Deep Learning 101

Recurrent Neural Network (RNN)

Schedule

week	Date	Topic
9	10.27	Environment setup, python, Jupyter, PyCharm, TensorFlow, & regression
10	11.03	Training and testing
11	11.10	CNN
12	11.17	RNN
13	11.24	Autoencoder & GAN

Today’s Class

- Recap
- Confusion matrix
- What is RNN?
  - How do you predict the next token?
  - Vanilla RNN
  - LSTM
- Lab time

Recap – neural network

- Neural network as a function
  - y = f(x)
- Perceptron
  - Y = WX + b
  - Two inputs: x1, x2
  - One output: y
  - Linear regression
- XOR problem
  - Linear regression can’t solve the XOR problem
  - Require multivariate regression

Recap - Training

- What it take to train a neural network:
  - Hypothesis: H = WX + b
  - Activation function: Sigmoid, tanh, ReLU, LeakyReLU, Softmax, etc.
  - Cost function: MSE, cross entropy, etc.
  - Gradient descent: backpropagation
- Training a neural network is basically the problem of minimizing the cost function: minimize cost(W, b)
- Gradient descent is the most popular optimizer.
- Training a neural network is NOT easy!
  - Finding hyperparameters, random initial weights, local minima, vanishing/exploding gradients, overfitting/underfitting, etc.

Recap - CNN

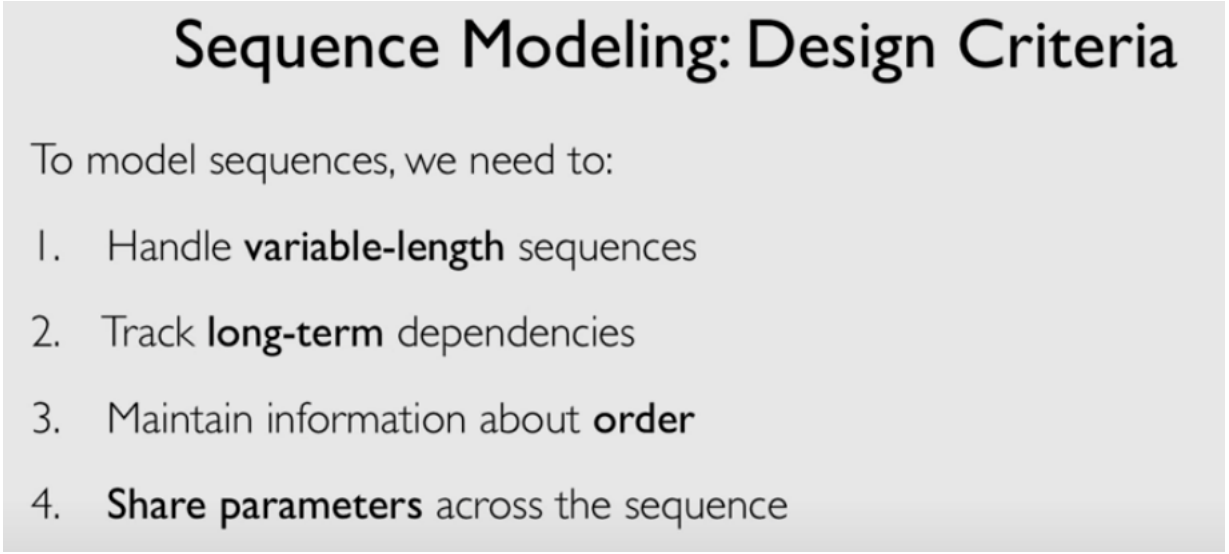
- Stride: step size
- Padding: putting zeroes around the outer edge of the input data. “same” means the output size will be the same as the input size when stride = 1.
- Kernel: the filter that extracts features
- Max pooling: means pooling the maximum value from the filtered feature map. The result is a down-sampling image (reduced dimensionality)
- Batch normalization: normalizing
- Dropout: regularization technique to prevent overfitting

How to model sequence?

