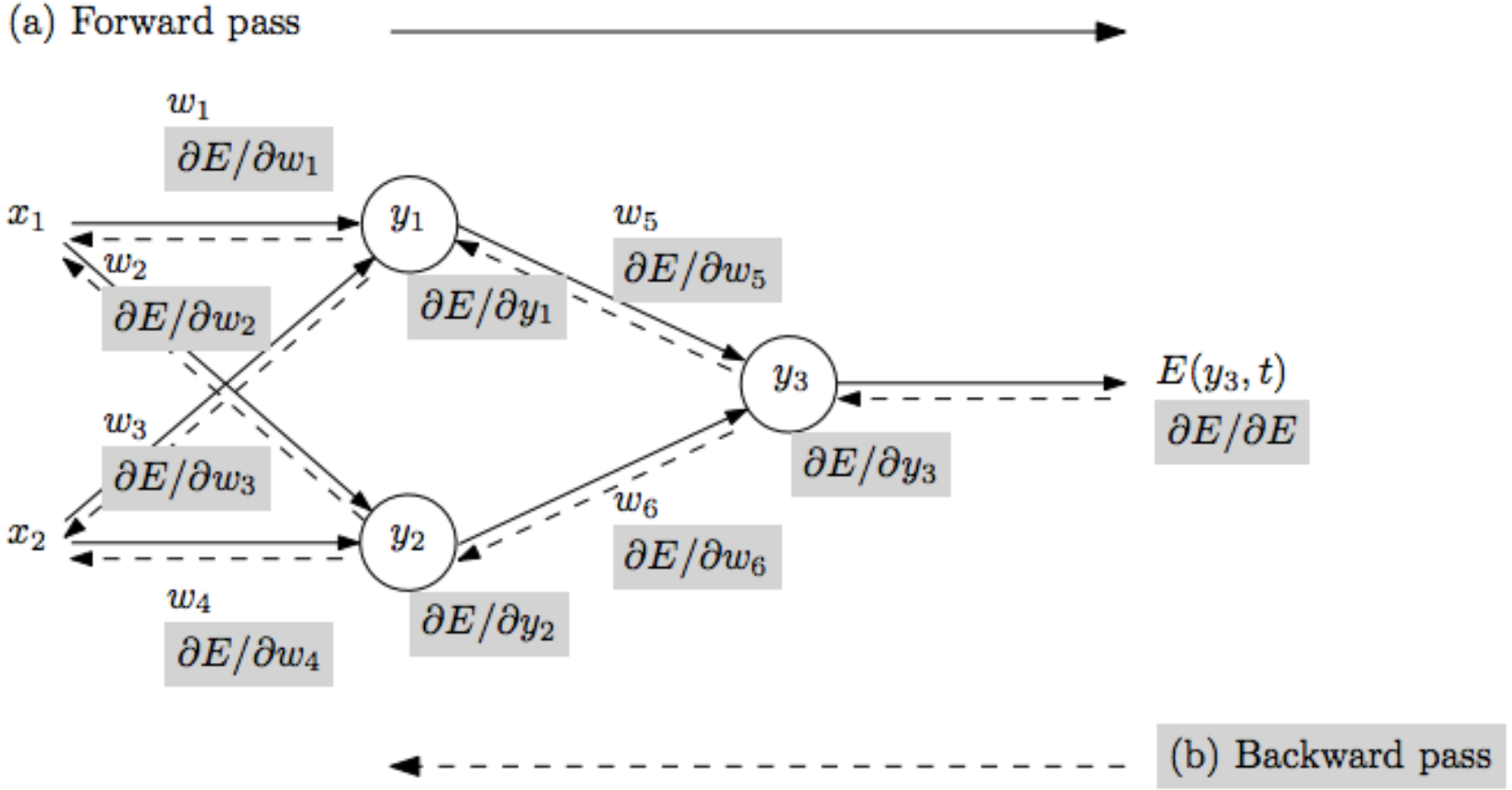
$$\tanh(z) = \frac{\exp(z)-\exp(-z)}{\exp(z)+\exp(-z)}$$

ReLU



$$\mathrm{ReLU}(z) = \max(0, z)$$

Training Approach



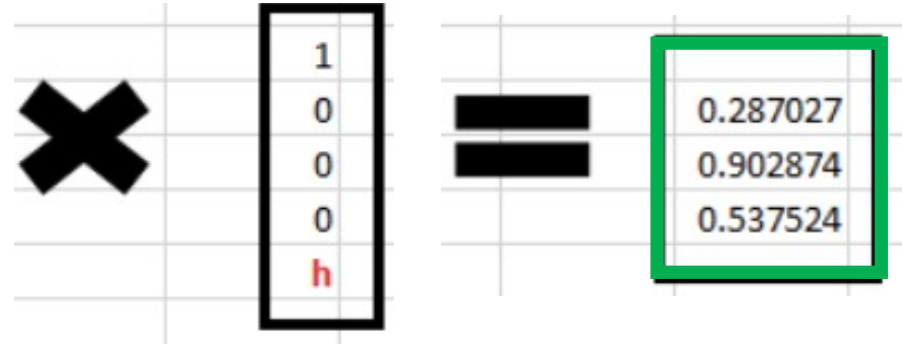
- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through unfolded network to predict
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter; weight sharing accounted for by summing gradients for all time steps
  4. Update each parameter using calculated gradients

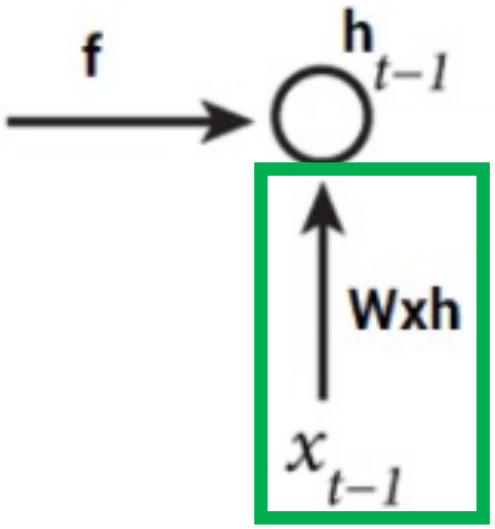
Figure from: Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, Jeffrey Mark Siskind; Automatic Differentiation in Machine Learning: a Survey; 2018

Example: Forward Pass

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + \text{bias})$$

wxh			
0.287027	0.84606	0.572392	0.486813
0.902874	0.871522	0.691079	0.18998
0.537524	0.09224	0.558159	0.491528





Why do we use three rows instead of 1 row for the weight matrix?

- provides more model parameters and so can generate a more complex function; the number of rows is a hyperparameter set by the developer

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

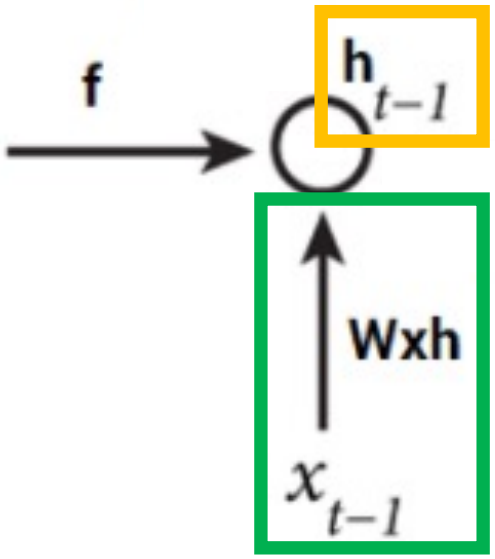
Example: Forward Pass

Initialize to random value: 0.567001

$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t + \text{bias})$

Initialize to 0

Input at next time step



wxh			
0.287027	0.84606	0.572392	0.486813
0.902874	0.871522	0.691079	0.18998
0.537524	0.09224	0.558159	0.491528

