Recurrent Neural Networks

Deep Learning

Brad Quinton, Scott Chin

Back to our hypothetical case study….

- 5 months into your very exciting role as Director of AI for the Toronto Raptors, you get a call from the CEO… (time to panic!)

- He calls you in: *“I’ve been hearing great things about your work…. Now, winning games is one thing, but what we really care about is making sure we have lots and lots of fans!”*

- *“That’s why I need your help. Toronto is a very multi-cultural city. And we need those new-comers to become Raptors fans!”*

Back to our hypothetical case study….

- *“… what we need is to simultaneously translate all Raptors broadcasts into 24 languages! AI can do that, right?”*

- You: *“Umm, uh…, oh, ya, sure. I’ll get right on it!”*

- And off you go on your next assignment….

Back to our hypothetical case study….

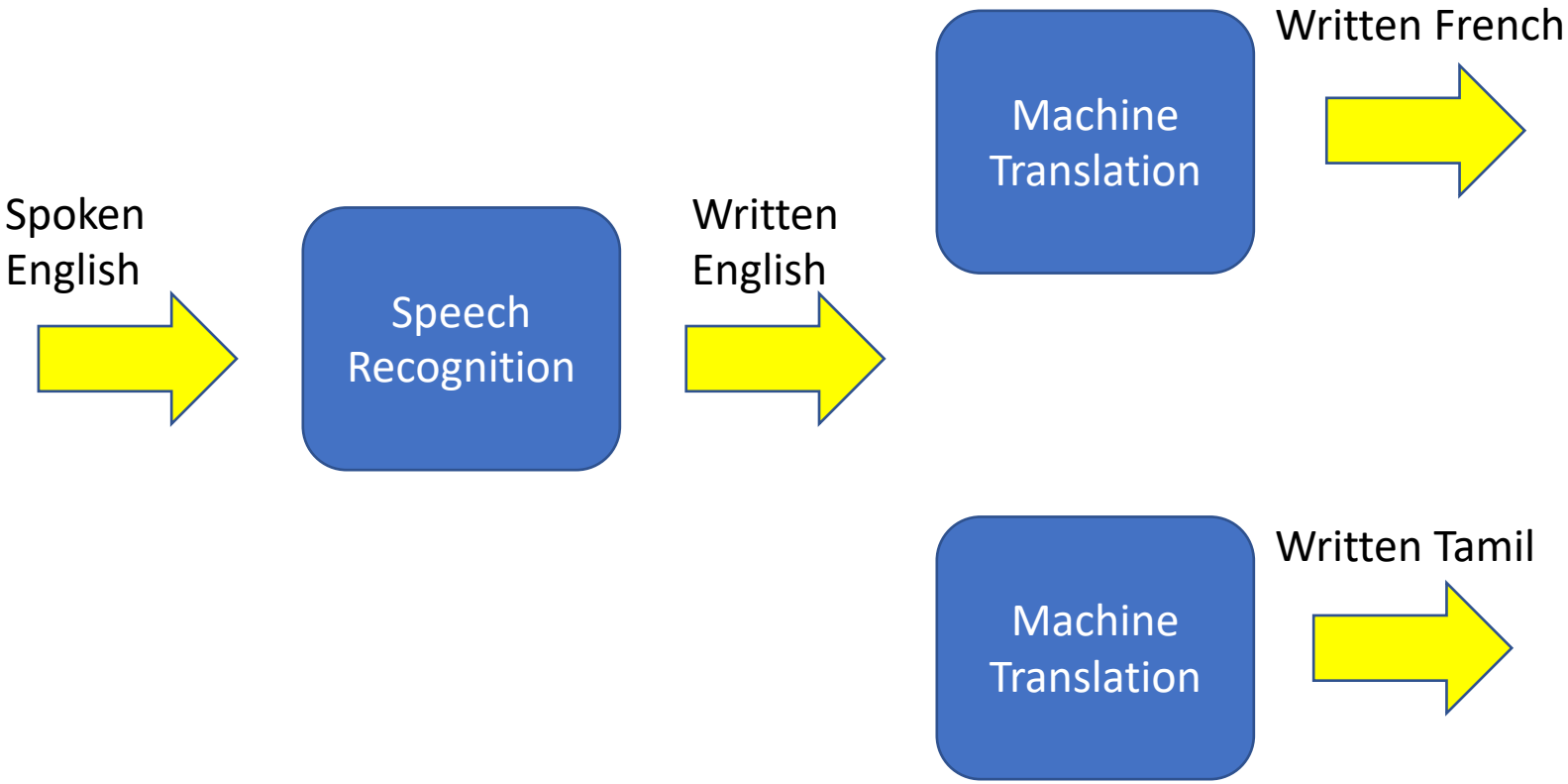
- As per usual, lets take stock… What do we need to do?

Back to our hypothetical case study….

- As per usual, lets take stock… What do we need to do?

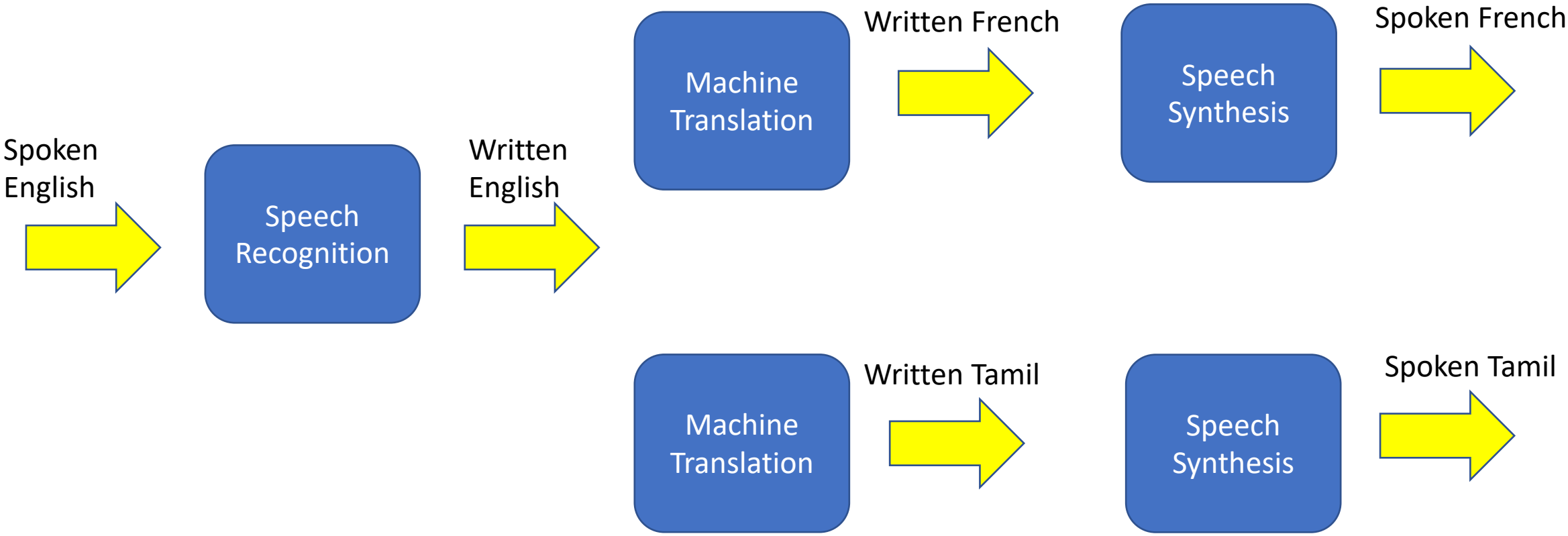
Back to our hypothetical case study….

- As per usual, lets take stock… What do we need to do?



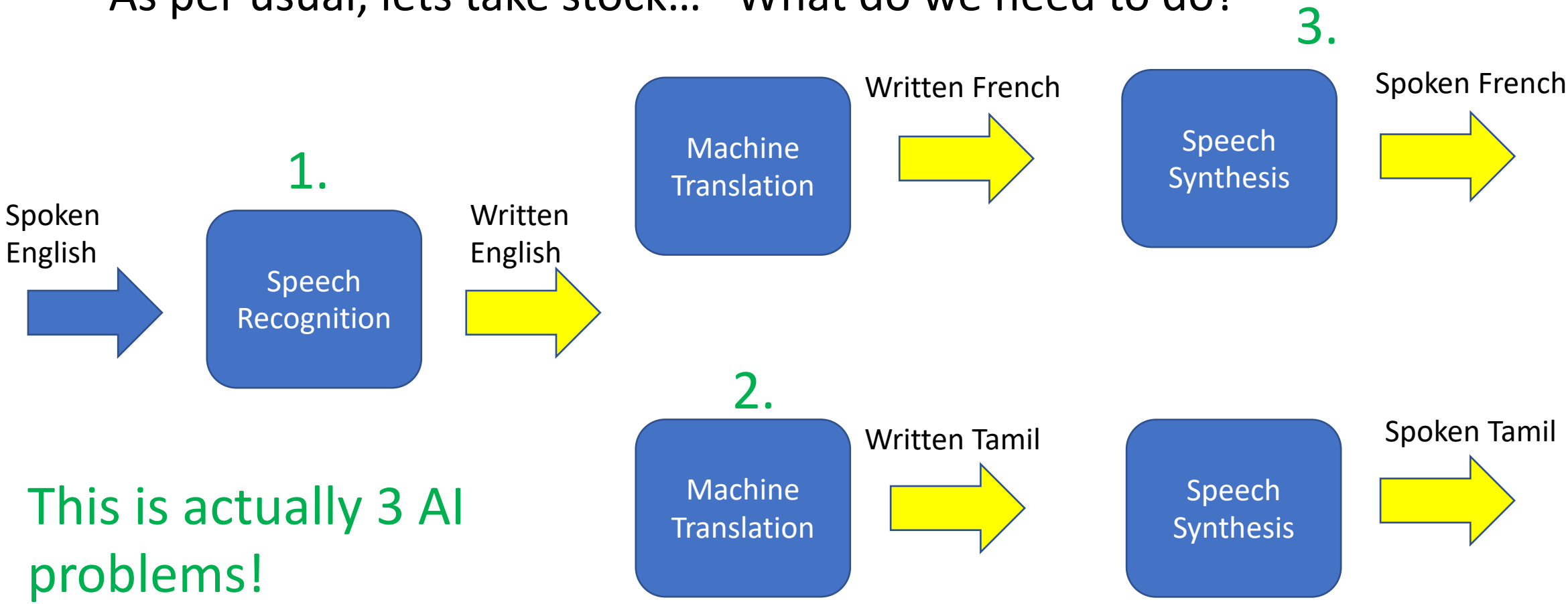
Back to our hypothetical case study….

- As per usual, lets take stock… What do we need to do?



Back to our hypothetical case study….

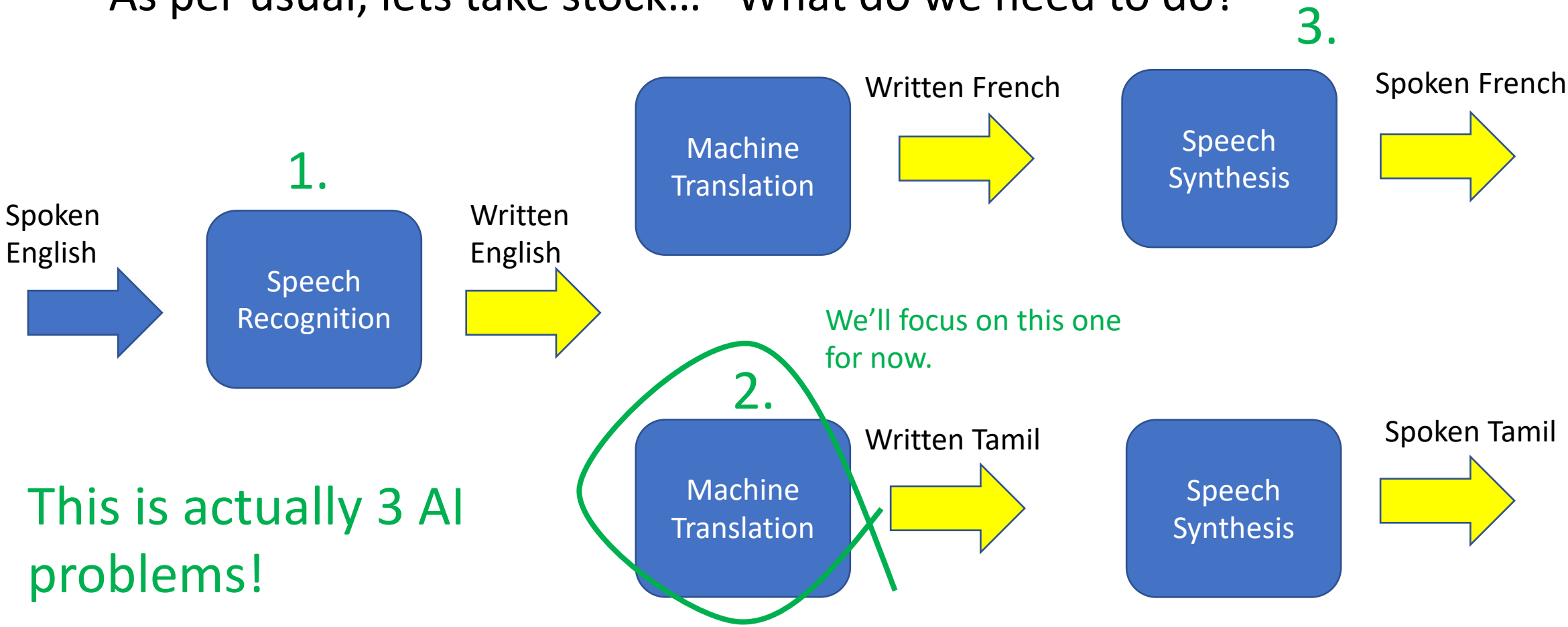
- As per usual, lets take stock… What do we need to do?



This is actually 3 AI problems!

Back to our hypothetical case study….

- As per usual, lets take stock… What do we need to do?

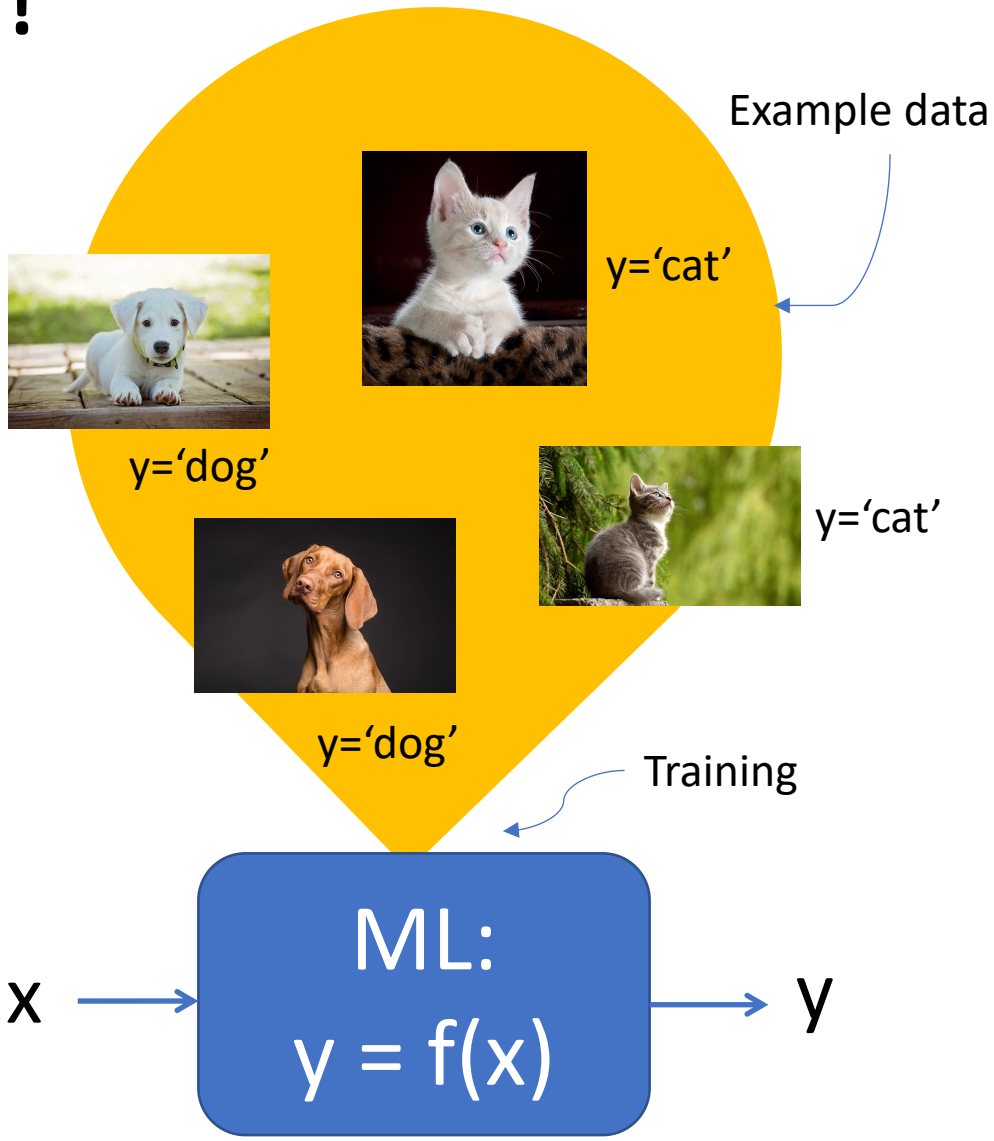


This is actually 3 AI problems!

Where are we with ML?

