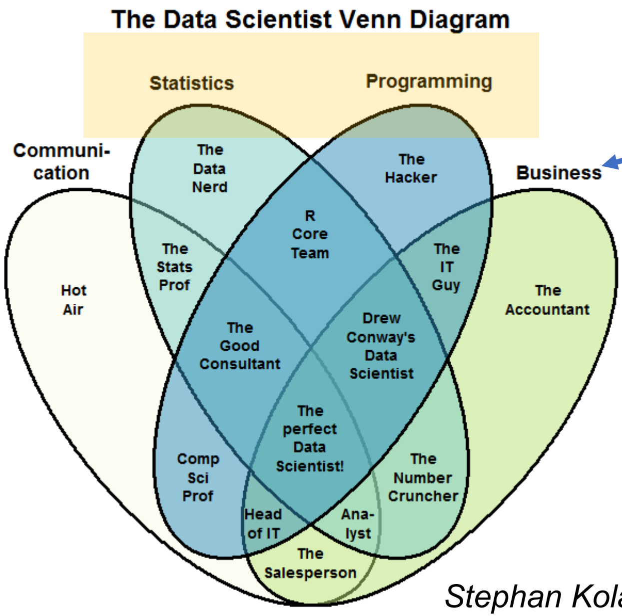
ISEA Week 6 – Machine Learning Applications

Lovenoor (Lavi) Aulck

IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
O UNIVERSITY OF OREGON

Recap

A Data Scientist



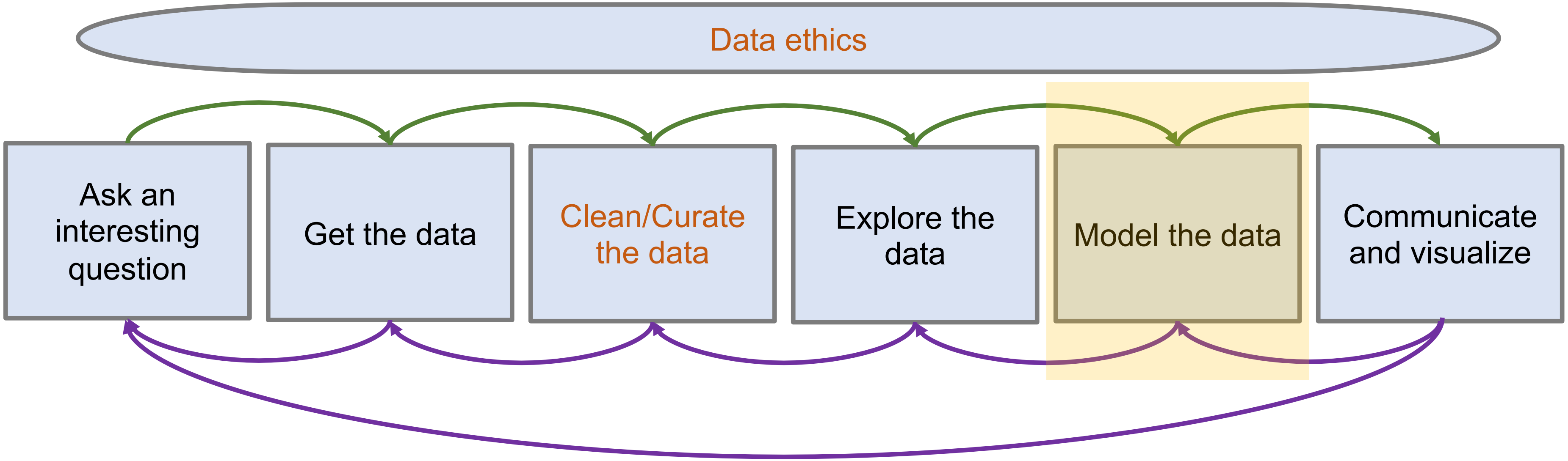
Domain expertise

Critical thinking, creativity, and experimental design?

Stephan Kolassa

Institute of Education Sciences · AmplifyLearn.AI · University of Washington eScience Institute: Advancing Data-Intensive Discovery in All Fields · University of Oregon

A Data Scientist



IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
O UNIVERSITY OF OREGON

Models for Predictions

> All models are, in small or large part, wrong
  - Miss features, omit dependencies, make assumptions, etc

> Models provide a simplification
  - Require assumptions
  - Require some understanding/intuition

> But… what do we need for prediction?

Models for Predictions

- But… what do we need for prediction?
  - If all we care about is predicting some target variable, maybe we can just ignore some of the messy assumptions and focus on specific metrics?

- A critical concern of machine learning is the ability to build models that accurately generalize while a critical concern of econometrics is the ability to build models that capture relationships

“Flavors” of Machine Learning

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

Designing for Prediction: Key Ideas

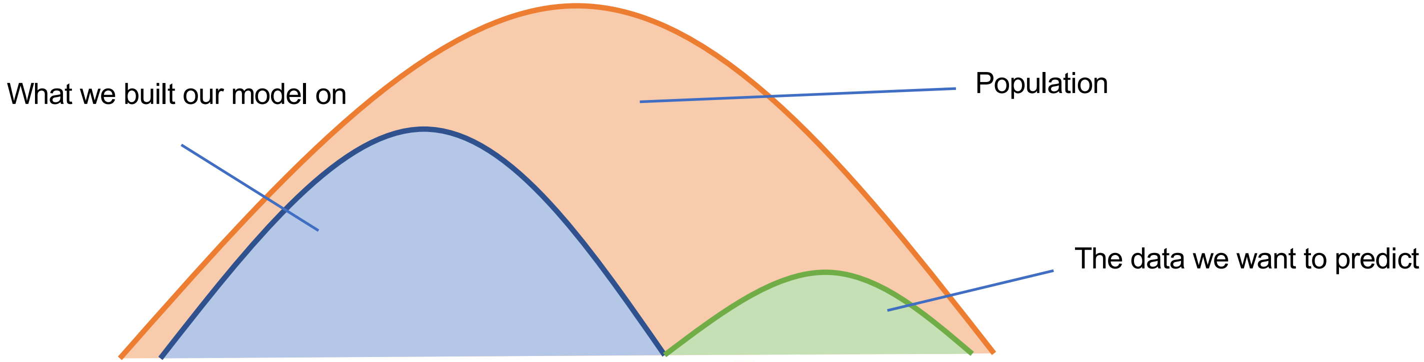
> Generalization and overfitting
> Training, validation, and test data
> Evaluation metrics
> Baselines
> Error analysis

The ML Dilemma

variance

> We want to find a balance between modeling the data we have while also being able to generalize to unseen data

bias



Training, Validation, and Test Data

- **Training set**: a set of examples used for learning, to which we fit parameters of a model
- **Validation set**: a set of examples used to tune model hyperparameters of a model (often a subset of the training set)
- **Test set**: a set of examples used only to assess the performance of a fully-specified model. DO NOT look at the test data to guide model design!!

