BONUS LECTURES

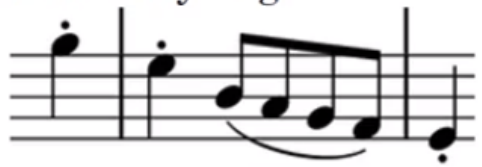
Make the most out of it!!!

RECURRENT NEURAL NETWORK

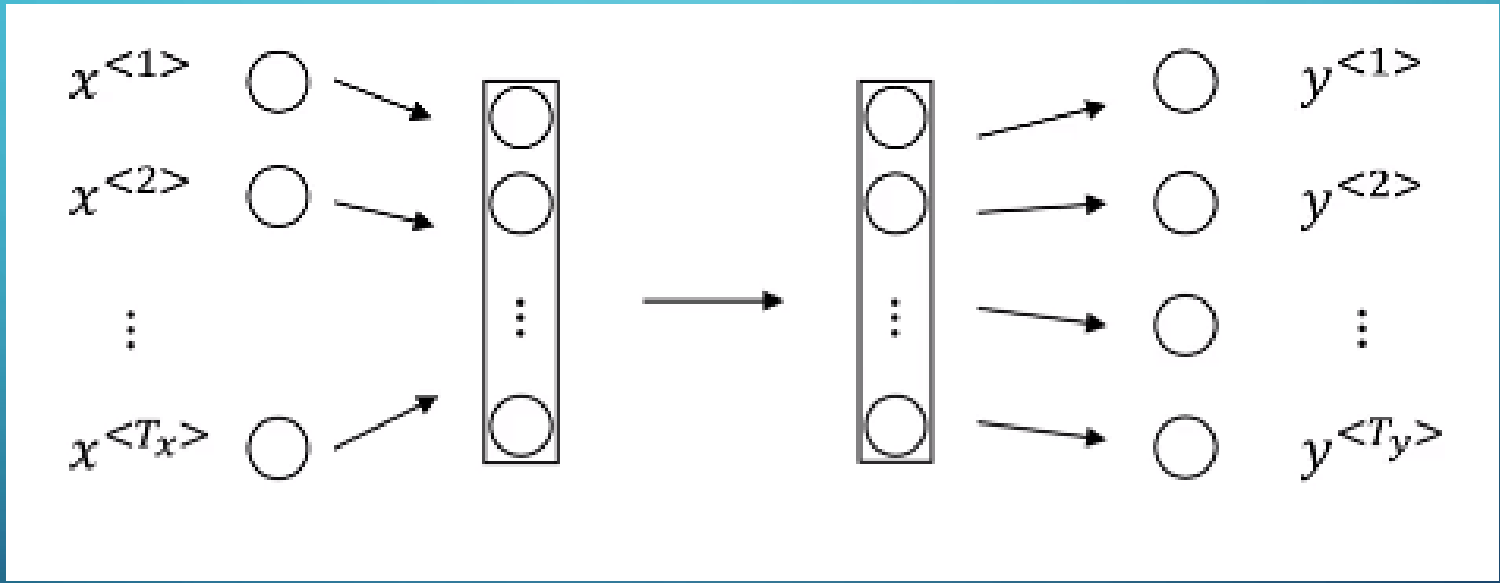
RNN

- It is used to build sequence models
- Applications in time series data
- Eg: Speech recognition, Natural Language Processing, data from Stock Market

EXAMPLE OF SEQUENCE MODELS

Speech recognition		→	"The quick brown fox jumped over the lazy dog."
Music generation	∅	→	
Sentiment classification	"There is nothing to like in this movie."	→	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	→	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	→	Do you want to sing with me?
Video activity recognition		→	Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	→	Yesterday, Harry Potter met Hermione Granger.