Predicting the next word

“The quick brown fox jumped over the lazy ___”

1. Fixed window
2. Bag of words
3. Big fixed window?
   - Long term dependency



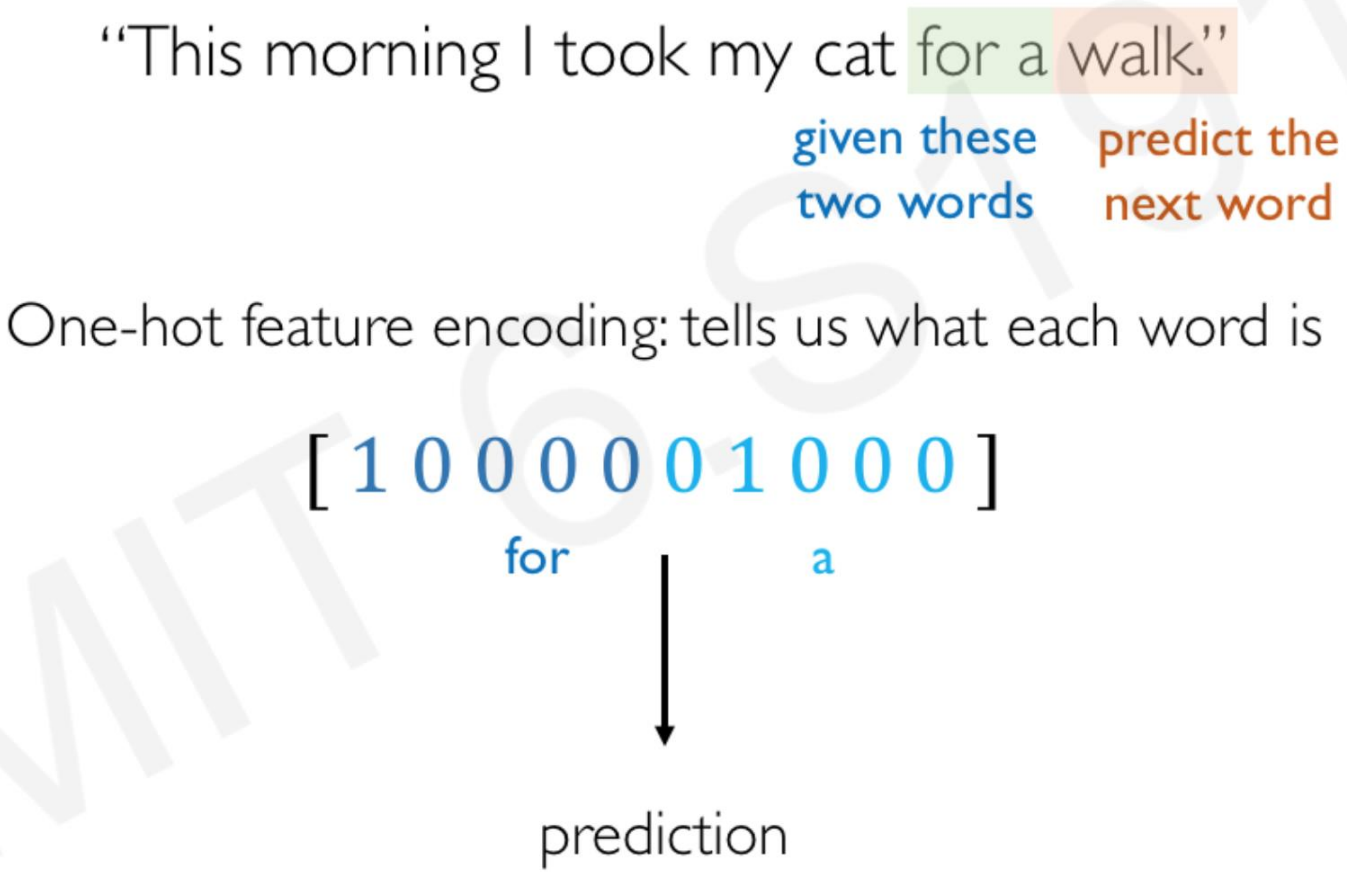
Next word prediction

"This morning I took my cat for a walk."

