

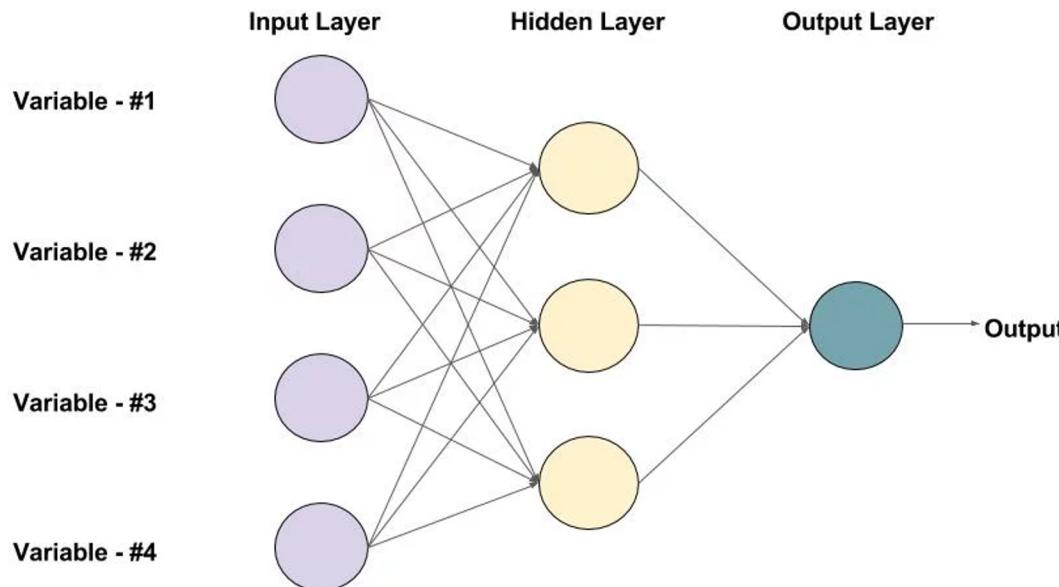
Recurrent Neural Networks and Long Short Term Memory

MIE1624 - Group 22

Aaron Hao Tan,
Alex Kwan, Jacob Bulir, Gurtej Bhasin, Elise Lagace,

Artificial Neural Nets

$$f(x) = \sigma(wx+b)$$

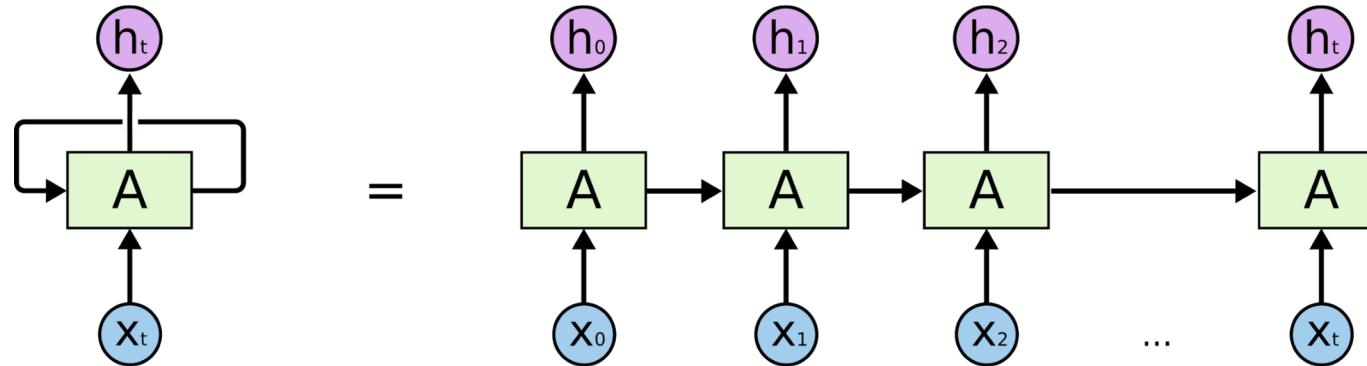


An example of a Feed-forward Neural Network with one hidden layer (with 3 neurons)

Recurrent Neural Network(RNN)

- Reuse previous hidden layers to assist in future predictions
- Useful for sequence predictions: text analysis, voice recognition
- Chatbot example:

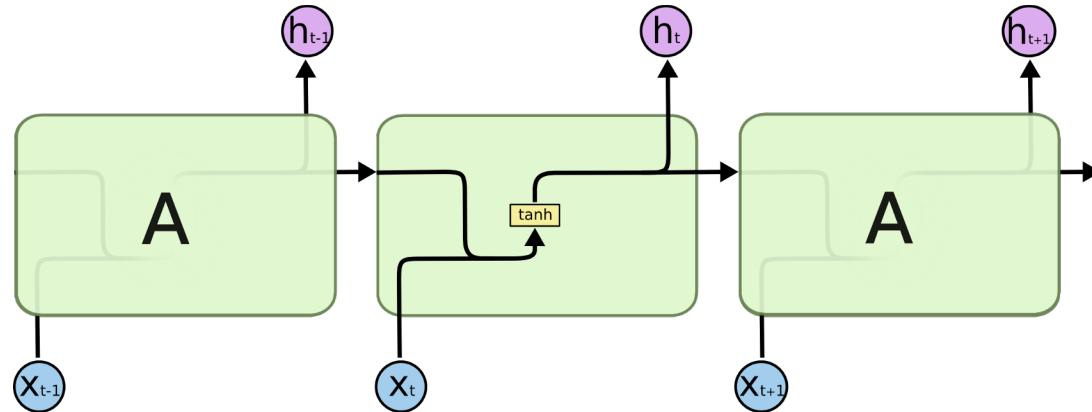
What is the weather going to be like tonight?



Recurrent Neural Network(RNN)

- Reuses the hidden layer through the tanh function
- Combines and compresses the previous hidden layer and new input

What is the weather going to be like tonight?

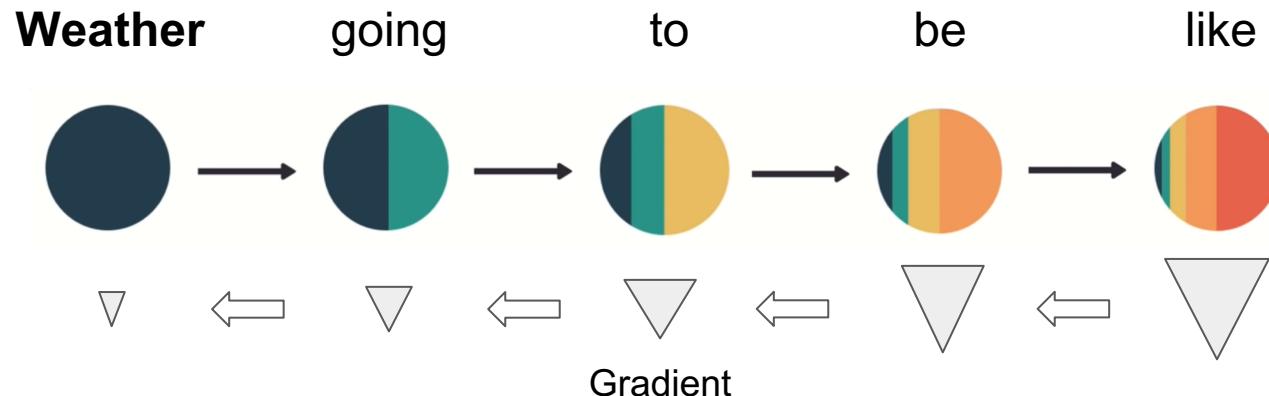


Problem: The Vanishing Gradient

- Use back-propagation through time to calculate gradients
- Previous examples can be thought of as their own layer