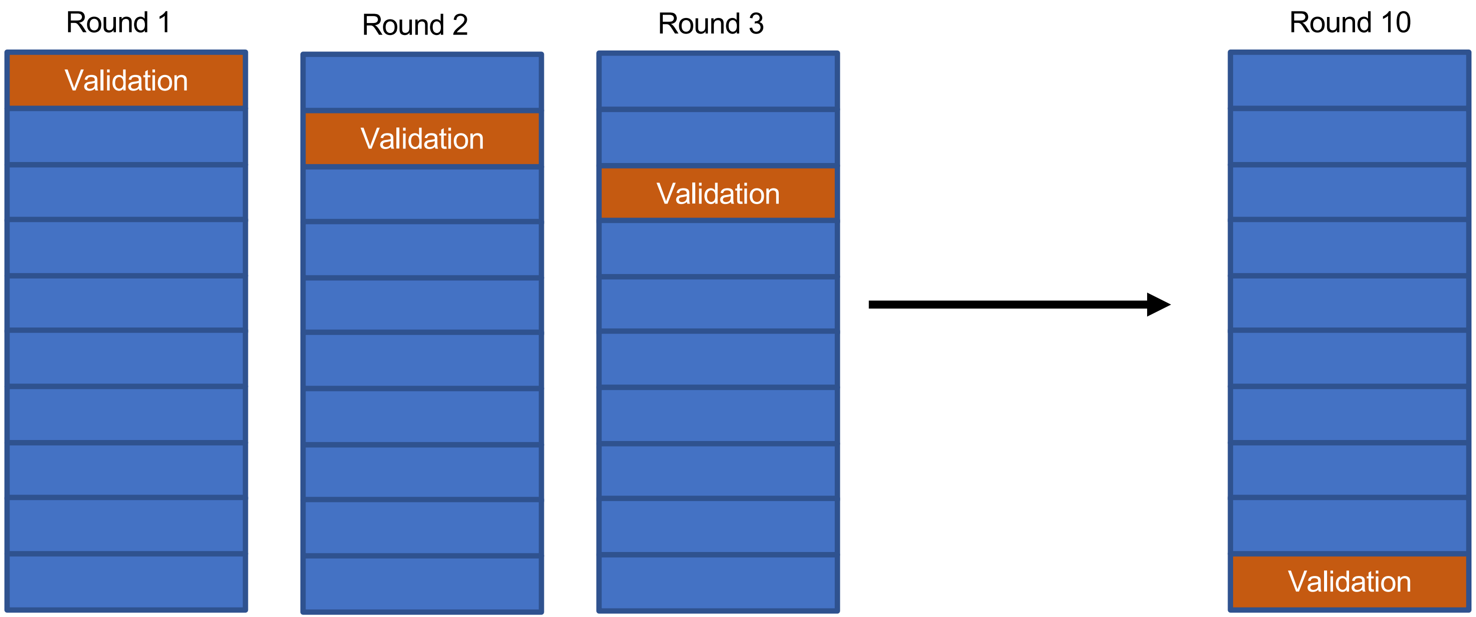
Bootstrapping

- Given unlimited data, it's easy to get new test data (assuming IID). What if you have limited data?

- Your "random" sample of training data may still not be representative
  - Remember: GIGO! (Garbage in, garbage out)

- Bootstrapping can help us recycle and maximize data usage
  - It also allows us an easy avenue to tune hyperparameters

K-Fold Cross Validation



Evaluation Metrics

- Figuring out which to use can be critical for evaluation
  - Can also weigh multiple metrics at once
- Classification:
  - Accuracy
  - Precision/recall
  - AUROC
- Regression:
  - MSE/RMSE
- Log loss for either
- Many, many more

Baselines

> We also often need to quantify progress (accuracy, correctness) relative to something meaningful. The following are often used:
>
> – vs random guessing
> – vs most likely label
> – vs state of the art
> – vs something else simple/intuitive

> Example: if our dataset has 90% of students retained, is an accuracy of 90% when predicting retention really all that great?

Ethics

- Beware of black boxes
- Pay attention to “explainable” ML
  - Just because something predicts well, doesn’t mean we can’t understand it
- Pay attention to biases in your data!
  - Your models will always reflect the biases they inherit
  - This can also relate to blindspots in your data

In The News

'Embarrassing and wrong': Google admits it lost control of image-generating AI

Devin Coldewey @techcrunch / 1:01 PM PST • February 23, 2024

Comment

In The News

'Embarrassing and wrong': Google admits it lost control of image-generating AI

Devin Coldewey @techcrunch / 1:01 PM PST • February 23, 2024 Comment

Certainly! Here is a portrait of a Founding Father of America:

Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

Outline

> Recap
> Continuing Modeling Exercise
> Break/Discussion
> Neural Networks
> Break
> Neural Network Flavors
> Programming Exercise (time permitting)

You Will Need...

- You will need to have the following python packages installed and working for the demos:
  - pandas
  - numpy
  - seaborn (and/or matplotlib)
  - sklearn
  - statsmodels
  - random
  - torch
  - pytorch_tabnet

Notebook

Break

- During break, think about an application of machine learning in your work
  - What would you try to predict/group/optimize?
  - What are some potential gains from this work?
  - What are some roadblocks?

Break

- During break, think about an application of machine learning in your work
  - What would you try to predict/group/optimize?
  - What are some potential gains from this work?
  - What are some roadblocks?

  - For Lavi:
    - Predict students' major enrollment

Break

- During break, think about an application of machine learning in your work
  - What would you try to predict/group/optimize?
  - What are some potential gains from this work?
  - What are some roadblocks?

  - For Lavi:
    - Predict students' major enrollment
    - Allows for better resource allocation and planning

Break

- During break, think about an application of machine learning in your work
  - What would you try to predict/group/optimize?
  - What are some potential gains from this work?
  - What are some roadblocks?

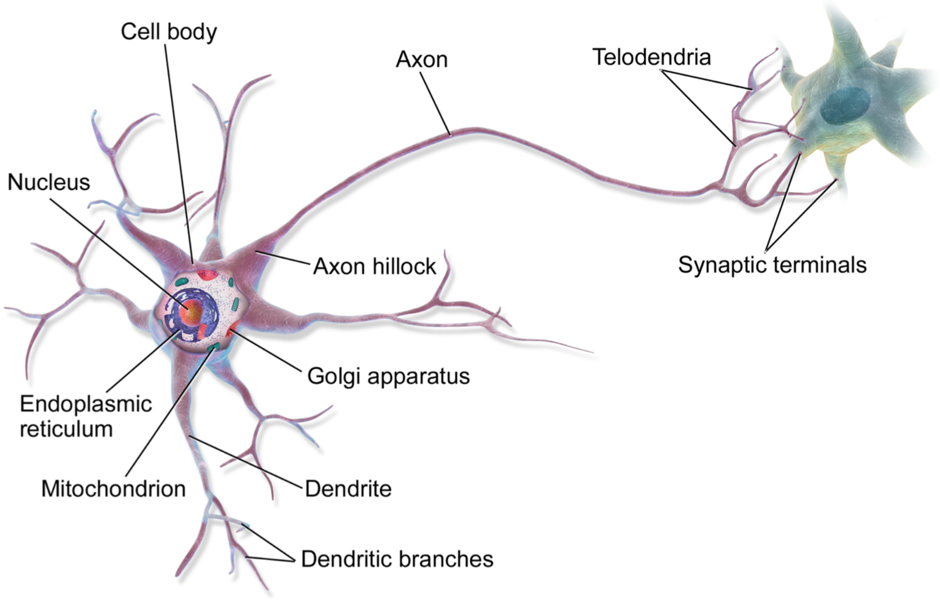
  - For Lavi:
    - Predict students' major enrollment
    - Allows for better resource allocation and planning
    - Resources. Also, students are (very) unpredictable!

Neurons

"I could not help but wonder if this new 'Deep Learning' was anything fancy or just a scaled up version of the 'artificial neural nets' that were already developed by the late 80s. And let me tell you, the answer is quite a story — the story of not just neural nets, not just of a sequence of research breakthroughs that make Deep Learning somewhat more interesting than 'big neural nets', but most of all of how several unyielding researchers made it through dark decades of banishment to finally redeem neural nets and achieve the dream of Deep Learning."

- Andrey Kurenkov

Neurons



Blausen Medical

Neurons

> All neurons are electrically excitable, maintaining voltage gradients across their membranes via metabolically-driven ion pumps

Neurons

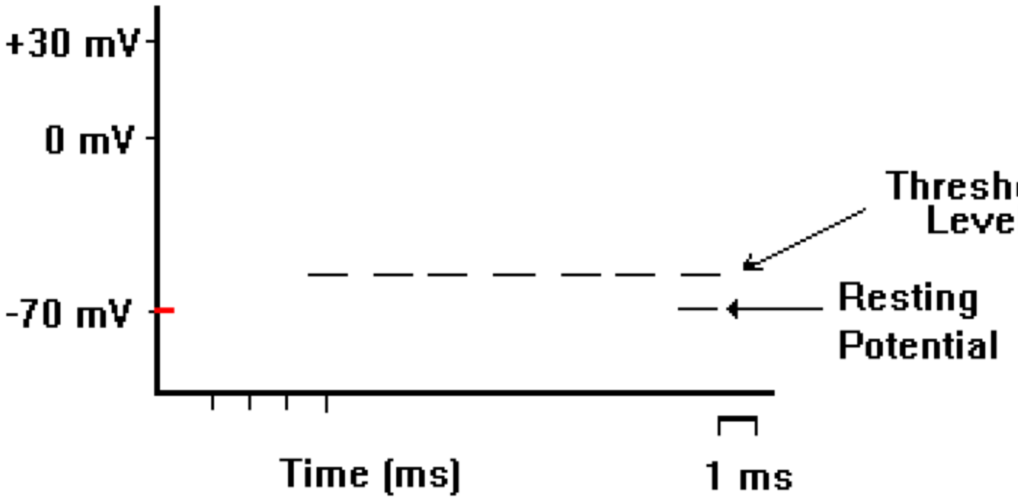
> All neurons are electrically excitable, maintaining voltage gradients across their membranes via metabolically-driven ion pumps