given these two words

predict the next word

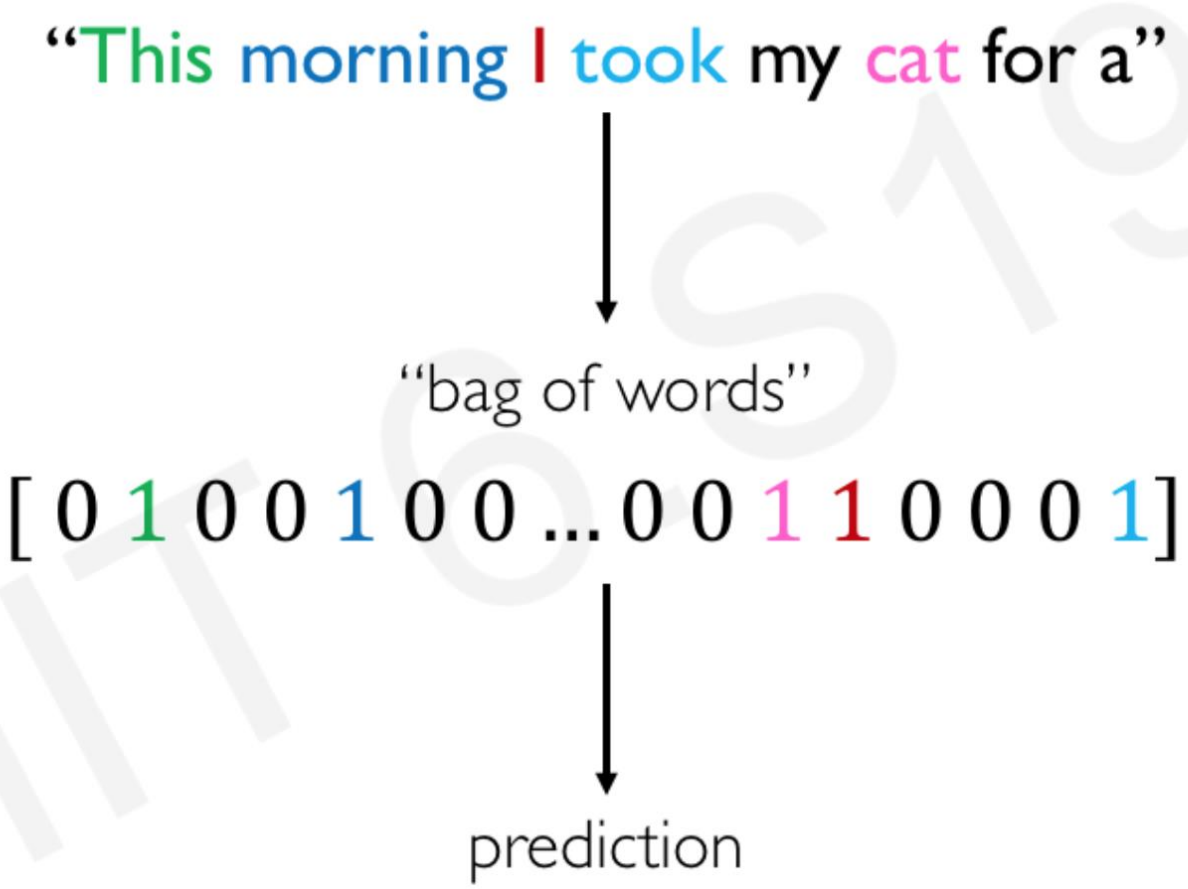
Fixed Window?



