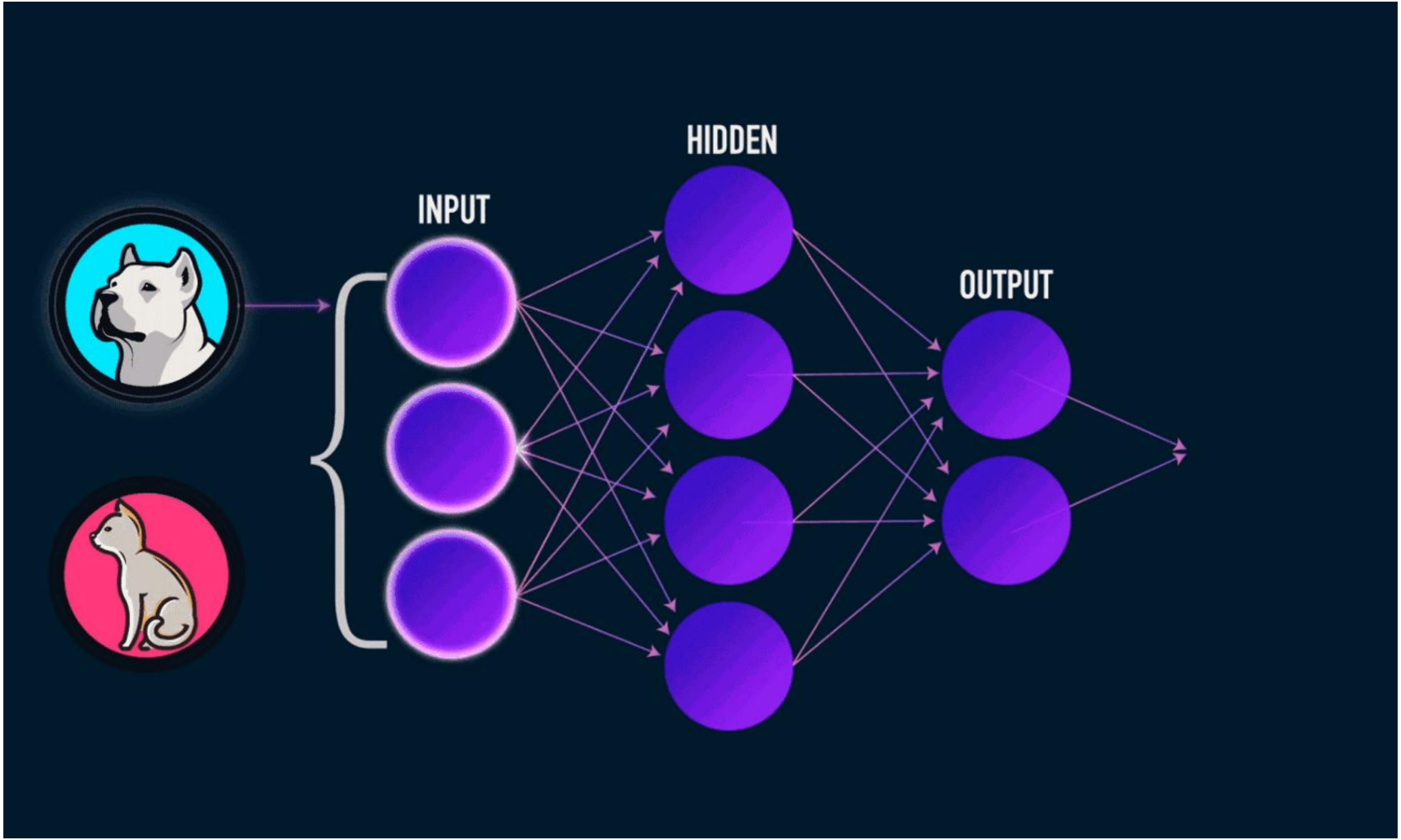
Neural Networks, etc.

Denis Peskov dpeskov@cs.umd.edu

LMU 5.26.20

Topics

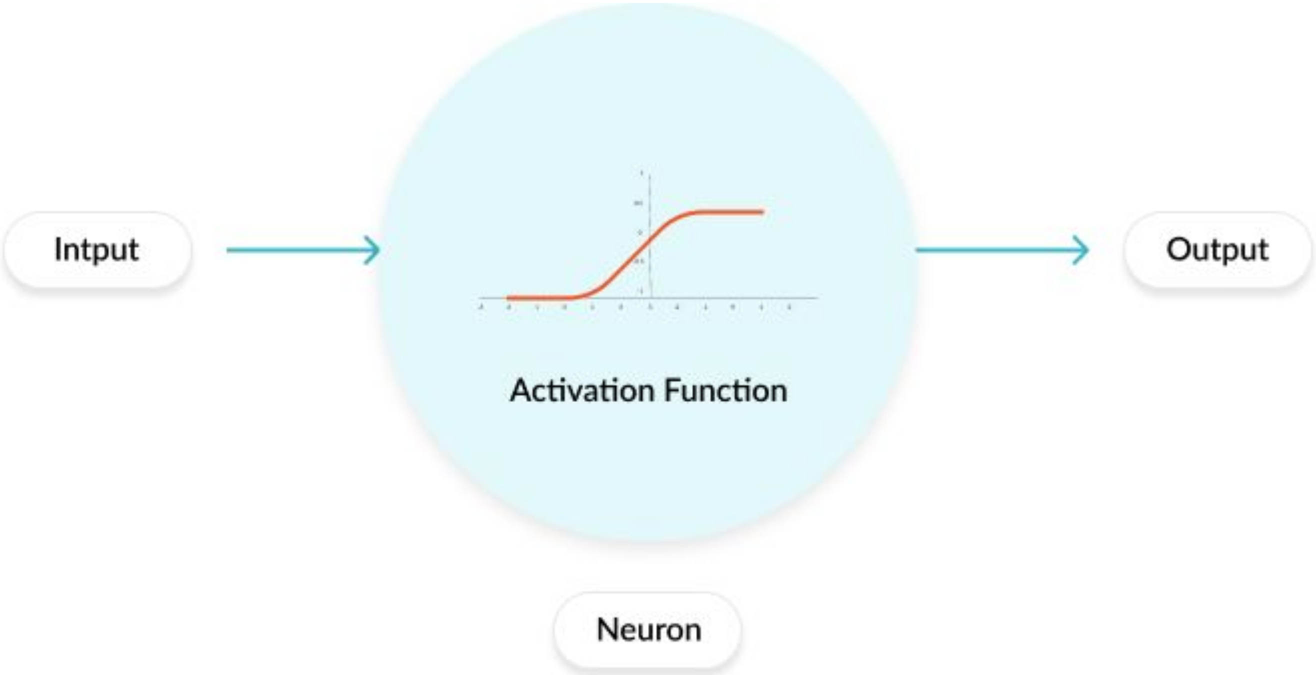
1. Activation functions
2. Training of neural networks
3. Recurrent networks
4. LSTM