> If the voltage changes by a large enough amount, an action potential is generated
  – Action potentials are all-or-nothing responses

Neurons

- All neurons are electrically excitable, maintaining voltage gradients across their membranes via metabolically-driven ion pumps
- If the voltage changes by a large enough amount, an action potential is generated
  - Action potentials are all-or-nothing responses
- These action potentials move along axons and activate synaptic connections with other neurons

Neurons



Eric Chudler

The Perceptron

- Rosenblatt’s (a psychologist’s) attempt at trying to mathematically model a neuron

The Perceptron

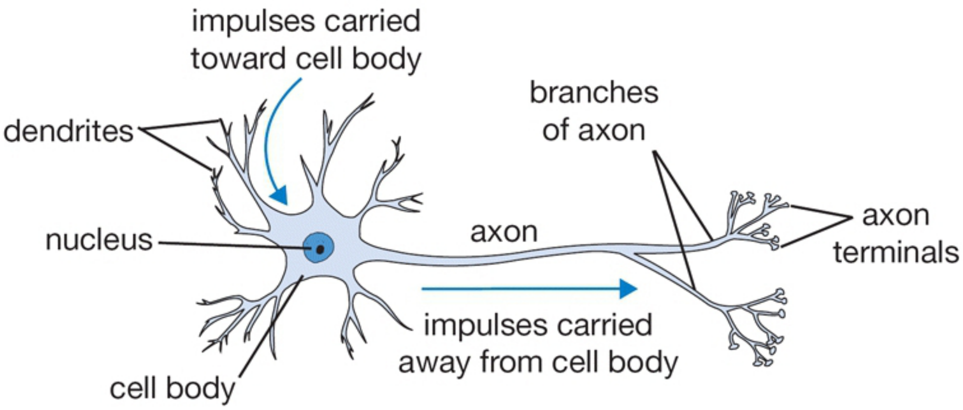
- Rosenblatt's (a psychologist's) attempt at trying to mathematically model a neuron

- The perceptron takes in a set of inputs, multiplies them by a weight, and then aggregates
  - Keep this process in mind! It is central to how neural networks work!

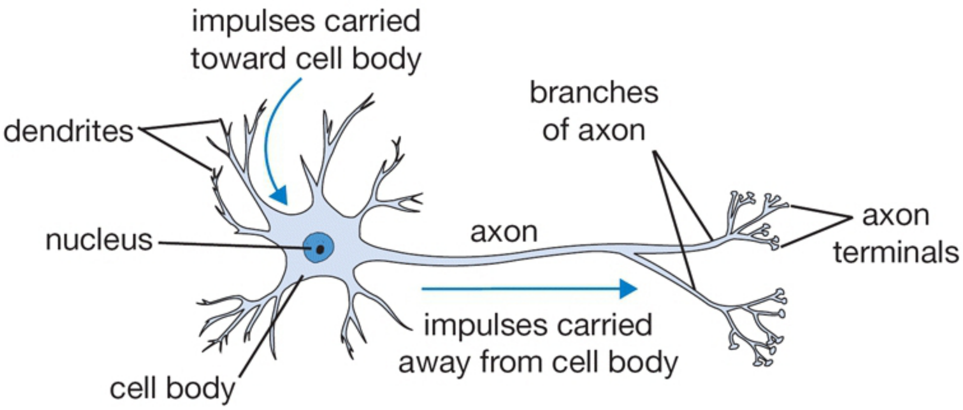
The Perceptron

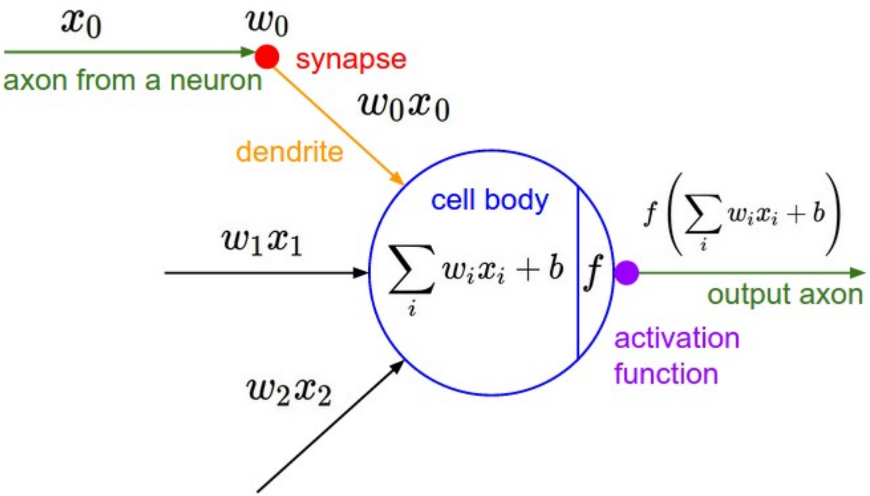
- Rosenblatt’s (a psychologist’s) attempt at trying to mathematically model a neuron
- The perceptron takes in a set of inputs, multiplies them by a weight, and then aggregates
  - Keep this process in mind! It is central to how neural networks work!
- Perceptron “fires” if the aggregate reaches some threshold
  - Very much like the all-or-nothing response of a neuron