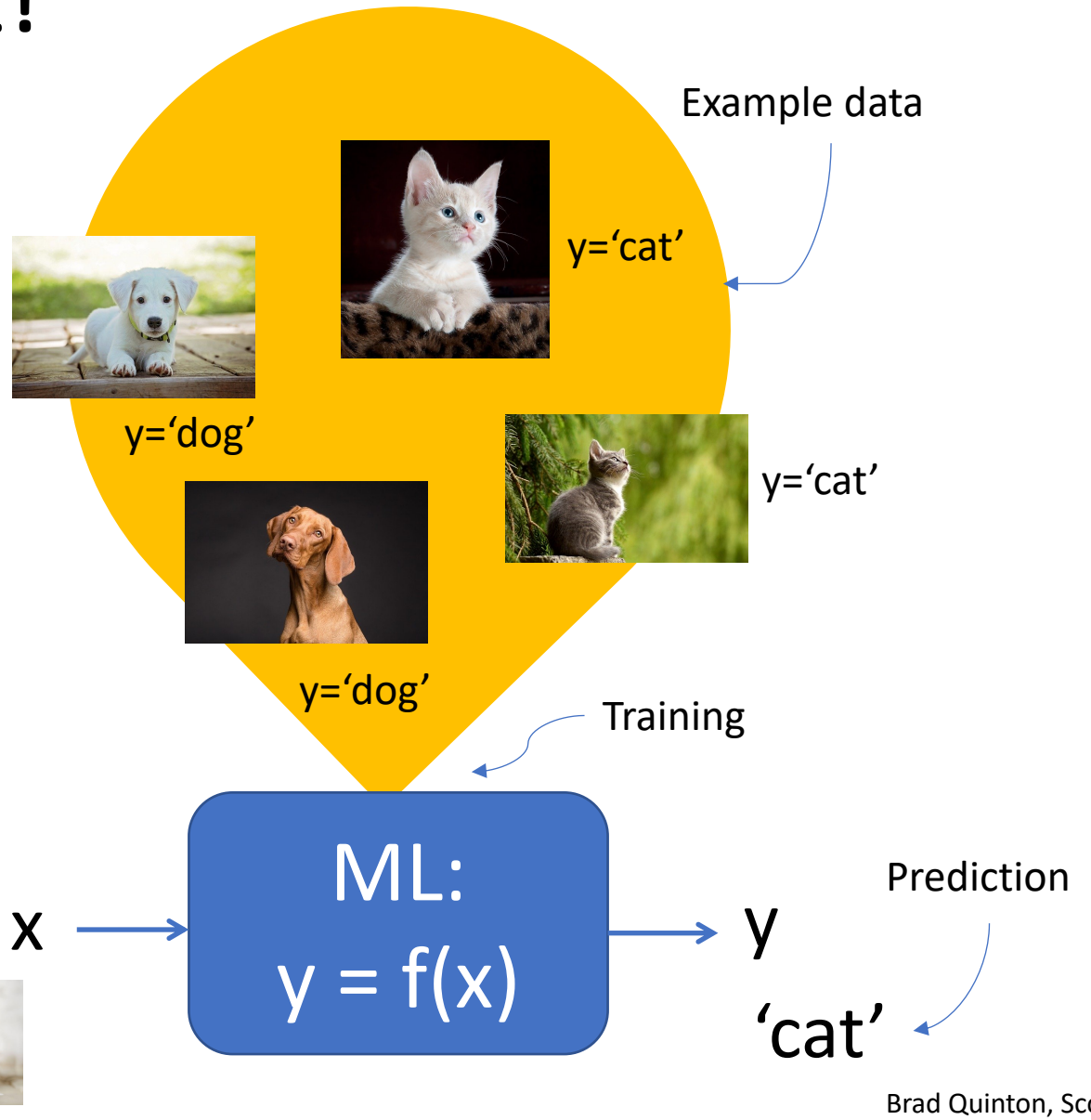
- Machine Learning algorithms allow you to create **approximate** functions using a set of example data

- The resulting approximate function can be applied to **new data**



Where are we with ML?

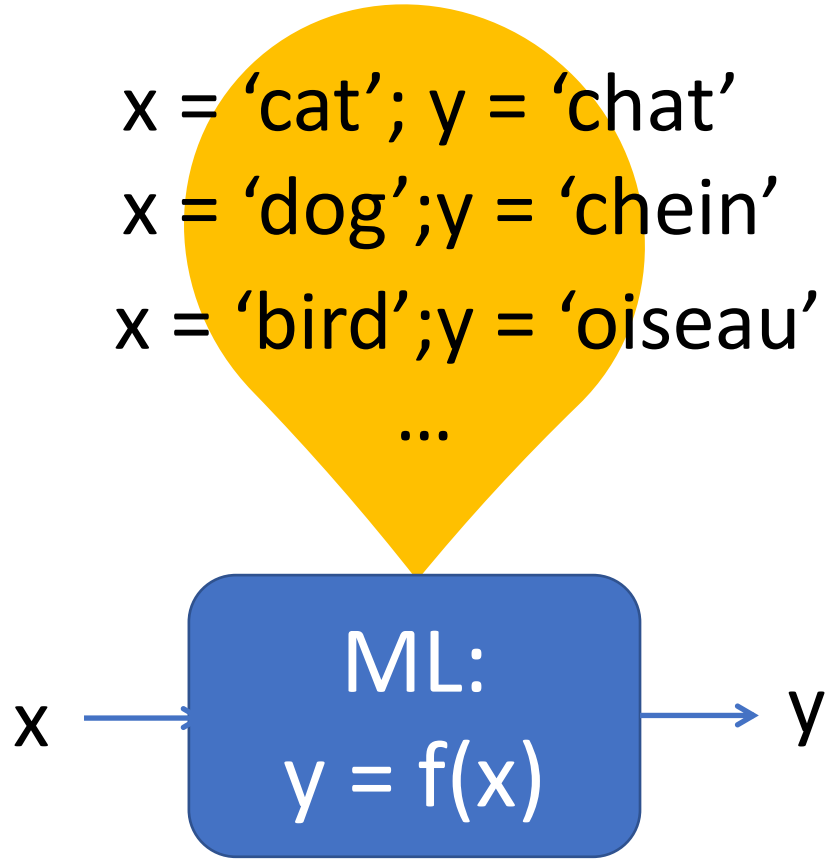
- Machine Learning algorithms allow you to create **approximate** functions using a set of example data
- The resulting approximate function can be applied to **new data**



Brad Quinton, Scott Chin

So translation with ML should be easy, right…?

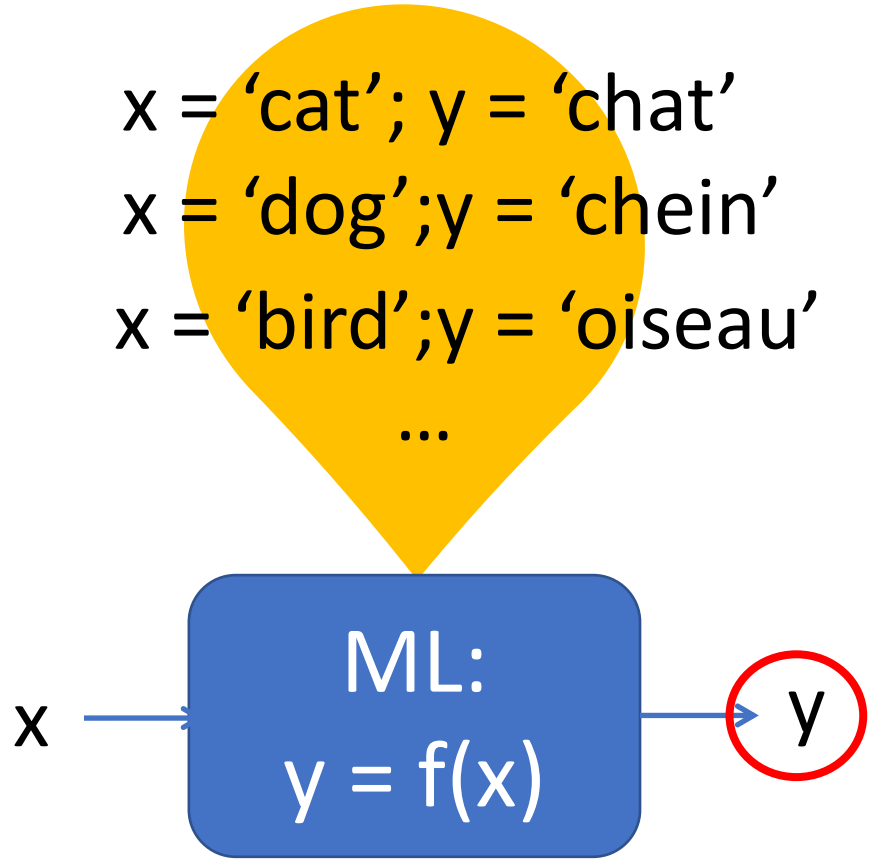
- Lots of examples: French/English dictionary!



Brad Quinton, Scott Chin

So translation with ML should be easy, right…?

- Lots of examples: French/English dictionary!



This will **not** produce a very high quality result!

Brad Quinton, Scott Chin

In the Real World Data Unfolds Over Time…

- Sound, Video, Natural language, Online interactions, Music, Sports, Real-time Navigating, Radar Tracking, etc.

- In these cases, there is *information* in both individual **components** of the data and their **ordering** with respect to other components

- Said another way, "For many tasks we would like to learn, we need to consider the **context.**"

Language Context

- Consider this sentence (that would be very relevant to our case study):

Language Context

- Consider this sentence (that would be very relevant to our case study):

“Michael Jordan made a basket.”

Language Context

- Consider this sentence (that would be very relevant to our case study):

“Michael Jordan made a basket.”

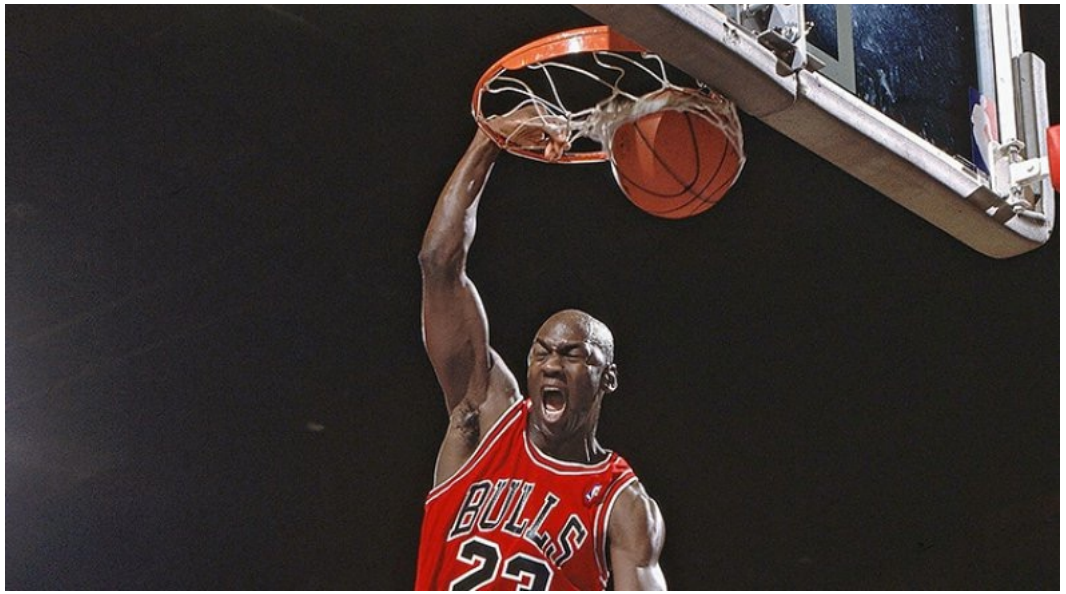
Did he really make a basket?

Language Context

- Consider this sentence (that would be very relevant to our case study):

“Michael Jordan made a basket.”

Did he really make a basket?



Or, score?

Brad Quinton, Scott Chin

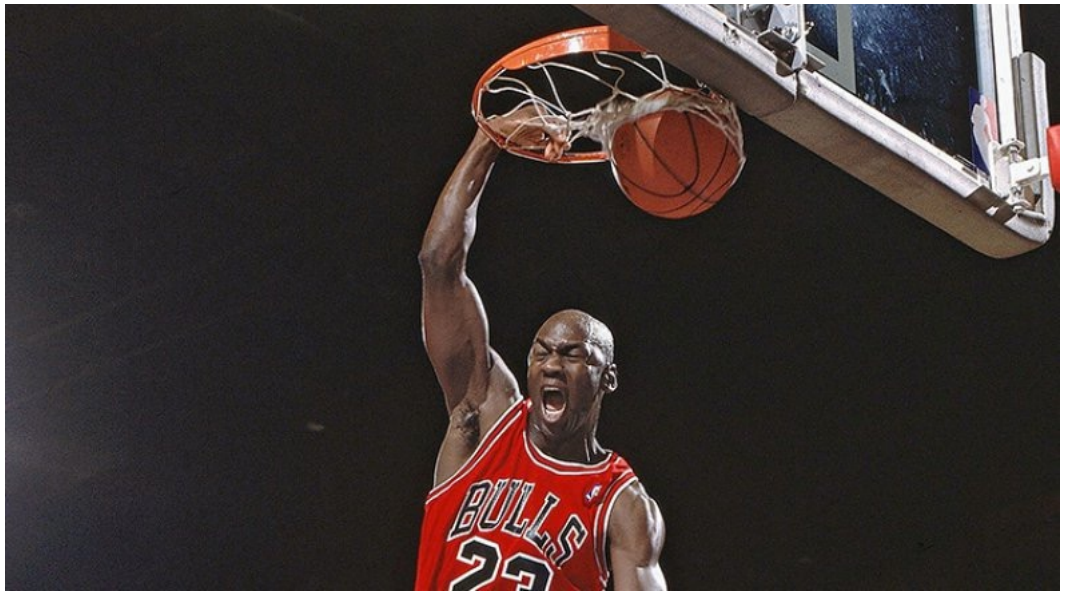
Language Context

- Consider this sentence (that would be very relevant to our case study):

"Michael Jordan made a basket"

The correct understanding is based on the context.

Did he really make a basket?



Or, score?

Brad Quinton, Scott Chin

Context, Context, Context

- Natural language is a just one example of how people use context to understand new information.

Is that a Bear or a big black dog?

Brad Quinton, Scott Chin

Context, Context, Context

- Natural language is a just one example of how people use context to understand new information.

Is that a Bear or a big black dog?

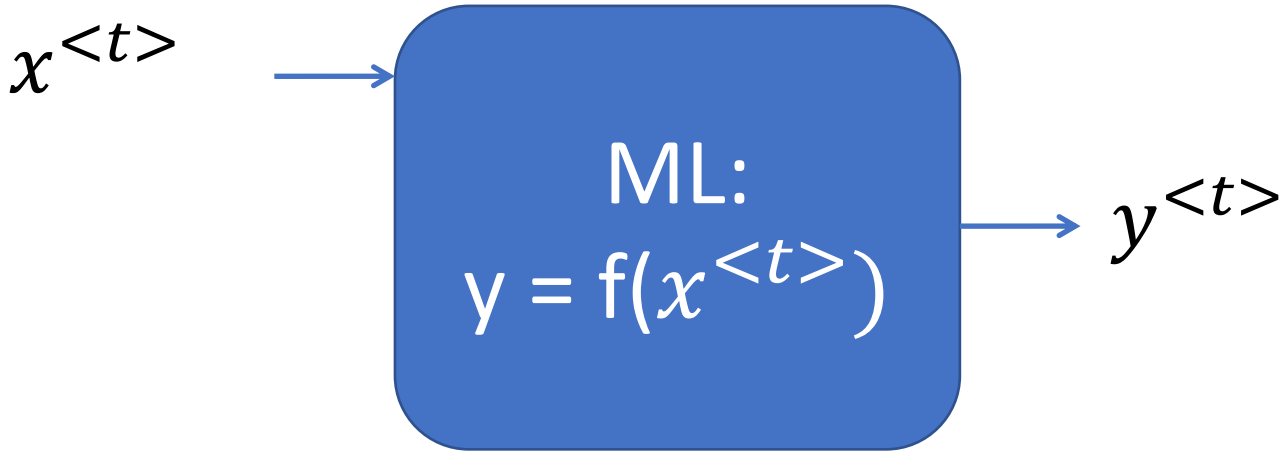
Depends where you are… If you are in Whistler, bear is a good guess, but in Stanley Park it is a dog.

How can we give our ML System Context?

- We could try to increase the inputs to our system to reflect the context

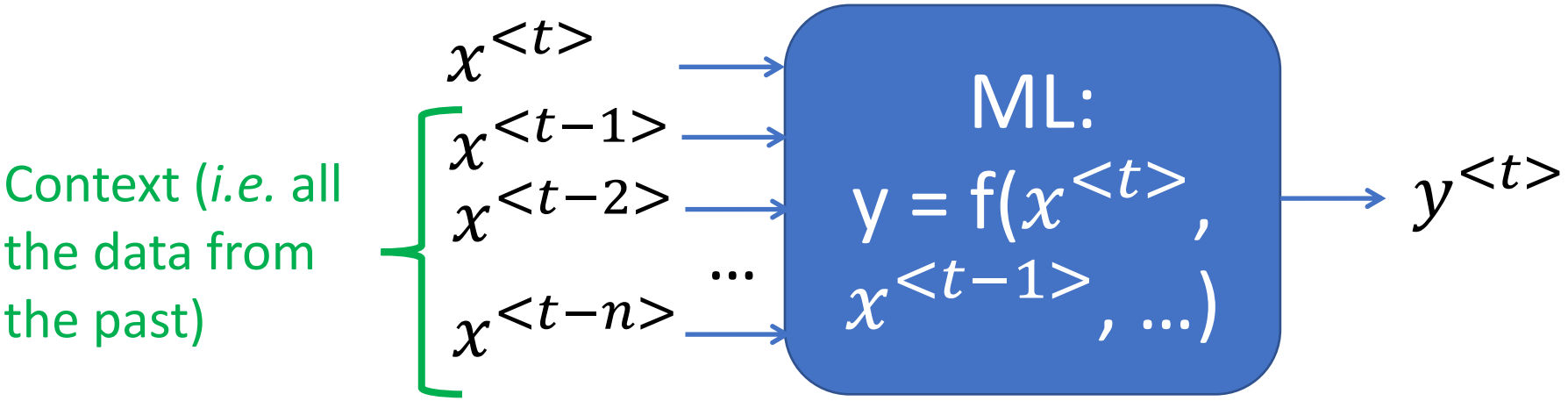
How can we give our ML System Context?