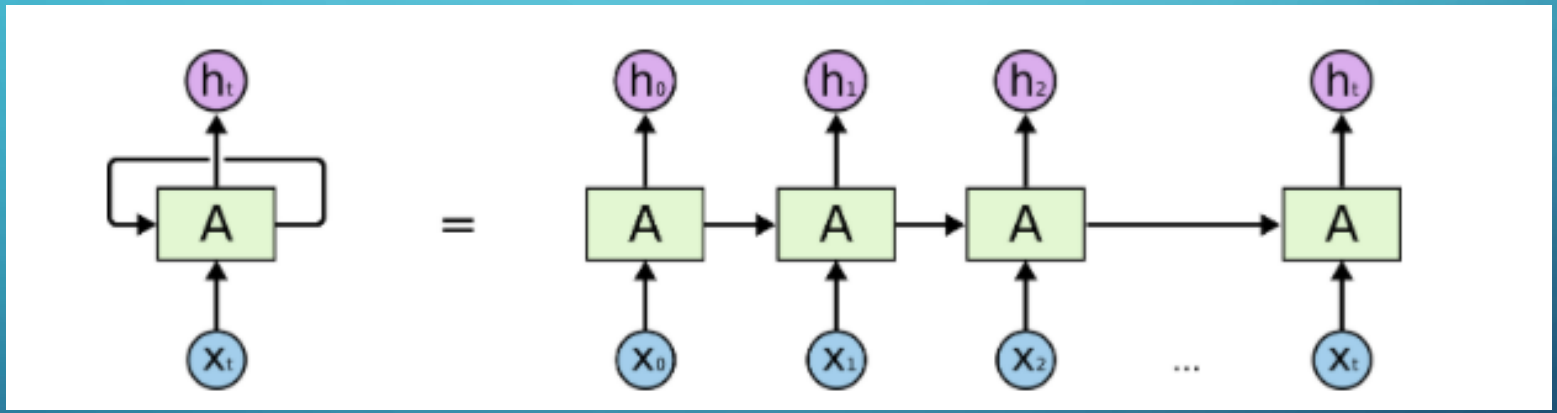
WHY NOT STANDARD NEURAL NETWORK??



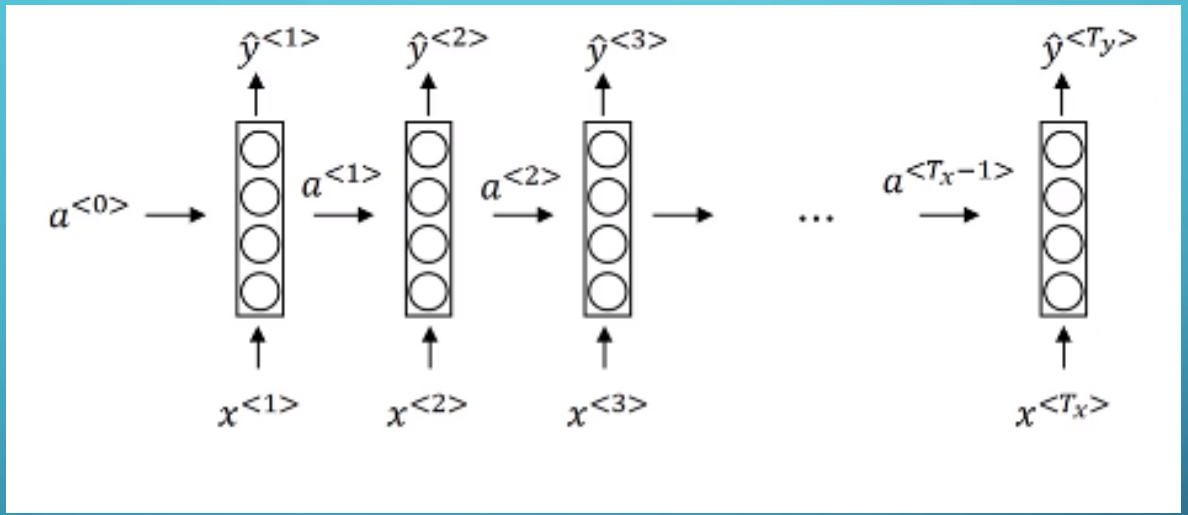
Problems:

- Inputs, outputs can be different lengths in different examples.
- Doesn't share features learned across different positions of text.