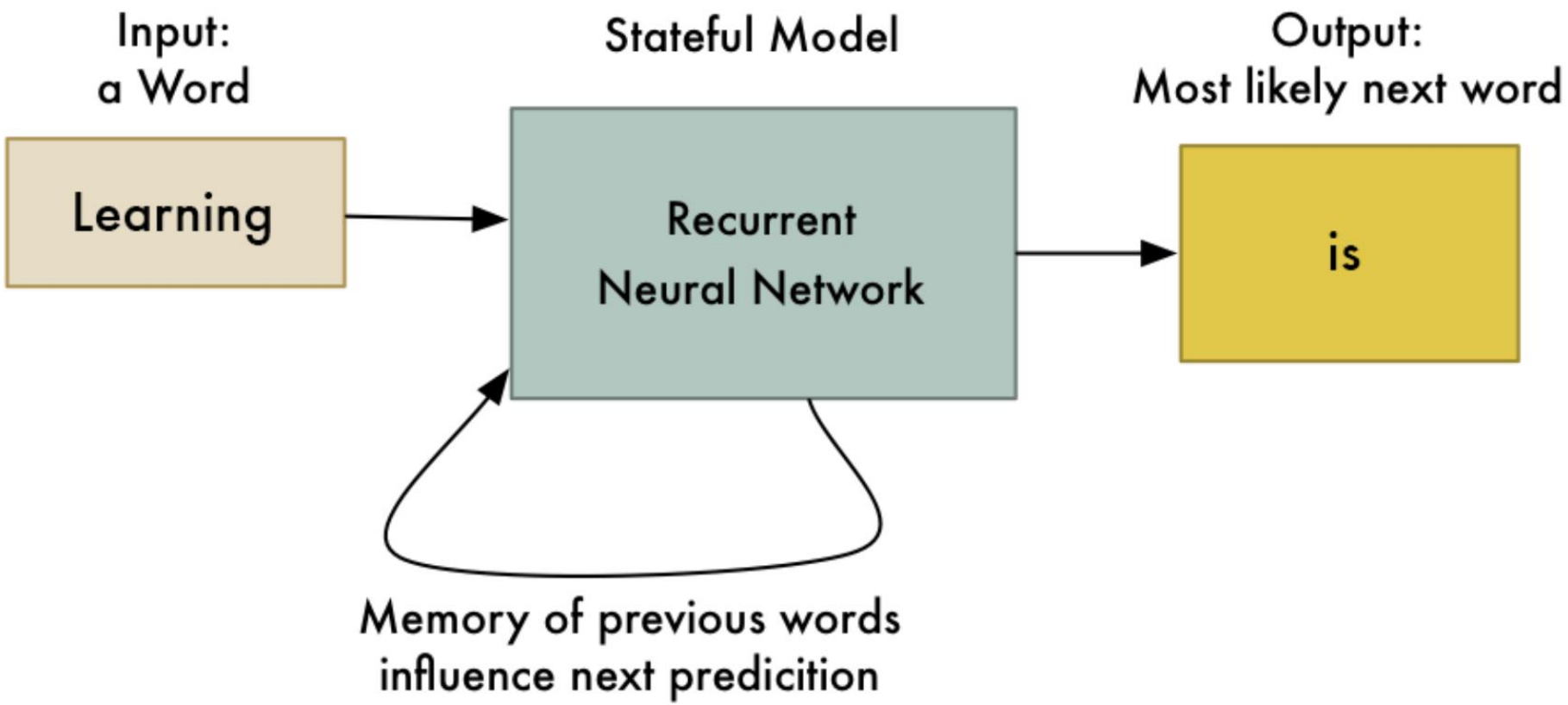
Long term dependency

- “Korea is where I was born and raised. I moved to the US when I was twenty. I am fluent in ____”
- How big of a window do you need?

Bag of words?