Input: “tonight” and previous hidden layer

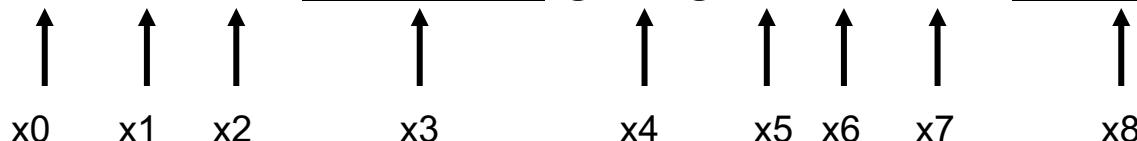
Output: Label



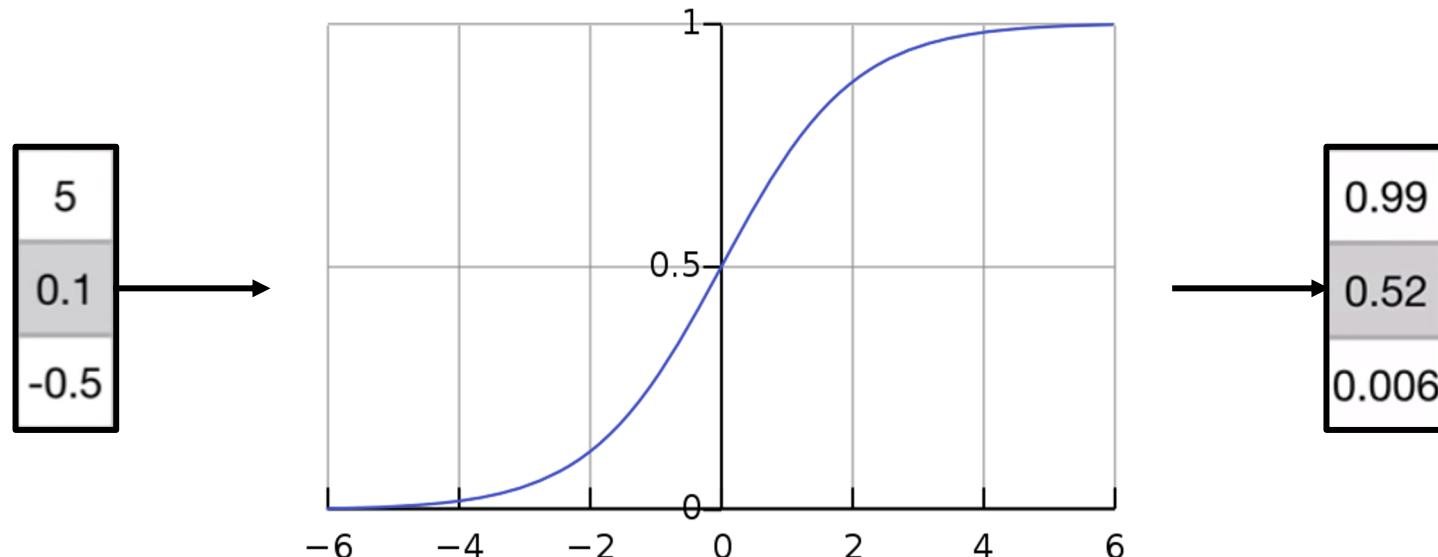
Long Short Term Memory (LSTM)

- A more evolved form of RNN
- Developed to solves the vanishing gradient problem by retaining and passing relevant information down the chain of connected NNs
- Has the ability to learn which data in a sequence is important to keep or throw away

What is the weather going to be like tonight?