RNNS

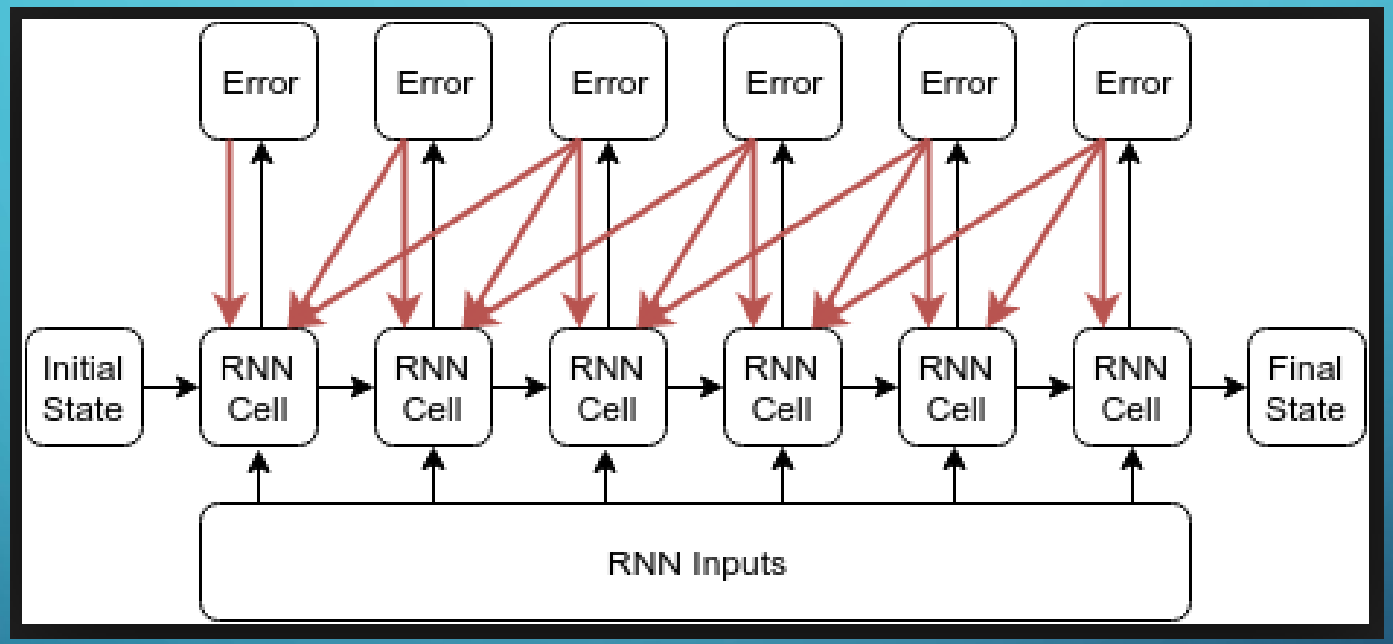


FORWARD PROPAGATION

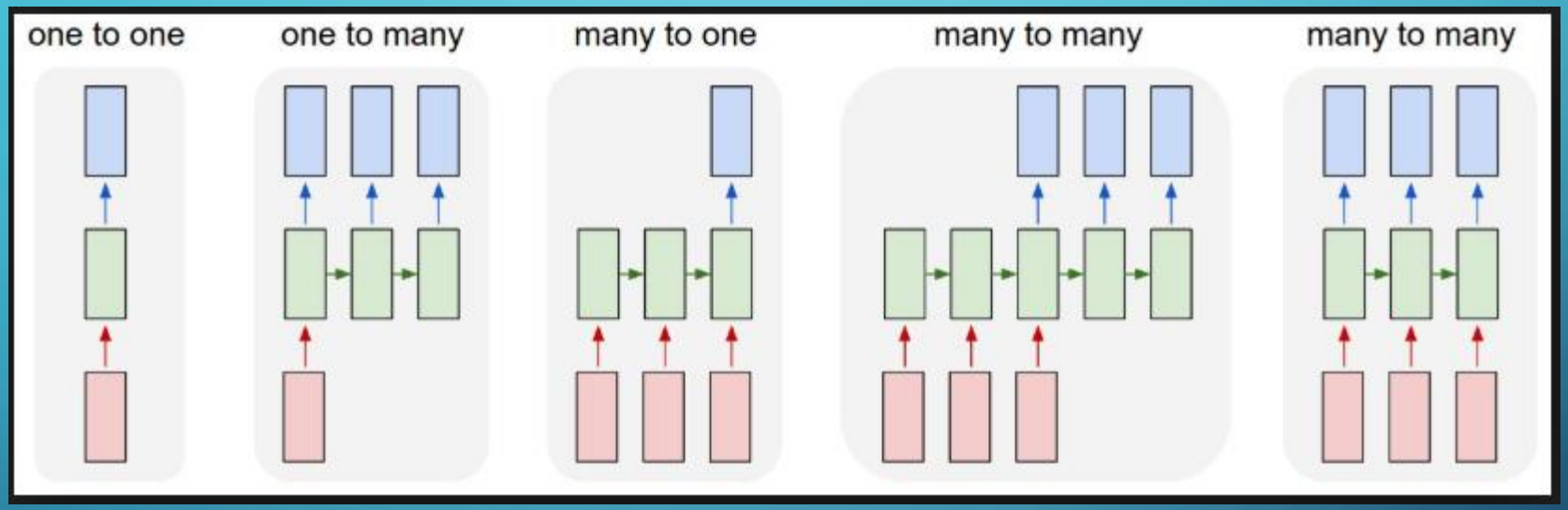


BACKPROPAGATION

- Also known as propagation through time