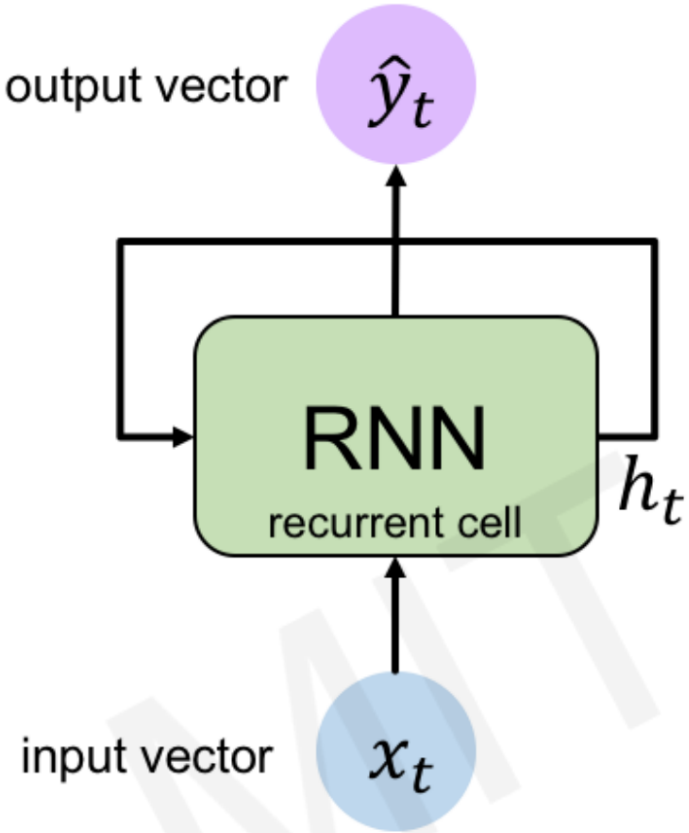

"This morning I took my cat for a"
"bag of words"
[0 1 0 0 1 0 0 ... 0 0 1 1 0 0 0 1]
prediction

RNN (Intuition)



https://wikidocs.net/35011

RNN in a nutshell



Apply a **recurrence relation** at every time step to process a sequence:

$$h_t = f_W(h_{t-1}, x_t)$$

cell state

function parameterized by W

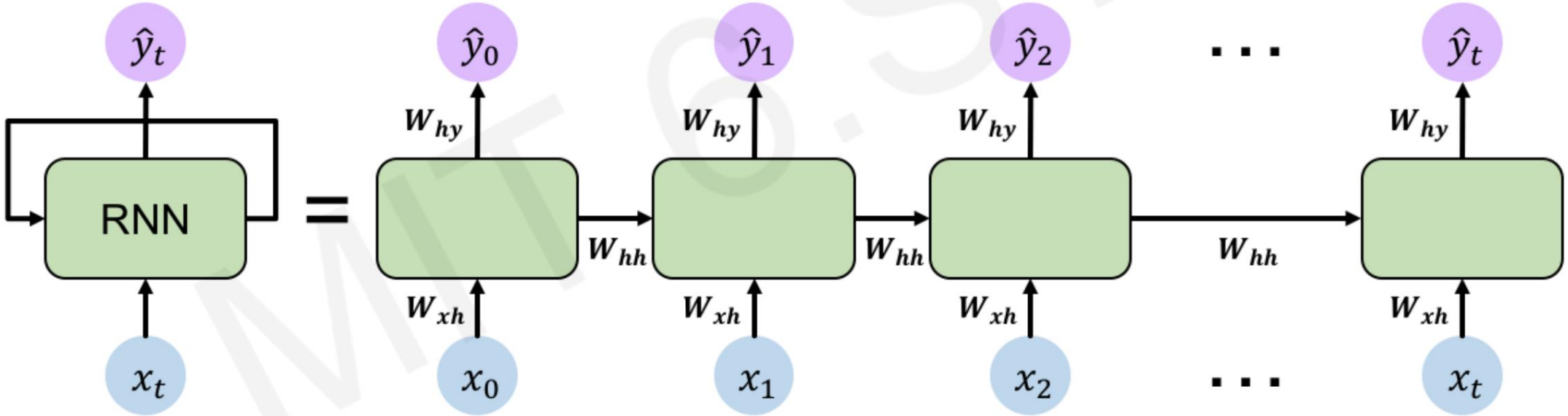
old state

input vector at time step *t*

Note: the same function and set of parameters are used at every time step

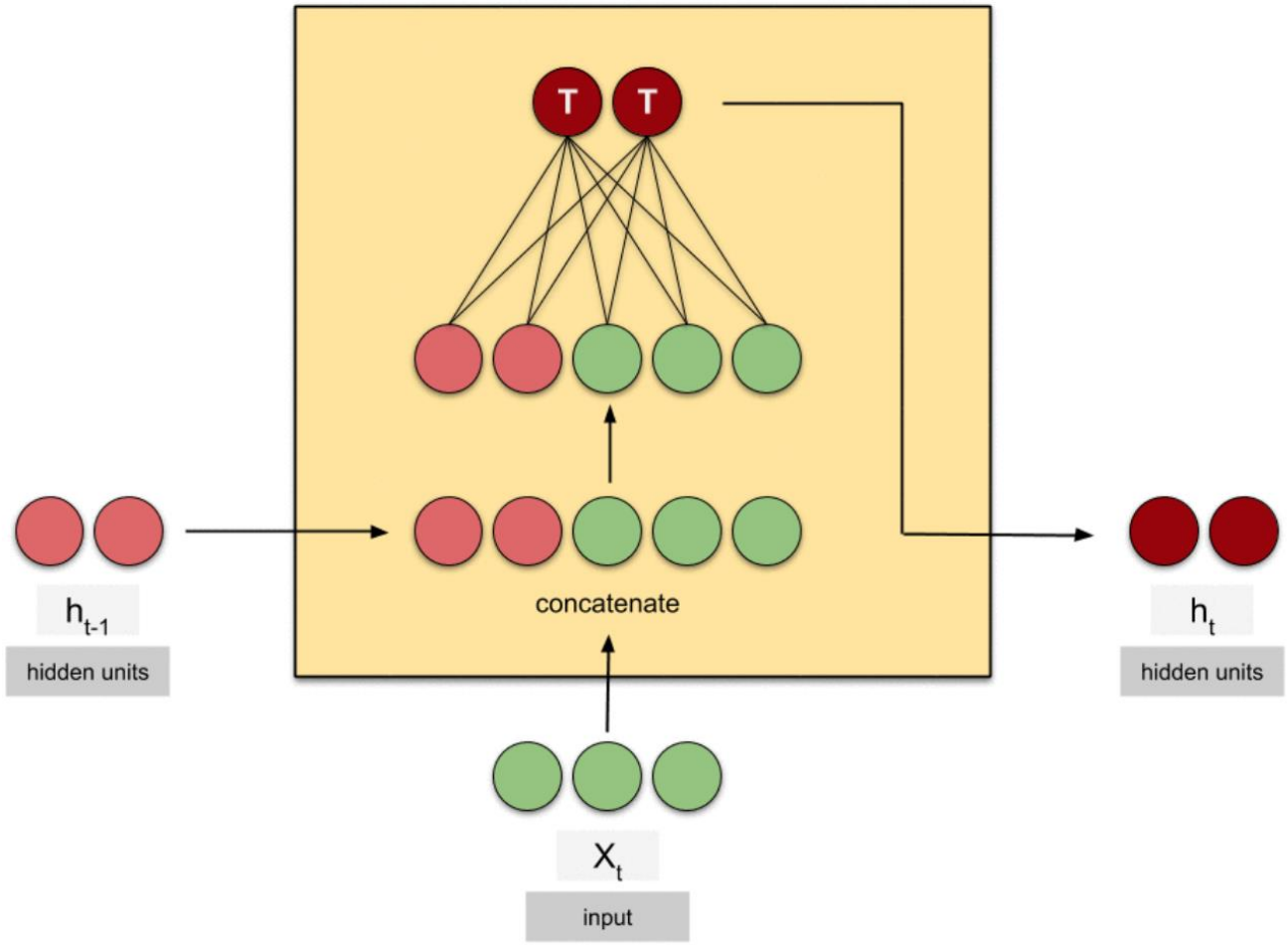
http://introtodeeplearning.com

RNN – Across Time



http://introtodeeplearning.com