HIDDEN
INPUT
OUTPUT

Training a Neural Network

- Neurons
- Loss Function
- Backpropogation
- Pragmatics

The Base Level: Neuron



Activation Function

Obvious:

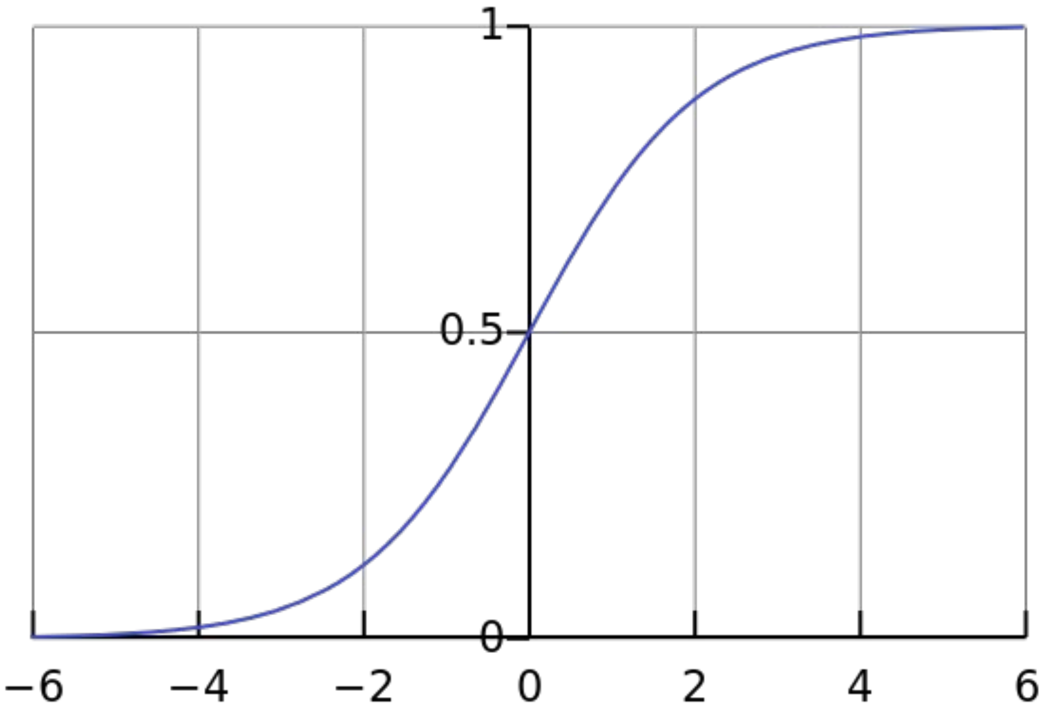
Linear: A = cx

But:

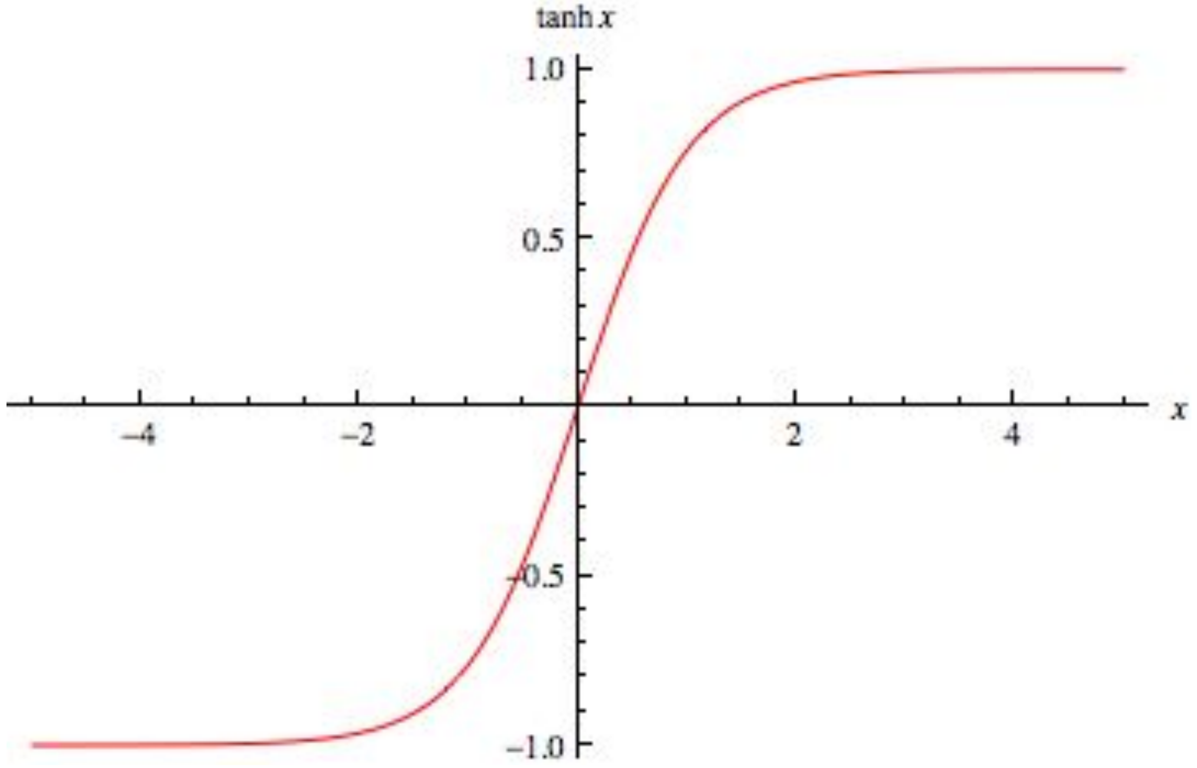
1. Derivative is constant
2. Stacking layers no longer works