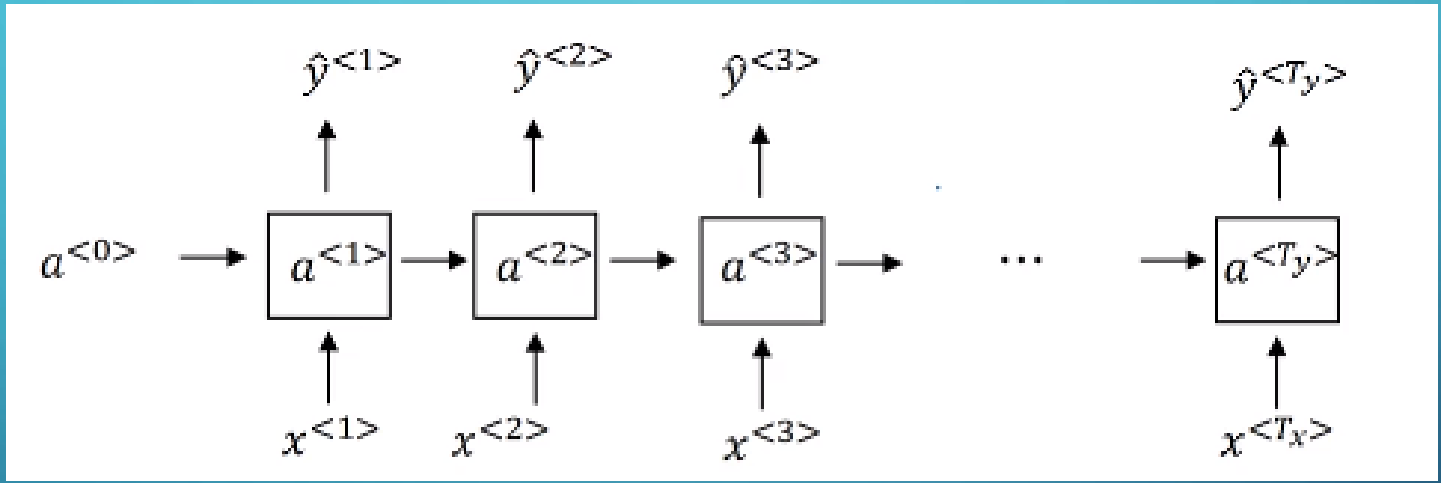

Error
Error
Error
Error
Error
Error
Initial State
RNN Cell
RNN Cell
RNN Cell
RNN Cell
RNN Cell
RNN Cell
Final State
RNN Inputs