RNN

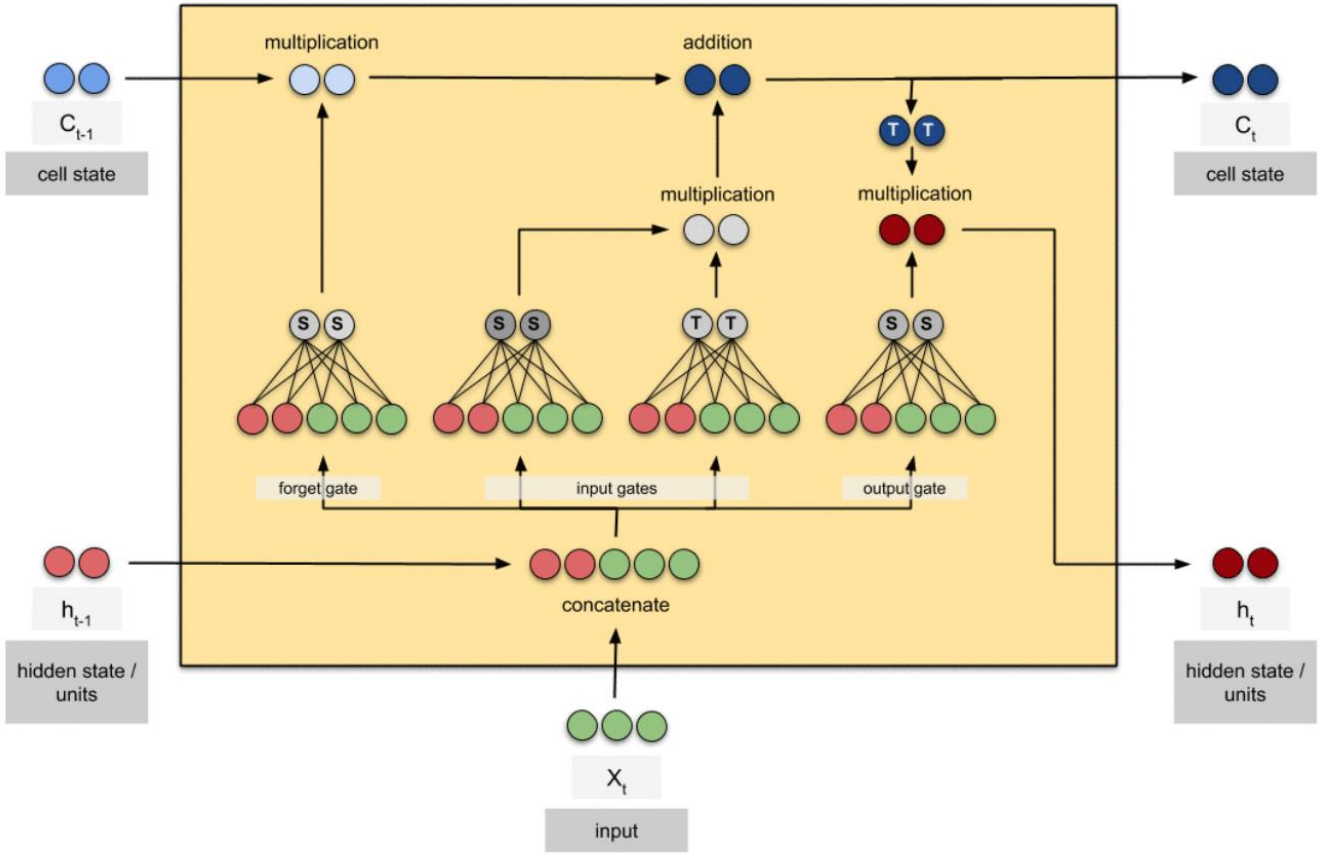


https://towardsdatascience.com/animated-rnn-lstm-and-gru-ef124d06cf45

Problems with RNN

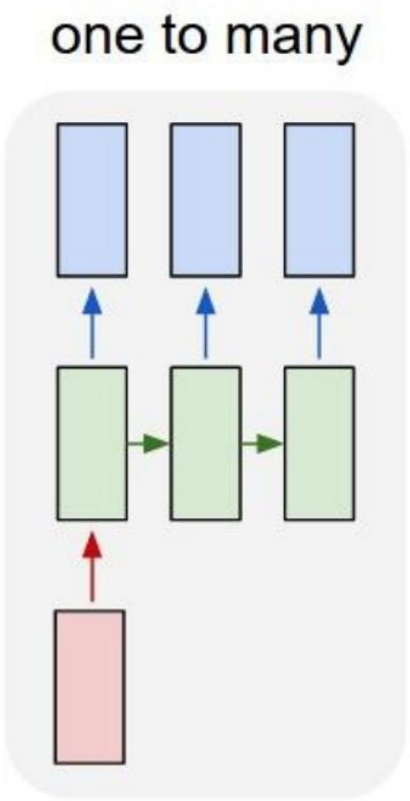
- To model long-term dependencies, you have to multiply many small numbers
  - ➔Vanishing gradients
- Use “gated cells”
  - ➔ LSTM (Long Short-Term Memory)