So we need **nonlinear** activation functions.

sigmoid() (and softmax)



tanh() (zero-centered)



Vanishing Gradient

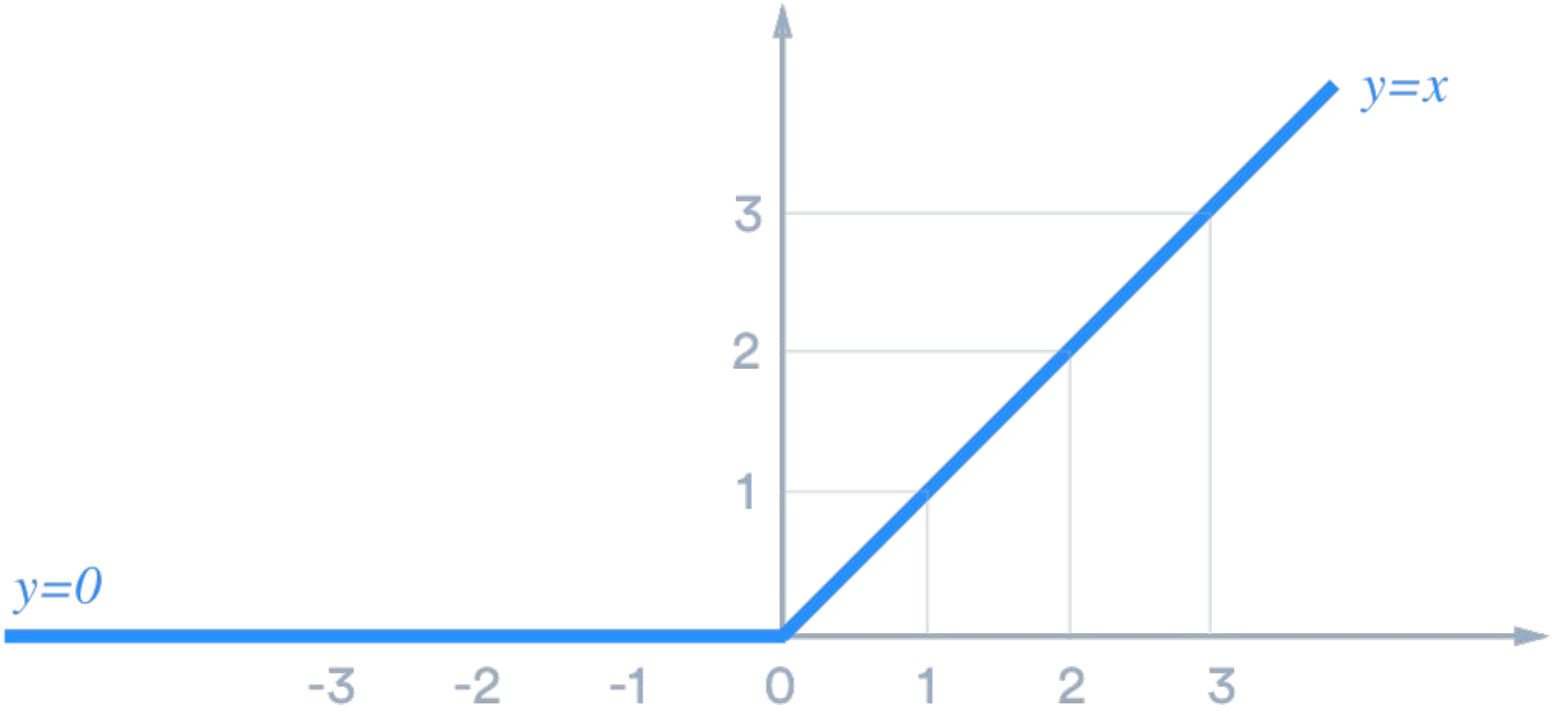
Gradient Descent is used for training Neural Networks