- We could try to increase the inputs to our system to reflect the context



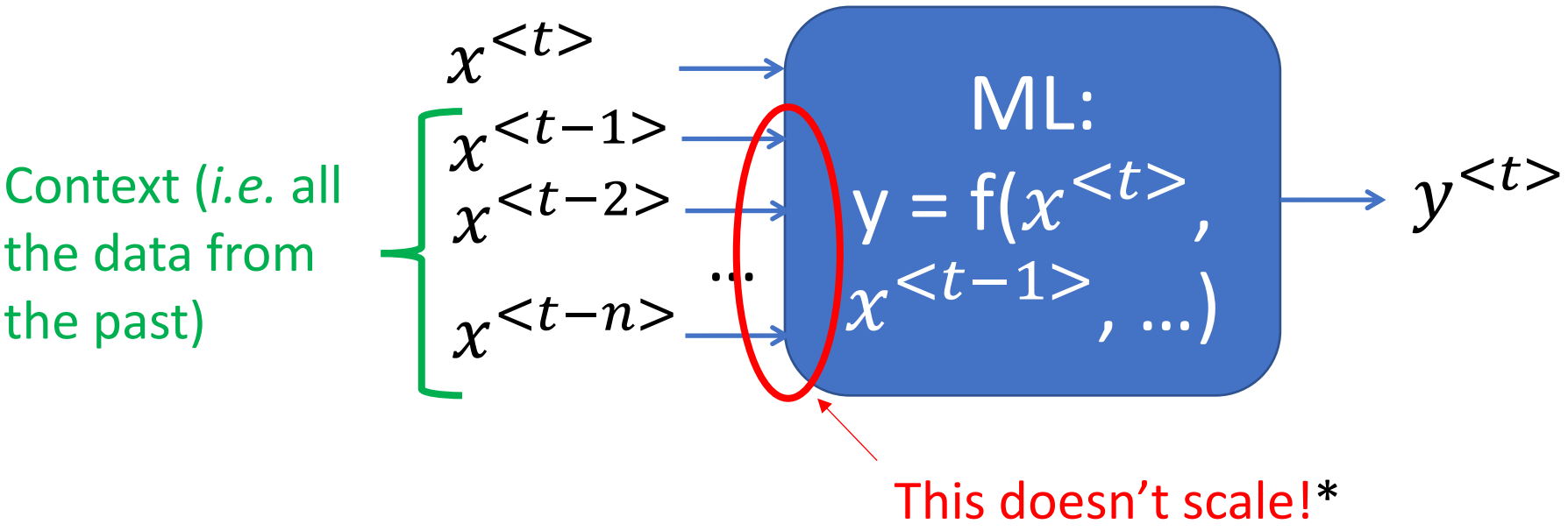
How can we give our ML System Context?

- We could try to increase the inputs to our system to reflect the context



How can we give our ML System Context?

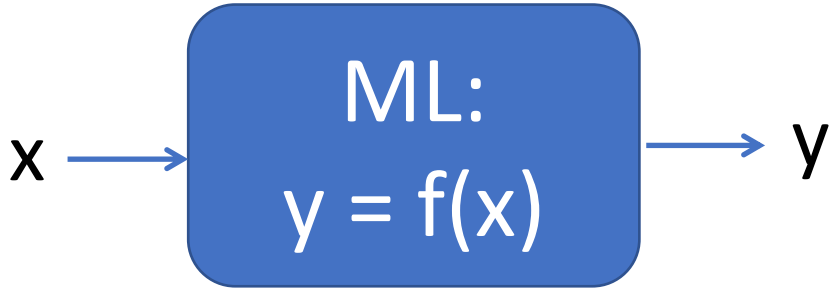
- We could try to increase the inputs to our system to reflect the context



* For example, the first sentence of a *Tale of Two Cities*: "It was the best of times…" has 119 words!

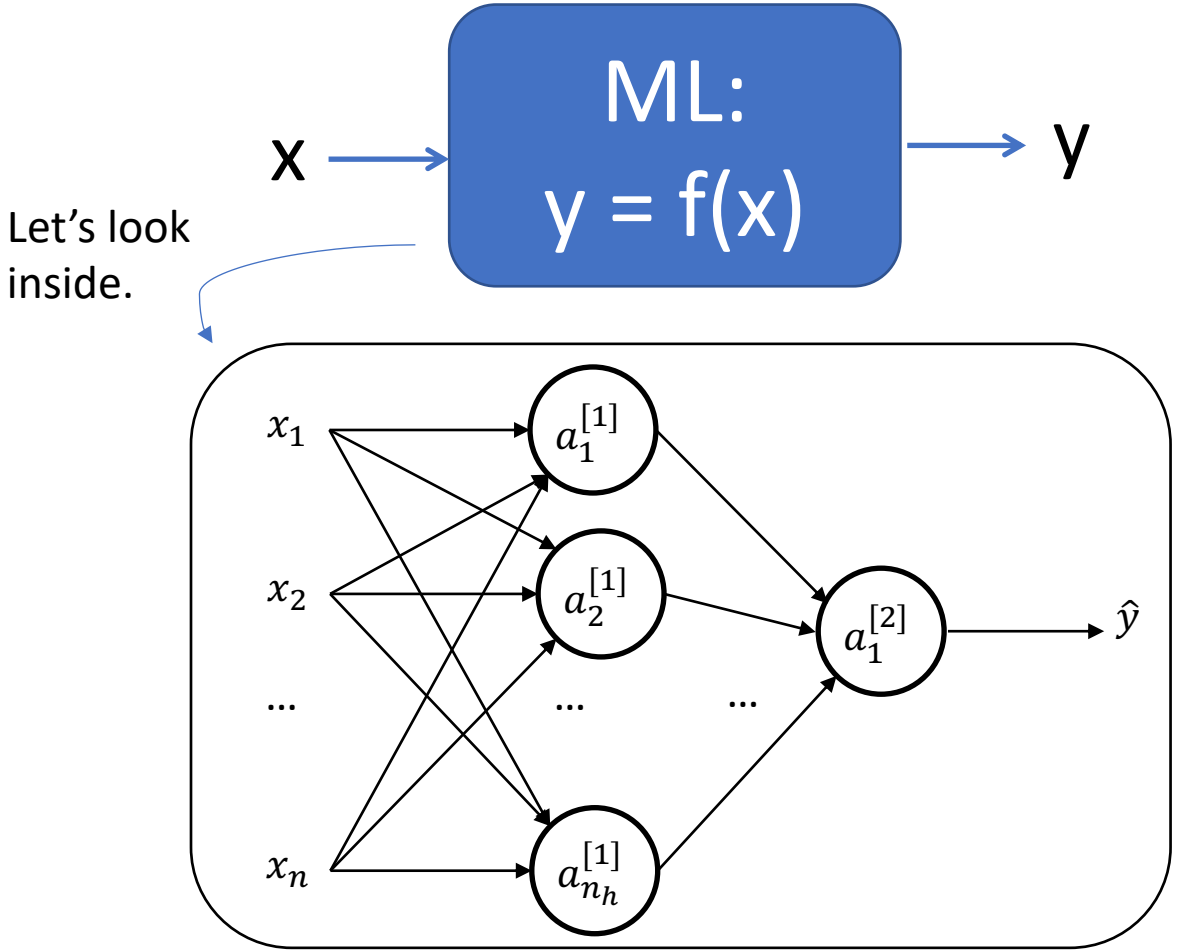
How can we give our ML System Context?

- What else do we have?



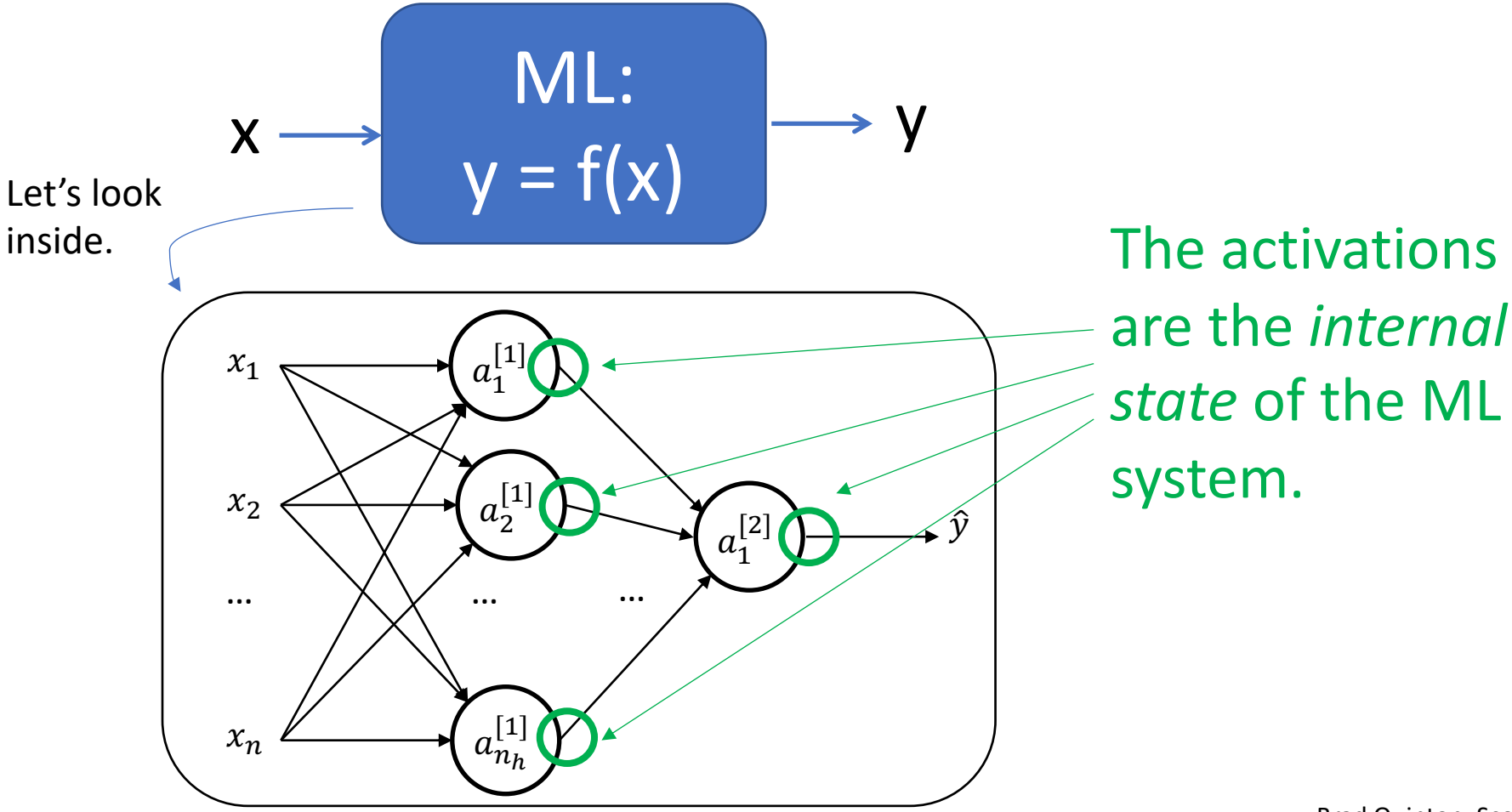
How can we give our ML System Context?

- What else do we have?



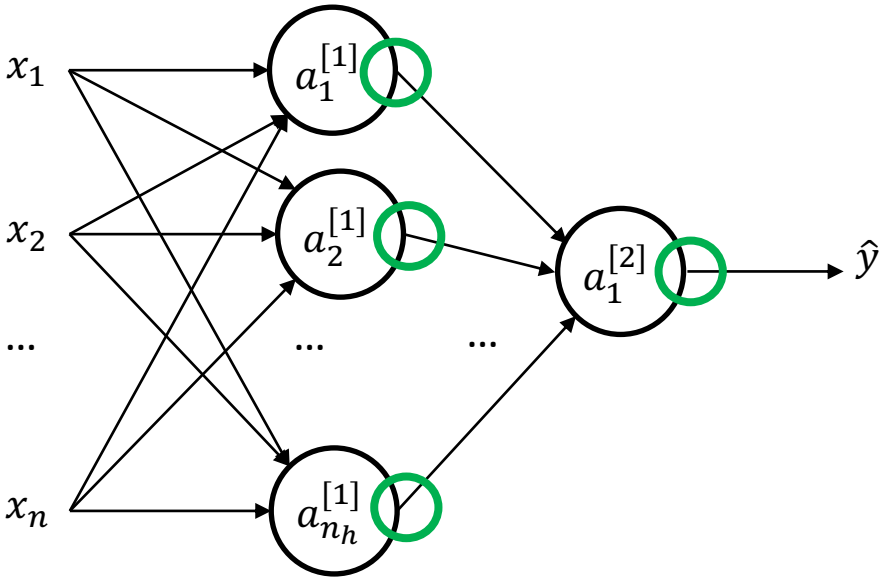
How can we give our ML System Context?

- What else do we have?



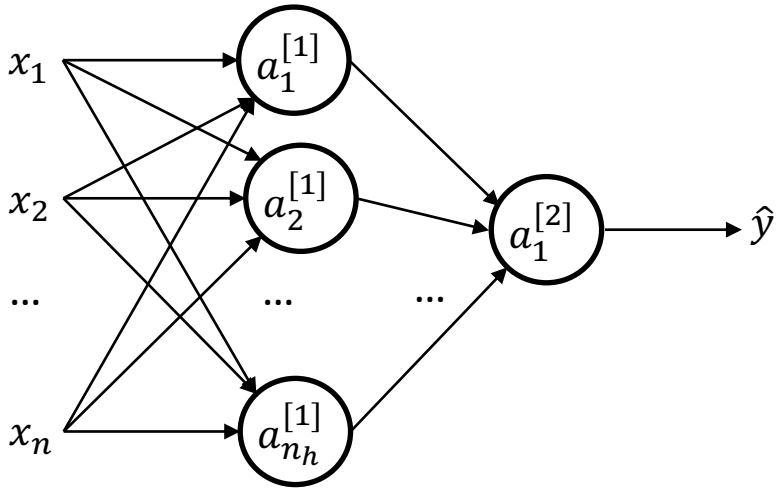
How can we give our ML System Context?

- The activations can be viewed as an *encoding* of the input, *x*
- Further, that encoding is one that was *learned* be useful to predict, *y*
- We would like to change the system so that the output, y can be influenced by a context…. (*i.e.* We want to bias the prediction based on what we have seen before!)



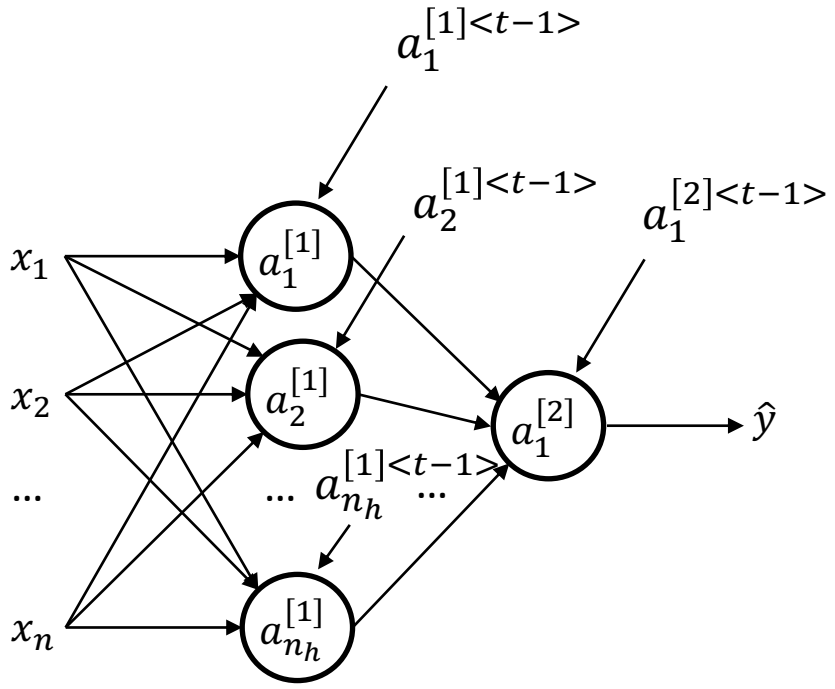
How can we give our ML System Context?

- The activations from the previous step in the sequence can be used to "bias" the activations on the next step!



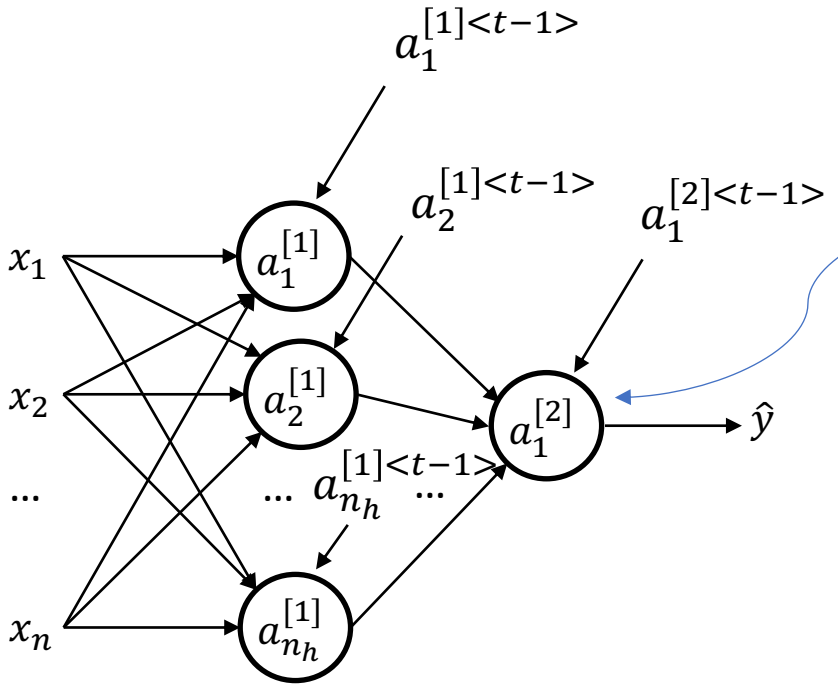
How can we give our ML System Context?

- The activations from the previous step in the sequence can be used to "bias" the activations on the next step!



How can we give our ML System Context?

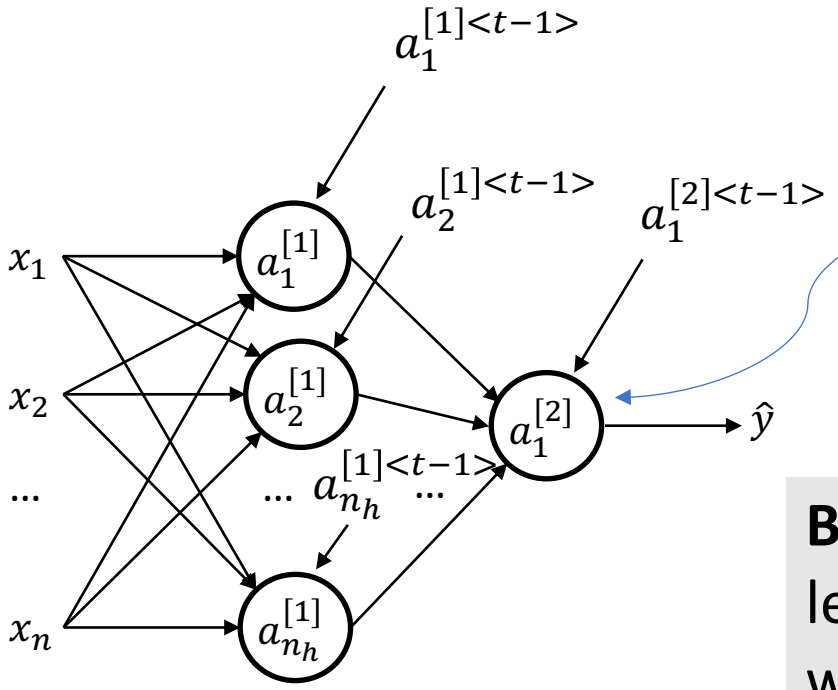
- The activations from the previous step in the sequence can be used to "bias" the activations on the next step!