1			0.287027
0			0.902874
0			0.537524
0			
h			

Ht	=	TANH	0.287027359 0.902874425 0.537523791	+ 0.567001 0.567001 0.567001
	=		0.693168 0.899554 0.802118	

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

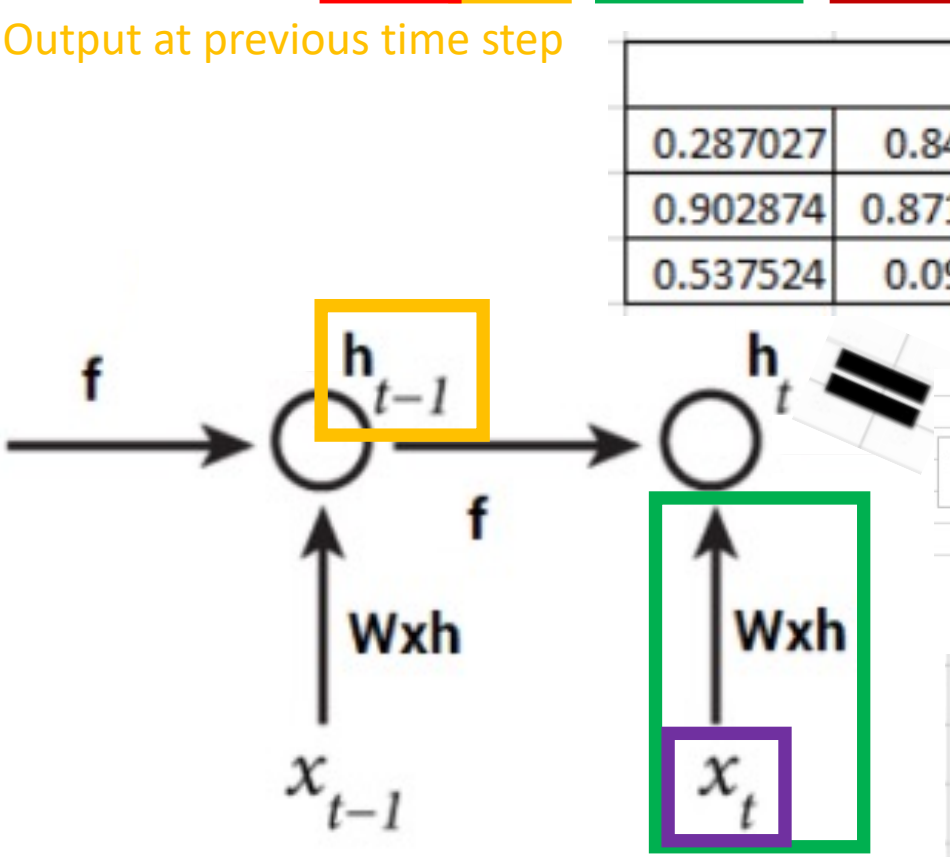
Example: Forward Pass

Initialize to random value: 0.427043

Recall: Initialized to random value: 0.567001

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t + \text{bias})$$

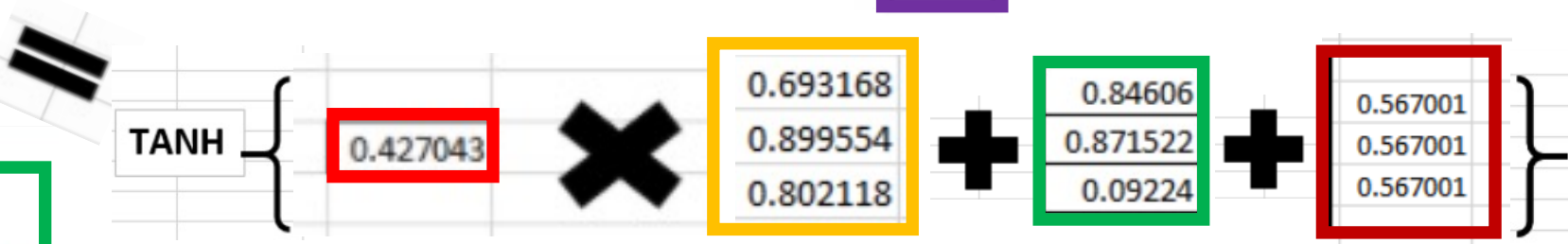
Output at previous time step

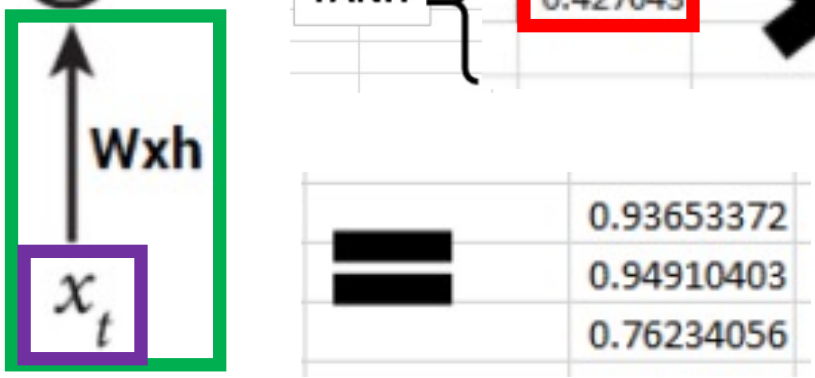


wxh			
0.287027	0.84606	0.572392	0.486813
0.902874	0.871522	0.691079	0.18998
0.537524	0.09224	0.558159	0.491528

0
1
0
0
e

0.84606
0.871522
0.09224

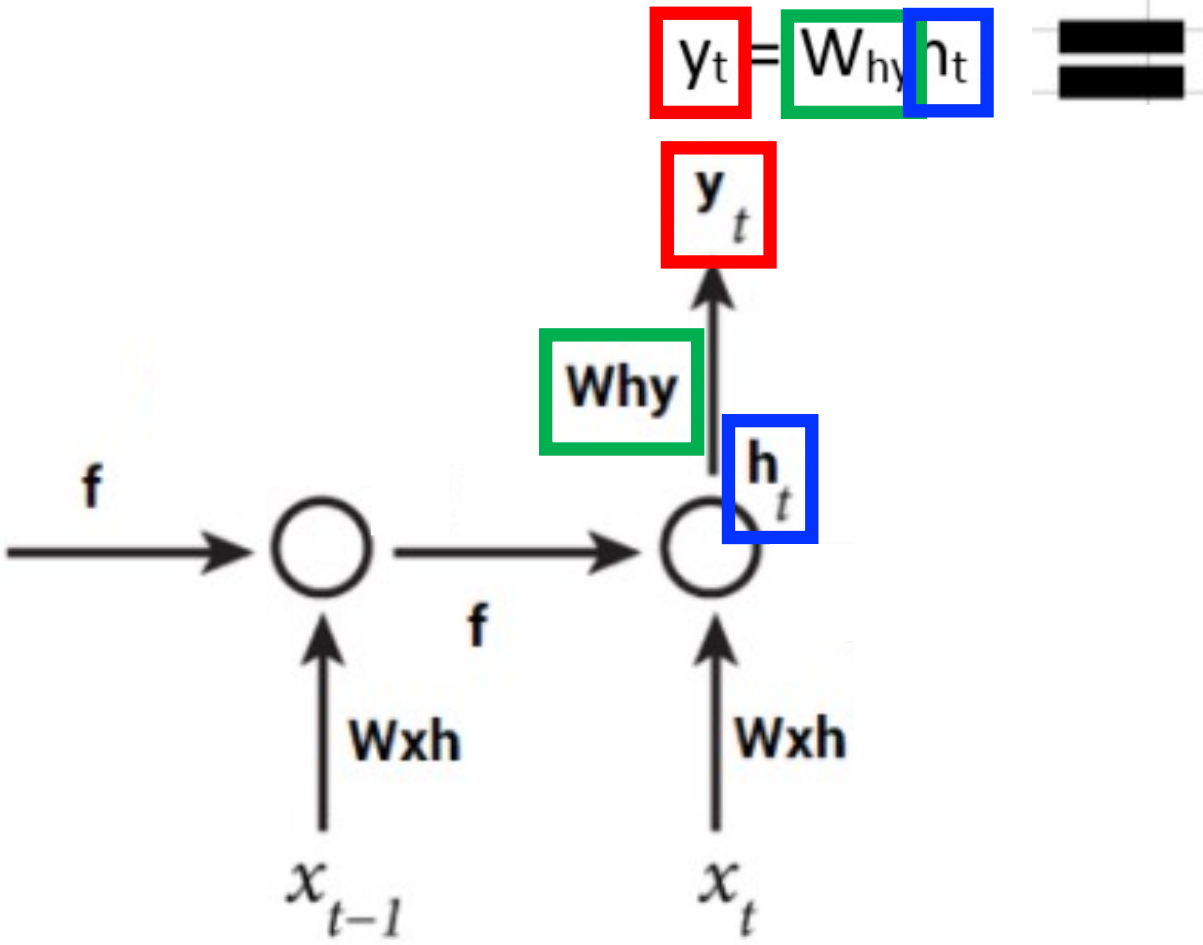




https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

Example: Forward Pass

(What Character Follows “he”?)



(assume bias term is 0 for this example)

why		
0.37168	0.974829459	0.830034886
0.39141	0.282585823	0.659835709
0.64985	0.09821557	0.334287084
0.91266	0.32581642	0.144630018

×

Ht
0.936534
0.949104
0.762341

=

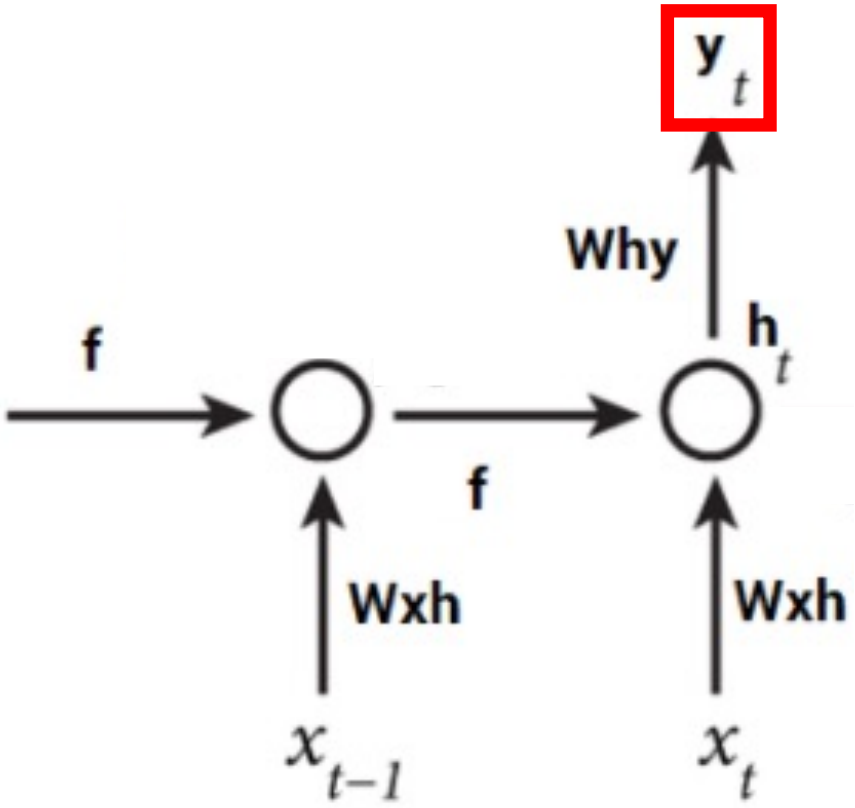
yt
1.90607732
1.13779113
0.95666016
1.27422602

Applying softmax, to compute letter probabilities:

0.419748
0.194682
0.162429
0.223141

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

Example: Forward Pass



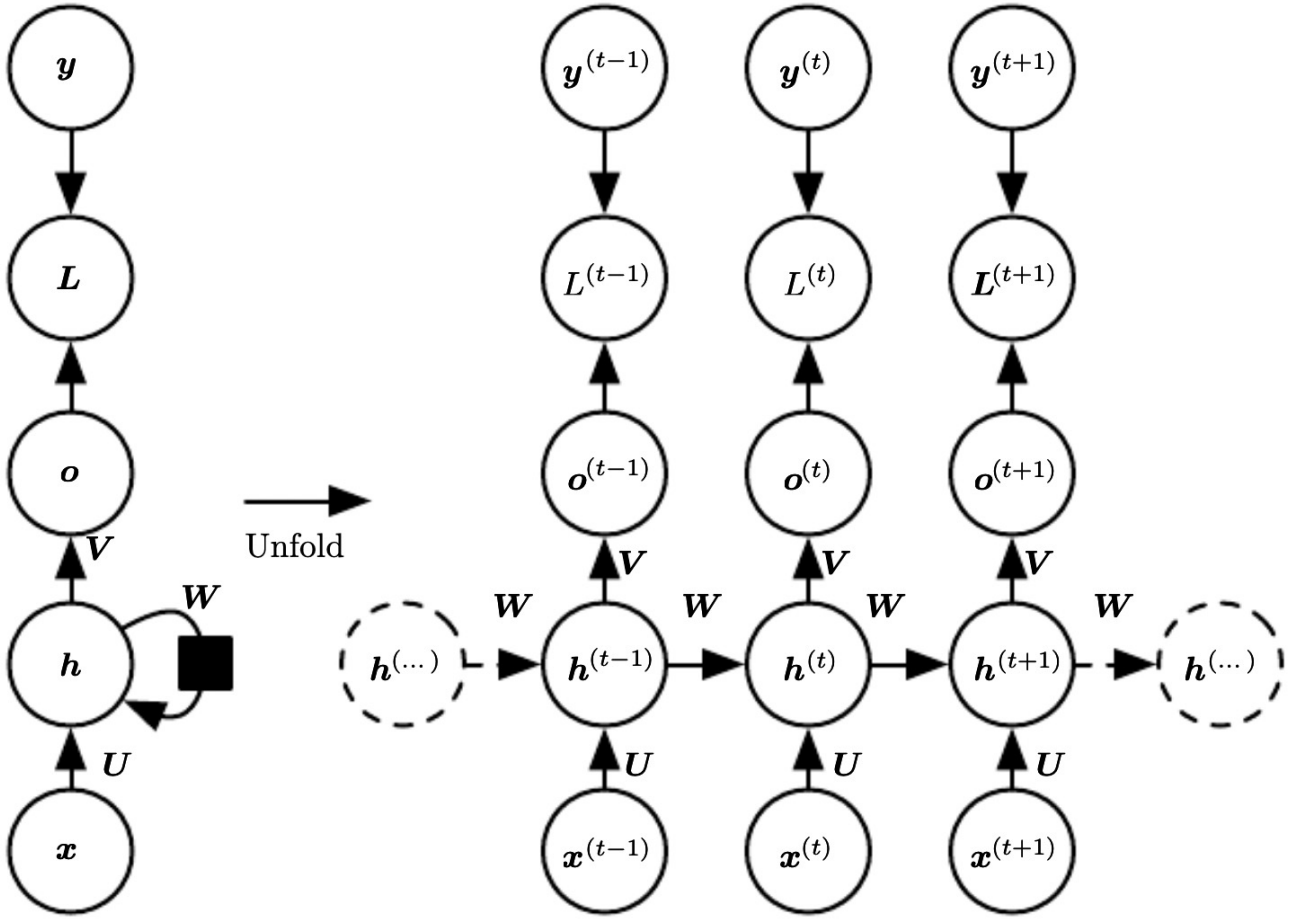
Given our vocabulary is {h, e, l, o}, what letter is predicted?