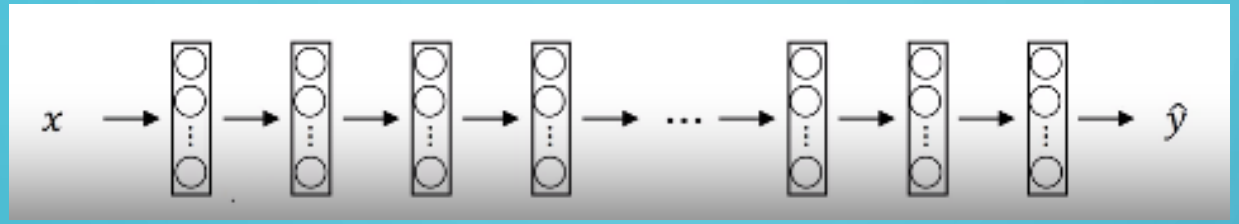
TYPES OF RNN



VANISHING GRADIENTS

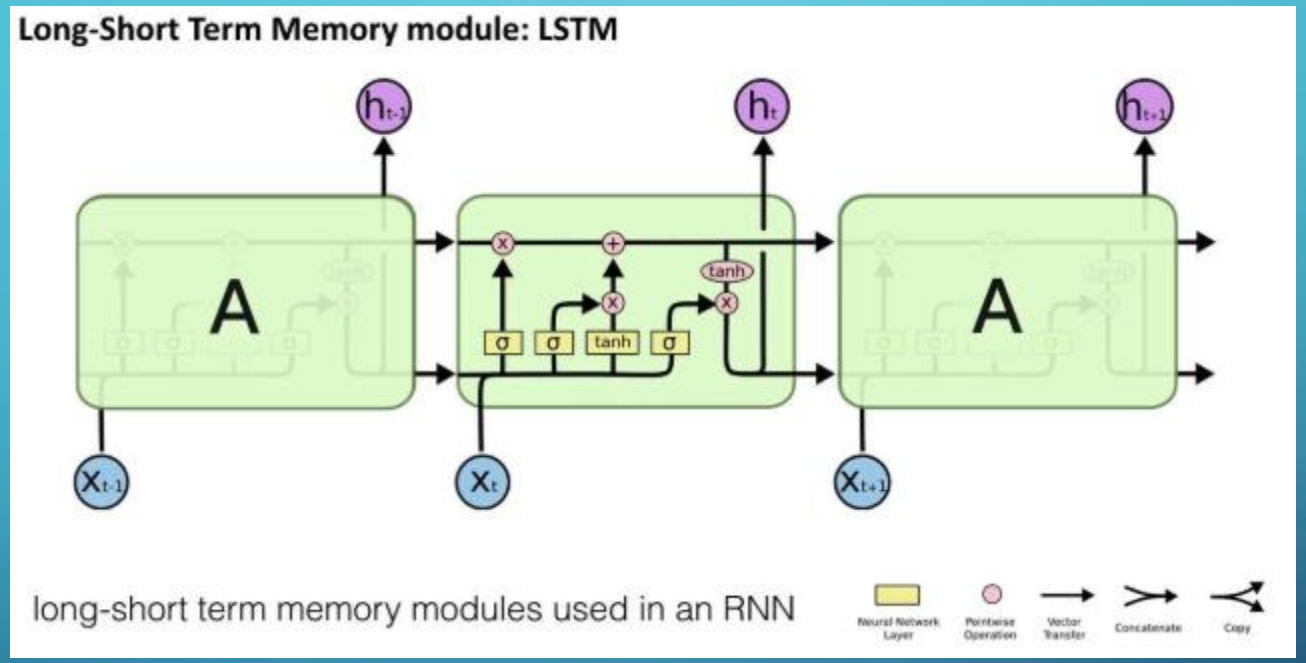
USUALLY RNNS ARE HUGE



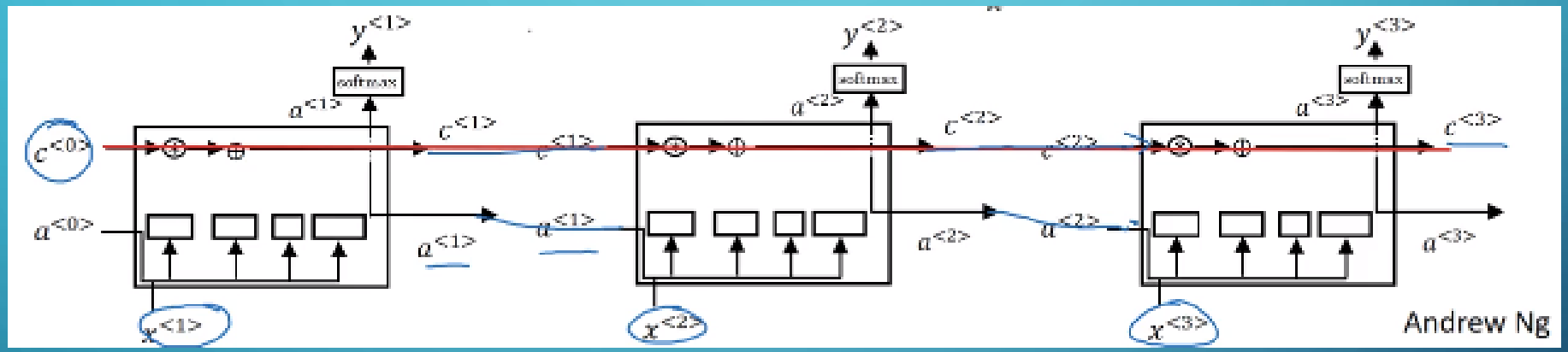


- Suppose this RNN has 100 recurrent units
- It will be very difficult to propagate the loss from $\hat{y}$ to the first recurrent unit
- Activation functions are of no help either!! Lol!!

LSTM TO THE RESCUE!!