But, by how much should the context bias the current prediction?

How can we give our ML System Context?

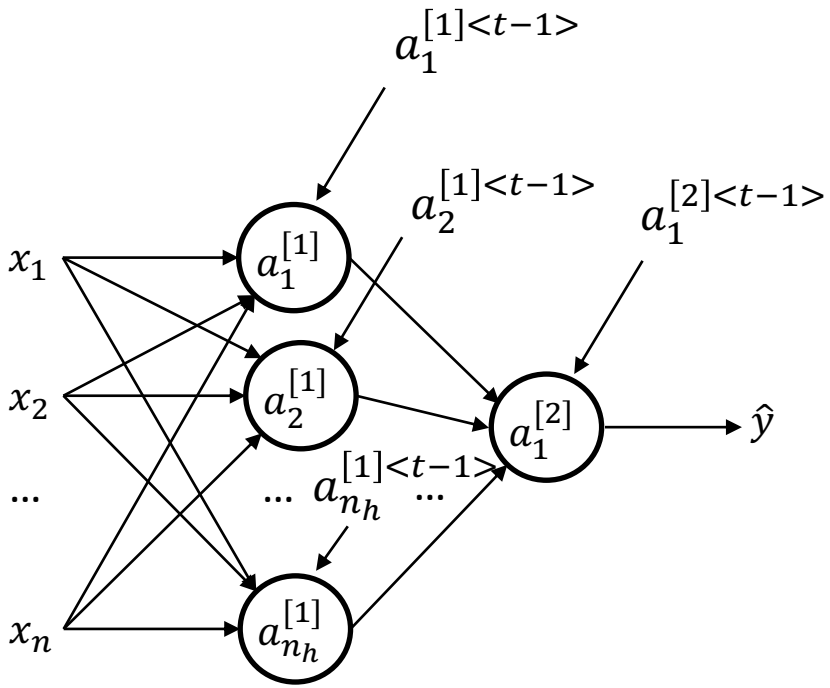
- The activations from the previous step in the sequence can be used to "bias" the activations on the next step!



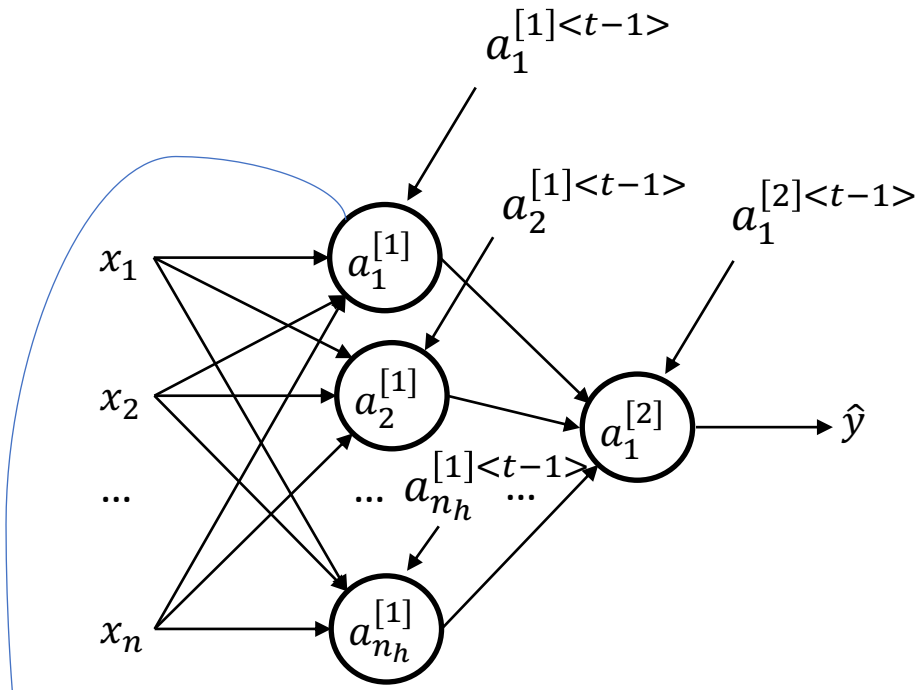
But, by how much should the context bias the current prediction?

BIG IDEA: We could simultaneously learn the amount of context required <u>while</u> we learn the input to output mappings.

Recurrent Neural Networks (RNNs)

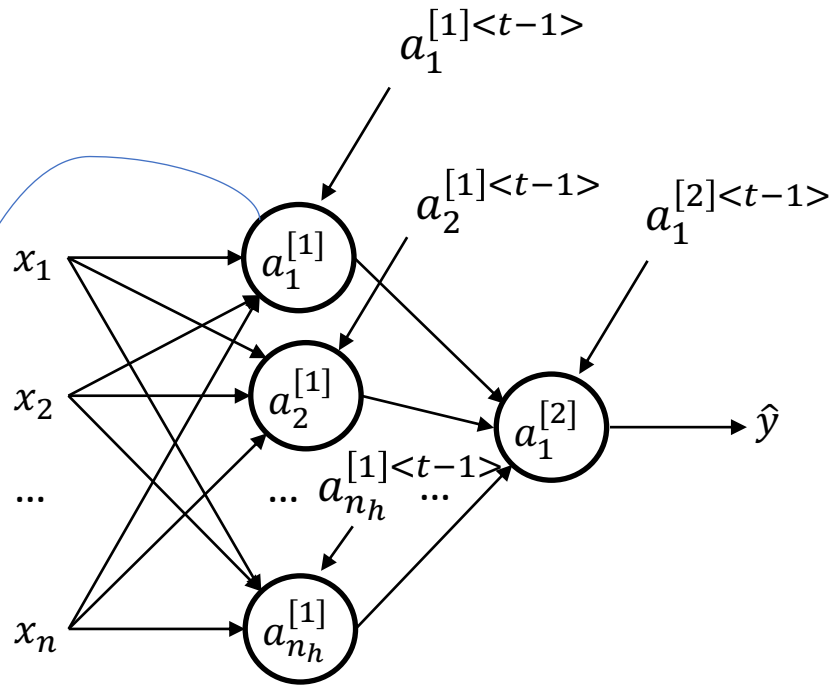


Recurrent Neural Networks (RNNs)



$$a_1^{[1]} = g(w_{ax}X_1 + w_{aa}a_1^{[1]<t-1>} + b)$$

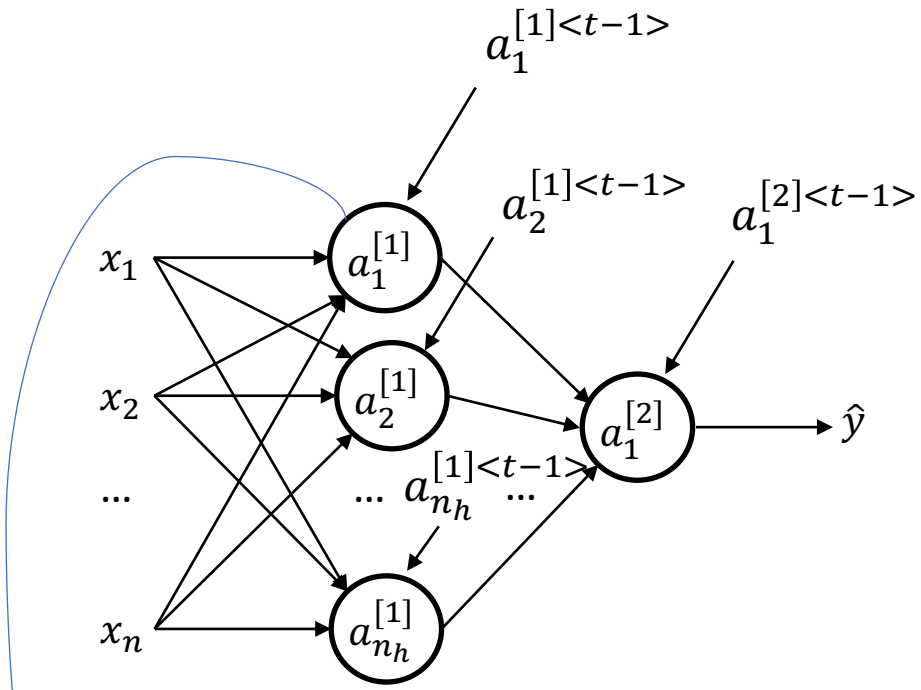
Recurrent Neural Networks (RNNs)



$$a_1^{[1]} = g(w_{ax}X_1 + w_{aa}a_1^{[1]<t-1>} + b)$$

Contribution from current input

Recurrent Neural Networks (RNNs)

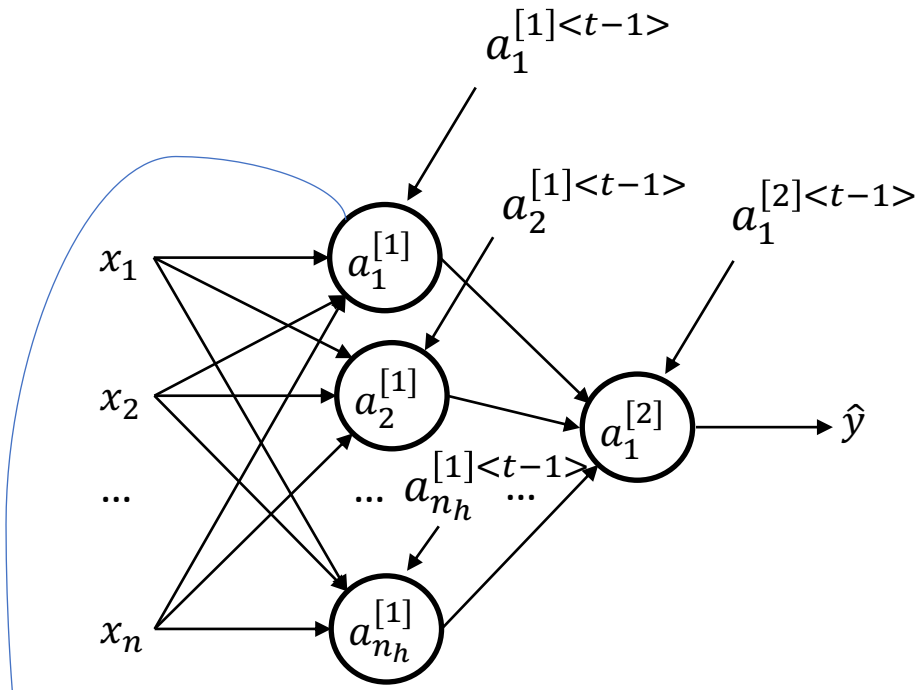


Contribution from current context
(encoding of previous inputs over time)

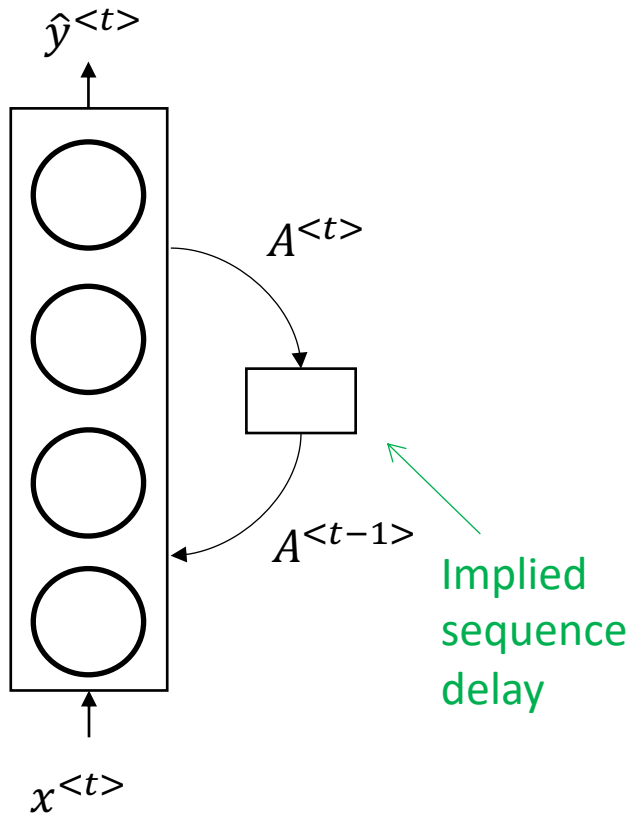
$$a_1^{[1]} = g(w_{ax}X_1 + w_{aa}a_1^{[1]<t-1>} + b)$$

Contribution from current input

Recurrent Neural Networks (RNNs)



drawing abstraction



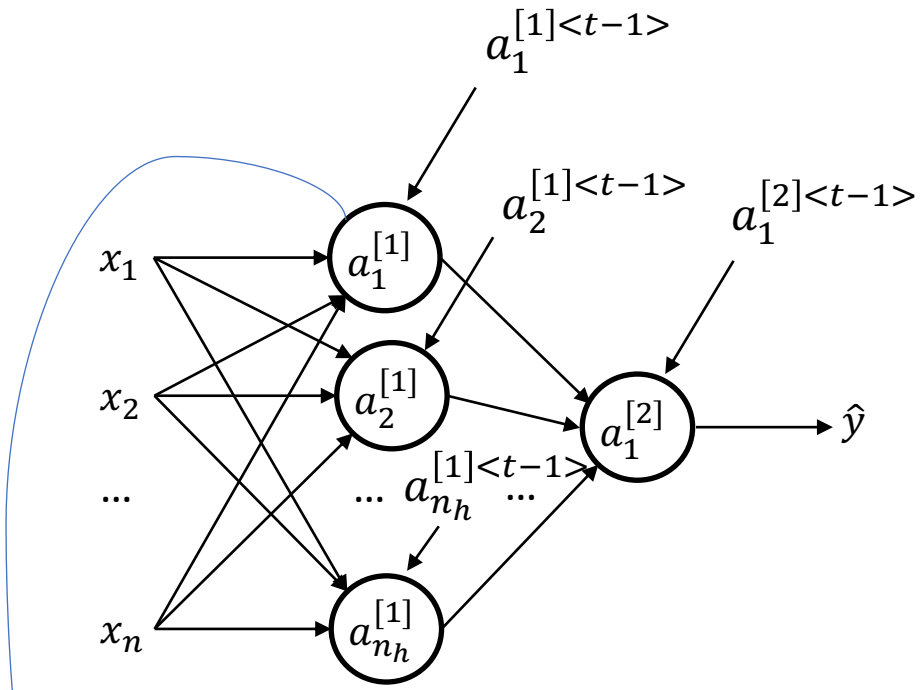
Contribution from current context (encoding of previous inputs over time)

$$a_1^{[1]} = g(w_{ax}X_1 + w_{aa}a_1^{[1]<t-1>} + b)$$

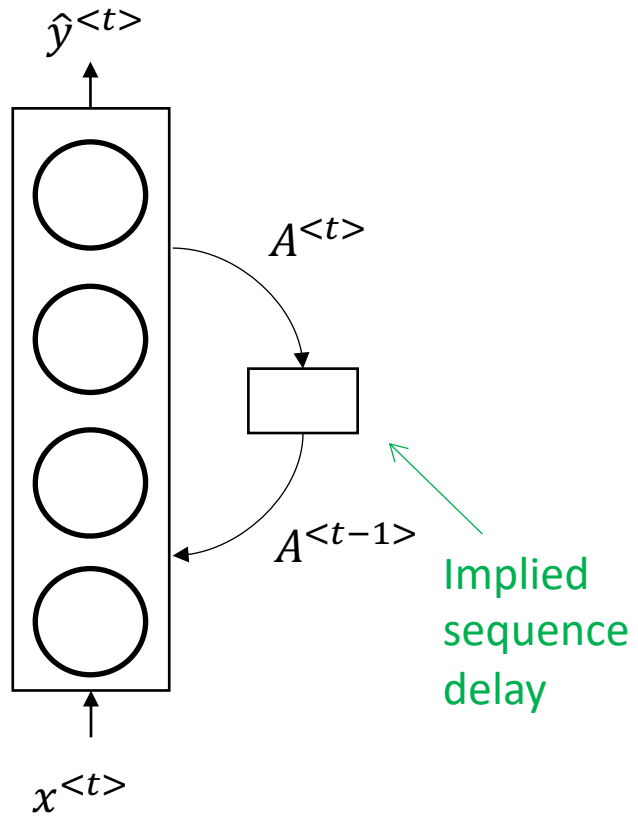
Contribution from current input

Brad Quinton, Scott Chin

Recurrent Neural Networks (RNNs)



drawing abstraction



Contribution from current context (encoding of previous inputs over time)

$$a_1^{[1]} = g(w_{ax}X_1 + w_{aa}a_1^{[1]<t-1>} + b)$$

Contribution from current input

Learned Parameters:

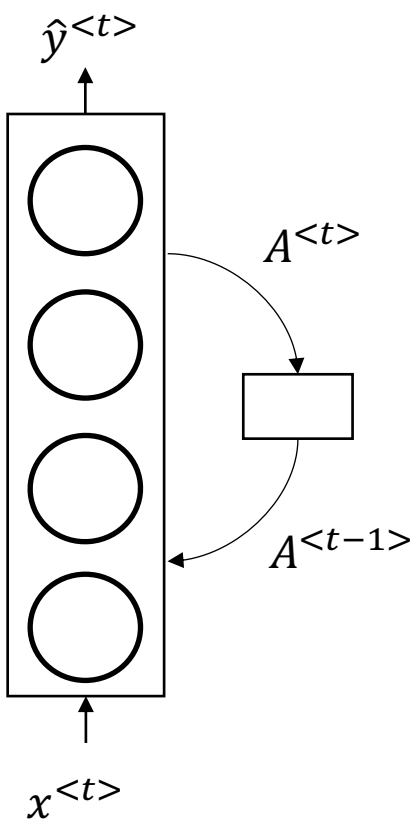
W_{ax} - "regular" NN parameters

W_{aa}- previous activation parameters

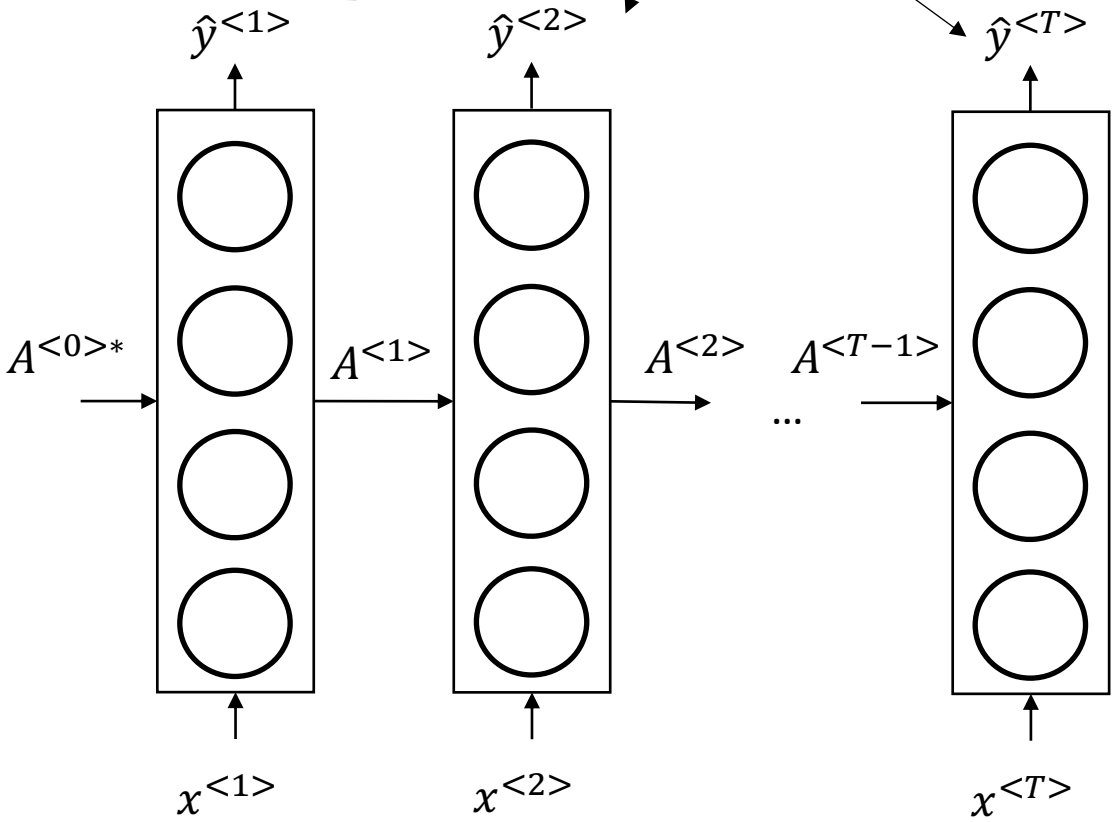
Brad Quinton, Scott Chin

RNNs

Note: This is the same NN with the same parameters W_{ax} and W_{aa}

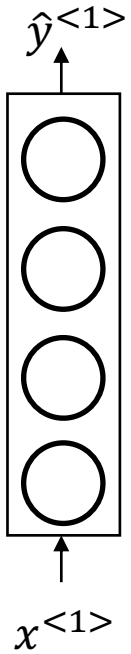


Unrolled for sequence t=[1:T]



*By convention $A^{<0>}$ is normally set equal to '0'.