$$o(x) = f_n(f_{n-1}(....f_1(x))$$

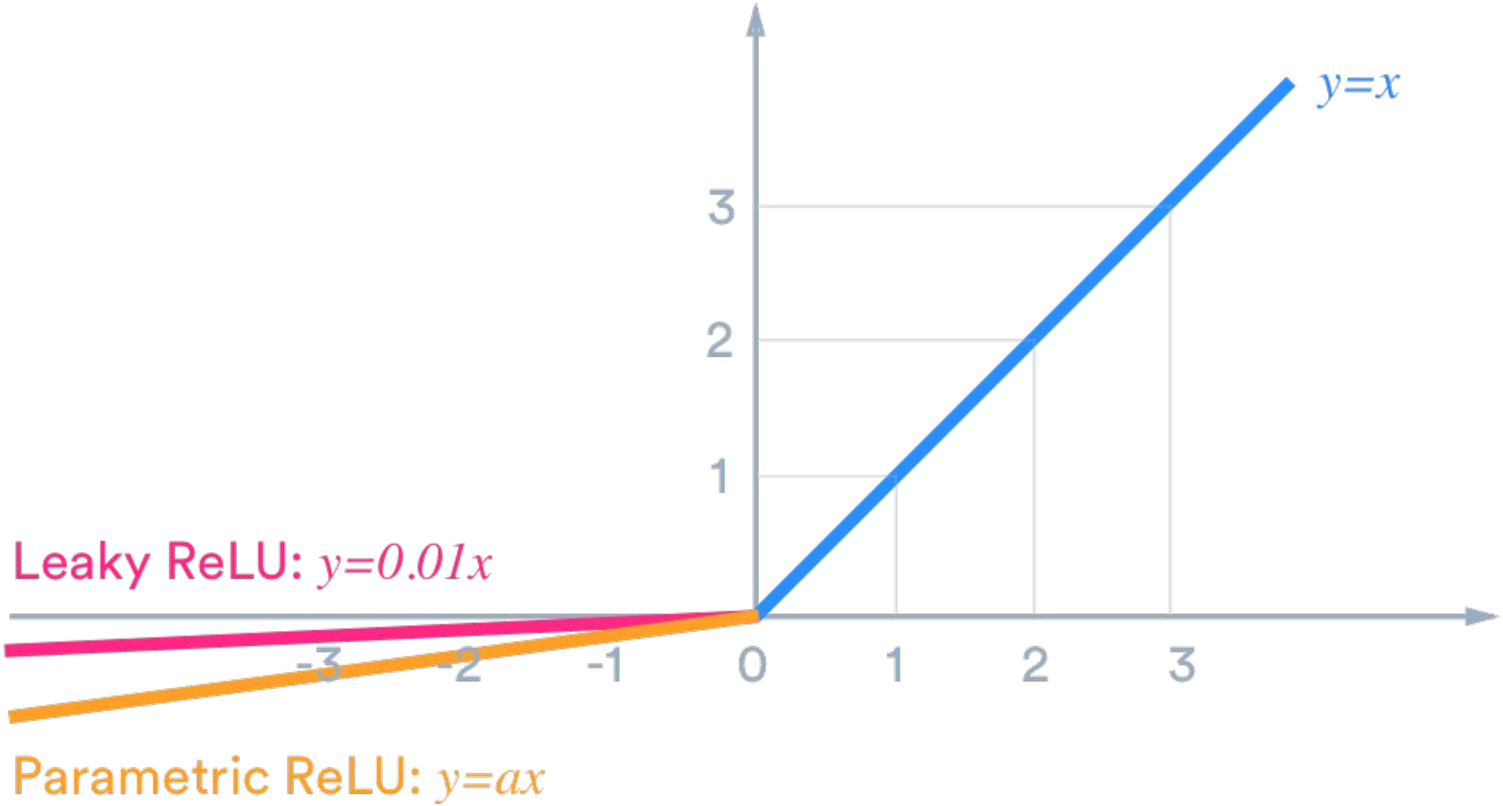
Multiplying values < 1 across multiple layers causes

VANISHING GRADIENT

Avoid Vanishing Gradient with ReLu



Avoiding “dying neurons”: Leaky ReLU

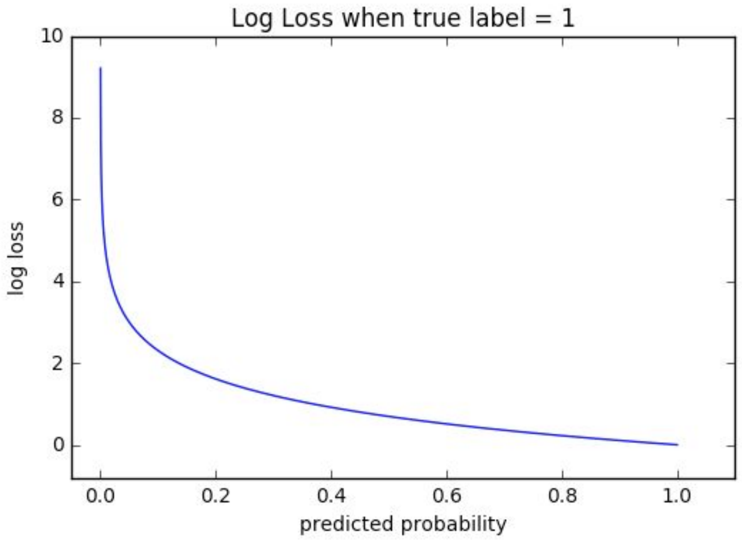


Loss Function

Mean Squared Error:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

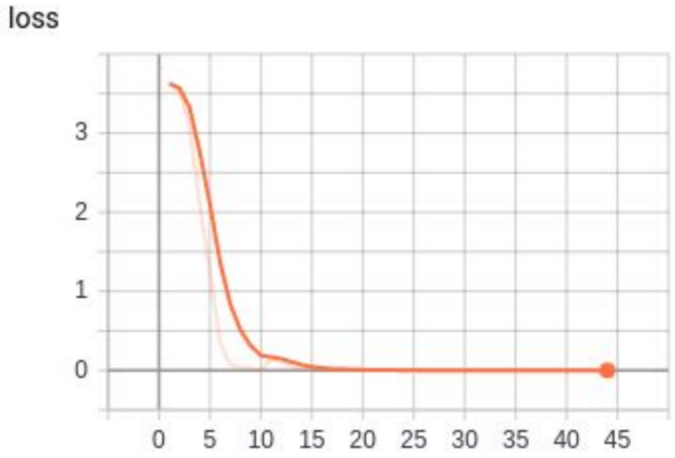
Cross Entropy Loss:



How to use loss?

Train your network while loss is decreasing.

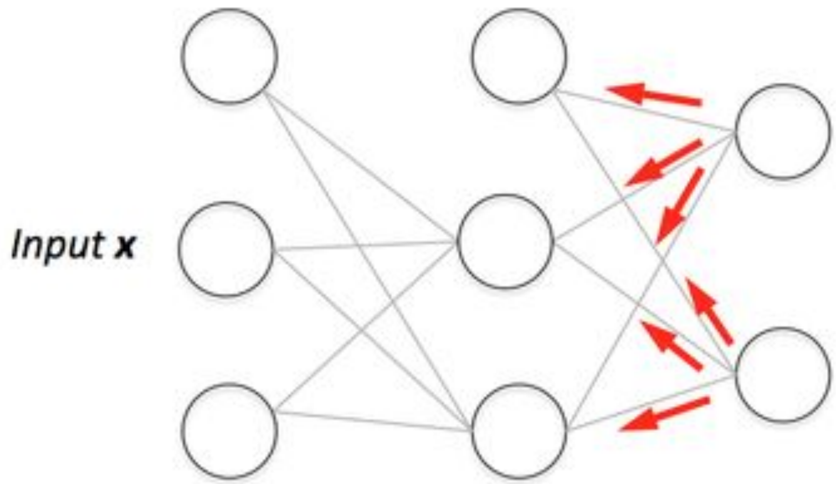
Perfect probability = loss of 0.