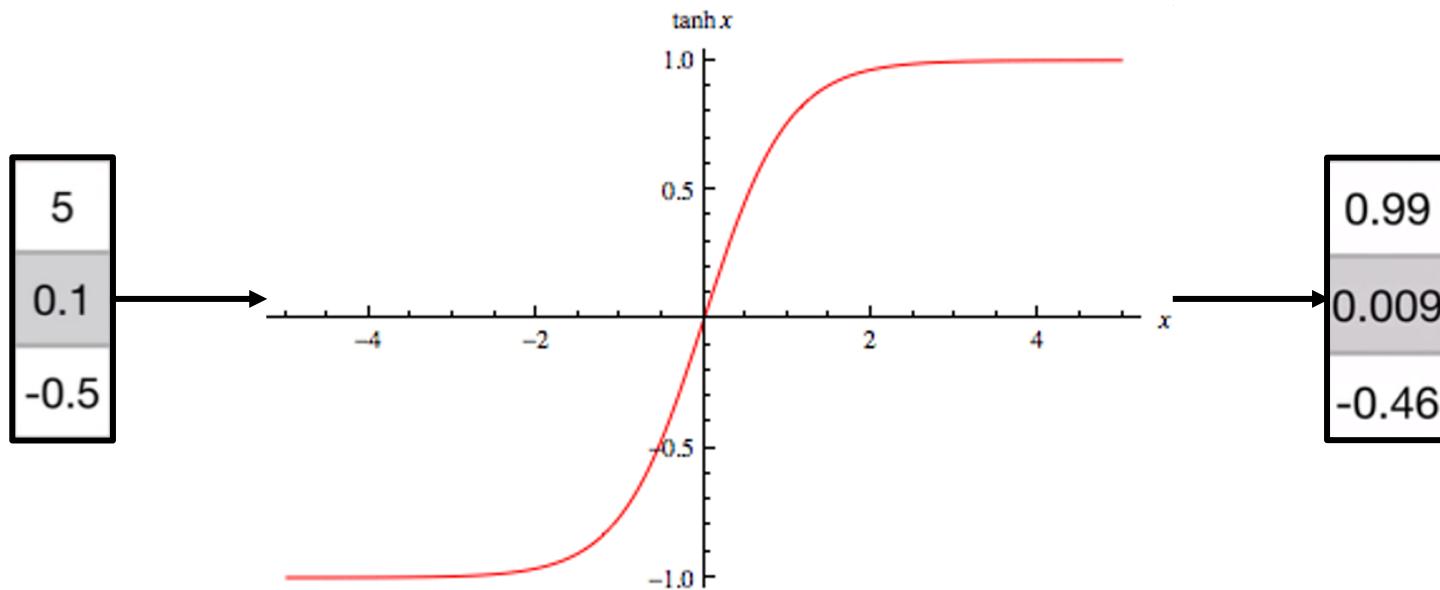


Sigmoid Function



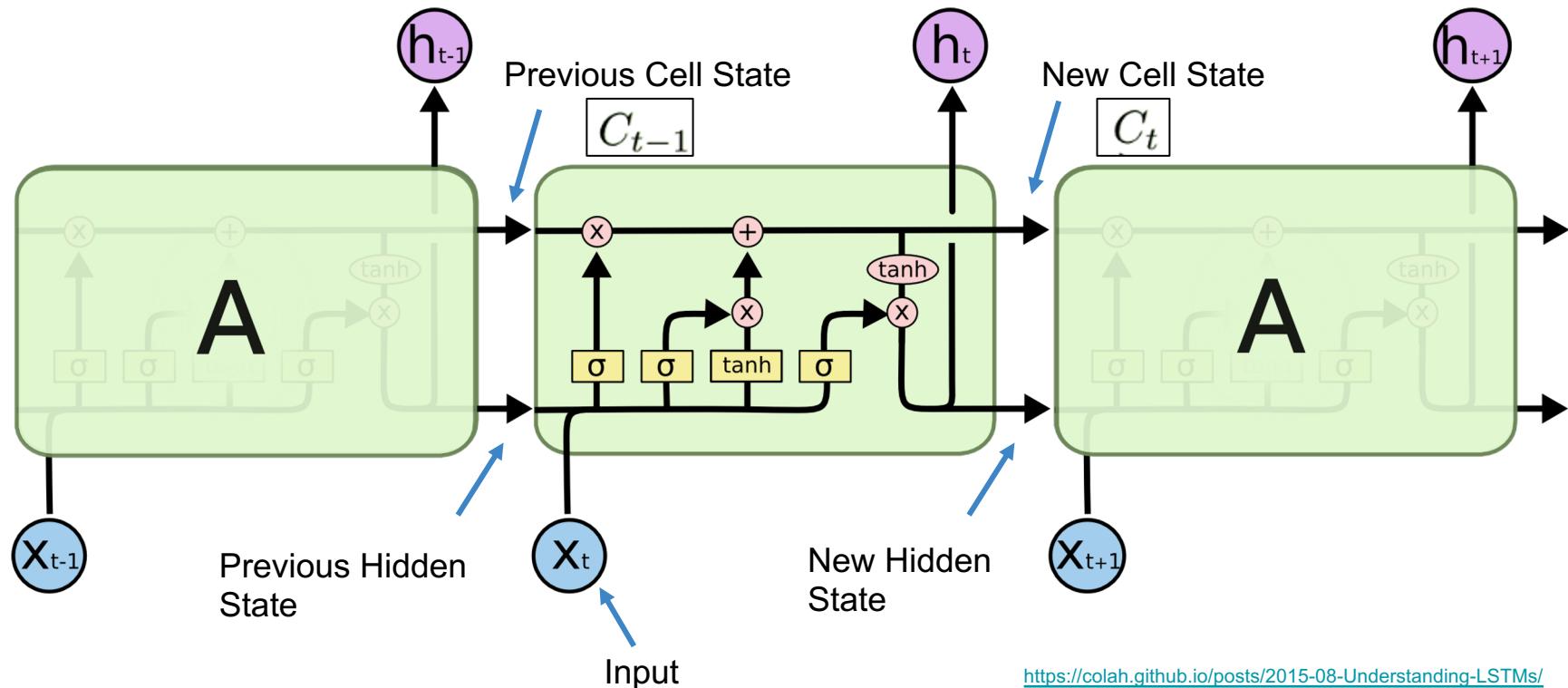
https://en.wikipedia.org/wiki/Sigmoid_function

Tanh Function



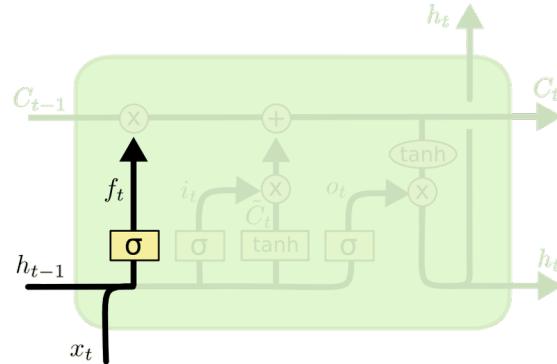
<http://mathworld.wolfram.com/HyperbolicTangent.htm>

Long Short Term Memory (LSTM)



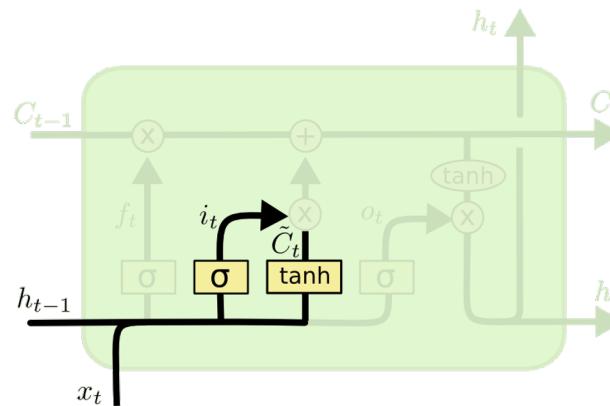
Long Short Term Memory (LSTM)