Applying softmax, to compute letter probabilities:

0.419748
0.194682
0.162429
0.223141

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

Training Approach; e.g., Many-to-Many Loss Using Cross Entropy Loss



- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through unfolded network to predict
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter; weight sharing accounted for by summing gradients for all time steps
  4. Update each parameter using calculated gradients

Goodfellow, Bengio, and Courville, Deep Learning.

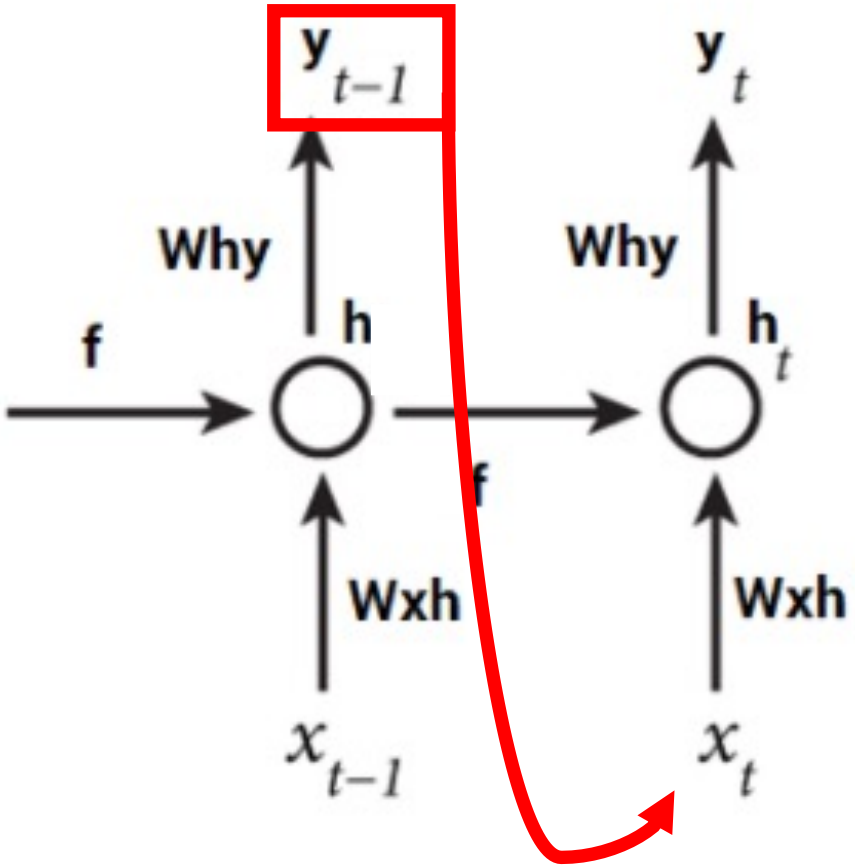
Training Approach

Called back-propagation through time (BPTT), it is the same generalized gradient derivation process we already have covered

- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through unfolded network to predict
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter; weight sharing accounted for by summing gradients for all time steps
  4. Update each parameter using calculated gradients

Figure from: Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, Jeffrey Mark Siskind; Automatic Differentiation in Machine Learning: a Survey; 2018

Test Time: Predict Sequence of Characters

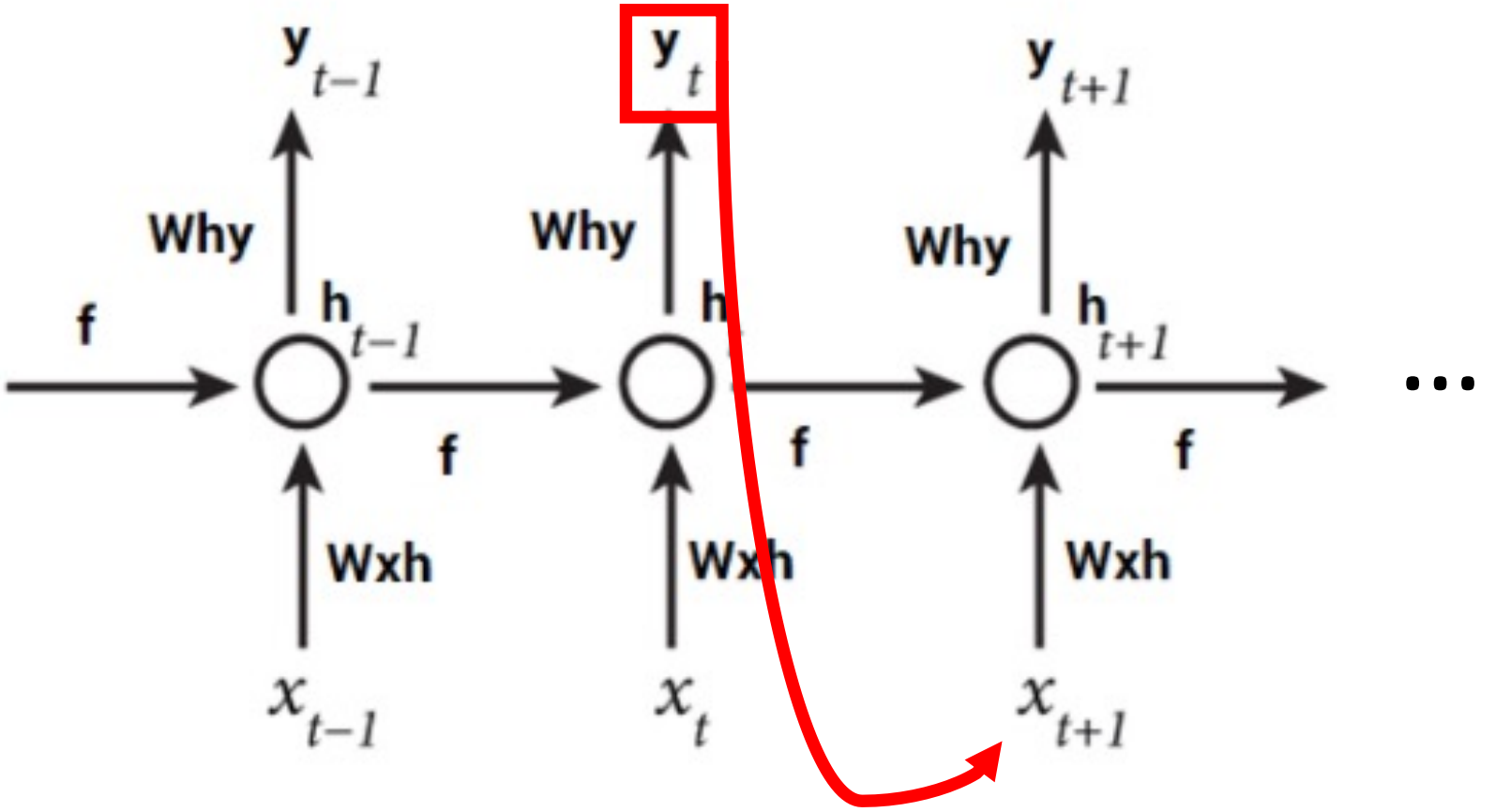


…

Feed previous prediction as input for next time step

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

Test Time: Predict Sequence of Characters

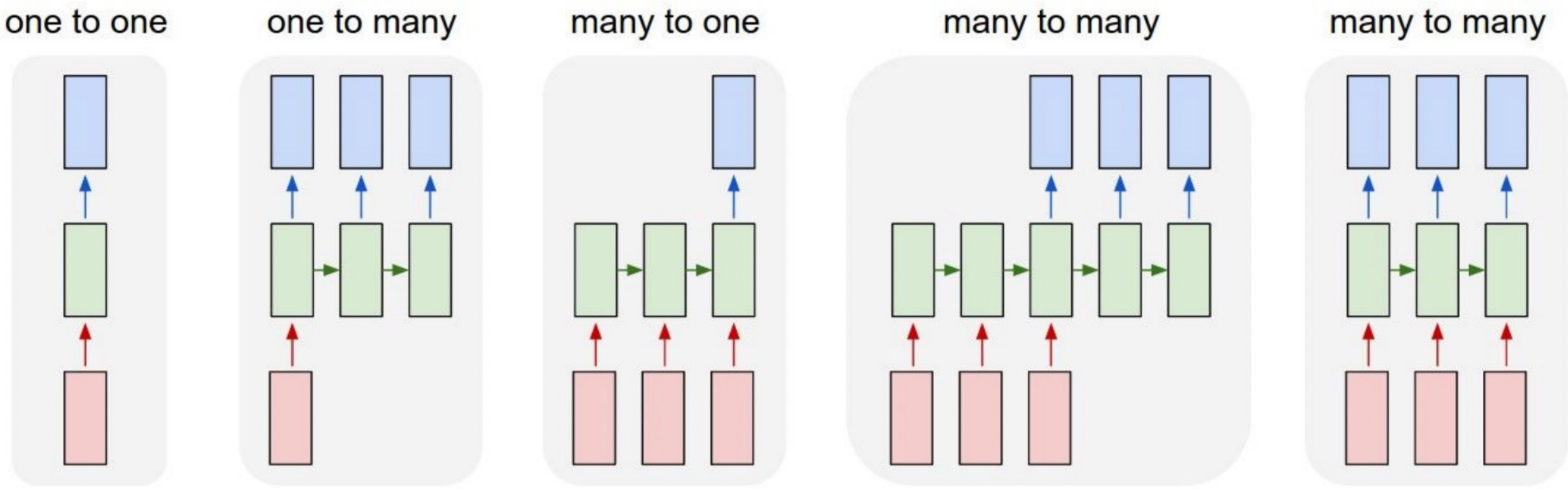


...