Loss isn't very intuitive.   Better to use accuracy or other metric.

Backpropogation

1. Forward pass
2. Error Calculation
3. Backwards Pass



How are Neural Networks Trained in Practice?

Data

- Large!

Tools

- PyTorch
- AllenNLP

Hardware

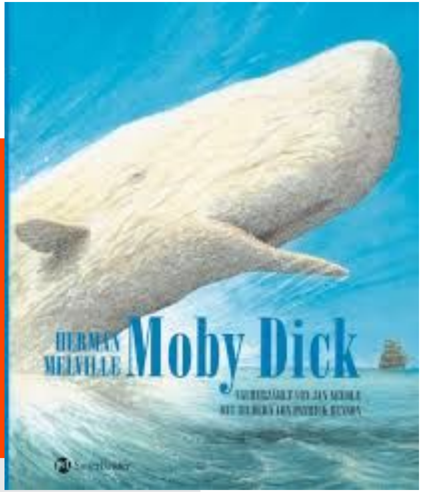
- GPUs (and even TPUs)

Neural Network Data

1. Neural networks often have thousands of parameters.
2. Law of large numbers avoids data inconsistency.
3. Beware of biases.
4. On **small** datasets in my own work, *I've* personally had close results with logistic regression models.

Where do you get data?

- Internet
- Books
- Crowd-Sourcing
- Artificial modifications
- Specialized Communities



But how do you guarantee quality?

- Interannotator Agreement
- Think about biases:
  - Label: you only learn what’s in the training data
  - Language: skewed towards popular languages
  - Text: text data requires less space than audio/video data and can be older
- Visualization

But what about language?

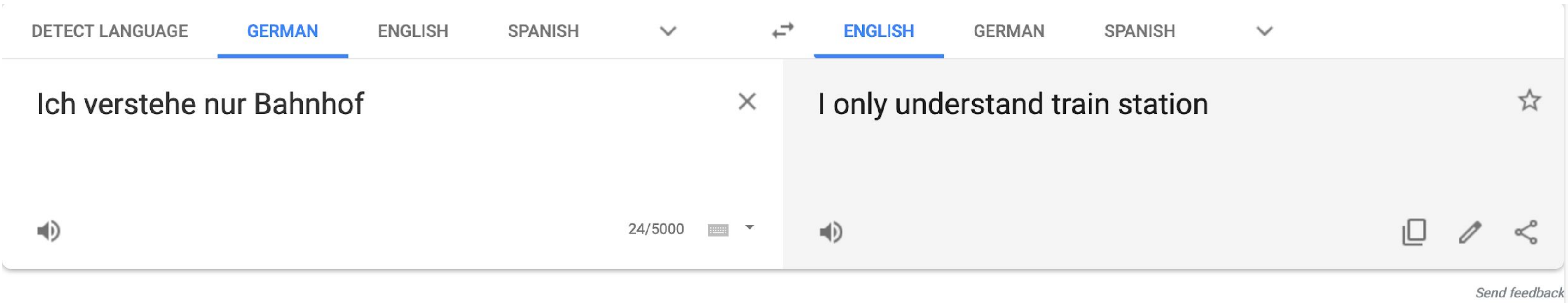
Neural Networks were a big leap in accuracy for **VISION.** Pixels were high in dimensionality, and difficult to interpret.

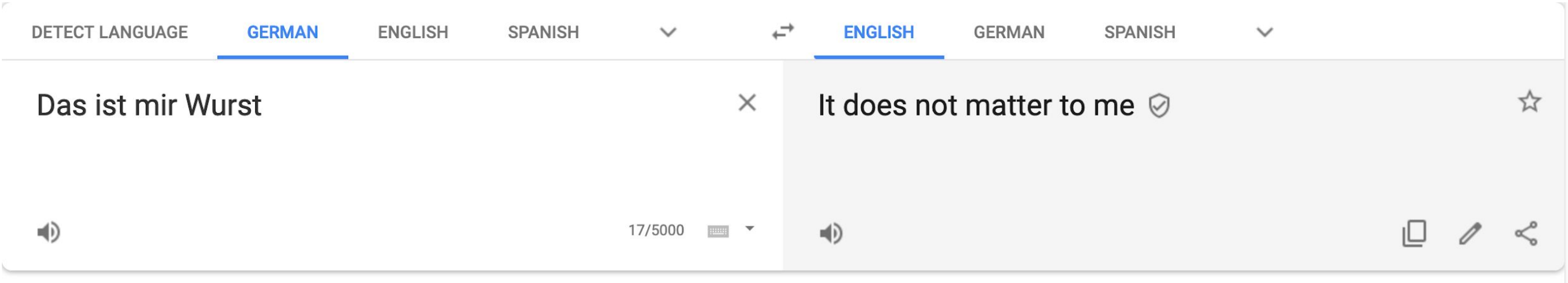
Human interpretable

Levels:

1. Character
2. Word
3. Phrase/Sentence
4. Document

Context Matters

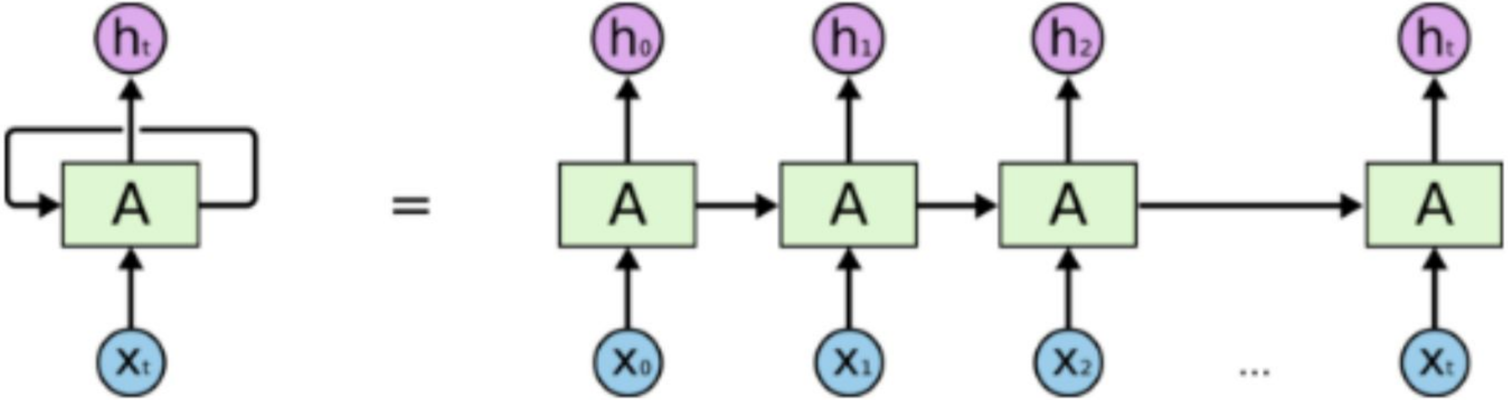




Pragmatics

1. Train/ Development/ Test splits
2. Batching
3. Random seed
4. Reasonable Significant Digits
5. Drop out data during training
6. Initialization
7. Human baselines & common sense
8. Monitor training loss

Recurrent Neural Networks (RNN)



An unrolled recurrent neural network.

Limitations

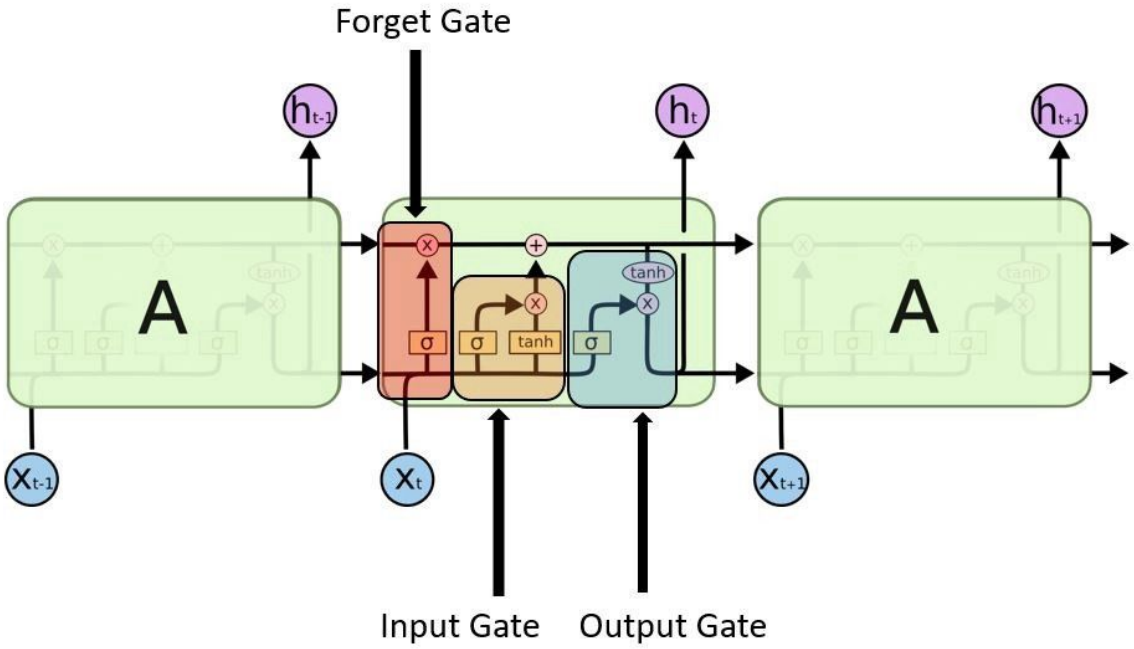
$$h_t = f(h_{t-1}, x_t)$$

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

W is *weight*, **h** is the *single hidden vector,* **Whh** is *the weight at previous hidden state,* **Whx** is the *weight at current input state,* **tanh** is the *Activation funtion,* that implements a Non-linearity that squashes the activations to the range[-1.1]

- Vanishing Gradient Losing Long Term Information
- Computation

Add Gates: Long Short-Term Memory (LSTM)



Input: Is this relevant?

Cell State: What to add?

Output: Where to send next?

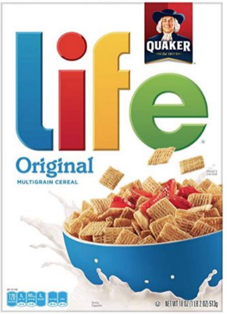
Customers Review 2,491

Thanos

September 2018

Verified Purchase

Amazing! This box of cereal gave me a perfectly balanced breakfast, as all things should be. I only ate half of it but will definitely be buying again!



A Box of Cereal
$3.99

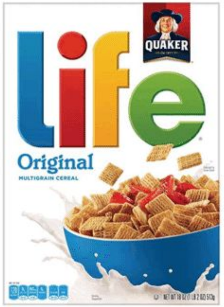
Customers Review 2,491

Thanos

September 2018

Verified Purchase

Amazing! This box of cereal gave me a **perfectly balanced breakfast**, as all things should be. I only ate half of it but **will definitely be buying again!**