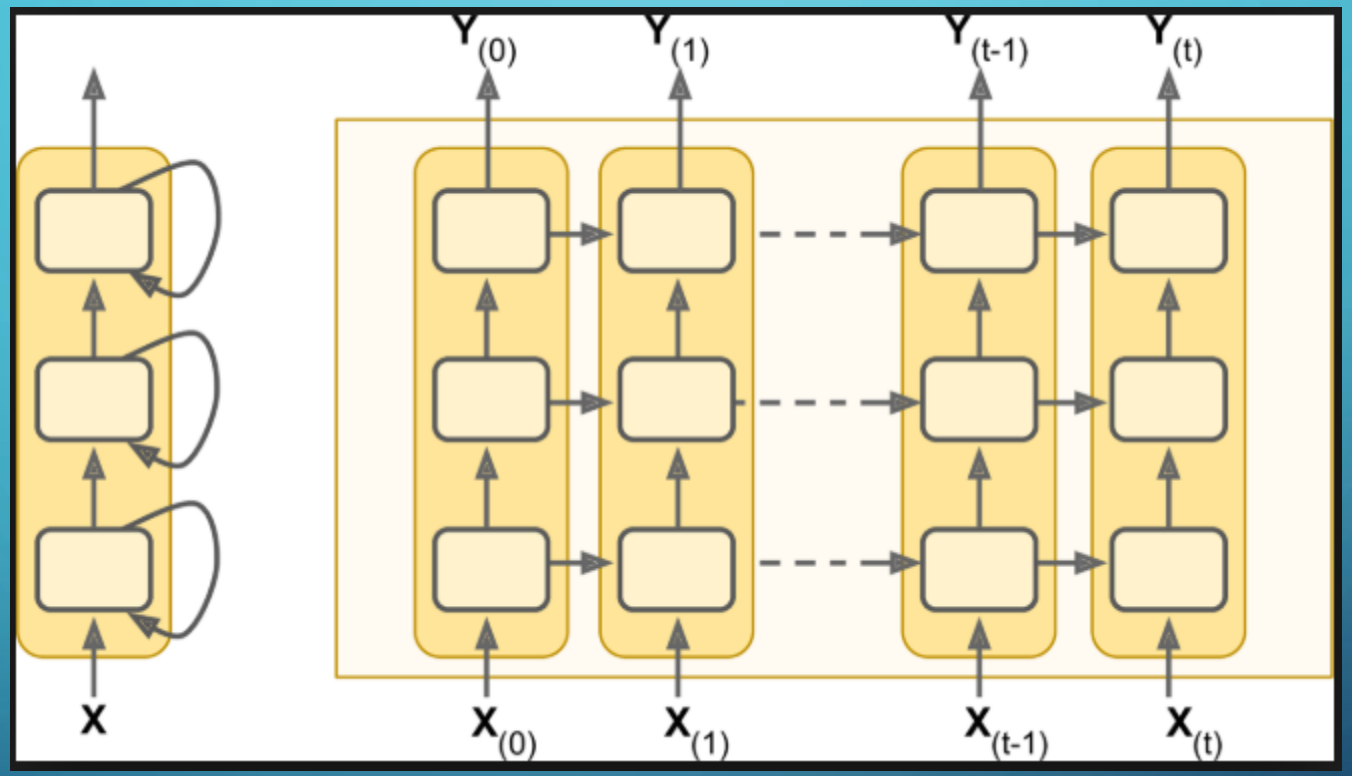


Notice the short circuit line at the top


$y^{<1>}$
$y^{<2>}$
$y^{<3>}$
softmax
softmax
softmax
$a^{<1>}$
$a^{<2>}$
$a^{<3>}$
$c^{<0>}$
$c^{<1>}$
$c^{<2>}$
$c^{<3>}$
$a^{<0>}$
$a^{<1>}$
$a^{<2>}$
$x^{<1>}$
$x^{<2>}$
$x^{<3>}$
Andrew Ng

DEEP RNNS

- Earlier we saw single hidden layer
- Here, we can add more hidden layers like we did in Artificial Neural Network


Y(0)
Y(1)
Y(t-1)
Y(t)
X
X(0)
X(1)
X(t-1)
X(t)