Forget Gate:
Decides what information is relevant from prior cells



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Input Gate:
Decides what information is relevant to add from current cell

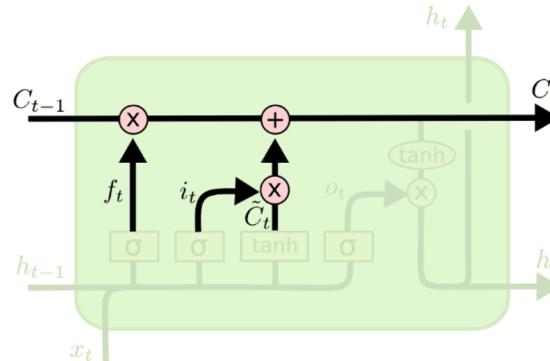


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

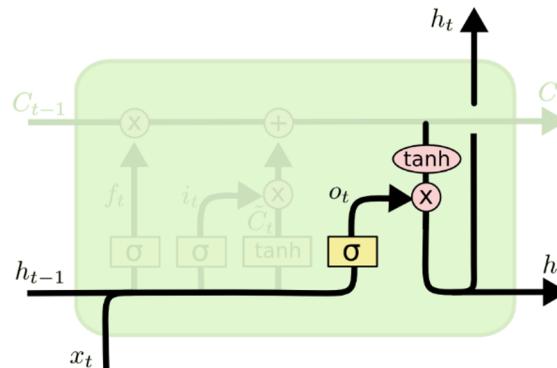
Long Short Term Memory (LSTM)

Cell State: Passage for information to pass through all the way through the chain



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Output Gate:
Determines what the next hidden state should be



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

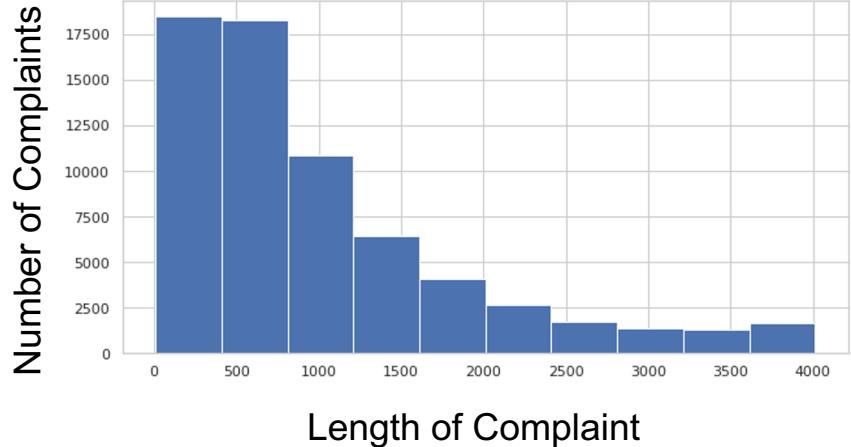
$$h_t = o_t * \tanh (C_t)$$

The Data

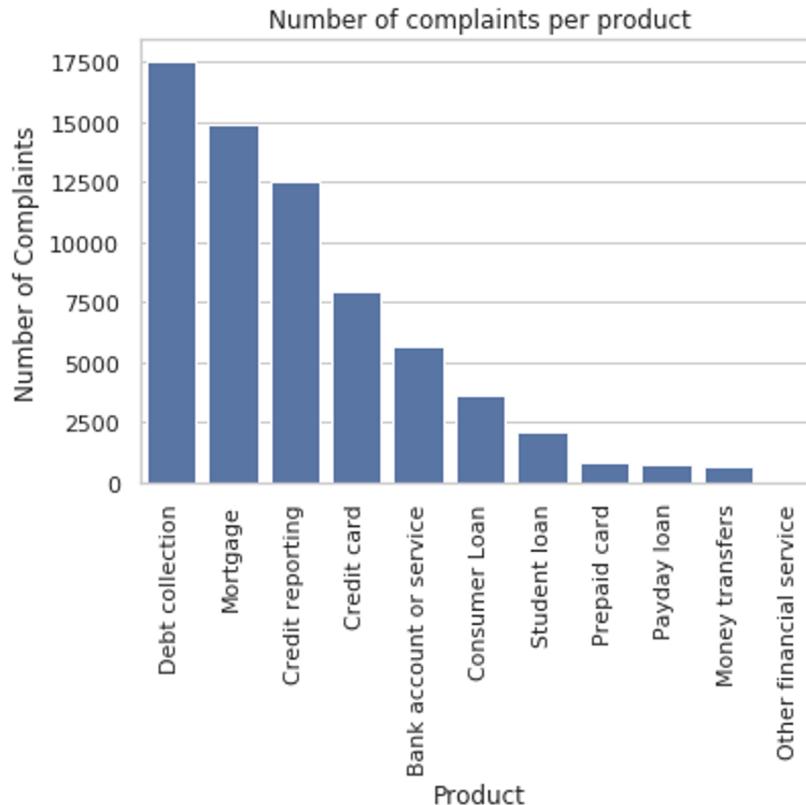
- Data: “US consumer complaints on financial products and company responses”
- Problem solved: Classify and predict product based on the complaint
- # samples = 67K