A Box of Cereal
$3.99

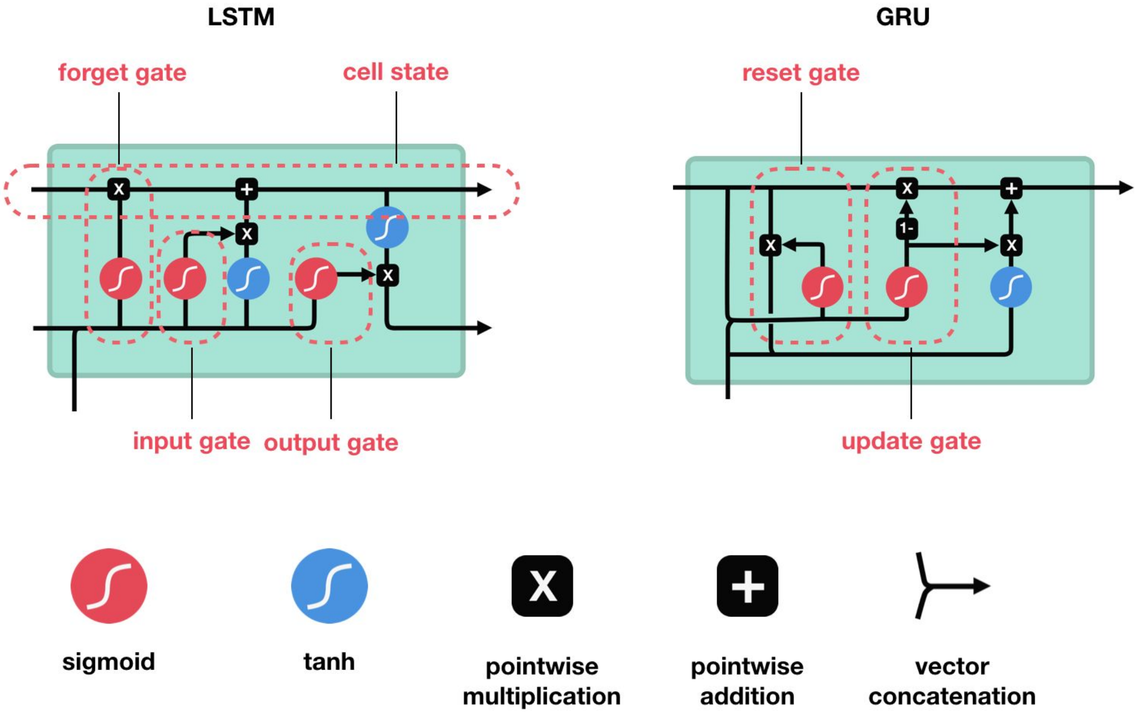
What’s the connection to LANGUAGE?

Language is:

- Dependent on overall context
- Often short-term sequential

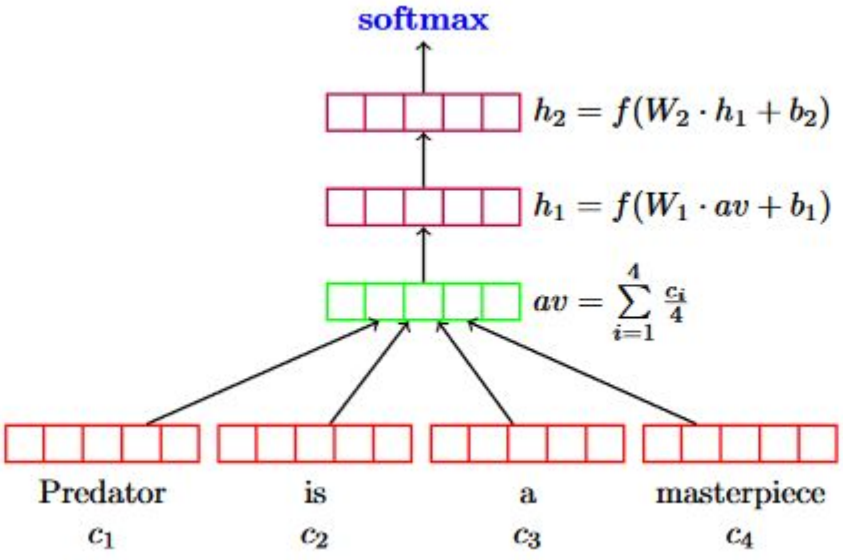
Other Variation: GRU

Optimizing the memory by forgetting leads to a Gradient Recurrent Unit (GRU)

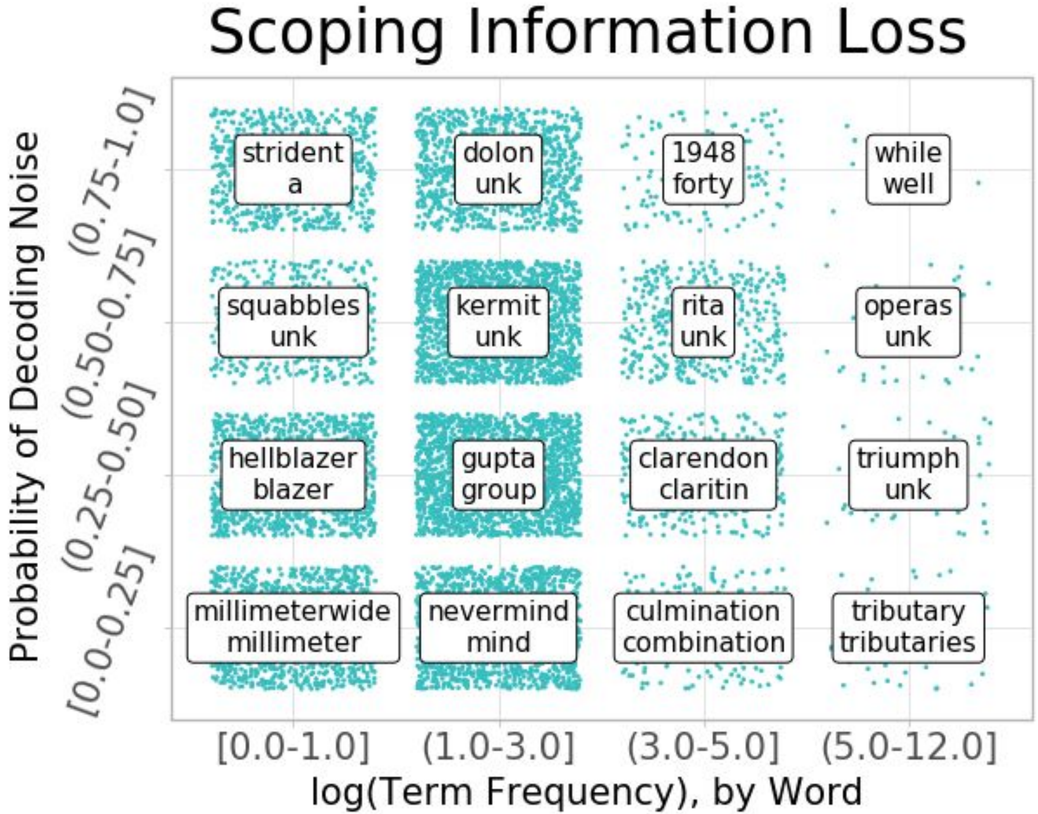


Other Variation: DAN

Dropping word order leads to a Deep Averaging Network (DAN)