The Perceptron



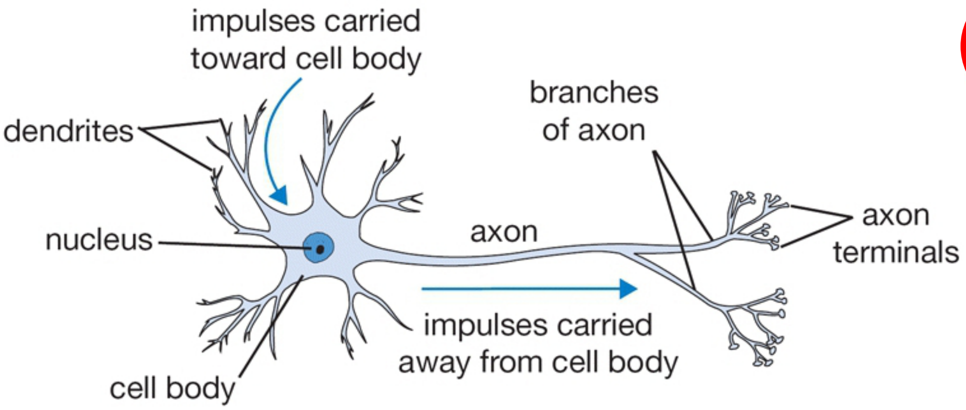
The Perceptron

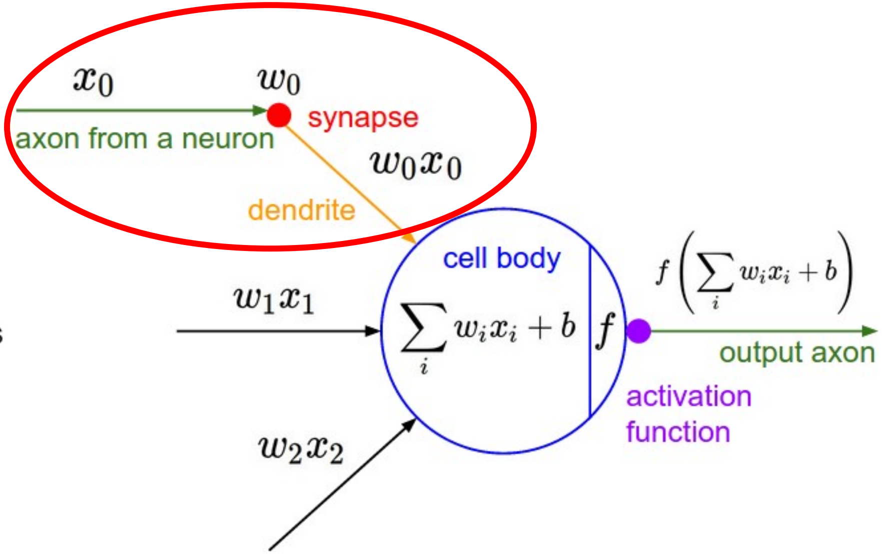




Stanford CS231n

The Perceptron





Stanford CS231n

Perceptrons For Logic

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	-1

Perceptrons For Logic

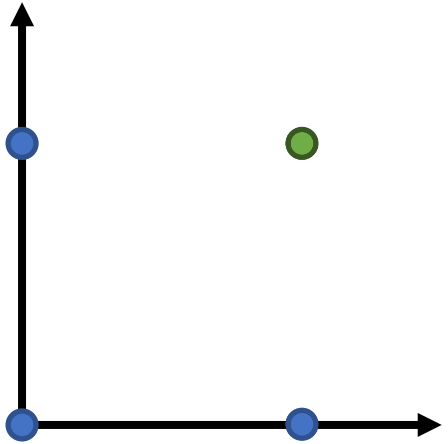
AND

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	-1

Perceptrons For Logic

AND

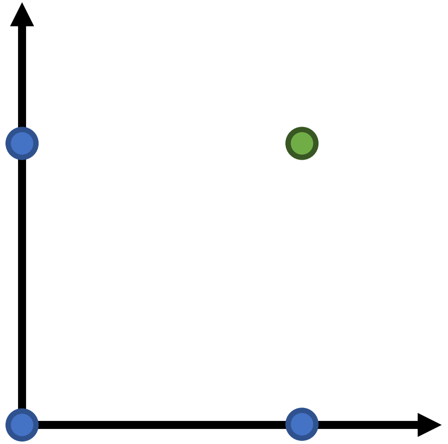
X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	-1

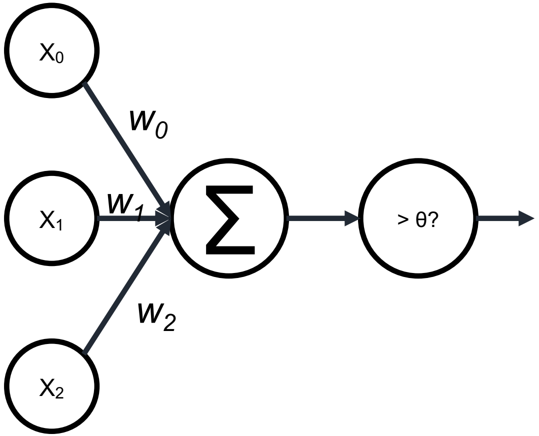


Perceptrons For Logic

AND

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	-1

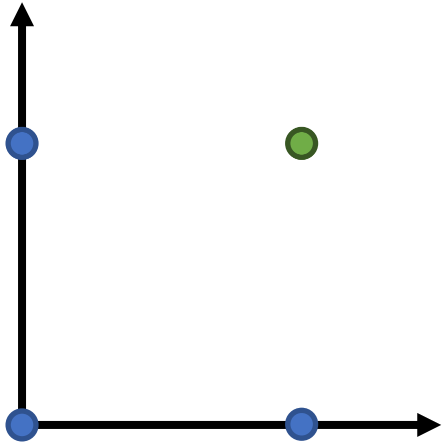


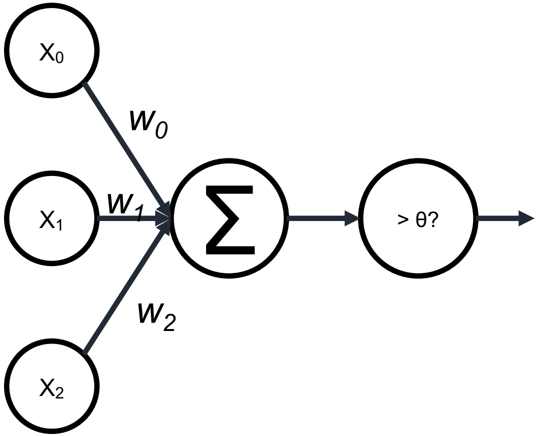


Perceptrons For Logic

AND

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	-1





(sum of our components >= threshold?)

$(X_0w_0 + X_1w_1 + X_2w_2) >= \text{threshold}$
$X_0w_0 + X_1w_1 + X_2w_2 >= \theta$ (T: y = 1; F: y = -1)

$w_0 + w_1 + w_2 >= \theta$
$w_0 + w_1 < \theta$
$w_0 + w_2 < \theta$
$w_0 < \theta$

How Perceptrons Learned

- In essence, the perceptron was a translation of inputs into outputs

How Perceptrons Learned

> In essence, the perceptron was a translation of inputs into outputs

> The weights needed to perform these translations could be learned based on available data

How Perceptrons Learned

- In essence, the perceptron was a translation of inputs into outputs

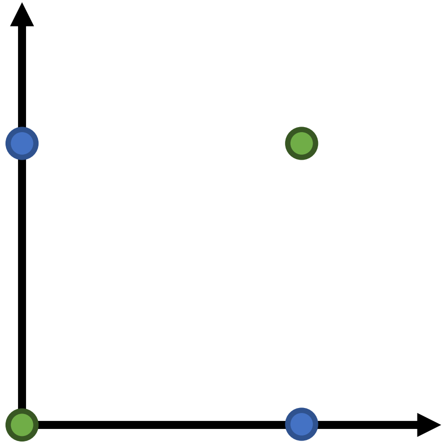
- The weights needed to perform these translations could be learned based on available data

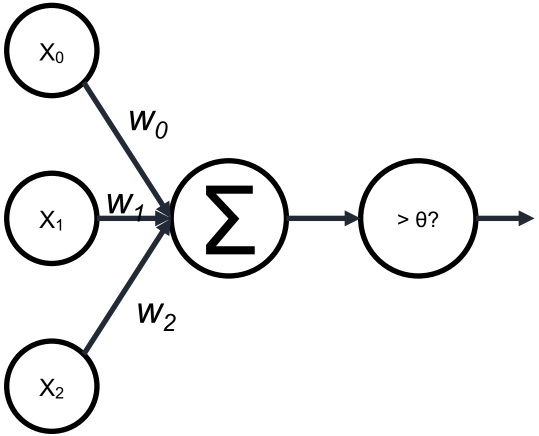
- One way: Rosenblatt's algorithm from the 50s
  - Showed convergence when two criteria are met: data is linearly separable and step size is sufficiently (in the proof, infinitely) small

Perceptrons For Logic

XOR

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	1

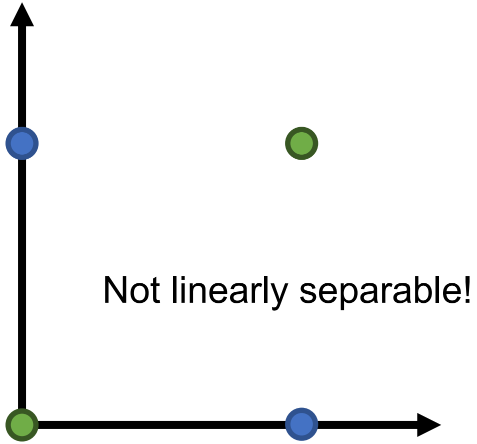


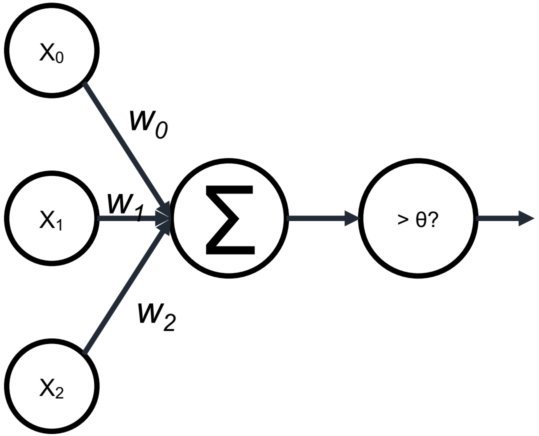


Perceptrons For Logic

XOR

X1	X2	Output
1	1	1
1	0	-1
0	1	-1
0	0	1





The First "AI Winter"