Feed previous prediction as input for next time step

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

RNN Variants: Variable Input/Output Lengths



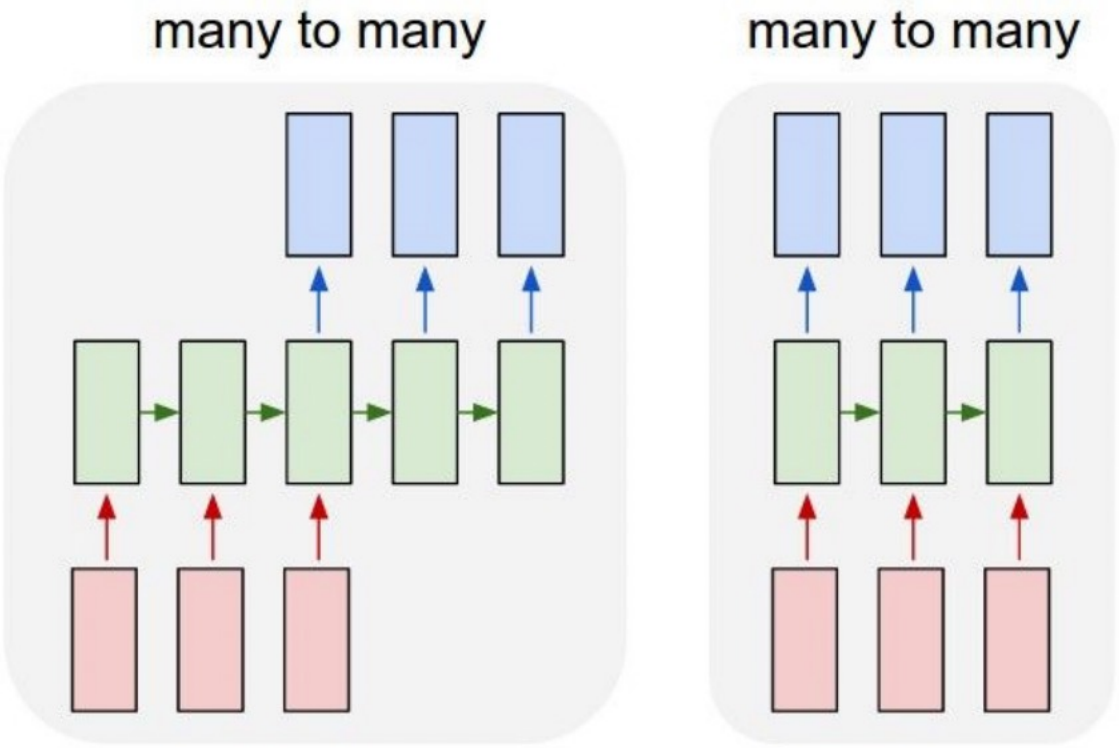
Which variant would you use for text classification?

http://cs231n.stanford.edu/slides/2016/winter1516_lecture10.pdf

RNN Variants: Variable Input/Output Lengths

Why would you choose the first variant of “many to many” over the second?

- Need full context before predicting; e.g., dialogue or Q&A

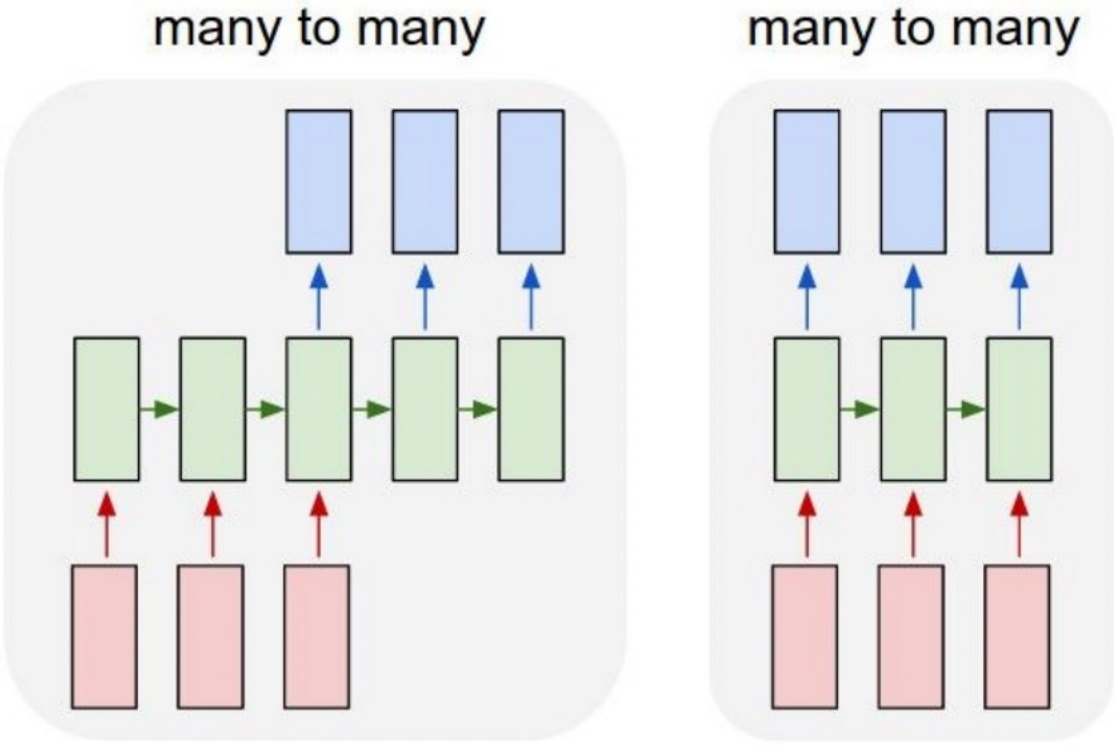


http://cs231n.stanford.edu/slides/2016/winter1516_lecture10.pdf

RNN Variants: Variable Input/Output Lengths

Why would you choose the latter variant of “many to many” over the former?

- Full context not needed to predict; e.g., named entity recognition

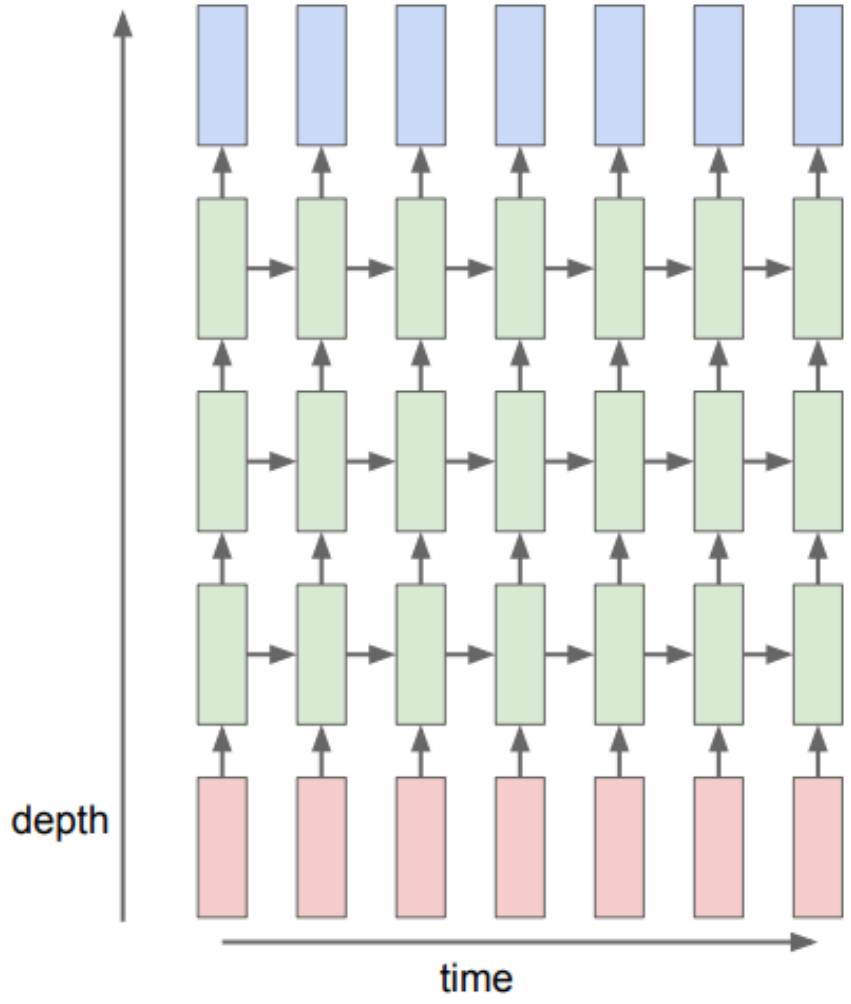


$[\text{Jim}]_{\text{Person}}$ bought 300 shares of $[\text{Acme Corp.}]_{\text{Organization}}$ in $[\text{2006}]_{\text{Time}}$

https://en.wikipedia.org/wiki/Named-entity_recognition

http://cs231n.stanford.edu/slides/2016/winter1516_lecture10.pdf

RNN Variants: Different Number of Hidden Layers and Number of Nodes Per Layer



Captures information about past, via hidden states, with more complex representations

Experimental evidence suggests deeper models can perform better:

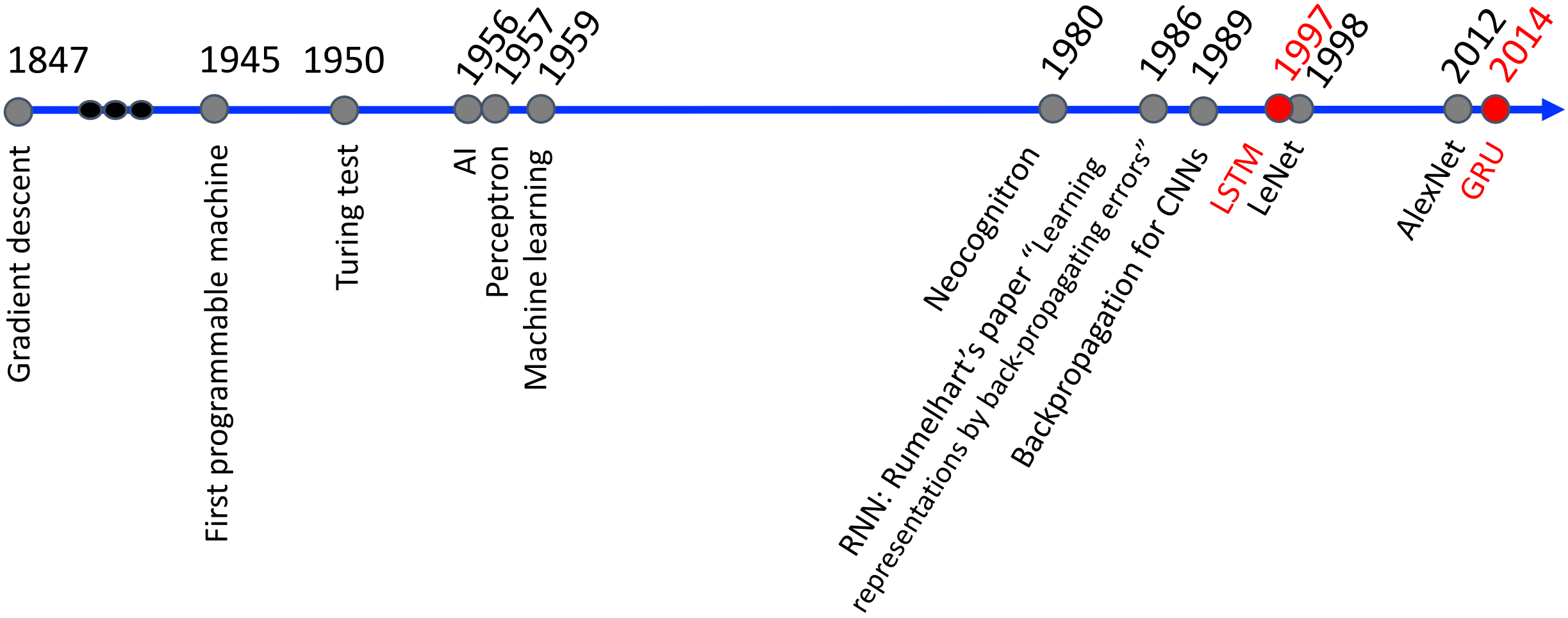
- Graves et al.; Speech Recognition with Deep Recurrent Neural Networks; 2013.
- Pascanu et al.; How to Construct Deep Recurrent Neural Networks; 2014.

http://cs231n.stanford.edu/slides/2016/winter1516_lecture10.pdf

Today's Topics

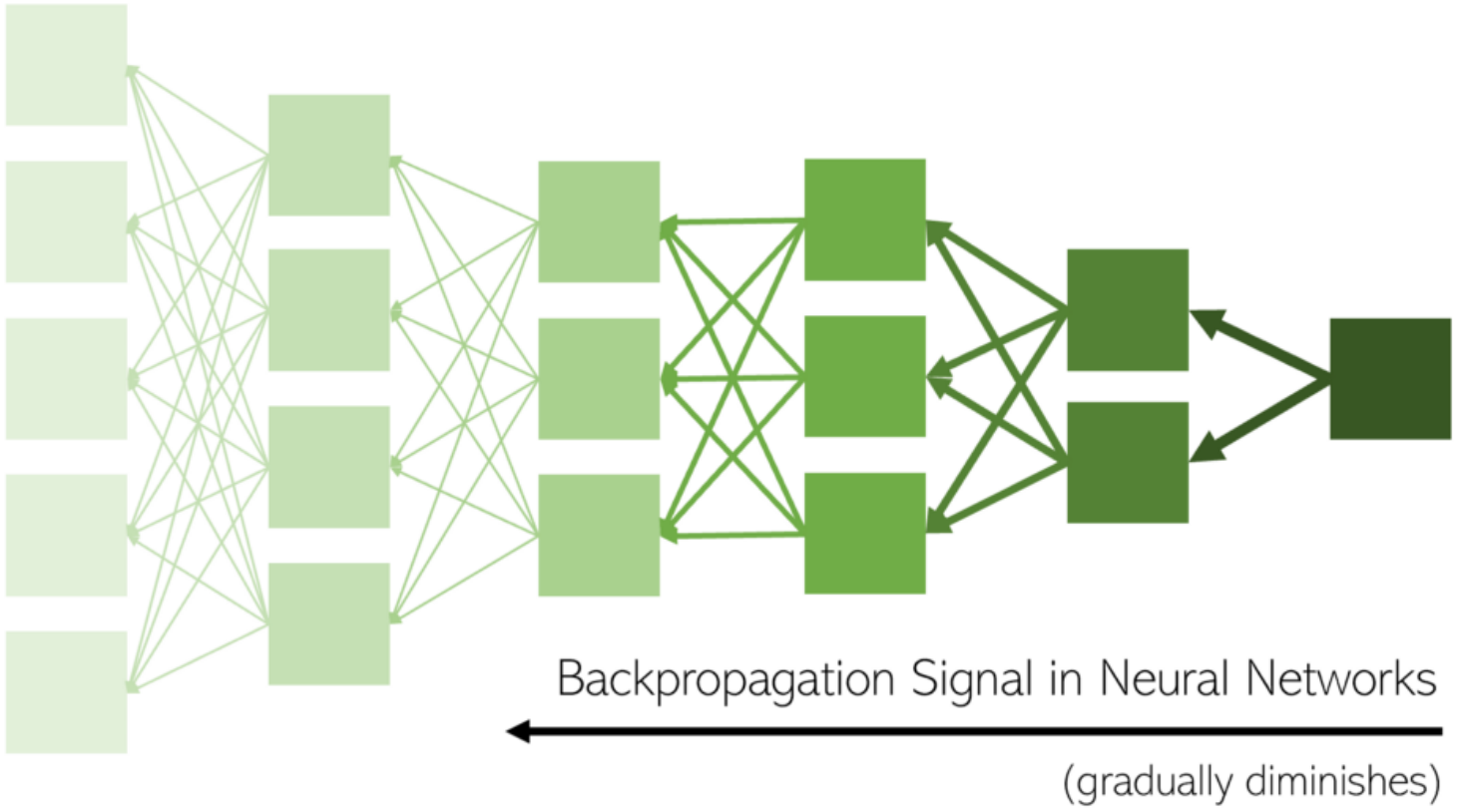
- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial

Historical Context



Motivation: Recall Vanishing Gradients Problem

Smallest gradients at **earliest layers make them slowest to train**, yet later layers depend on those earlier layers to do something useful; consequently, NNs struggle with garbage in means garbage out



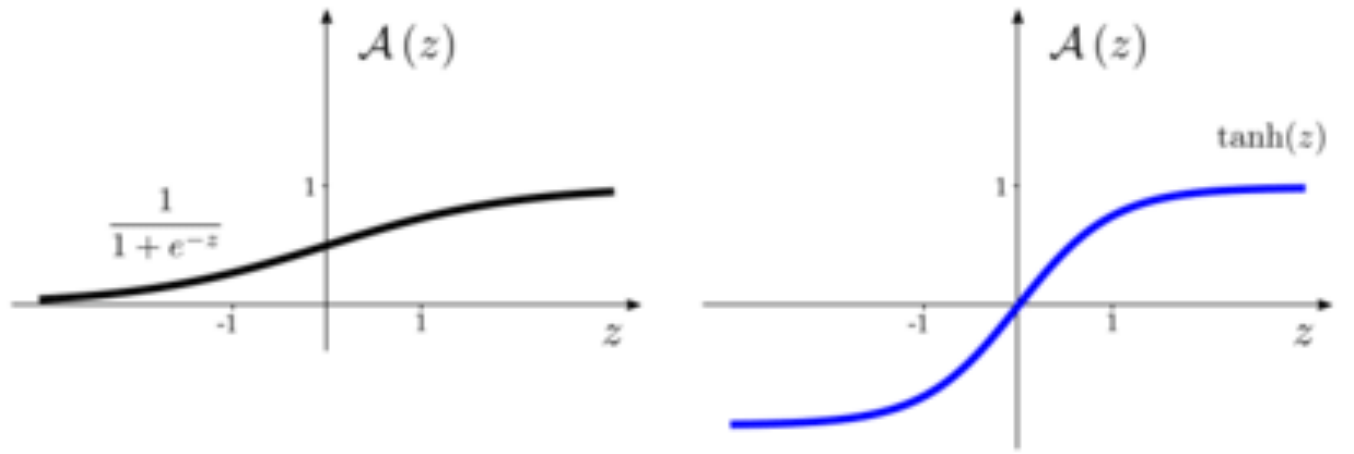
https://towardsdatascience.com/batch-normalization-the-greatest-breakthrough-in-deep-learning-77e64909d81d

Motivation: Recall Vanishing Gradients Problem

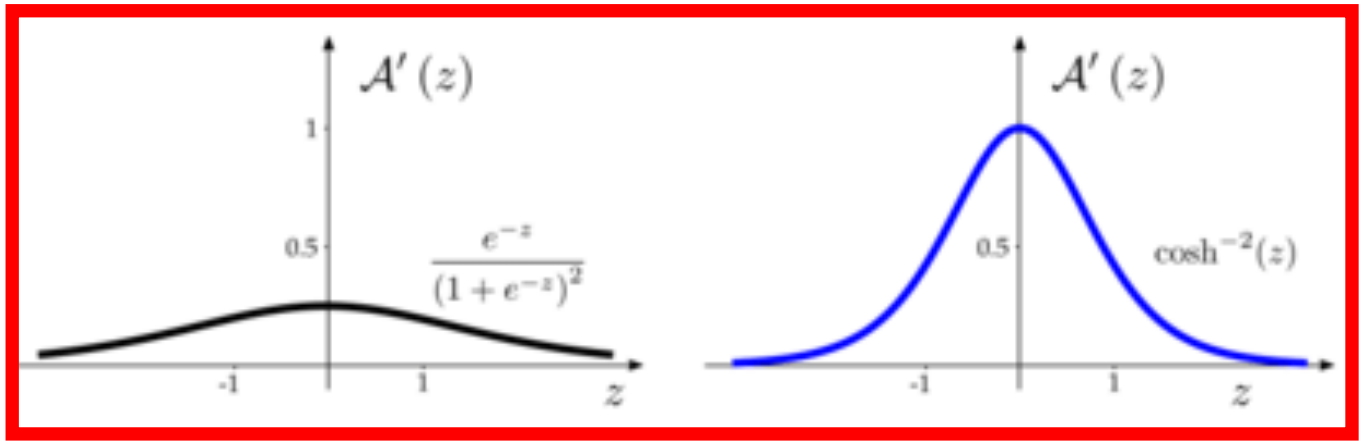
Sigmoid

Tanh

Main reason is activation functions and their derivatives:



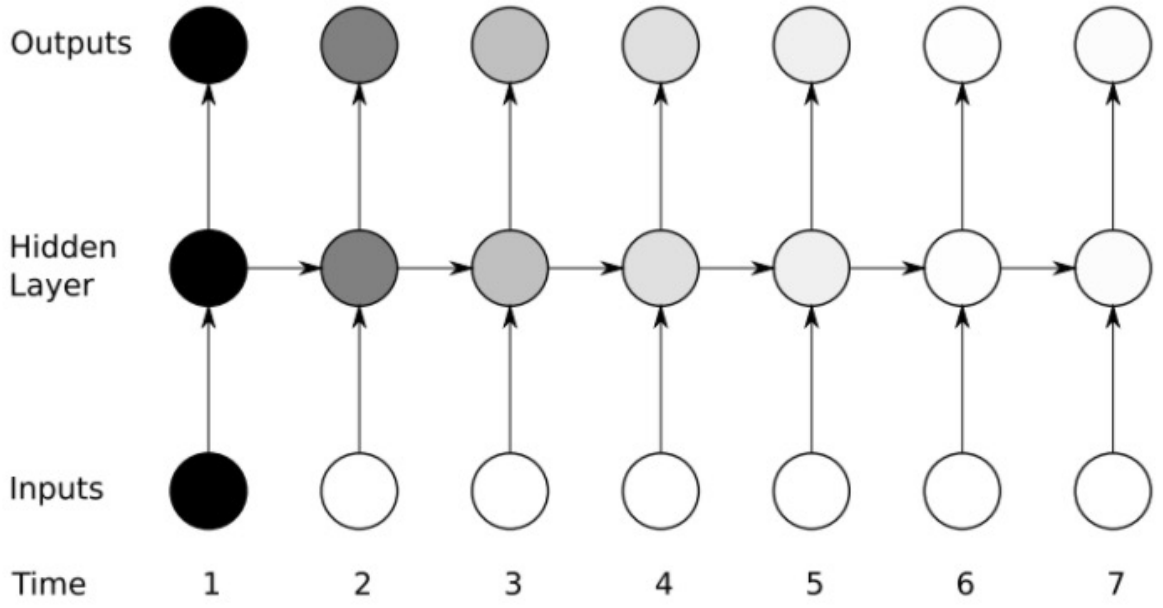
Ranges from 0 to 1, and typically less than 1:



Masi et al. Journal of the Mechanics and Physics of Solids. 2021

RNN Challenge: Learn Long-Term Dependencies

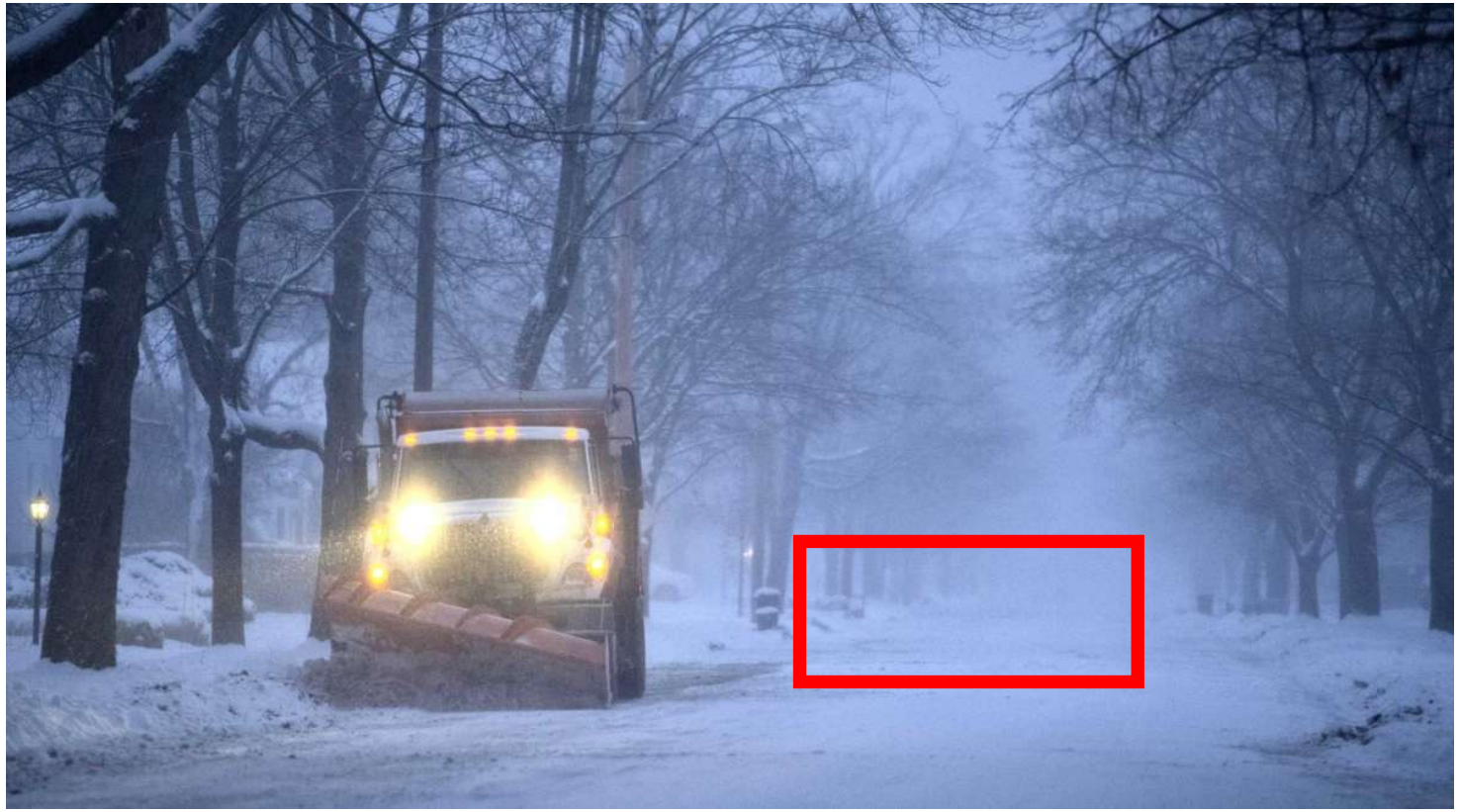
- e.g., language: “In 2004, I started college” vs “I started college in 2004”



- e.g., $\partial E/\partial W = \partial E/\partial y_3 * \partial y_3/\partial h_3 * \partial h_3/\partial y_2 * \partial y_2/\partial h_1$
  - Vanishing gradient: a product of numbers less than 1 shrinks to zero

https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/

RNN Challenge: Learn Long-Term Dependencies



Analogy: memory of past in RNN diminishes similar to how the road starts to disappear as the vehicle that plowed the snow moves farther down the road

Image: https://www.timesunion.com/news/article/albany-saratoga-snow-totals-16826908.php

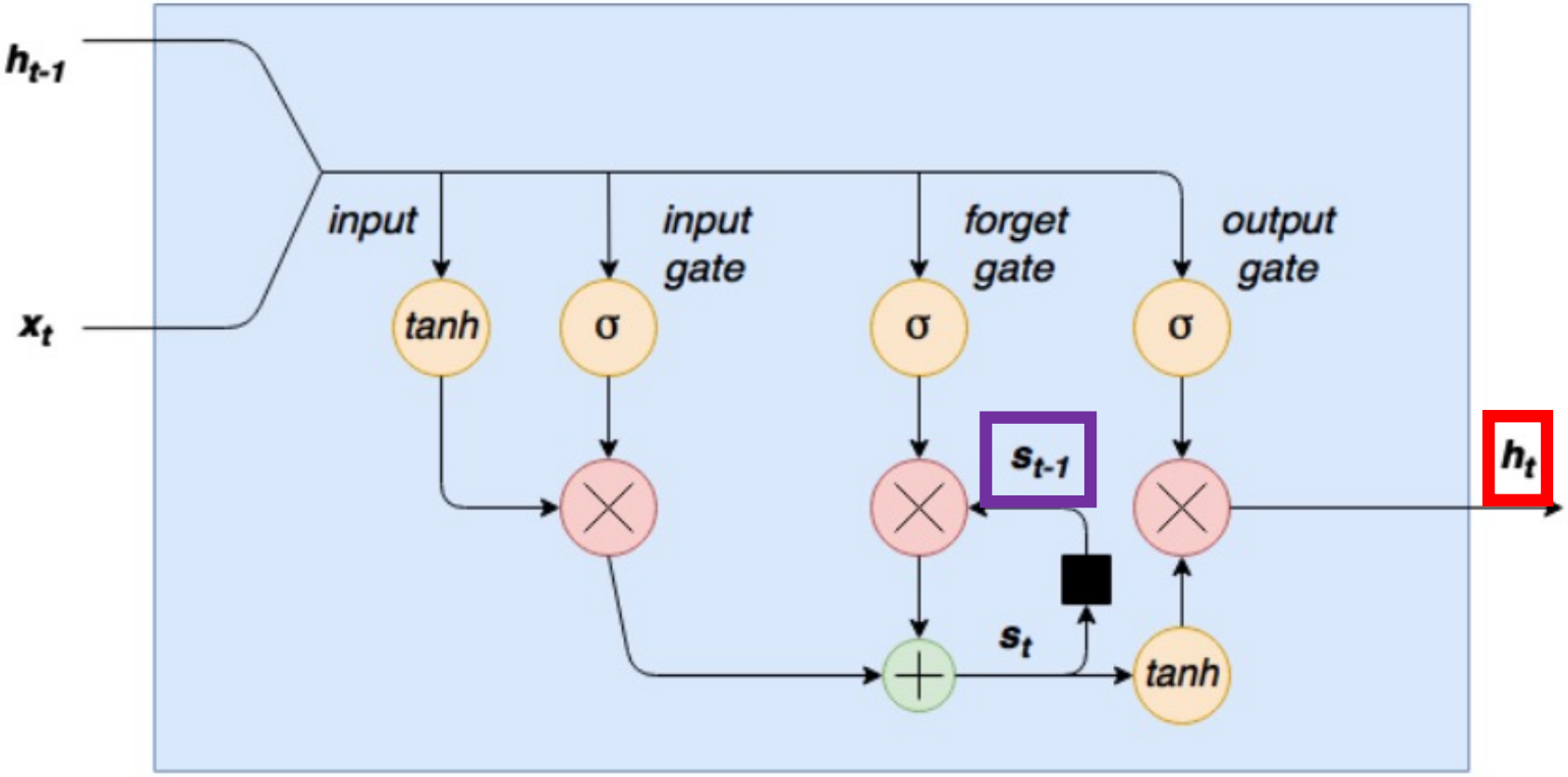
What Are Vanishing Gradient Remedies?