Label		Input Feature					
date_received	product	sub_product	issue	sub_issue	consumer_complaint_narrative	company_public_response	
08/12/2015	Credit reporting	NaN	Incorrect information on credit report	Account status	I am disputing incorrect information on my cre...	NaN	
06/26/2015	Bank account or service	Checking account	Problems caused by my funds being low	NaN	My account was overdrawn, however I made a ret...	Company chooses not to provide a public response	
12/16/2015	Debt collection	Credit card	Cont'd attempts collect debt not owed	Debt resulted from identity theft	This is a complaint against Encore Capital Gro...	NaN	
03/07/2016	Mortgage	Conventional fixed mortgage	Loan servicing, payments, escrow account	NaN	This is the final time going to let Bank of Am...	Company has responded to the consumer and the ...	
11/12/2015	Credit reporting	NaN	Incorrect information on credit report	Information is not mine	My FICO score has been dropping ever since the...	NaN	

Exploratory Data Analysis

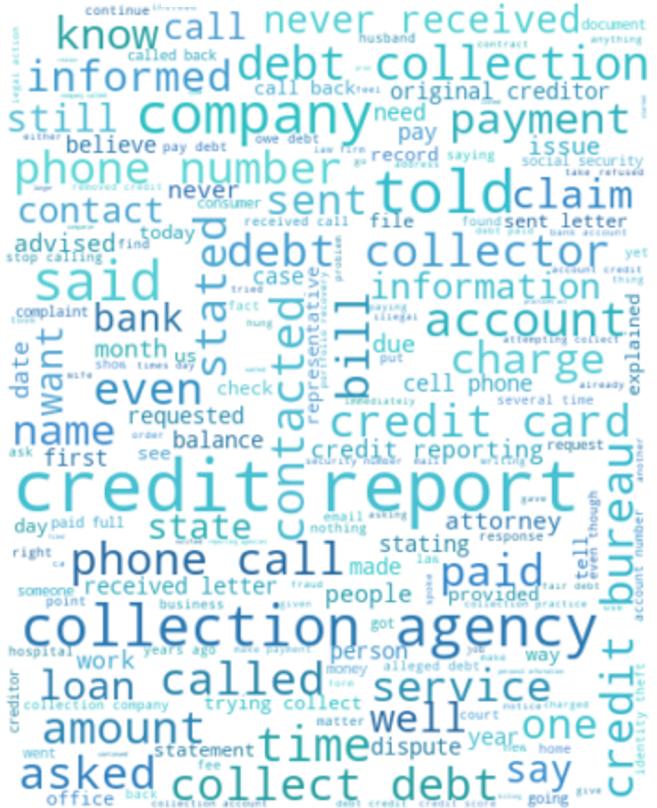


Customers have many issues with Debt Collection or Mortgages... no kidding!