- Minsky and Papert published *Perceptrons: an introduction to computational geometry* in 1969
  - Presented an analysis on opportunities and limitations of perceptrons (including, among other things, their inability to work on XOR logic)
- During the 1960s, the US government was also interested in machine translation (particularly language to language)
  - Why? It was the Cold War
- A confluence of circumstances led to a decreased interest (and funding) in AI and neural networks for much of the 70s

The Thaw of The First "AI Winter"

- The first AI winter eventually ended
  - Spurred by the development of expert systems (which weren't necessarily based neural nets)
- There was a rapid growth in the commercialization of AI-related products and conferences through the 1980s
  - Many of the top conferences today were established: AAAI and NeurIPS
- Through the first AI Winter and AI Spring, there were also developments in neural networks

Multi-Layer Networks

- Minskey and Papert argued that a multi-layered network was needed to capture XOR logic
  - Spurred by the development of expert systems (which weren't necessarily based neural nets)

Multi-Layer Networks

> Minskey and Papert argued that a multi-layered network was needed to capture XOR logic
  - Spurred by the development of expert systems (which weren't necessarily based neural nets)

> What is a multi-layered network?
  - Between the inputs and the aggregation, add a layer of neurons
  - This layer can be of arbitrary length but is connected to the input layer

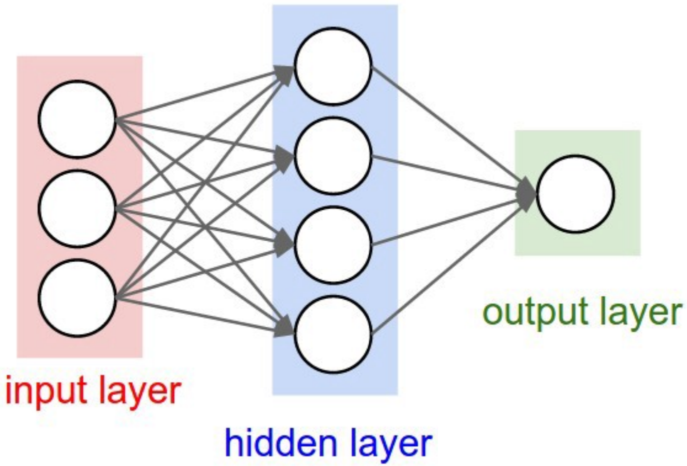
Multi-Layer Networks

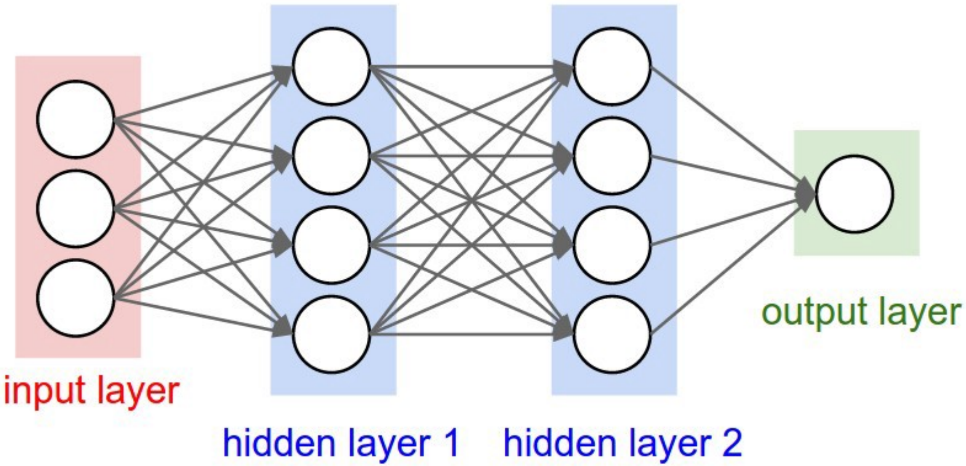
> Minskey and Papert argued that a multi-layered network was needed to capture XOR logic
  – Spurred by the development of expert systems (which weren't necessarily based neural nets)

> What is a multi-layered network?
  – Between the inputs and the aggregation, add a layer of neurons
  – This layer can be of arbitrary length but is connected to the input layer

> Neural networks with a single hidden layer are universal approximators
  – They can mimic any(!!!) function (provided enough neurons)

Multi-Layer Networks





Modified from Stanford CS231n

Taking A Step Back

- Neural networks are just that – a network of neurons
  - Neurons are represented as perceptrons
  - Sometimes also called artificial neural networks (ANNs)
  - Sometimes also called multi-layer perceptrons (MLPs; though MLPs are often vanilla architectures)
  - Neurons are also called “units”

Taking A Step Back

- Neural networks are just that – a network of neurons
  - Neurons are represented as perceptrons
  - Sometimes also called artificial neural networks (ANNs)
  - Sometimes also called multi-layer perceptrons (MLPs; though MLPs are often vanilla architectures)
  - Neurons are also called “units”
- What is deep learning?
  - Generally, neural networks with a large number of hidden layers! (thus, “deep”)

Training Multi-Layer Networks

- Forward propagation
  - Pass the input data through the network and calculate outputs ($\hat{y}$)
  - Basically "make predictions"

Training Multi-Layer Networks

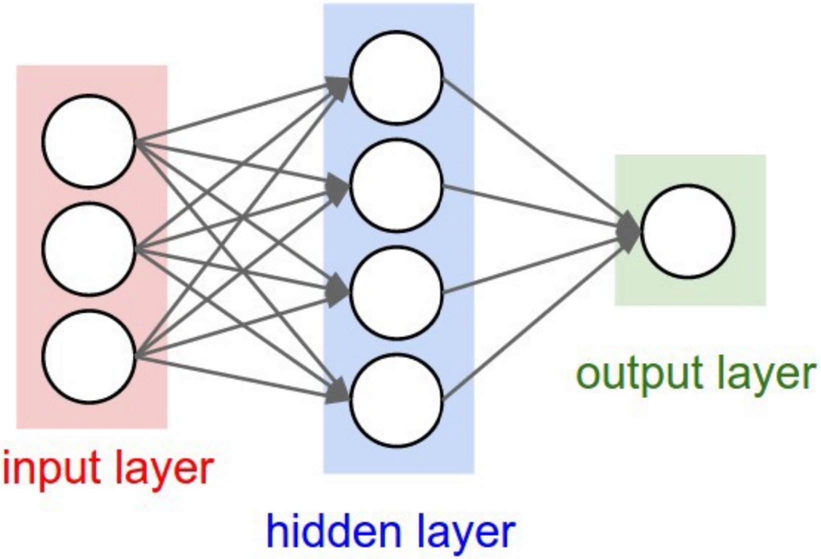
- Forward propagation
  - Pass the input data through the network and calculate outputs ($\hat{y}$)
  - Basically “make predictions”