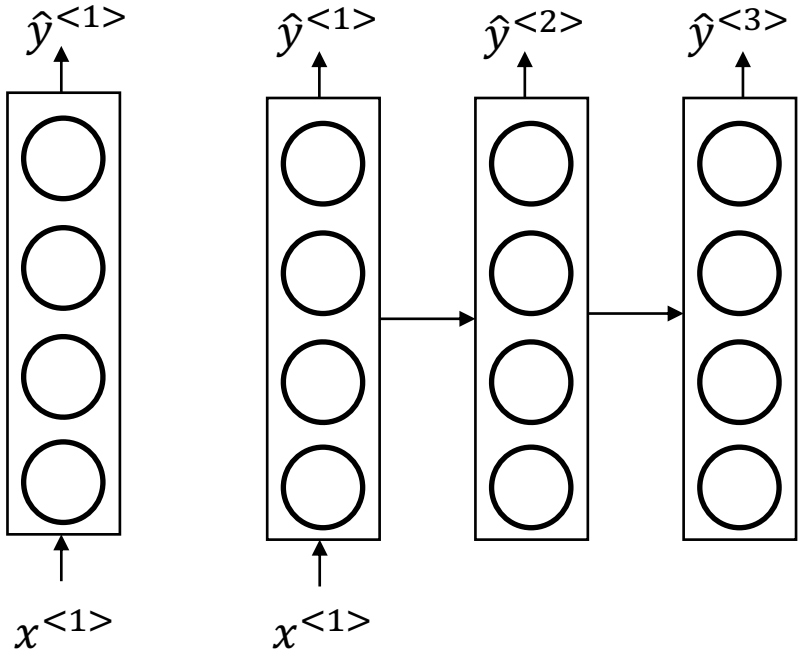
I/O Sequence Length Flexibility in RNNs



a. one-to-one

Image Classification

I/O Sequence Length Flexibility in RNNs



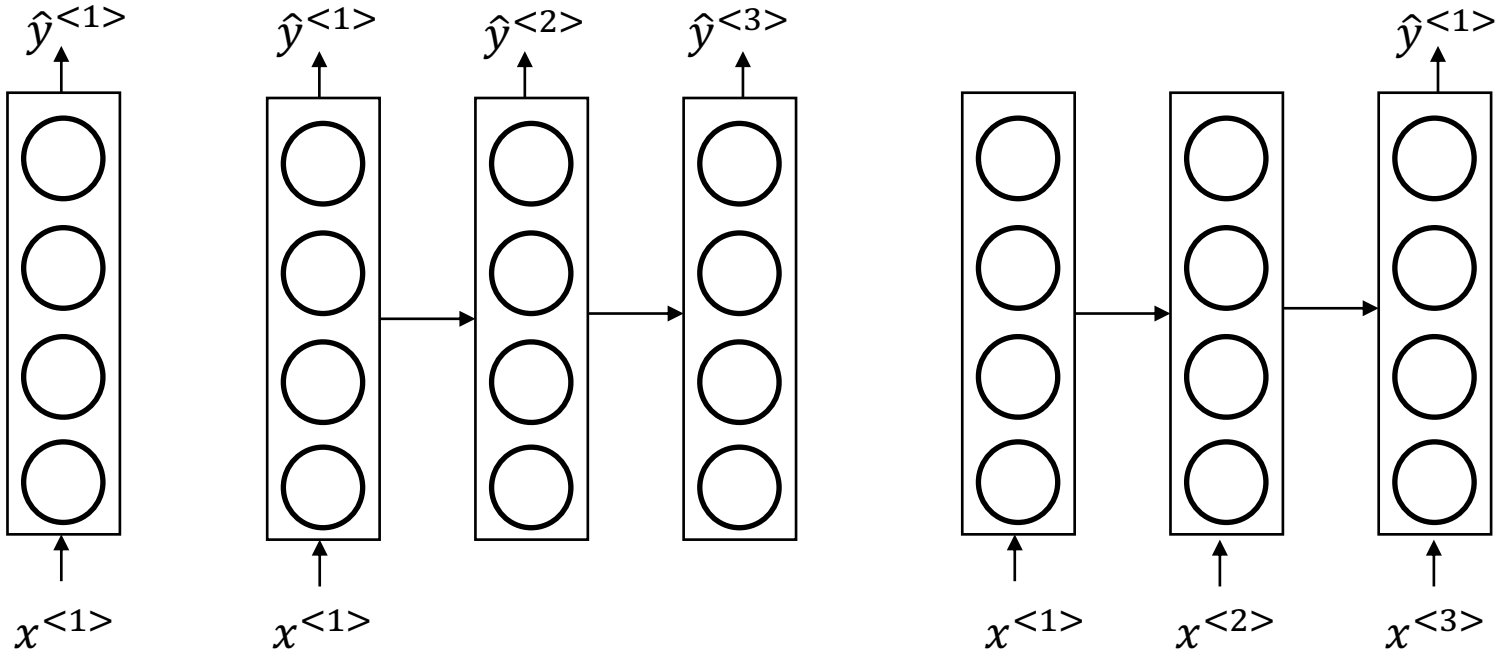
a. one-to-one

Image Classification

b. one-to-many

Image Captioning

I/O Sequence Length Flexibility in RNNs



a. one-to-one

Image Classification

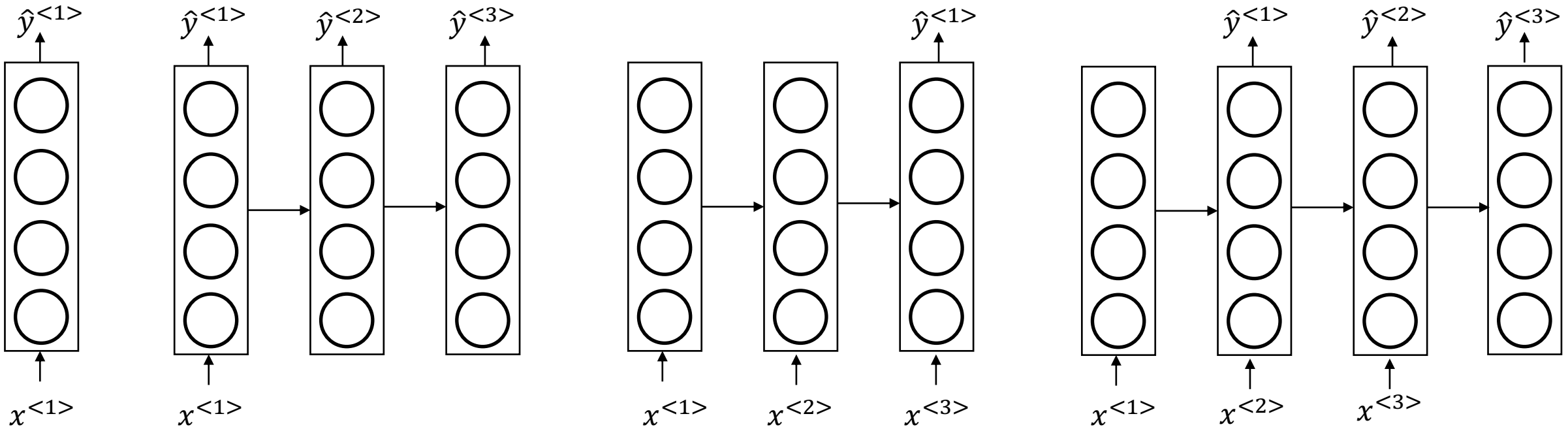
b. one-to-many

Image Captioning

c. one-to-many

Sentiment Classification

I/O Sequence Length Flexibility in RNNs



a. one-to-one

Image Classification

b. one-to-many

Image Captioning

c. one-to-many

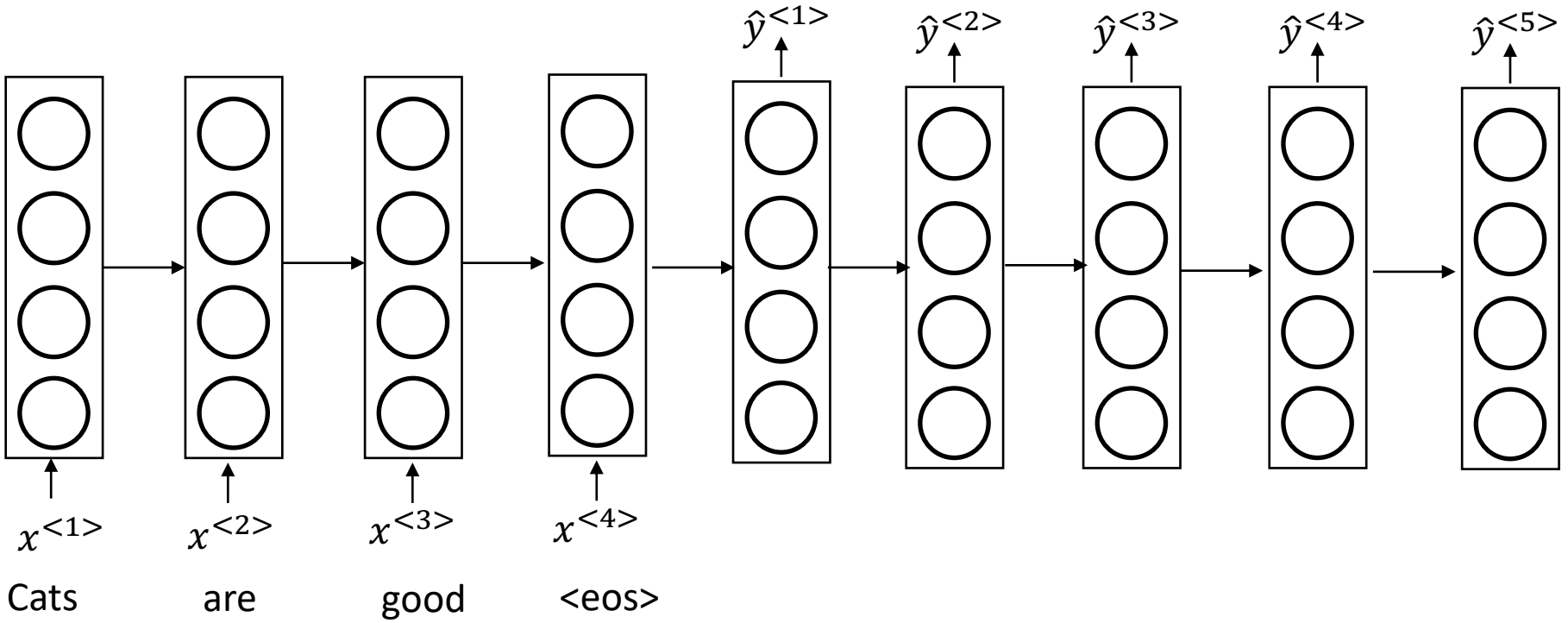
Sentiment Classification

d. many-to-many

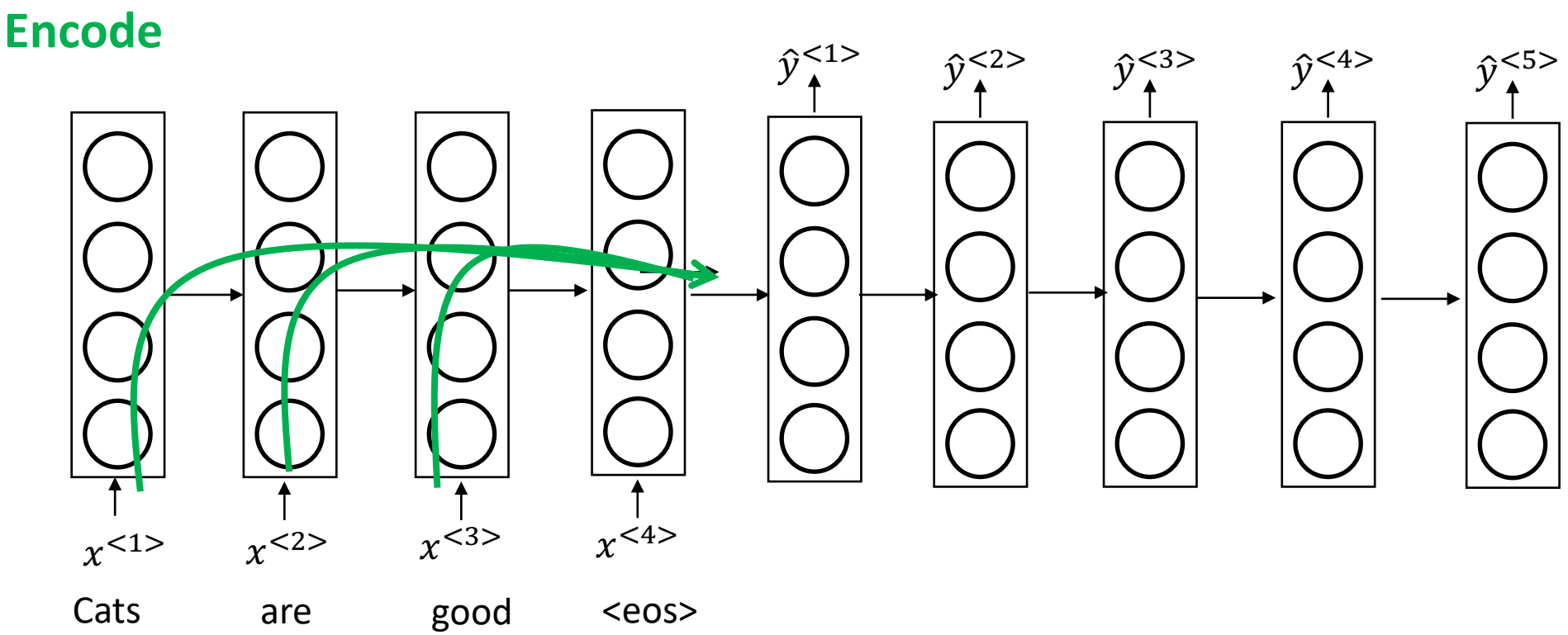
Machine Translation

Brad Quinton, Scott Chin

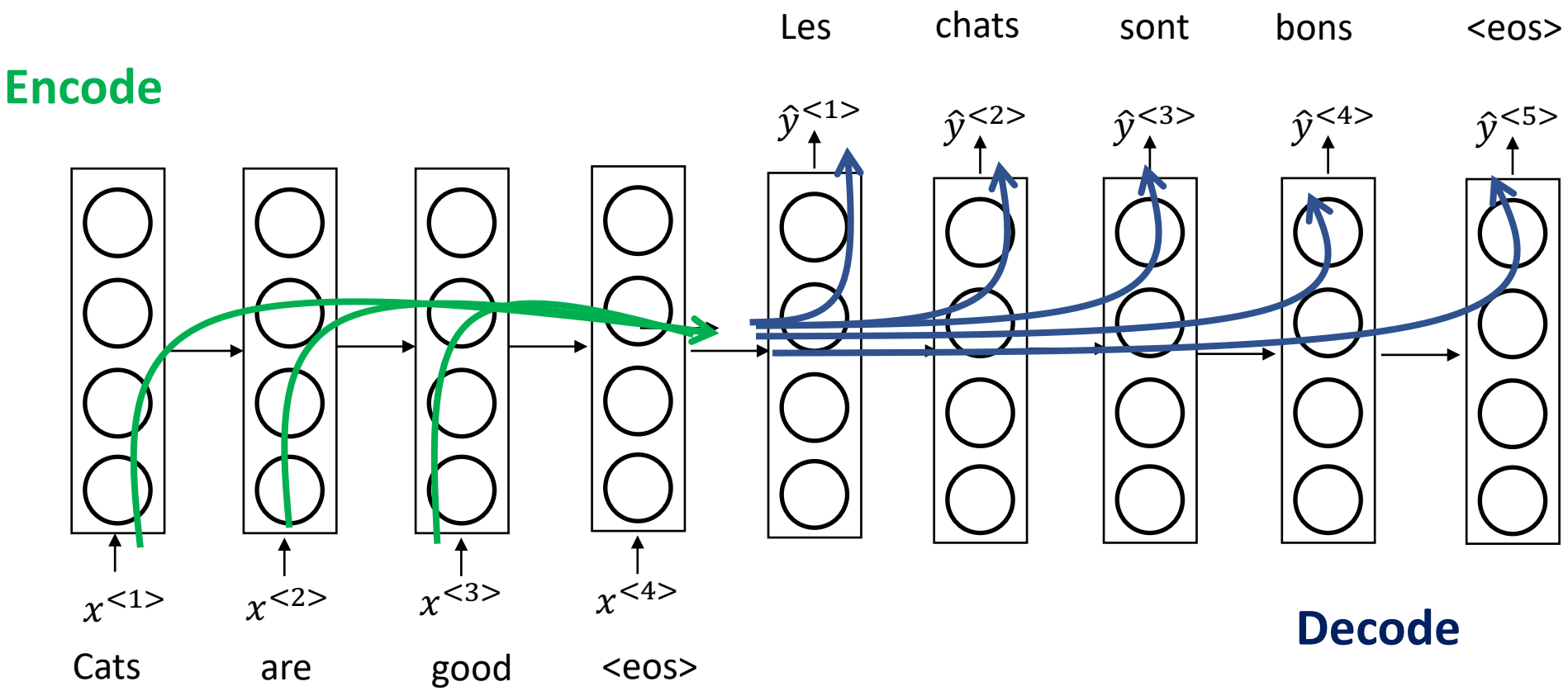
Machine Translation Example



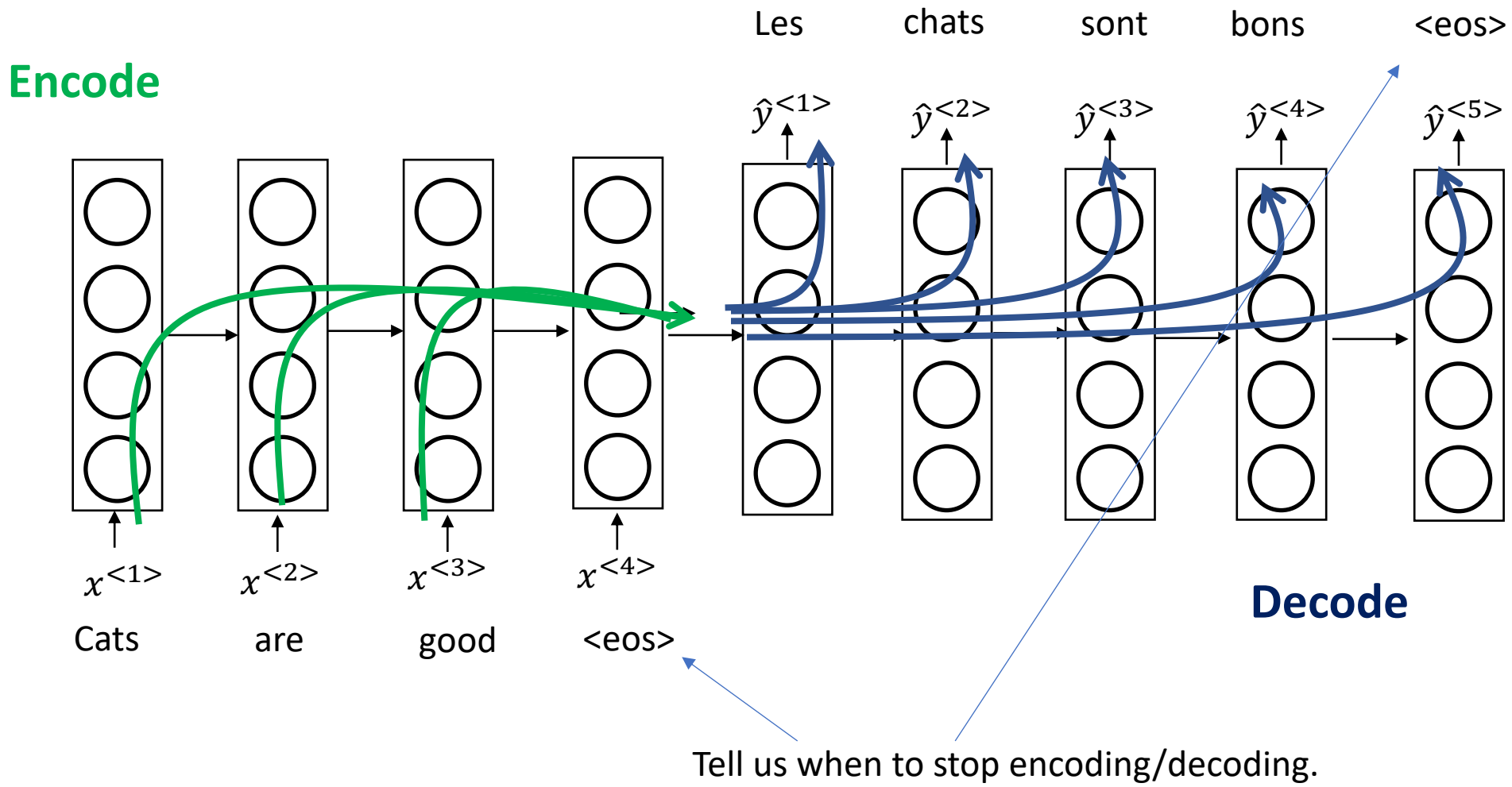
Machine Translation Example



Machine Translation Example



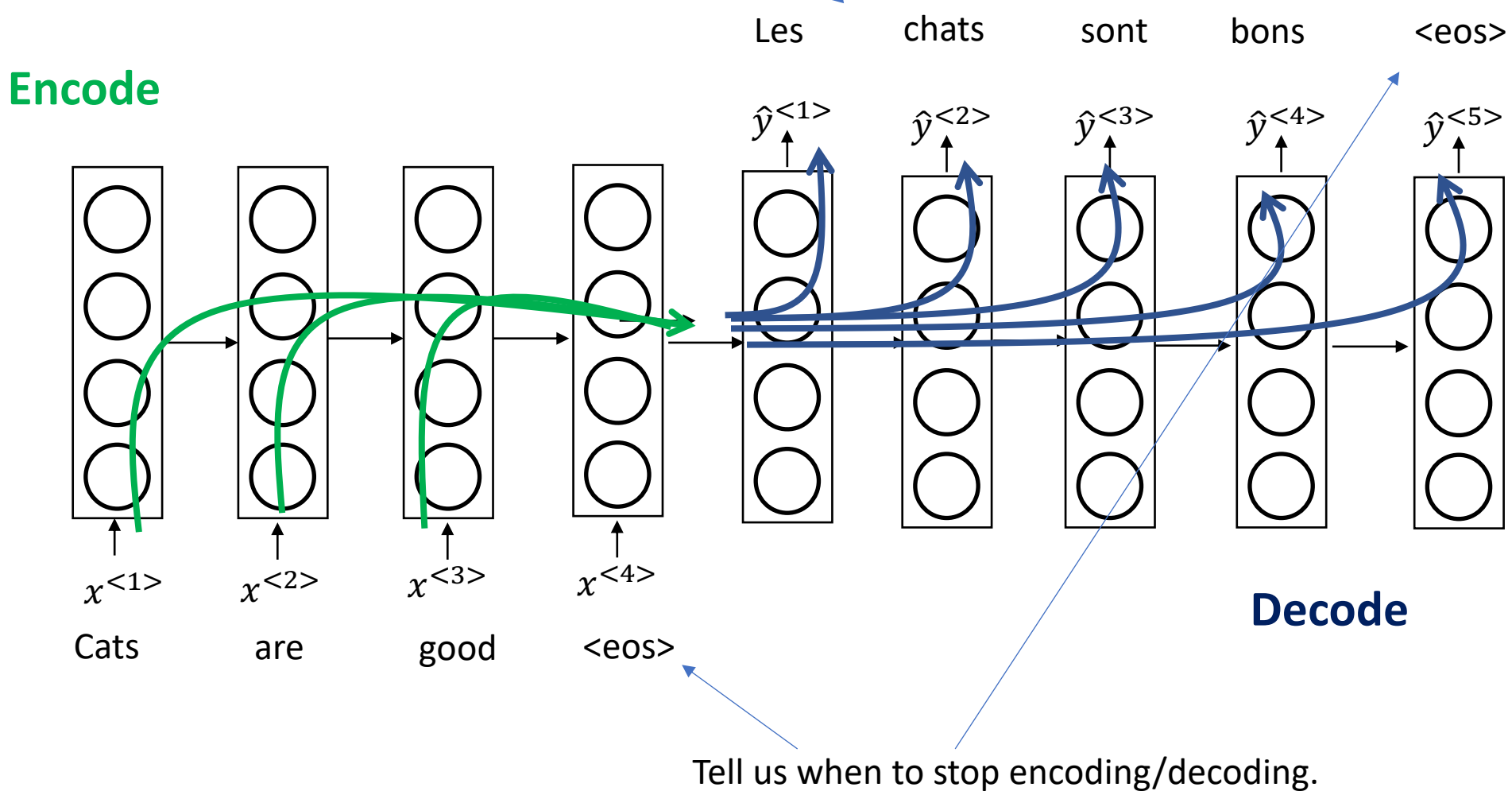
Machine Translation Example



Brad Quinton, Scott Chin

Machine Translation Example

Extra words are easily accommodated.



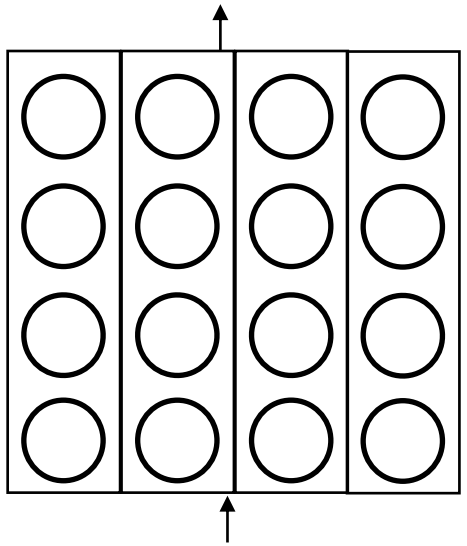
Tell us when to stop encoding/decoding.

Brad Quinton, Scott Chin

RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure does a form of feature extraction

$\hat{y}^{(i)} = [y_1\ y_2\ y_3\ y_4]$



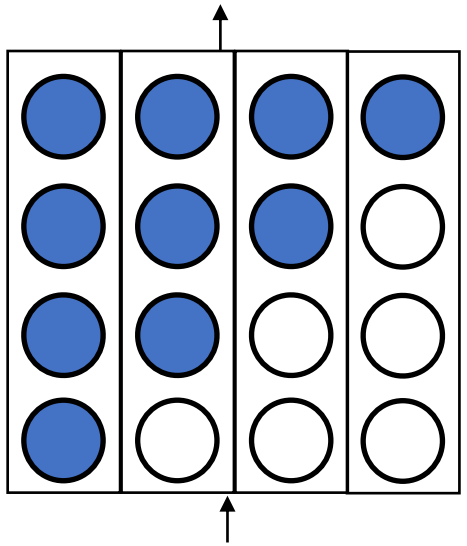
$x^{(1)} = [\text{“}Cats\ are\ really\ great\text{”}]$

$x^{(2)} = [\text{“}I\ really\ love\ cats\text{”}]$

RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure does a form of feature extraction

$\hat{y}^{(i)} = [y_1\ y_2\ y_3\ y_4]$



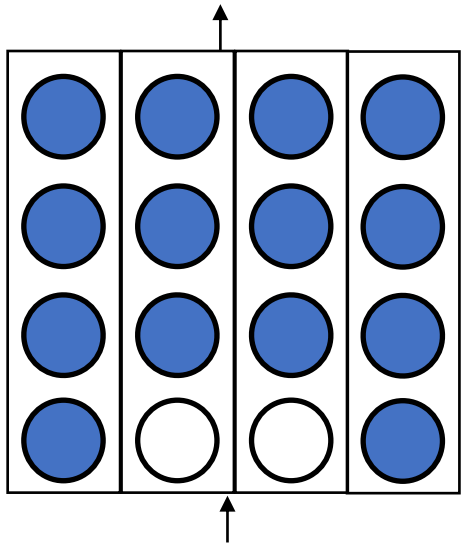
$x^{(1)} = [\text{“}\boldsymbol{Cats}\ are\ really\ great\text{”}]$

$x^{(2)} = [\text{“}I\ really\ love\ cats\text{”}]$

RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction

$\hat{y}^{(i)} = [y_1\ y_2\ y_3\ y_4]$



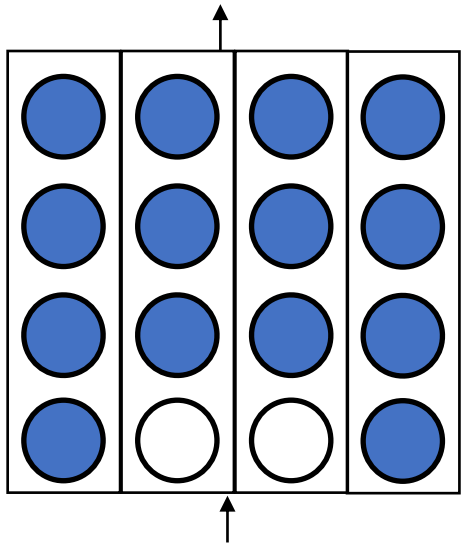
$x^{(1)} = [\text{“}\boldsymbol{Cats}\ are\ really\ great\text{”}]$

$x^{(2)} = [\text{“}I\ really\ love\ \boldsymbol{cats}\text{”}]$

RNN Feature Extraction

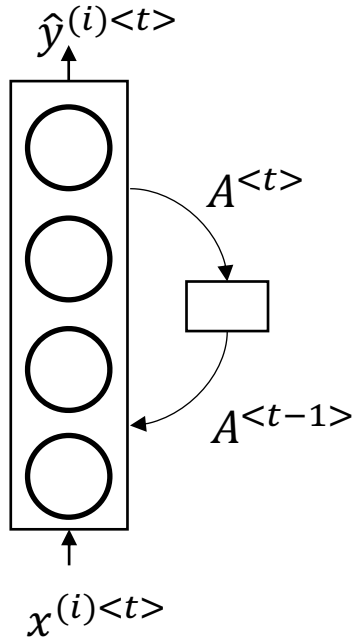
- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction

$\hat{y}^{(i)} = [y_1\ y_2\ y_3\ y_4]$



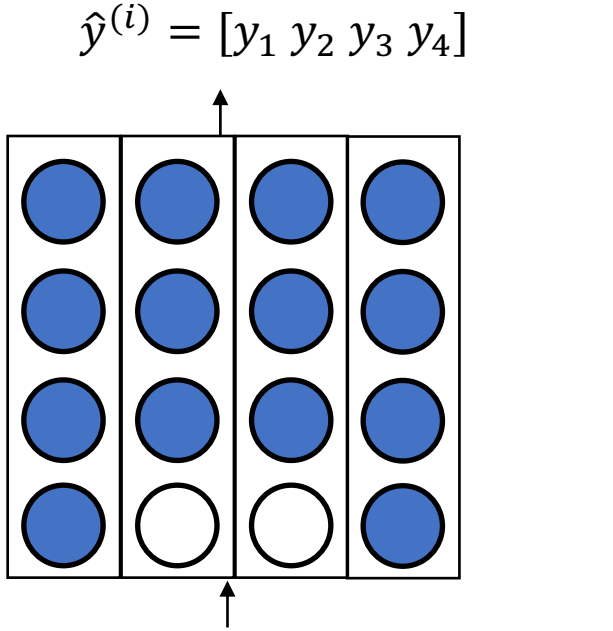
$x^{(1)} = [\text{“}\boldsymbol{Cats}\ are\ really\ great\text{”}]$

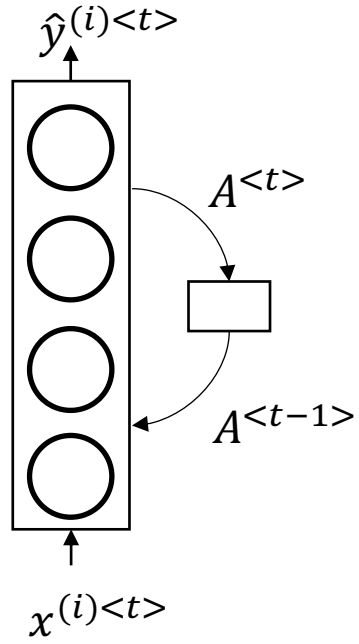
$x^{(2)} = [\text{“}I\ really\ love\ \boldsymbol{cats}\text{”}]$



RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction





$x^{(1)<1>} = Cats$

$x^{(1)<2>} = are$

$x^{(1)<3>} = really$

$x^{(1)<4>} = great$

$x^{(2)<1>} = I$

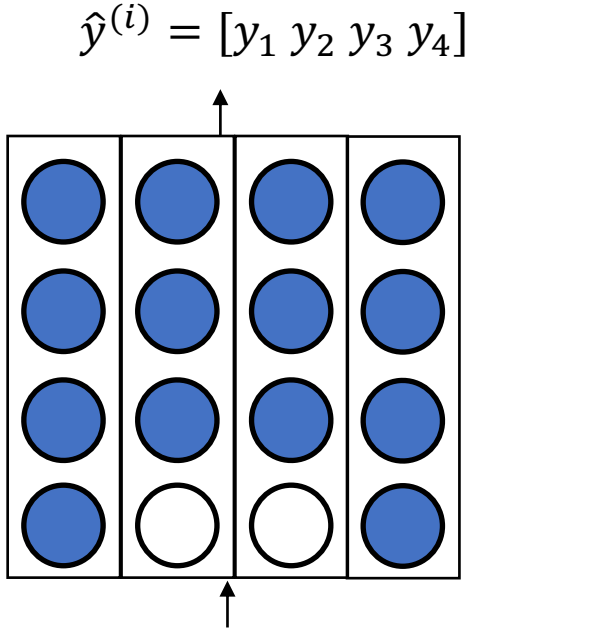
$x^{(2)<2>} = really$

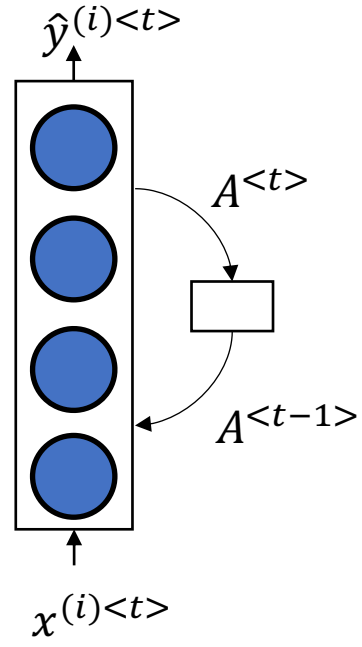
$x^{(2)<3>} = love$

$x^{(2)<4>} = cats$

RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction





$x^{(1)<1>} = \boldsymbol{Cats}$

$x^{(1)<2>} = are$

$x^{(1)<3>} = really$

$x^{(1)<4>} = great$

$x^{(2)<1>} = I$

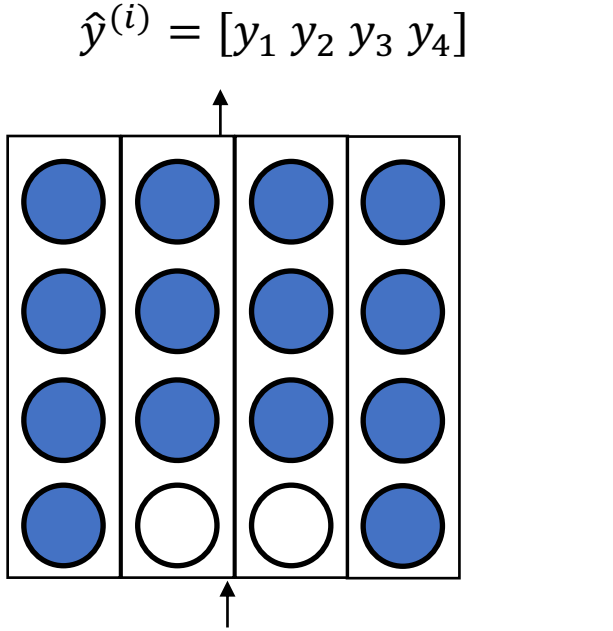
$x^{(2)<2>} = really$

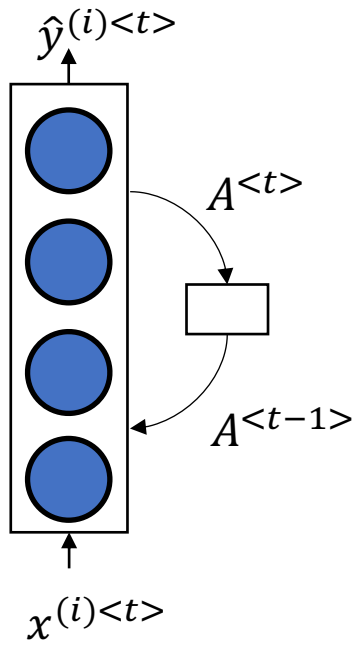
$x^{(2)<3>} = love$

$x^{(2)<4>} = cats$

RNN Feature Extraction

- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction





$x^{(1)<1>} = \boldsymbol{Cats}$

$x^{(1)<2>} = are$

$x^{(1)<3>} = really$

$x^{(1)<4>} = great$

$x^{(2)<1>} = I$

$x^{(2)<2>} = really$

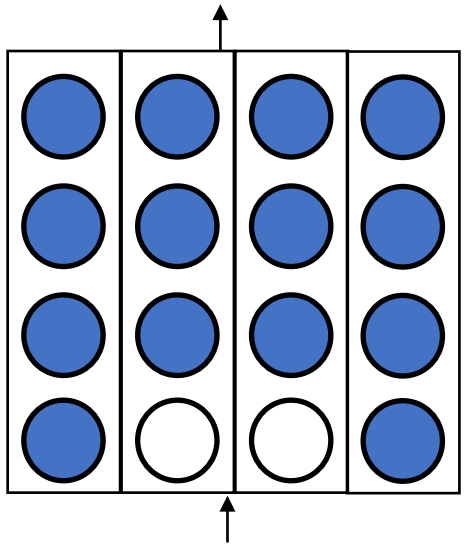
$x^{(2)<3>} = love$

$x^{(2)<4>} = cats$

RNN Feature Extraction

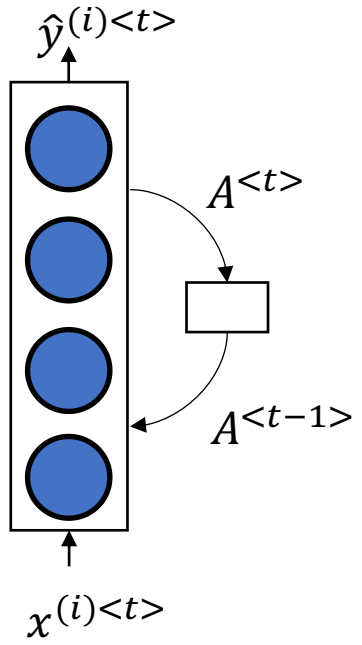
- Beyond providing flexibility around the size of the input and the output data, the RNN structure has does a form of feature extraction

$\hat{y}^{(i)} = [y_1\ y_2\ y_3\ y_4]$



$x^{(1)} = [\text{“}\boldsymbol{Cats}\ are\ really\ great\text{”}]$

$x^{(2)} = [\text{“}I\ really\ love\ \boldsymbol{cats}\text{”}]$



$x^{(1)<1>} = \boldsymbol{Cats}$

$x^{(1)<2>} = are$

$x^{(1)<3>} = really$

$x^{(1)<4>} = great$

$x^{(2)<1>} = I$

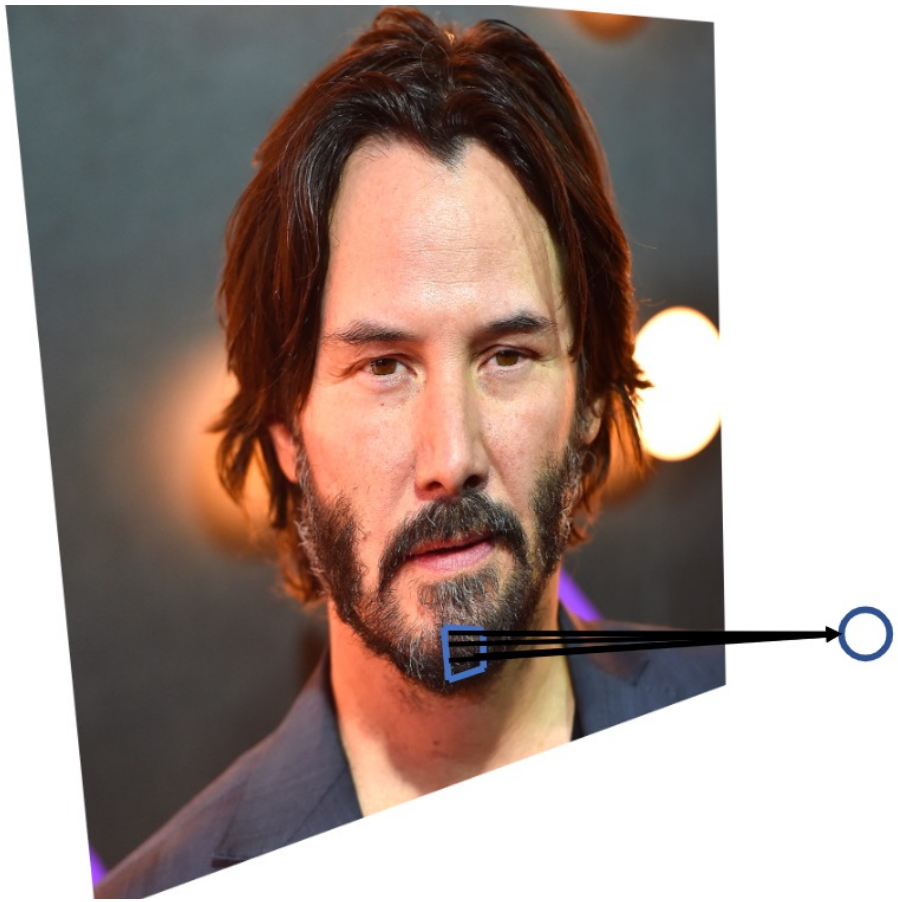
$x^{(2)<2>} = really$

$x^{(2)<3>} = love$

$x^{(2)<4>} = cats$

Analogy to CNNs

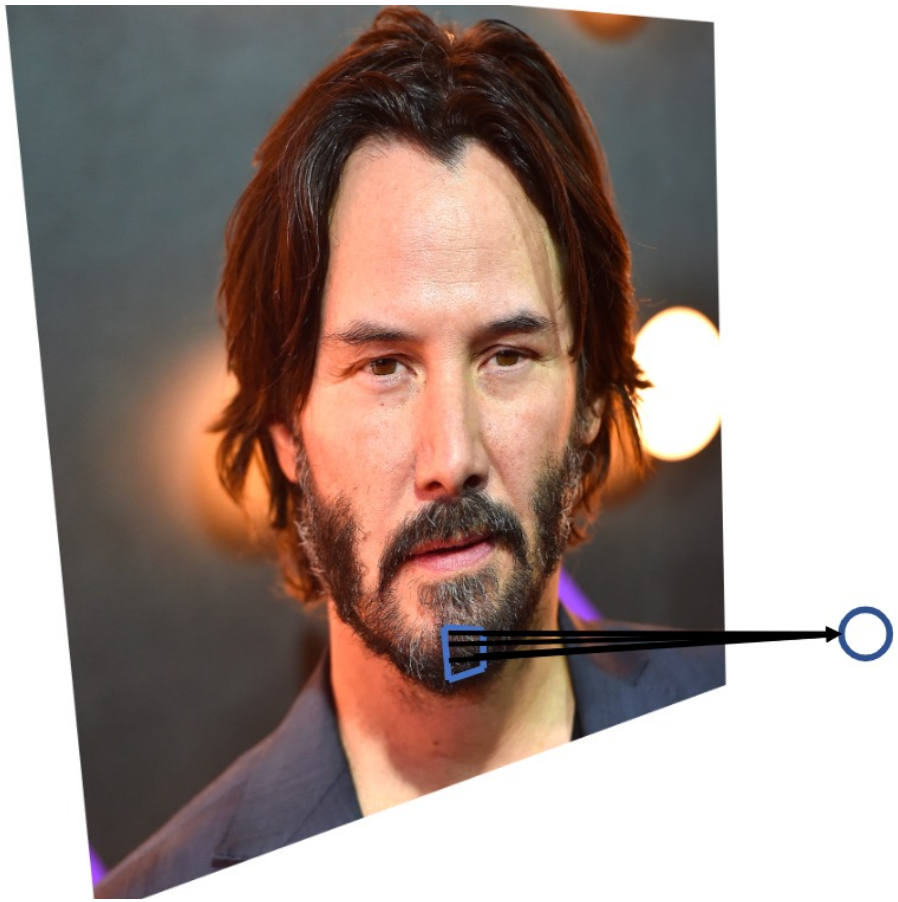
- Remember from Scott's lecture…



- For example, if a neuron learned to activate strongly to patterns that make up an 11x11 swatch of facial hair.

Analogy to CNNs

- Remember from Scott's lecture…



- For example, if a neuron learned to activate strongly to patterns that make up an 11x11 swatch of facial hair.

RNNs isolate elements of sequences like convolutional filters isolate regions of an image.

Context Doesn’t Only Flow One Way…

- Although data unfolds over time, once we have the data we can ”look forward and backward in time”

- Even when we deploy a system, we can often buffer the inputs long enough to consider context in two directions

- Consider these sentences:

> “The fair approach is to split the profits 50/50.”
>
> “The fair was in town for only one night.”

Brad Quinton, Scott Chin

Context Doesn’t Only Flow One Way…

- Although data unfolds over time, once we have the data we can ”look forward and backward in time”

- Even when we deploy a system, we can often buffer the inputs long enough to consider context in two directions

- Consider these sentences:

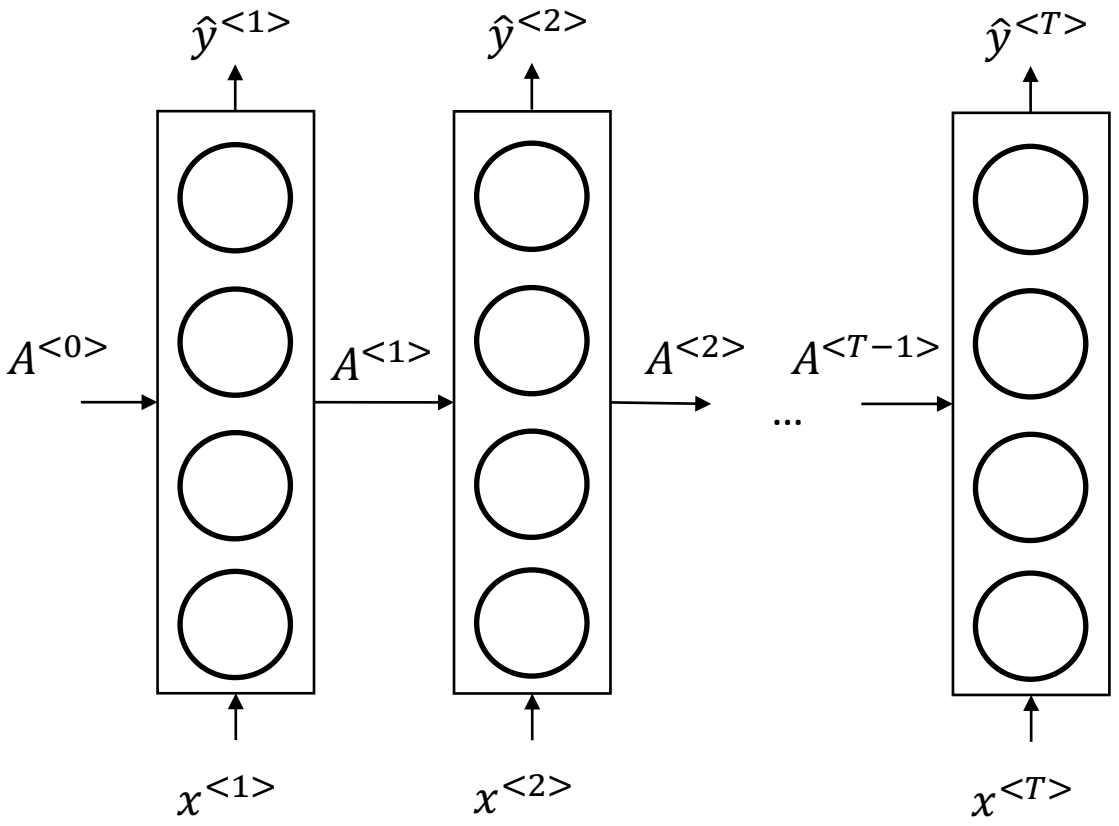
“The fair approach is to split the profits 50/50.”

“The fair was in town for only one night.”

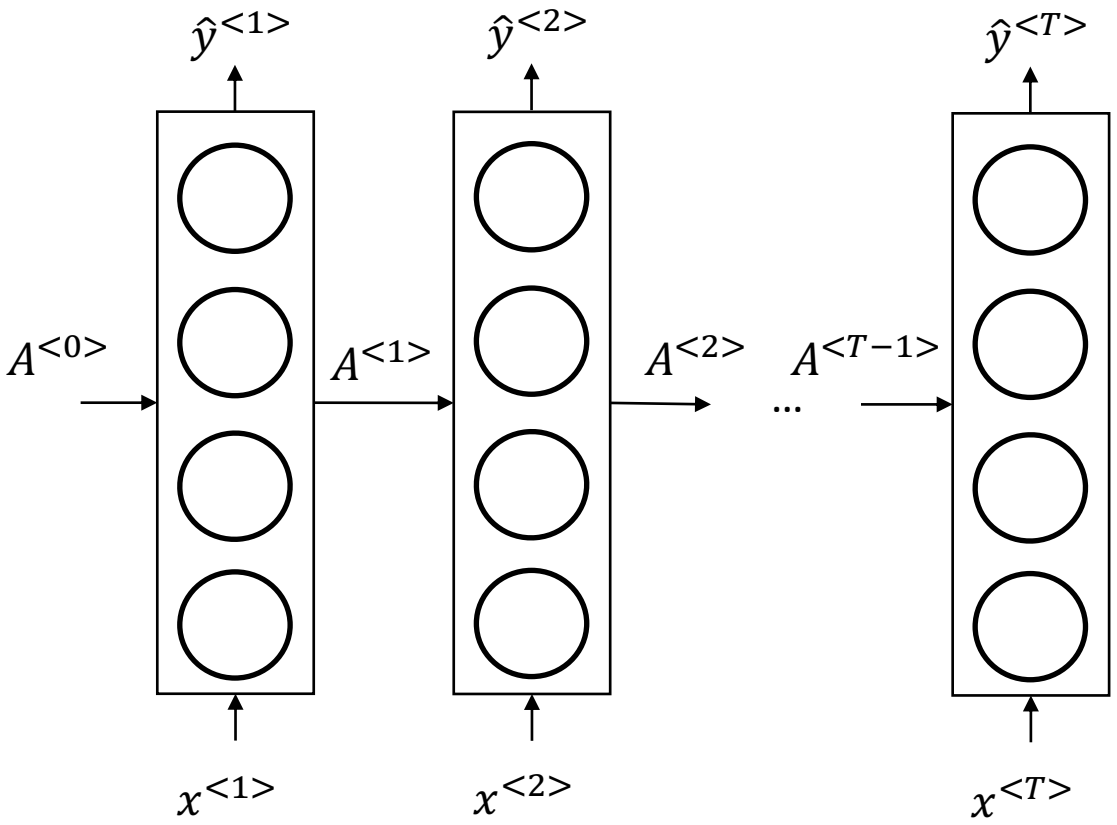
The context is defined later in the sentence.