LSTM (Long Short-Term Memory

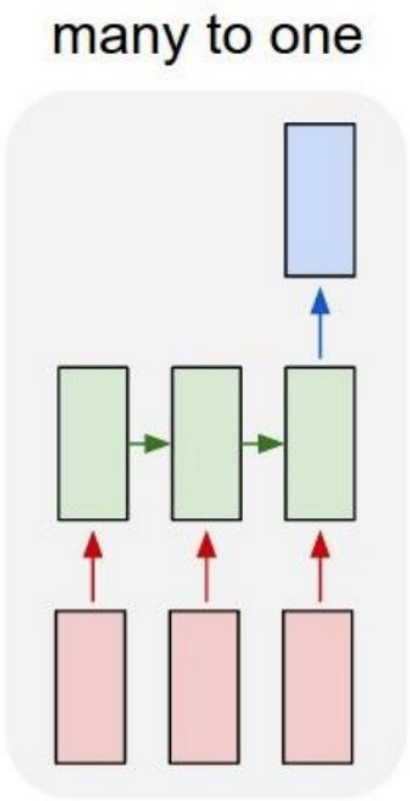


https://towardsdatascience.com/animated-rnn-lstm-and-gru-ef124d06cf45

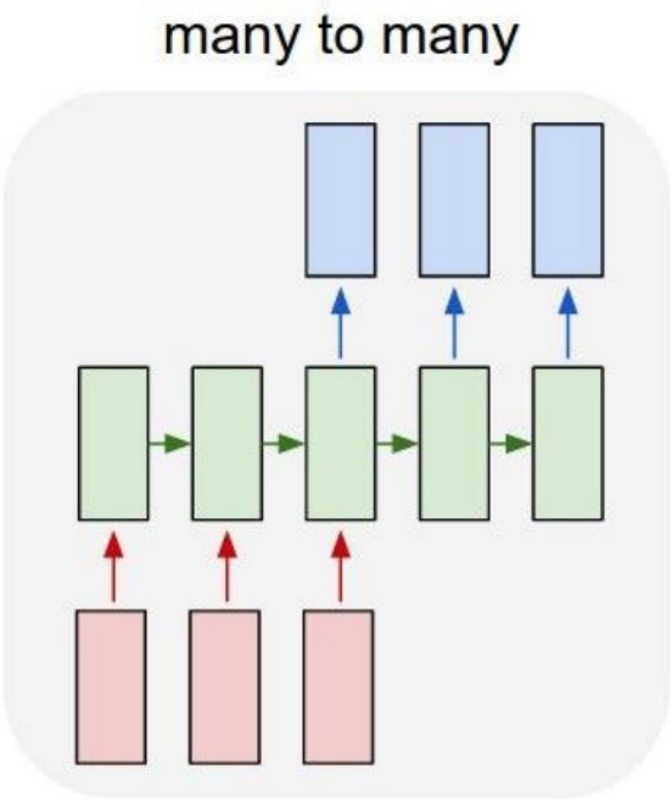
RNN Applications



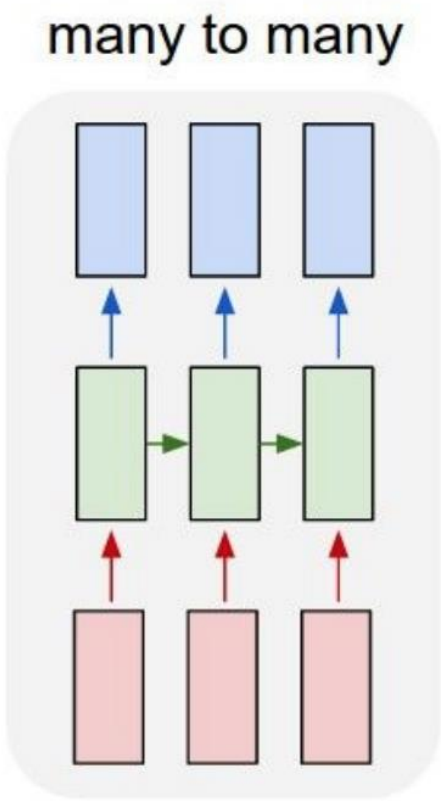
Captioning



Sentiment



Translation



Music Generation

Word Embeddings

- To turn words into vectors

king – man + woman ~= queen

https://jalammar.github.io/illustrated-word2vec/

Lab time

- To clone: from your terminal
  - >git clone https://github.com/changsin/DeepLearning-101.git
- Or use google colab to point to the git hub repository
- Git is an open source version control system
  - Github is a host service using git.