- Calculate error
  - Determine how far off from y we are with $\hat{y}$

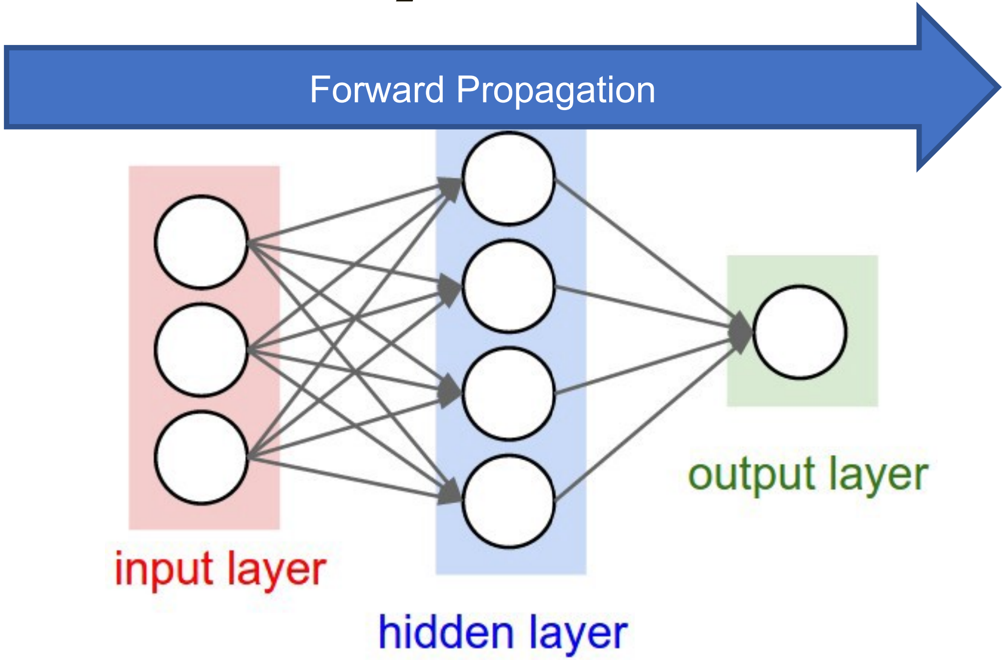
Training Multi-Layer Networks

- Forward propagation
  - Pass the input data through the network and calculate outputs (ŷ)
  - Basically “make predictions”
- Calculate error
  - Determine how far off from y we are with ŷ
- Backpropagation (Backward propagation)
  - Update weights/biases based on the error

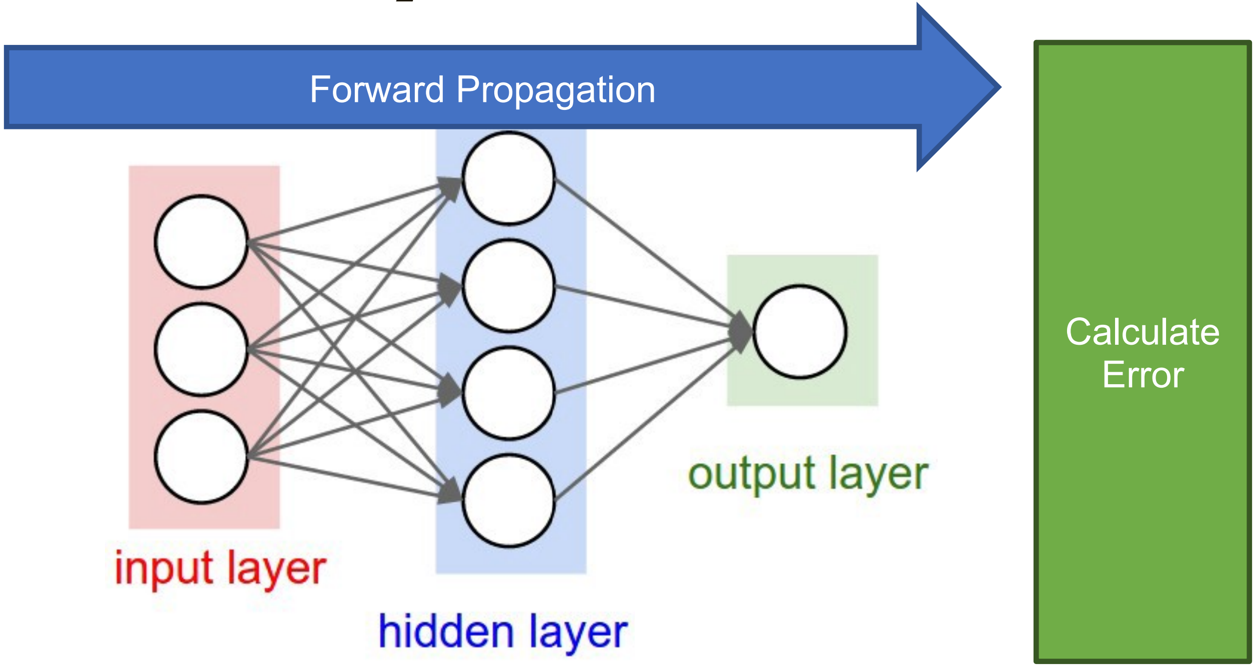
Training Multi-Layer Networks



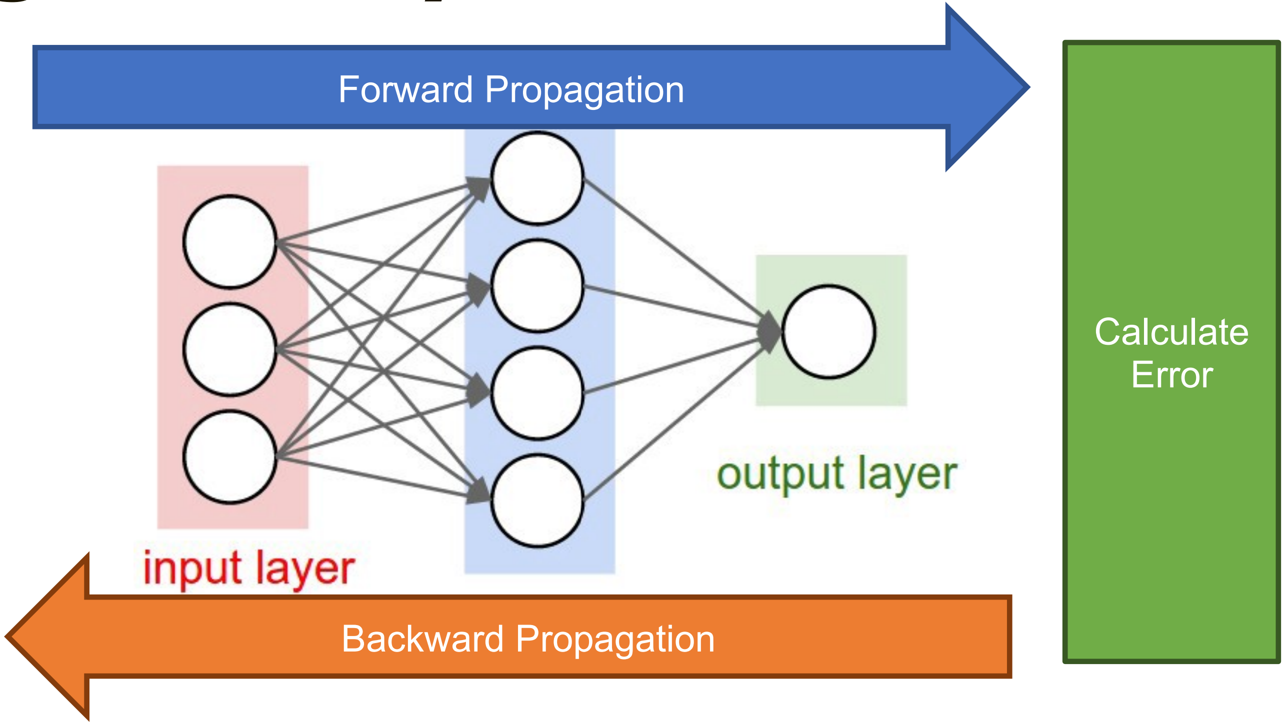
Training Multi-Layer Networks



Training Multi-Layer Networks



Training Multi-Layer Networks



The Second AI Winter

- Some time after the first AI Spring, a second AI winter came about
- Multi-layered networks could be trained (in theory); however, they took a lot of computational resources
  - Alternatively, there became a growing interest in SVMs and Bayesian approaches
  - At some point, NeurIPS (which was founded a conference on neural networks) was accepting more papers in non-neural domains than neural domains
- On the flip side, commercialization of expert systems stalled