WORD REPRESENTATION

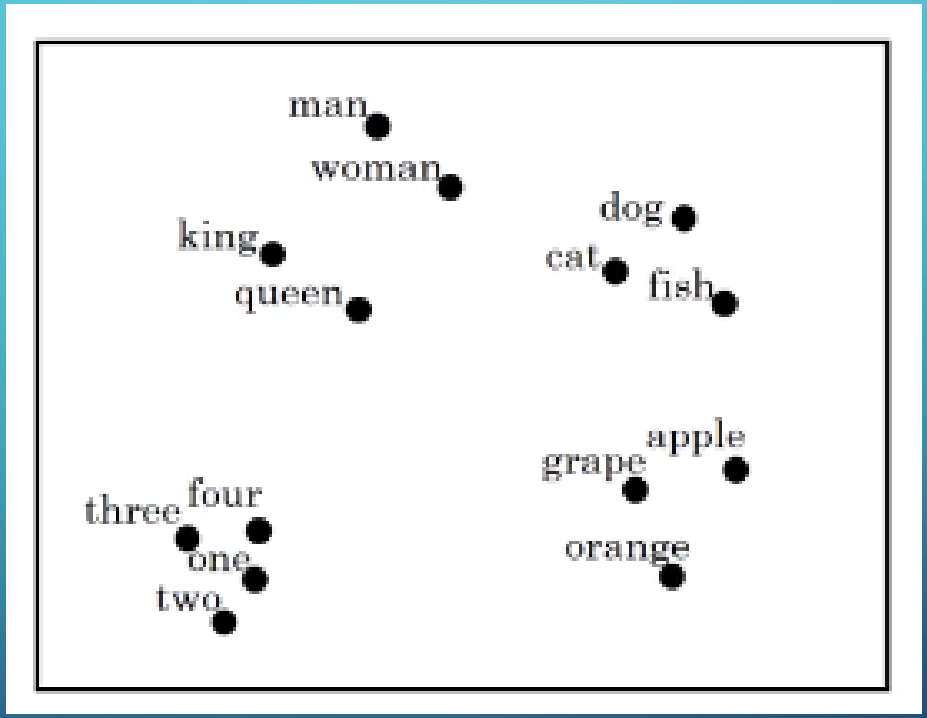
BASIC METHOD

- Prepare dictionary
- Use word count
- Also known as Bag of Words

WORD EMBEDDING

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.7	0.69	0.03	-0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97

VISUALIZING WORD EMBEDDINGS



PRE-TRAINED WORD EMBEDDINGS

1. Learn word embeddings from large text corpus. (1-100B words)

   (Or download pre-trained embedding online.)

2. Transfer embedding to new task with smaller training set. (say, 100k words)

ANALOGIES IN THE WORD EMBEDDINGS

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	−1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

MAN:WOMAN :: KING:QUEEN

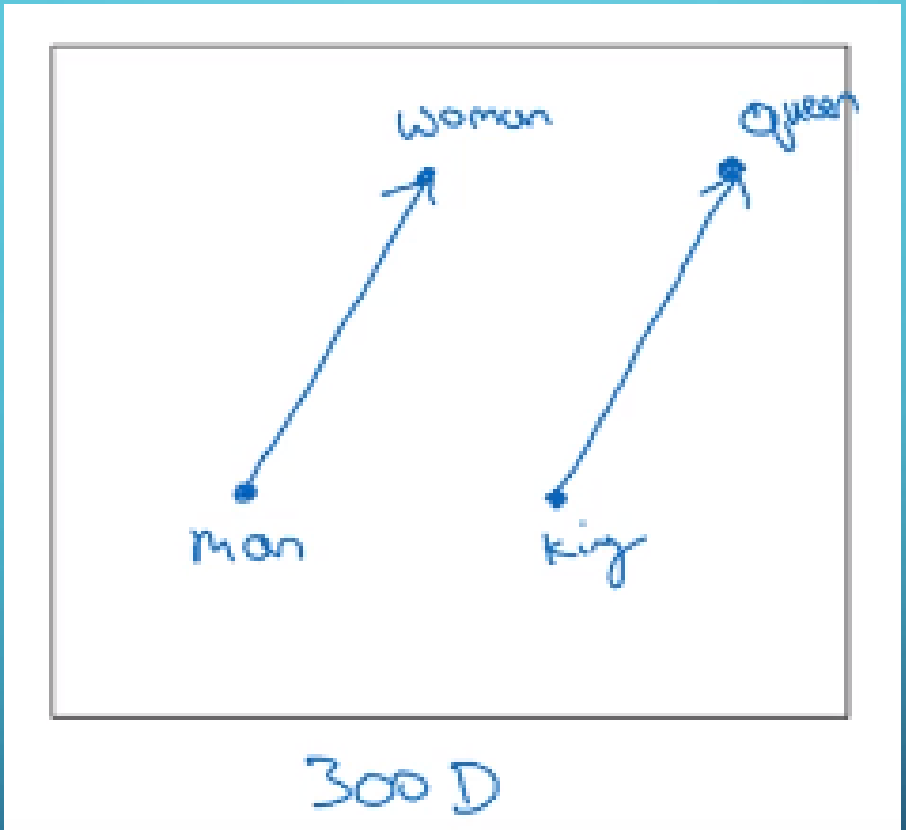
	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	−1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

e_{5391} / e_{man} e_{woman}

Man → Woman :: King → ?

$e_{man} - e_{woman} \approx \begin{bmatrix} -2 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$

$e_{king} - e_{queen} \approx \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$


Woman
Queen
Man
King
300 D

QUESTION:

What will be King – Man + Woman?

- Word Embedding gives variance to your dataset
- By training on a word, you are actually training on all the word vectors in its vicinity
- Mainly used when your dataset is small

CASE STUDY