Exploratory Data Analysis

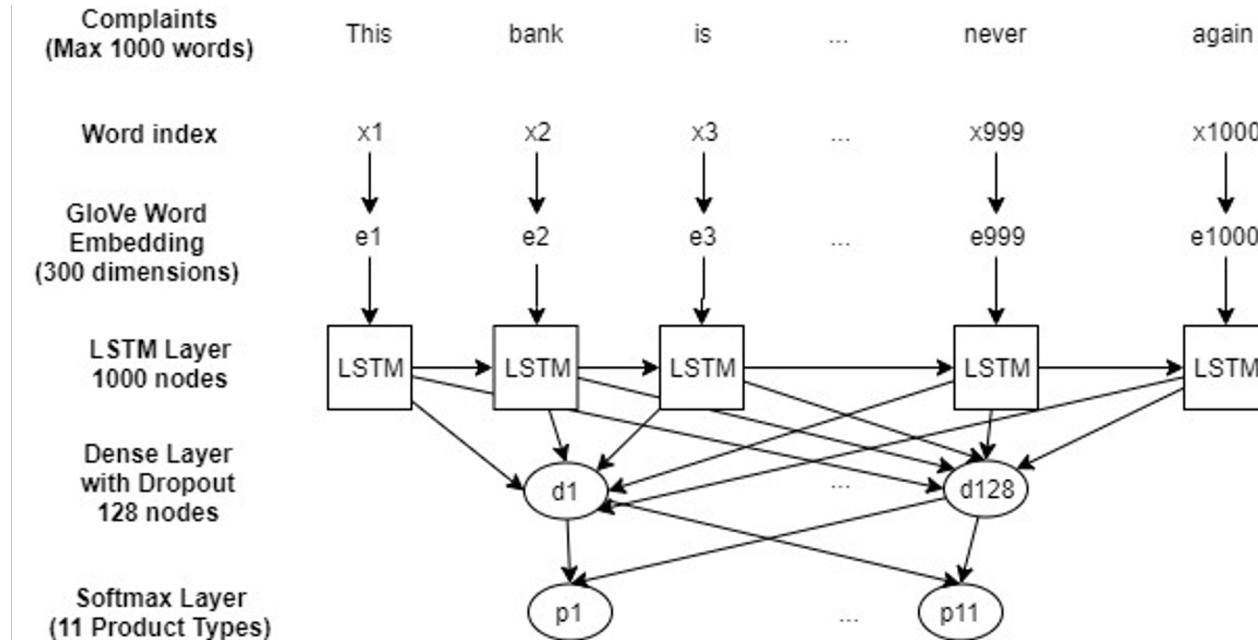
Debt Collection



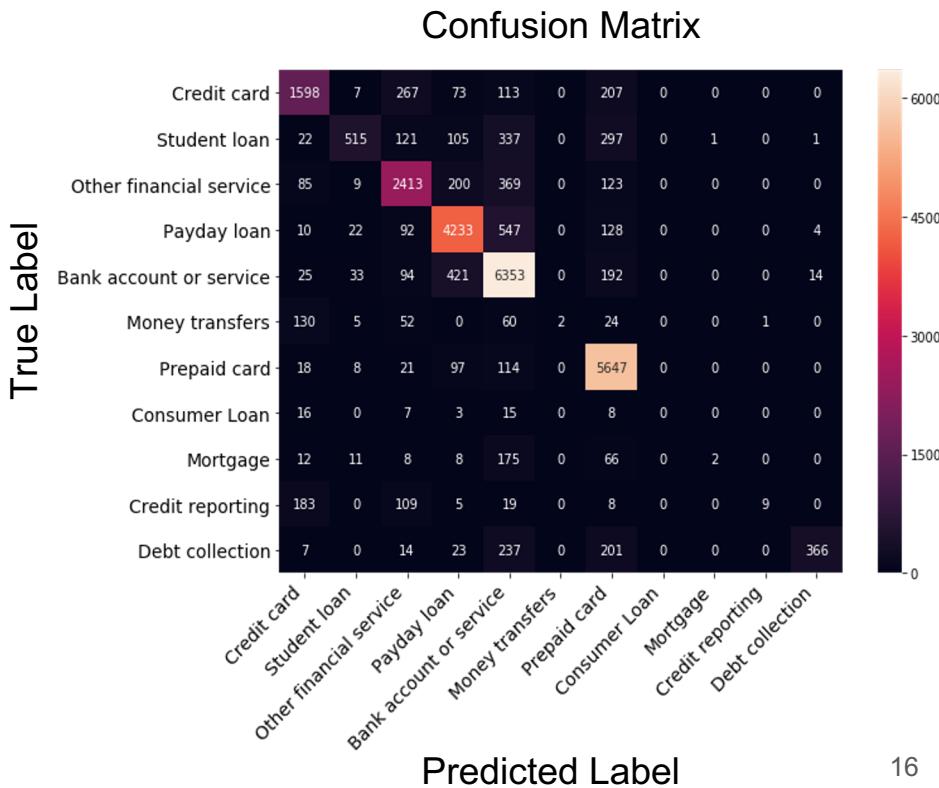
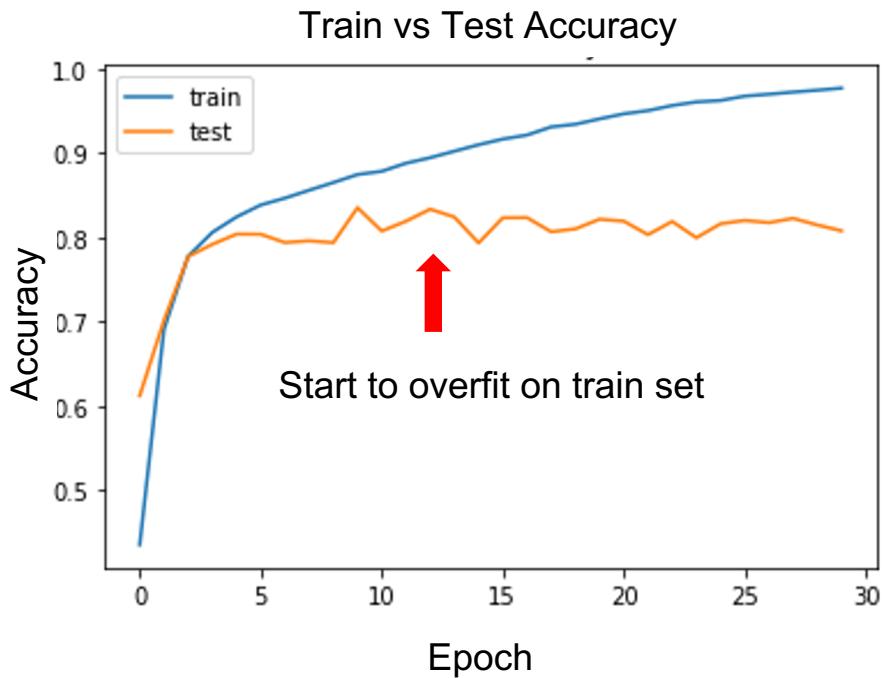
Mortgage