- Many remedies we have discussed can be used with RNNs; e.g.,
  - ReLU activation functions
  - Intelligent weight initialization
  - Skip connections

- Popular RNN Solution: gates

Popular RNN Variant #1: LSTM

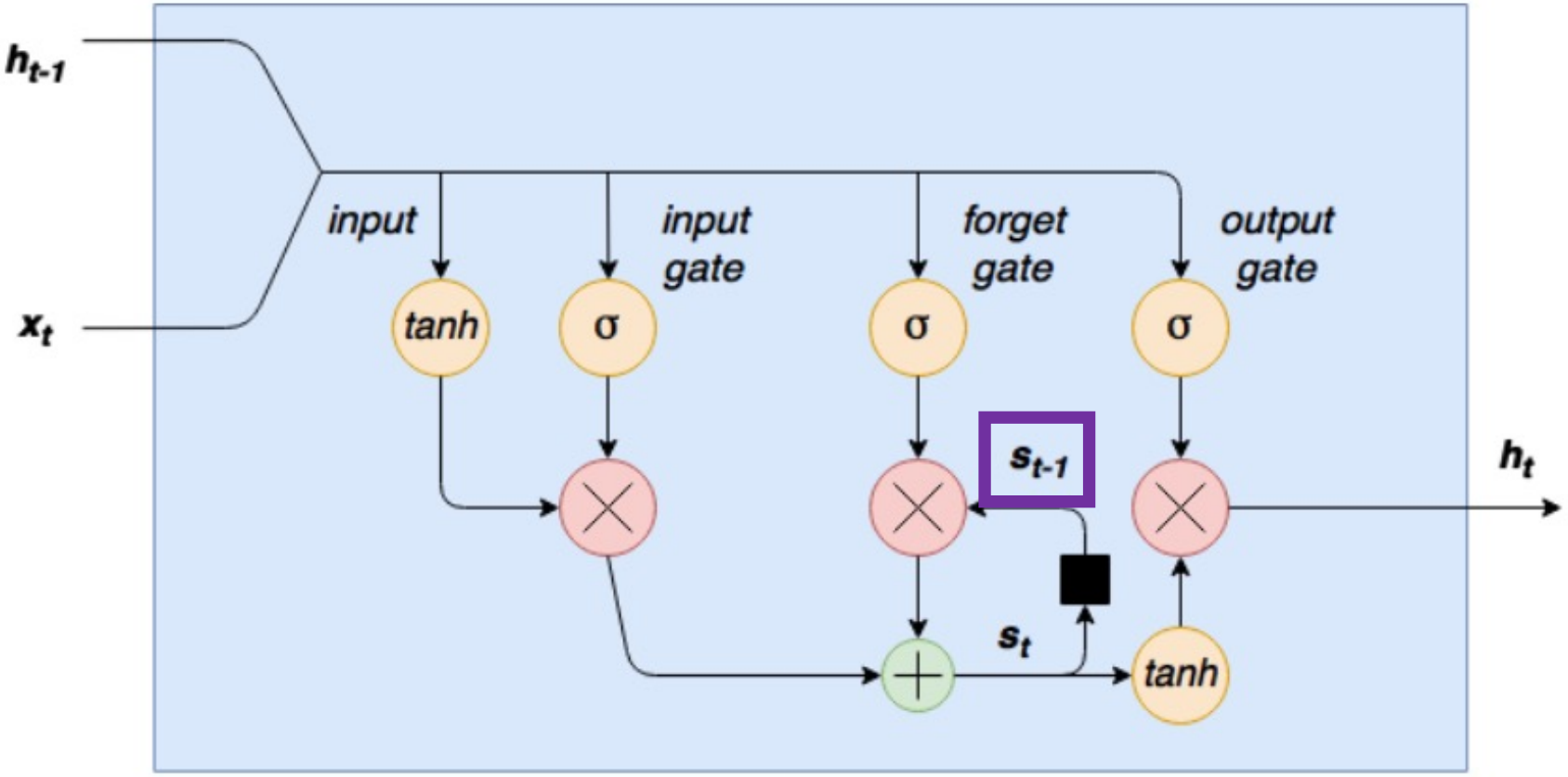
Both the previous hidden state and cell state are passed to next time steps



http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/

Popular RNN Variant #1: LSTM

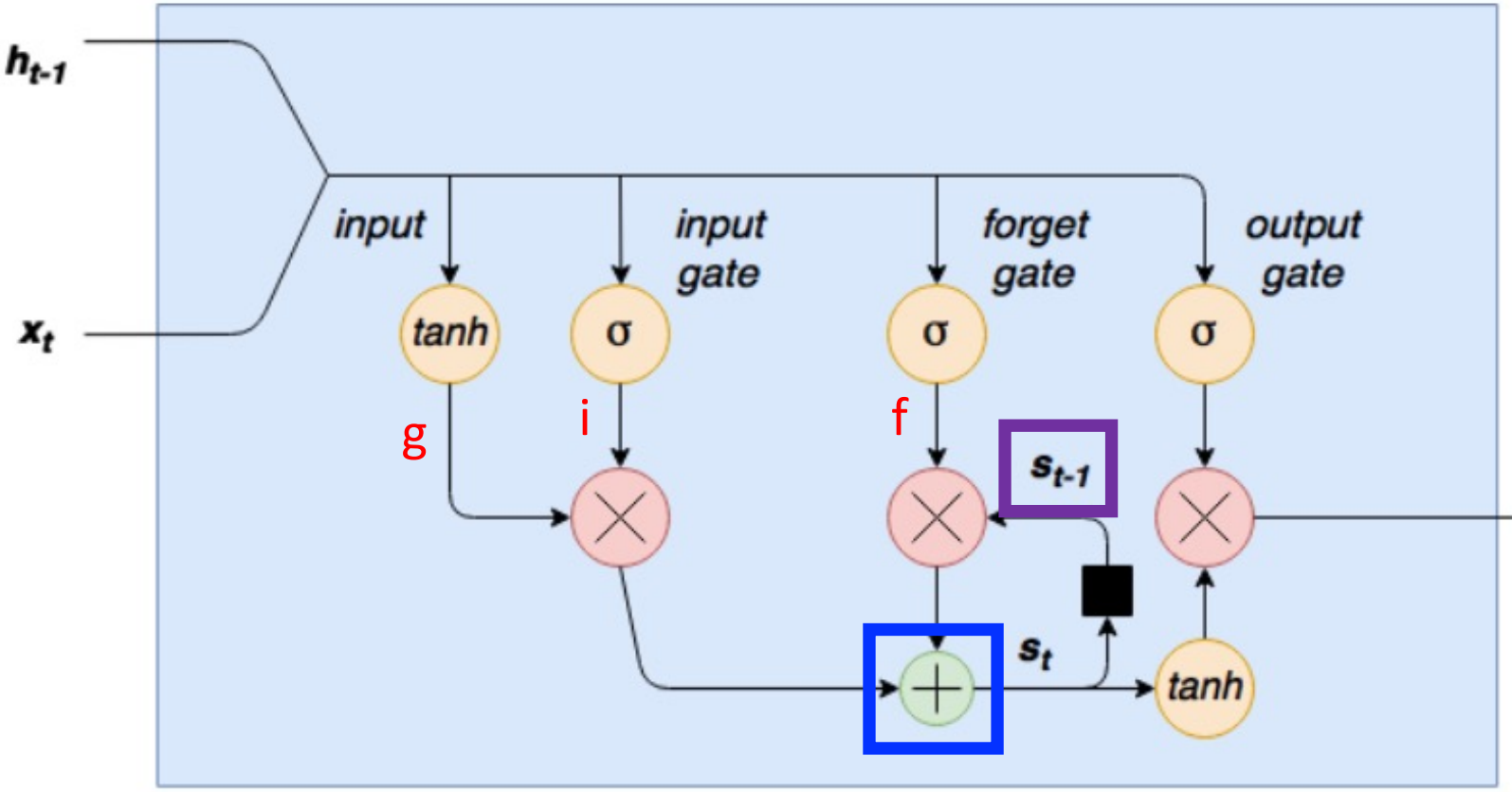
Cell state: can retain memory of the past for a long duration



http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/

Popular RNN Variant #1: LSTM

Cell state: can retain memory of the past for a long duration, based on the forget gate



$$s_t = s_{t-1} \circ f + g \circ i$$

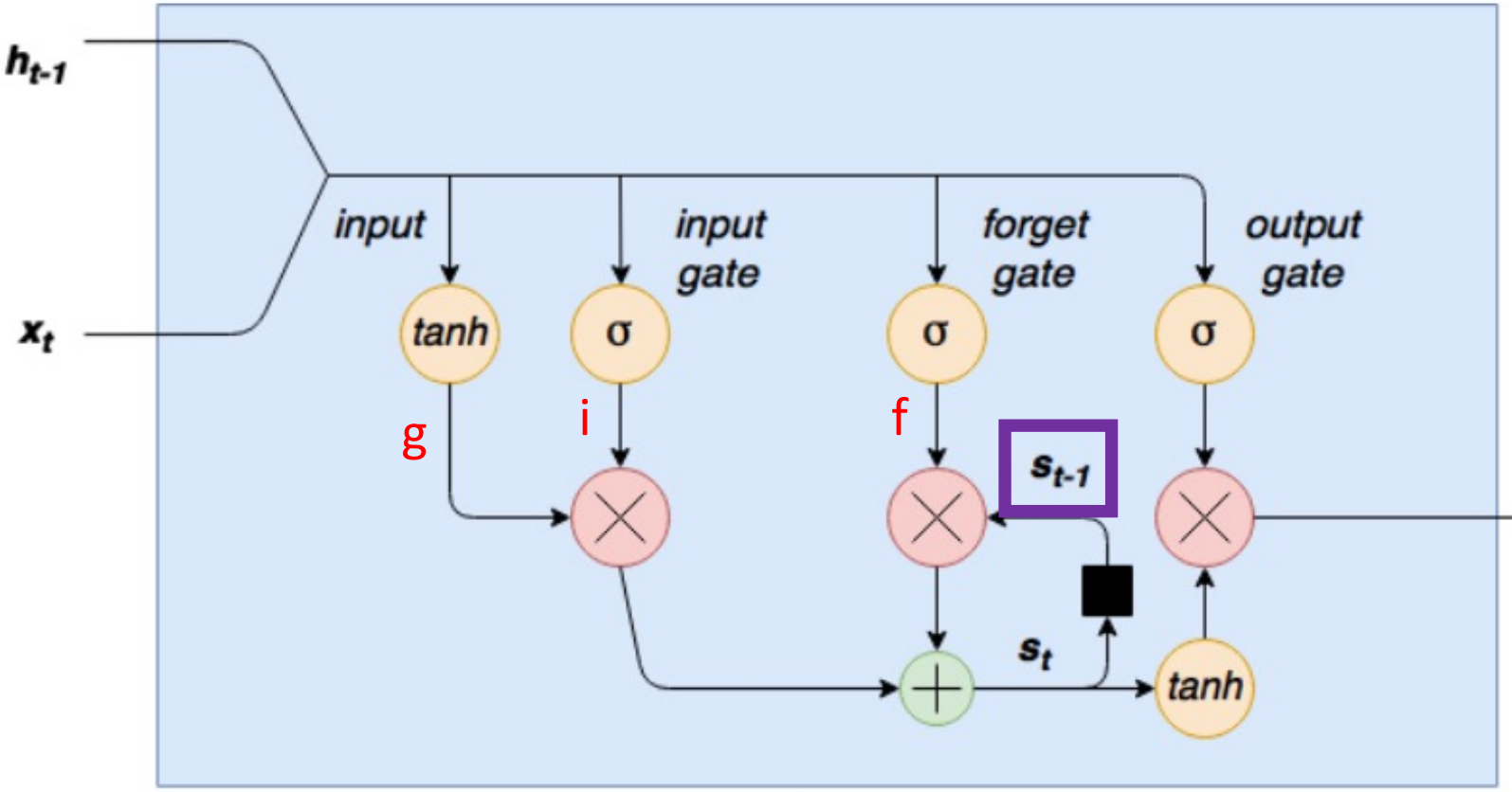
$$\frac{\partial s_t}{\partial s_{t-1}} = f$$