Brad Quinton, Scott Chin

Standard RNN

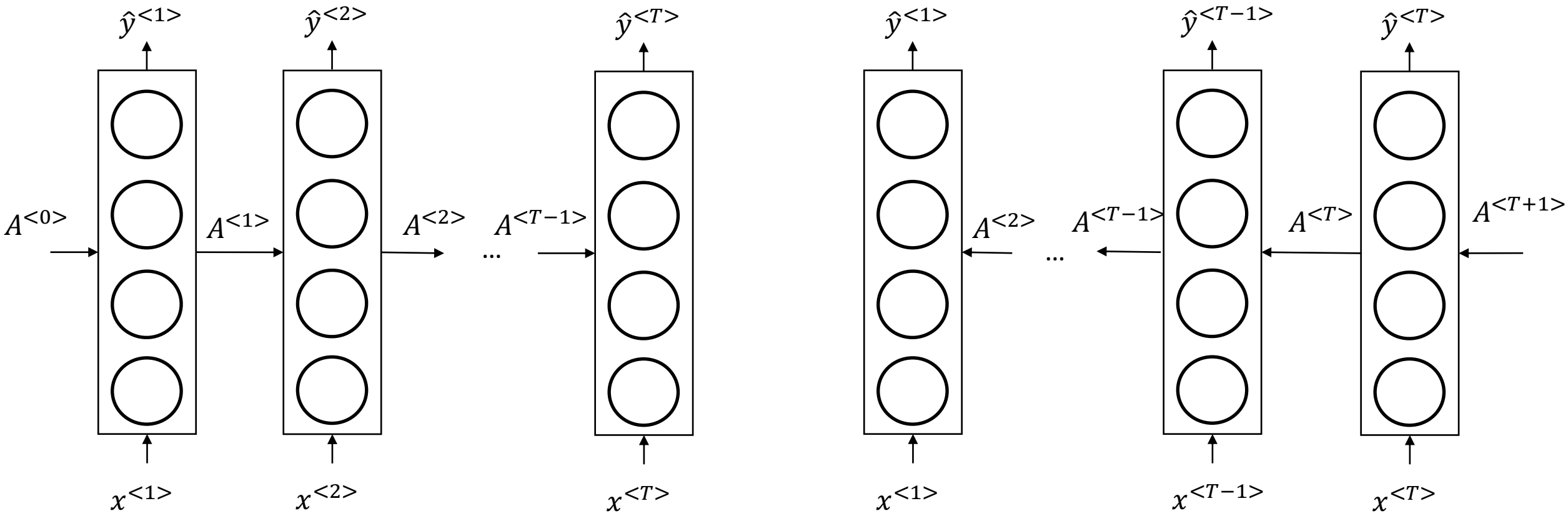


Creating a Bidirectional RNN



Forward flow RNN

Creating a Bidirectional RNN



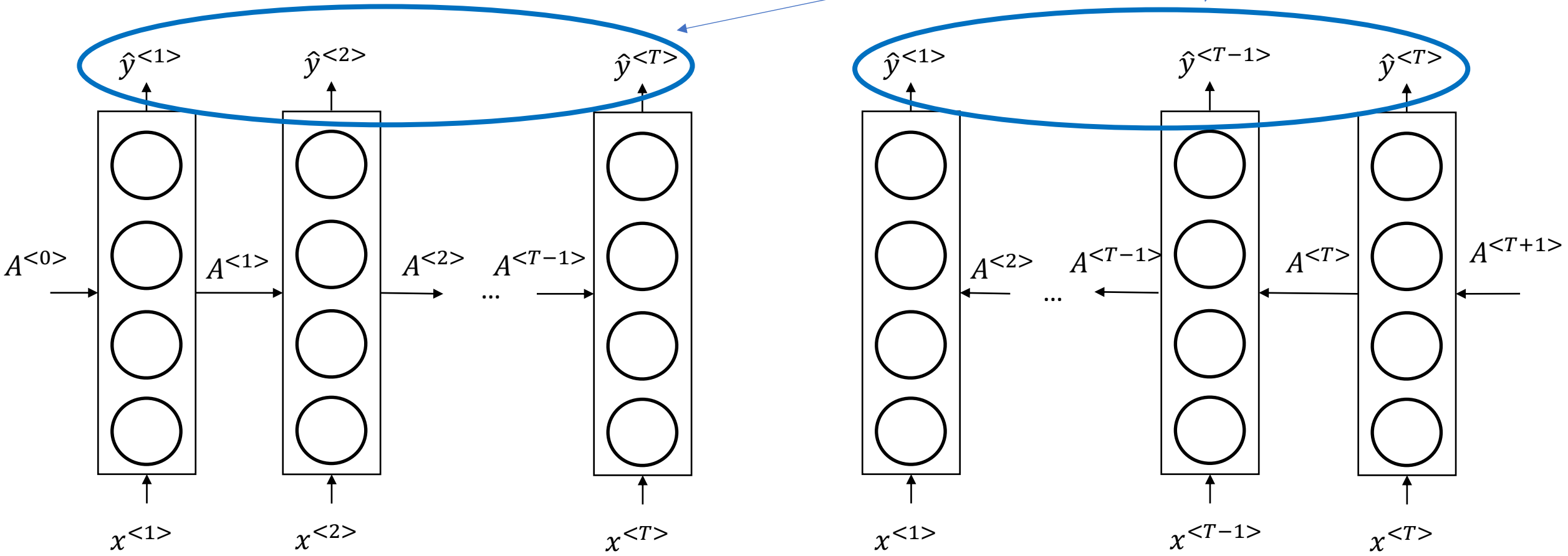
Forward flow RNN

Backward flow RNN

Brad Quinton, Scott Chin

Creating a Bidirectional RNN

Just need to combine the outputs…

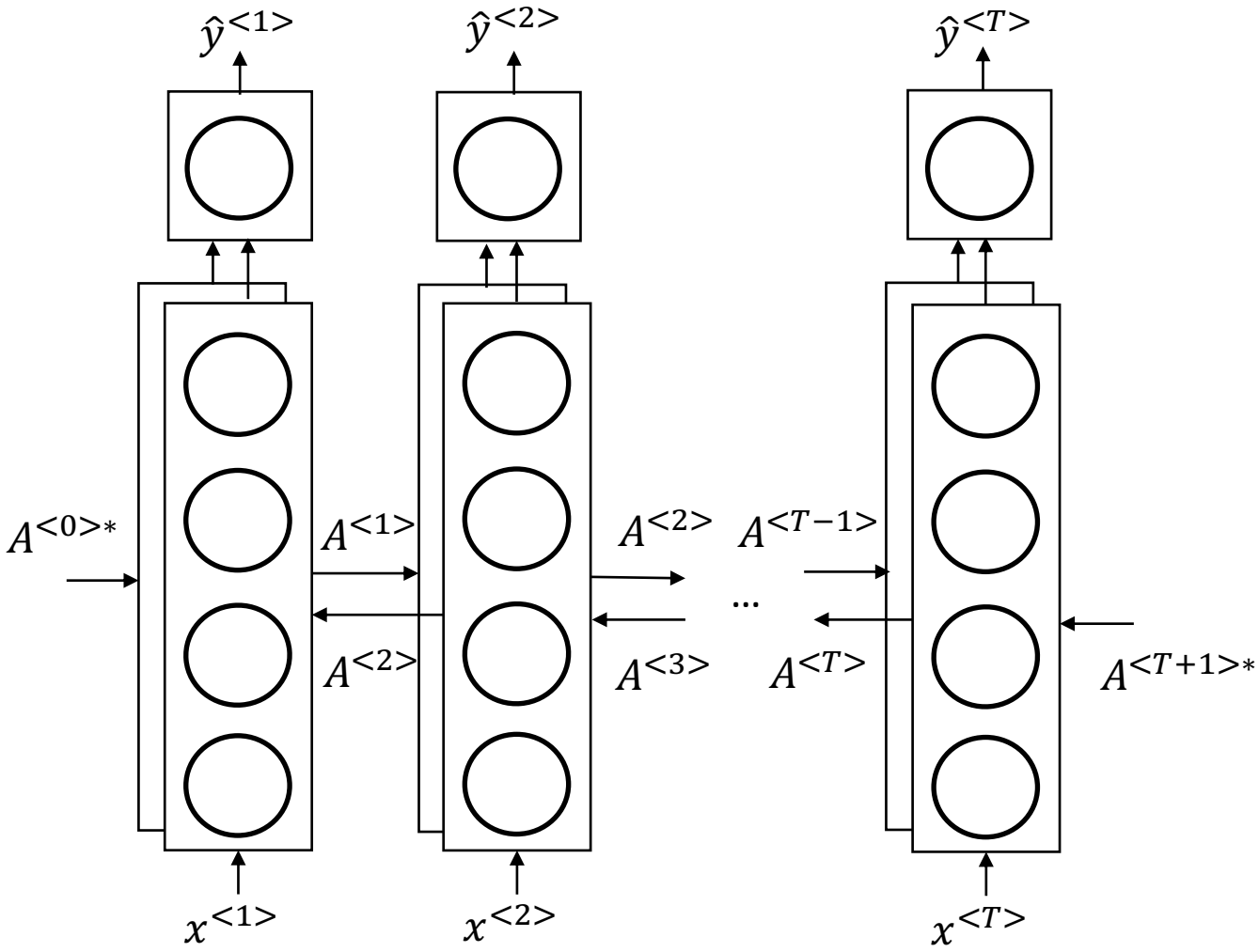


Forward flow RNN

Backward flow RNN

Brad Quinton, Scott Chin

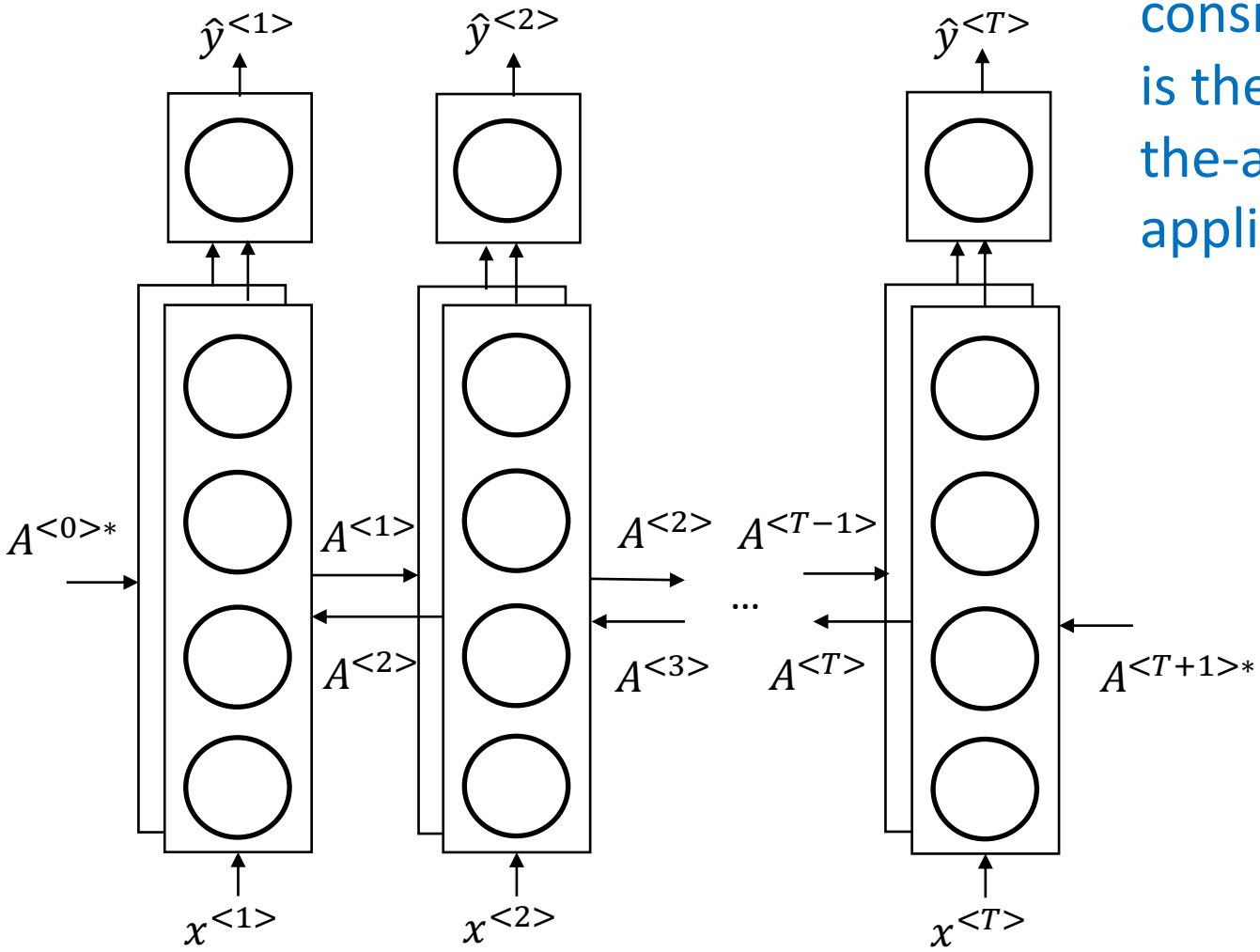
Bidirectional RNN



*By convention $A^{<0>}$ and $A^{<T+1>}$ are normally set equal to '0'.

Bidirectional RNN

Using BRNNs with each sentence considered a sequence is the current state-of-the-art for most NLP applications today.



*By convention $A^{<0>}$ and $A^{<T+1>}$ are normally set equal to '0'.

RNNs Applications

- As I pointed out before much of data in the real world is sequential: Sound, Video, Natural language, Online interactions, Music, Sports, Real-time Navigating, Radar Tracking, etc.

- RNNs expanded the applications of Deep Learning substantially!

Character-level RNN Generation

- Here is some fictious C++ written by a character-level RNN trained on the Linux Source

```
/*
 * 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++) {
```

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

OpenAI GPT-2 Natural Language Processing

Artificial intelligence (AI)

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

Alex Hern

@alexhern

Thu 14 Feb 2019
12.00 EST

https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction

OpenAI GPT-2 Natural Language Processing

SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

https://openai.com/blog/better-language-models/

Image Captioning

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."

"black ar jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

https://cs.stanford.edu/people/karpathy/deepimagesent/

Brad Quinton, Scott Chin

Aside - Predicting Human Behaviour?

- Consider this as a thought experiment: With what you know, could you build an AI to predict what you (or anyone else) will do next?

- How would you train it? How would it make predictions?

Aside - Predicting Human Behaviour?

- Consider this as a thought experiment: With what you know, could you build an AI to predict what you (or anyone else) will do next?

- How would you train it? How would it make predictions?

  Step 1: Record everything you see (video recording), hear (audio recording)

  Step 2: Record everywhere ever you move (GPS?) and whatever you say (audio recording) and type (keyboard capture)

Aside - Predicting Human Behaviour?

- Consider this as a thought experiment: With what you know, could you build an AI to predict what you (or anyone else) will do next?

- How would you train it? How would it make predictions?

  Step 1: Record everything you see (video recording), hear (audio recording)

  Step 2: Record everywhere ever you move (GPS?) and whatever you say (audio recording) and type (keyboard capture)

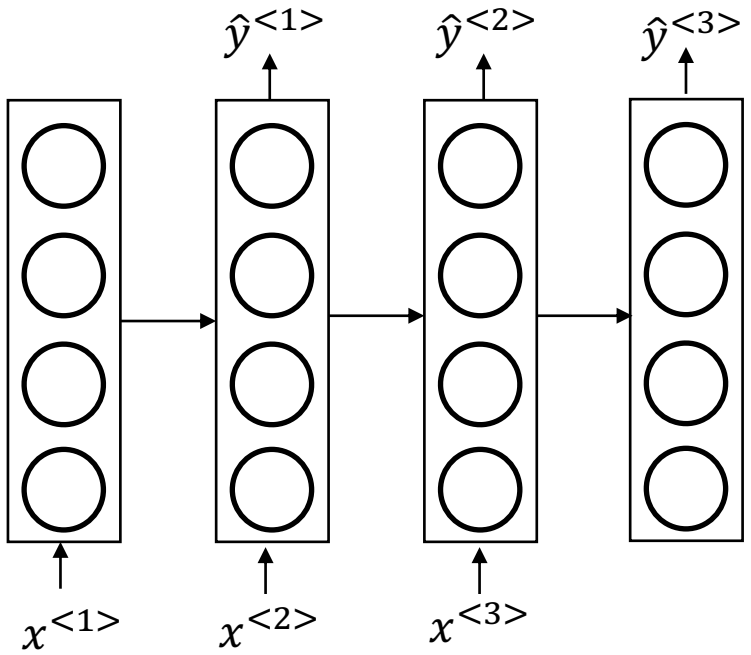
over let’s say a year…

Brad Quinton, Scott Chin

Aside - Predicting Human Behaviour?

- **Step 3:** Train our RNN

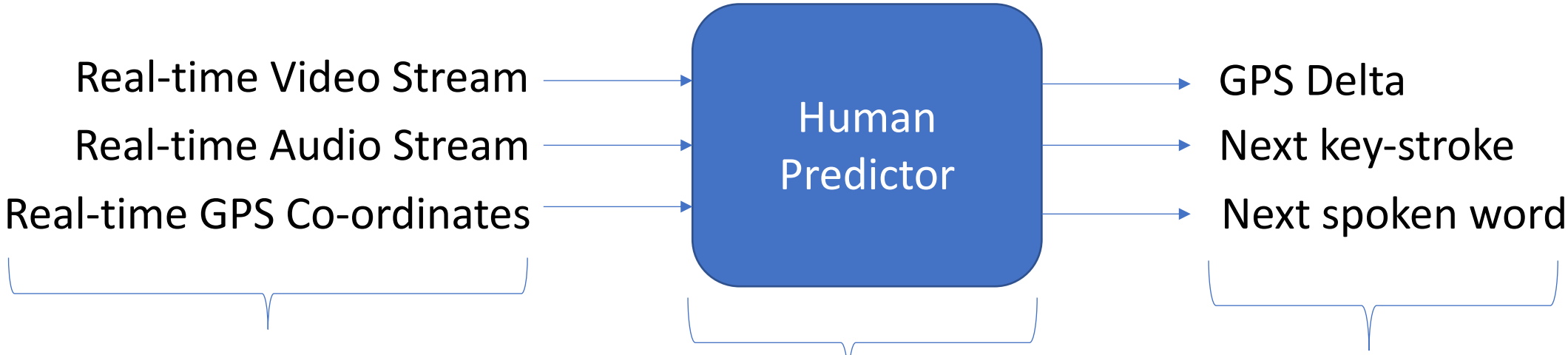
Y = Voice recording, typing record, GPS delta.



X = Video, Audio recording, GPS location.

Aside - Predicting Human Behaviour?

- **Step 4:** Deployment



Current stimulus

Next action prediction.

Learned Parameters – Encoded action probabilities based on what you have done in the past.

RNN feedback (internal state) – What you have experienced in the recent past (i.e. your *context*)

Aside - Predicting Human Behaviour?

- Would it work?
  - Your phone is certainly capable of the data recording/storage required
  - The biggest distributed RNNs would be able to process the data without much of a challenge
  - The only real question would be **how predictable** are you and **would it be worth** the time and effort to do that training.

Aside - Predicting Human Behaviour?

- Consider:

  If hear your alarm, what's the chances that you get up, say "good morning" and check your e-mail?

  If you are at the Canucks game what are the chances you are going to cheer?

  If you hear your favorite song when you wake-up, are you more likely to buy a more expensive coffee?

  You you see an article about obesity, will you click on the running shoe ad? And buy?

Brad Quinton, Scott Chin

Aside - Predicting Human Behaviour?

- Many human behaviors are predictable AND there is a huge $ motivation

- It is no co-incidence that the biggest on-line advertisers and retailers (Google, Facebook, Amazon) also have the biggest AI research budgets?