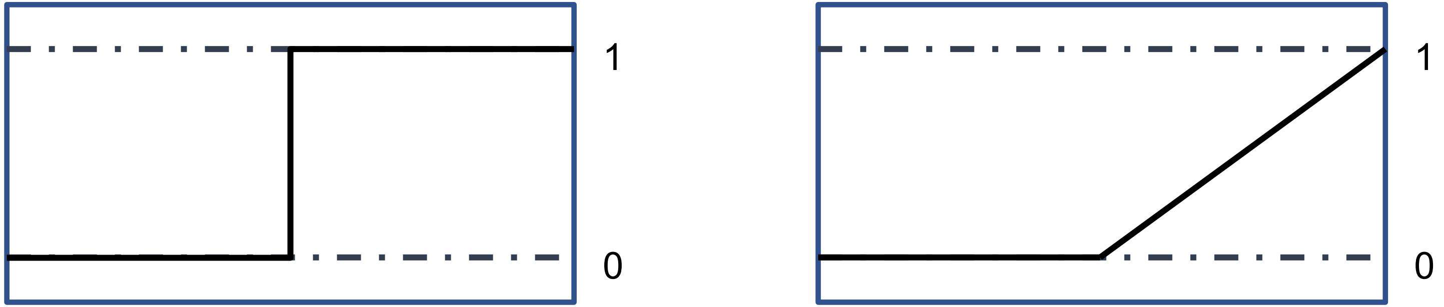
The Second “AI Spring”

> Several additional key developments brought about the growth of “Deep Learning”

The Second "AI Spring"

> Several additional key developments brought about the growth of "Deep Learning"

1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)



The Second "AI Spring"

- Several additional key developments brought about the growth of "Deep Learning"
  1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)
  2. Video games

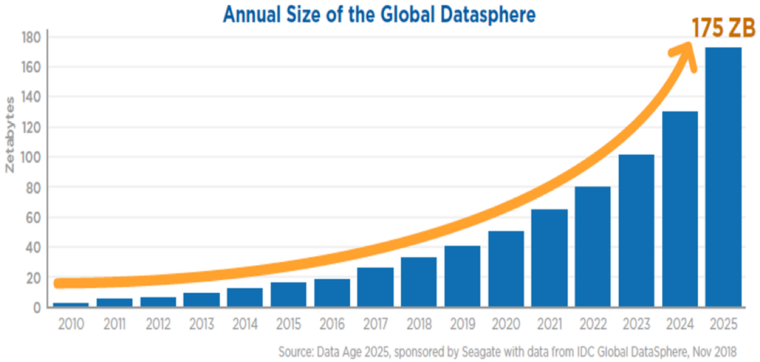
The Second "AI Spring"

- Several additional key developments brought about the growth of "Deep Learning"
  1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)
  2. ~~Video games~~ GPUs

The Second "AI Spring"

> Several additional key developments brought about the growth of "Deep Learning"

1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)
2. ~~Video games~~ GPUs
3. More data



IES Institute of Education Sciences | W AmplifyLearn.AI | UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS | UNIVERSITY OF OREGON

The Second "AI Spring"

- Several additional key developments brought about the growth of "Deep Learning"
  1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)
  2. ~~Video games~~ GPUs
  3. More data
  4. Rebranding

Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

The Second "AI Spring"

> Several additional key developments brought about the growth of "Deep Learning"

1. Rethinking activation functions (using ReLUs allowed for faster training/less leakage)
2. ~~Video games~~ GPUs
3. More data
4. Rebranding
5. Numerous other advancements in optimization theory and approaches

Break

> During break, ensure you have all packages installed locally and/or ready to go on colab

Flavors Of Neural Networks

> Autoencoders
> Convolutional Neural Networks
> Recurrent Neural Networks
> Generative Adversarial Networks
> Transformers

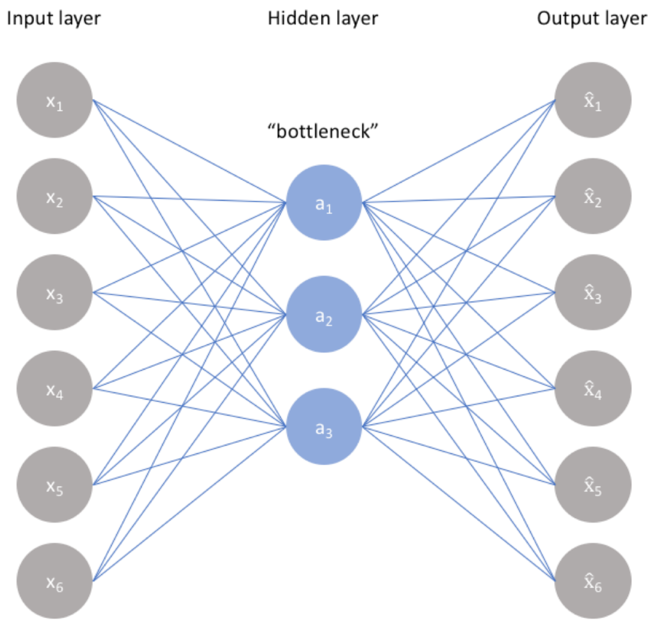
Autoencoders

> An autoencoder is a neural network whose output is the same as its input

Autoencoders

- An autoencoder is a neural network whose output is the same as its input
- What's the point?
  - Dimensionality reduction
    - Using a more constrained model to represent data inputs
    - Can be used for clustering
  - Anomaly detection
    - If the autoencoder can't reconstruct the output, maybe it's dissimilar to the training data?

Autoencoders



Jeremy Jordan

Convolutional Neural Networks

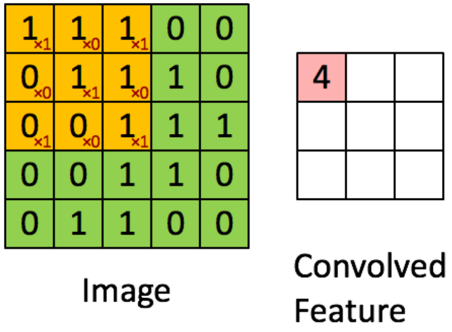
- Convolutional neural networks are networks that employ a convolution function at some point
  - Similar to the convolution used in signal processing
- Focus on small, simple patterns in the data
  - Thus, they are very effective for images and videos
  - Allows us to take into account spatial structures
- Generally involve successive convolution and pooling of features

Convolutional Layer

- The convolutional layer can be thought of as a filter that passes over the input
  - This zeros in on locality – we focus on small activation maps

Convolutional Layer

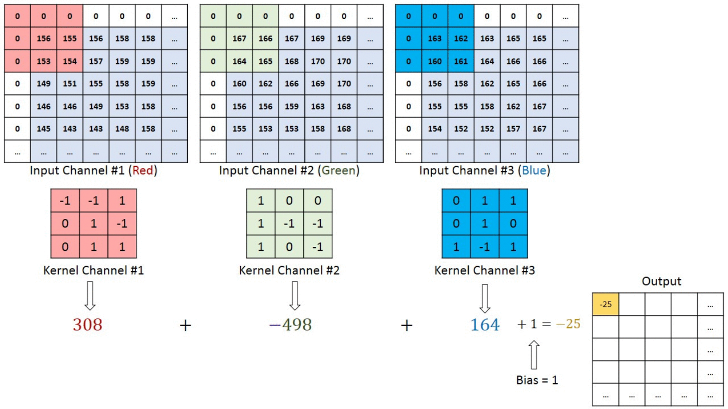
- The convolutional layer can be thought of as a filter that passes over the input
  - This zeros in on locality – we focus on small activation maps



Sumit Saha

Convolutional Layer

> The convolutional layer can be thought of as a filter that passes over the input
  - This zeros in on locality – we focus on small activation maps



Sumit Saha

Institute of Education Sciences
AmplifyLearn.AI
University of Washington eScience Institute
Advancing Data-Intensive Discovery in All Fields
University of Oregon

Convolutional Neural Networks

A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way

CS231n Convolutional Neural Networks for Visual Recognition

IES Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

Recurrent Neural Networks

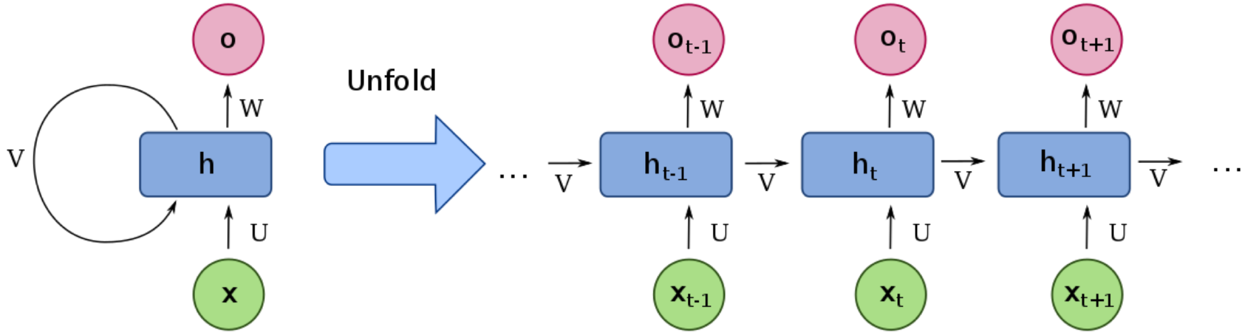
- Take in an input as a temporal sequence
- The order of the input matters and aids in predictions
  - Thus, they are very effective for tasks involving natural language