LSTM Architecture

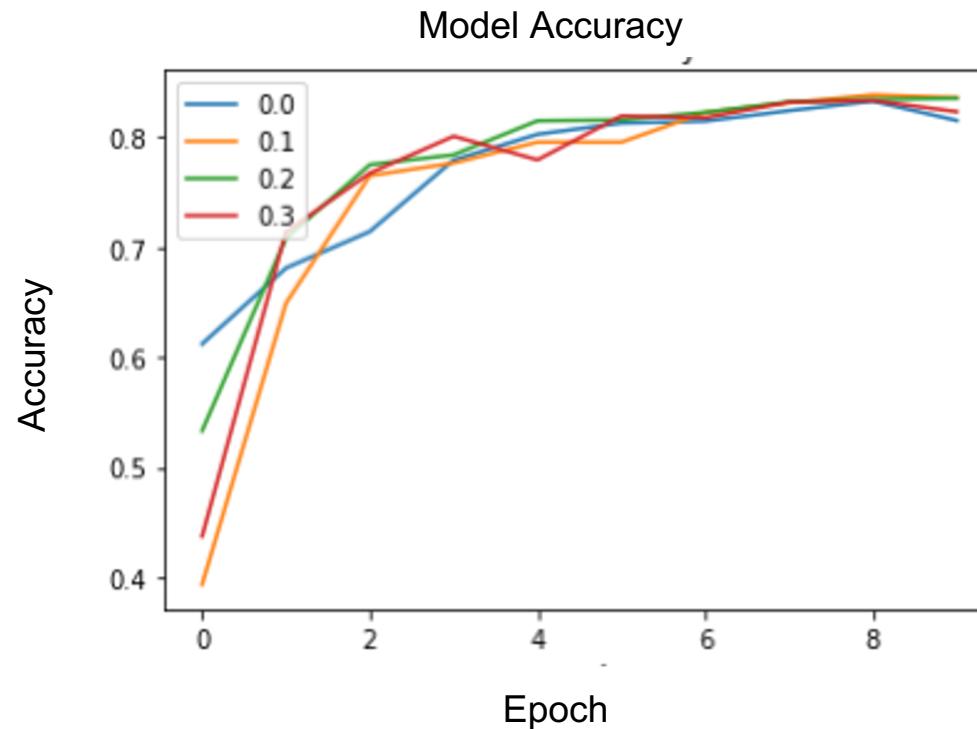


Results



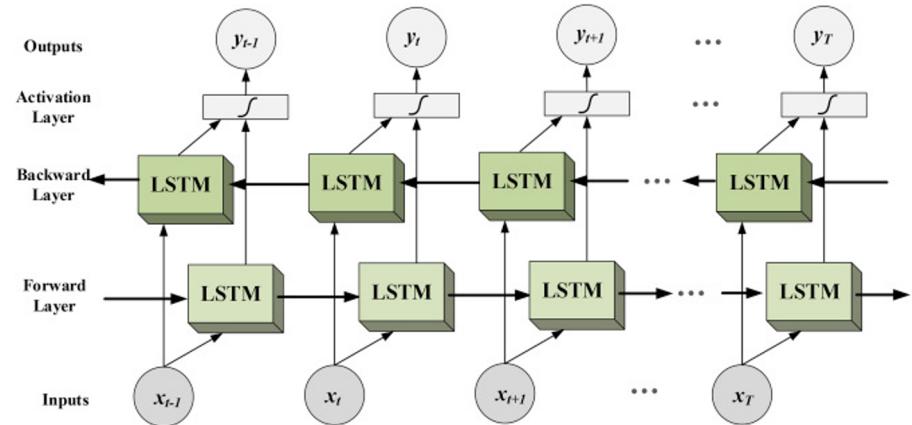
Hyperparameter Tuning

Dropout percentage



Future Work

Bi-directional LSTM
(read backwards and forwards)



Hyperparameter tuning:

- Larger batch size (256)
- Activation functions (leaky ReLU)
- Optimizer (Adam)
- # hidden units
- # layers
- Early stopping (# epochs)

Cross validation, but is computationally expensive

The End