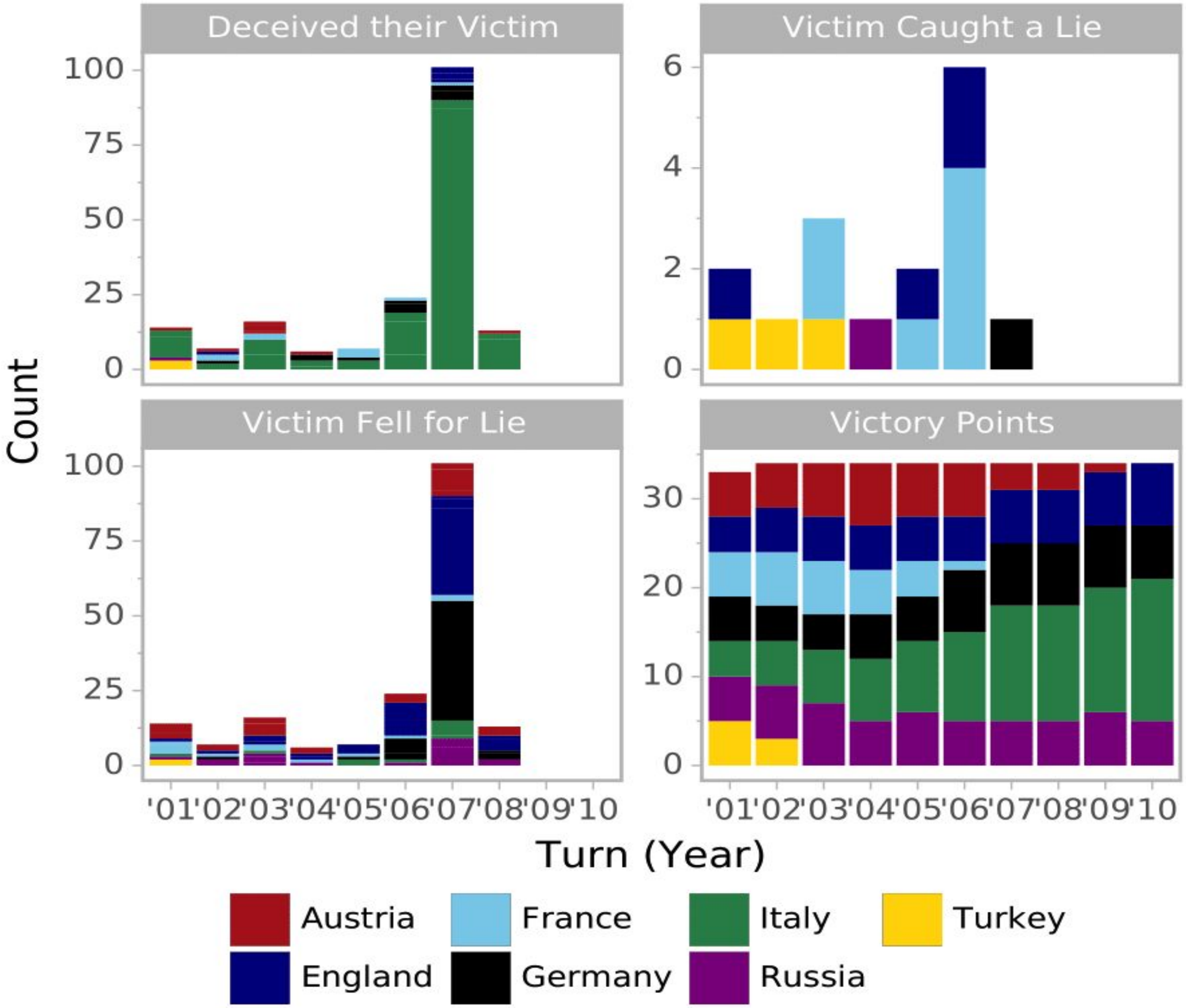


Anecdotal Applications of LSTMs



Role	Turn	Annotations
A	Hey there! Good morning. You're connected to LMT Airways. How may I help you?	DA = { elicitgoal }
C	Hi, I wonder if you can confirm my seat assignment on my flight tomorrow?	IC = { SeatAssignment }
A	Sure! I'd be glad to help you with that. May I know your last name please?	DA = { elicitslot }
C	My last name is Turker.	IC = { contentonly }, SL = {Name : Turker }
A	Alright Turker! Could you please share the booking confirmation number?	DA = { elicitslot }
C	I believe it's AMZ685.	IC = { contentonly }, SL = { Confirmation Number : AMZ685 }
...	...	...

Table 1: A segment of a dialogue from the airline domain annotated at the turn level. This data is annotated with agent dialogue acts (DA), customer intent classes (IC), and slot labels (SL). Roles C and A stand for "Customer" and "Agent", respectively.



Deceived their Victim
Victim Caught a Lie
Victim Fell for Lie
Victory Points
Count
0
25
50
75
100
0
2
4
6
0
10
20
30
'01'02'03'04'05'06'07'08'09'10
Turn (Year)
Austria
England
France
Germany
Italy
Russia
Turkey

High Level Questions?

????????????

Image citations

Neuron: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/

https://towardsdatascience.com/power-of-a-single-neuron-perceptron-c418ba445095

NN:https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a

Sigmoid: https://en.wikipedia.org/wiki/Sigmoid_function#/media/File:Logistic-curve.svg

Tanh: https://mathworld.wolfram.com/HyperbolicTangent.html

Relu: https://medium.com/@sonish.sivarajkumar/relu-most-popular-activation-function-for-deep-neural-networks-10160af37dda

Leaky Relu:
https://medium.com/@himanshuxd/activation-functions-sigmoid-relu-leaky-relu-and-softmax-basics-for-neural-networks-and-deep-8d9c70eed91e

RNNs, LSTMS: https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e

GRUs: https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

DAN: https://mlexplained.com/2018/05/11/paper-dissected-deep-unordered-composition-rivals-syntactic-methods-for-text-classification-explained/