The addition enables the cell state to flow back through time for long durations

http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/

Popular RNN Variant #1: LSTM

Cell state: can retain memory of the past for a long duration, based on the forget gate



$$s_t = s_{t-1} \circ f + g \circ i$$

$$\frac{\partial s_t}{\partial s_{t-1}} = f$$

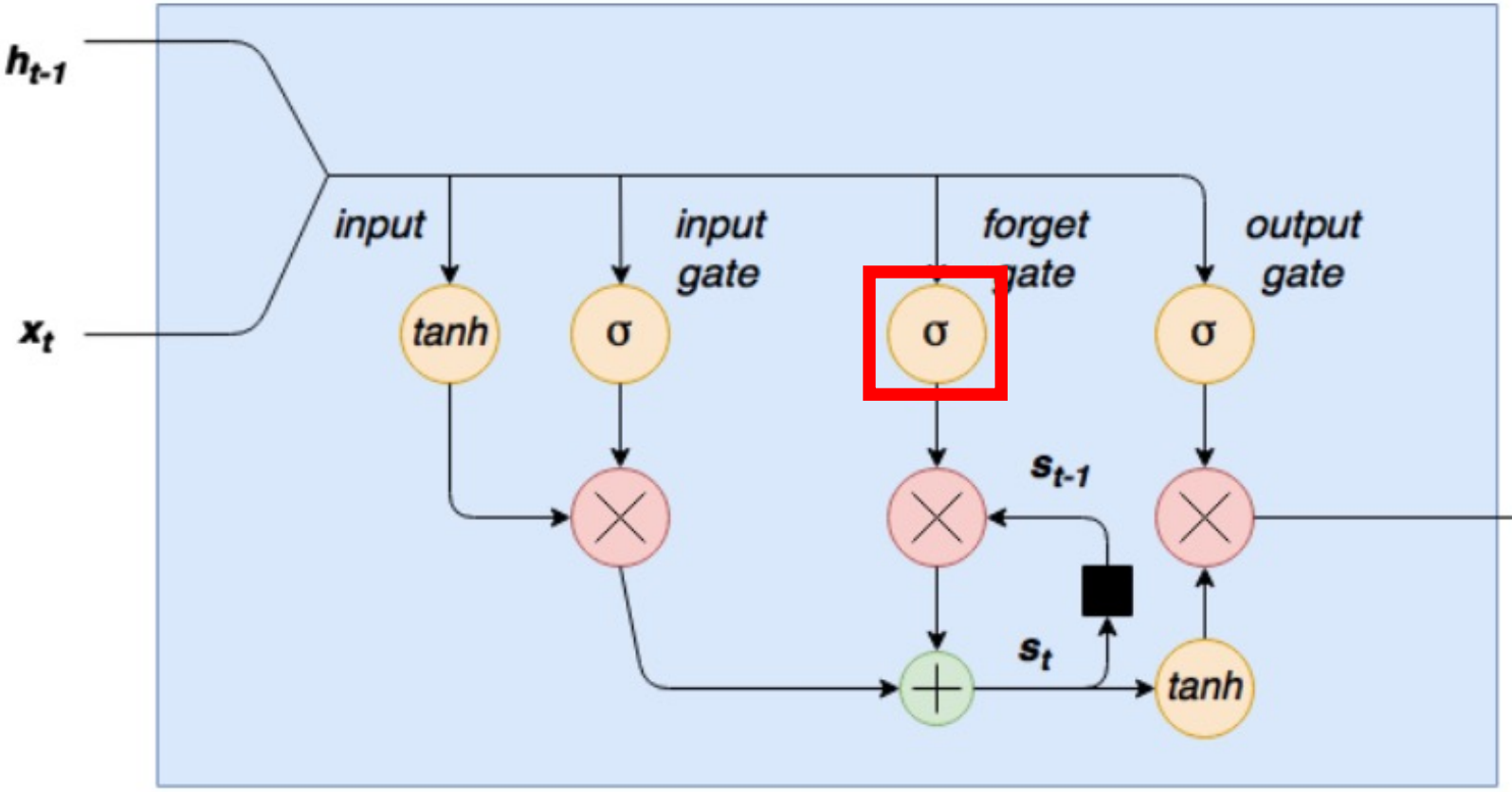
Forget gate range: 0 to 1

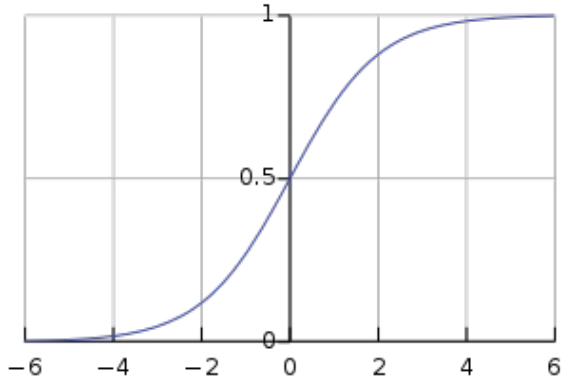
- What happens if 0?
  - Past discarded
- What happens if 1?
  - Past retained

http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/

Popular RNN Variant #1: LSTM

Cell state: can retain memory of the past for a long duration, based on the forget gate



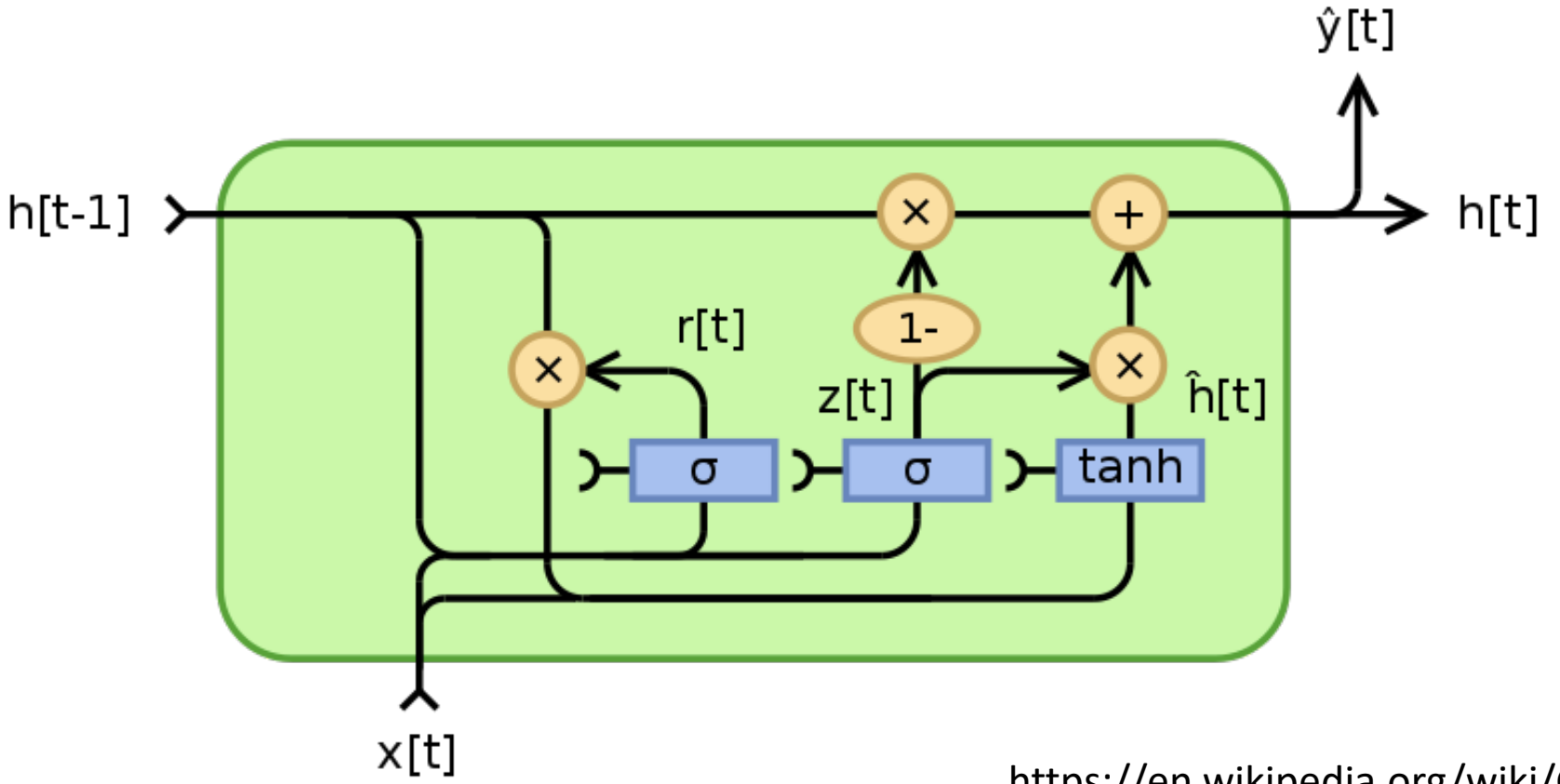


Called gate because sigmoid function mimics open (1) and closed (0)

http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/

Popular RNN Variant #1: Gated Recurrent Unit

- Simplifies LSTM by merging: (1) cell and hidden states and (2) forget and input gates



https://en.wikipedia.org/wiki/Gated_recurrent_unit

K. Chou et al; Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation; 2014

Today’s Topics

- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial

Today’s Topics

- Deep learning for sequential data
- Recurrent neural networks (RNNs)
- Gated RNNs
- Programming tutorial


The End