Recurrent Neural Networks

- Take in an input as a temporal sequence
- The order of the input matters and aids in predictions
  - Thus, they are very effective for tasks involving natural language

Michael Nguyen

Recurrent Neural Networks



fdeloche

Recurrent Neural Networks

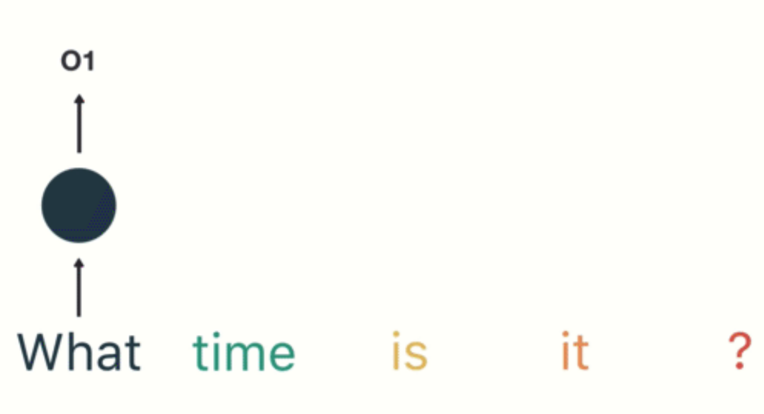
What time is it?

Michael Nguyen

Recurrent Neural Networks

Michael Nguyen

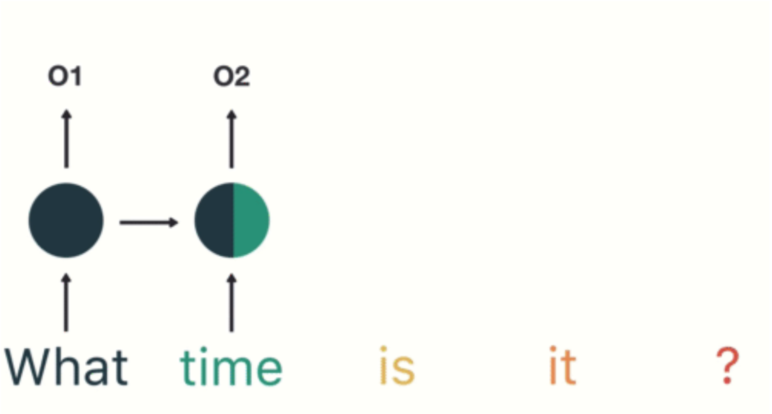
Recurrent Neural Networks



Michael Nguyen

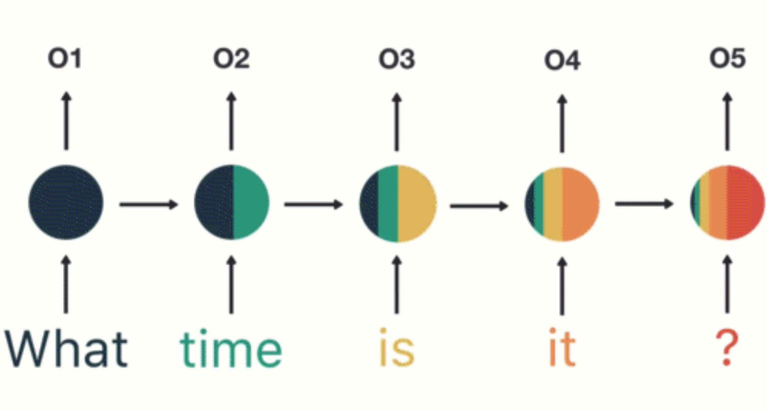
IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Recurrent Neural Networks



Michael Nguyen

Recurrent Neural Networks



Michael Nguyen

Institute of Education Sciences
AmplifyLearn.AI
University of Washington eScience Institute
Advancing Data-Intensive Discovery in All Fields
University of Oregon

Recurrent Neural Networks

Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs

Recurrent Neural Networks (RNNs) are popular models that have shown great promise in many NLP tasks. But despite their recent popularity I've only found a limited number of resources that throughly explain how RNNs work, and how to implement them. That's what this tutorial is about. It's a multi-part series in which I'm planning to cover the following:

colah's blog

Understanding LSTM Networks

towards
data science

Illustrated Guide to Recurrent Neural Networks

Understanding the Intuition

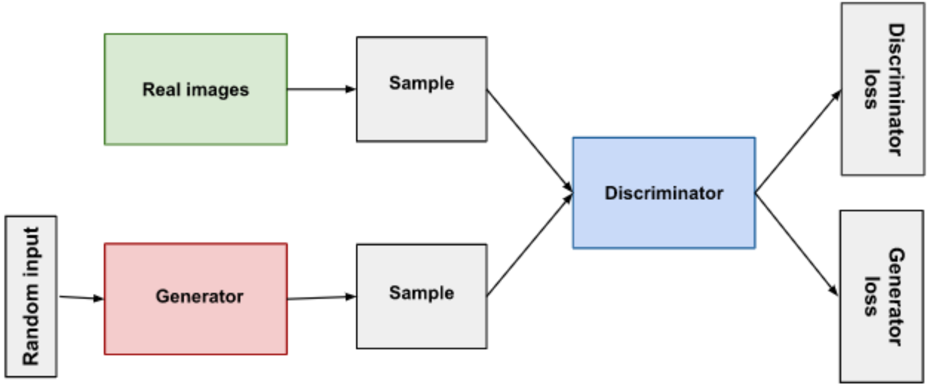
IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Generative Adversarial Networks

- Broadly: composed of two neural networks
  - A generator network which generates data
  - A discriminator network which classifies data

Generative Adversarial Networks

- Broadly: composed of two neural networks
  - A generator network which generates data
  - A discriminator network which classifies data



Generative Adversarial Networks

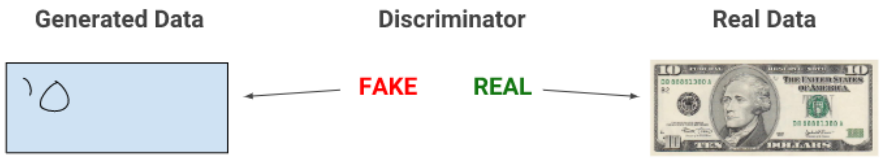
- The generator learns to generate plausible data
  - Its job is to fool the discriminator

Generative Adversarial Networks

> - The generator learns to generate plausible data
>   - Its job is to fool the discriminator
> - The discriminator learns to distinguish the generator's data from real data
>   - Its job is to detect data from the generator

Generative Adversarial Networks

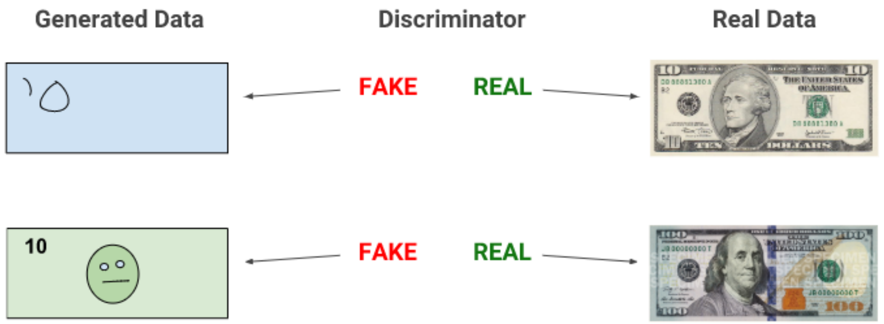
- The generator learns to generate plausible data
  - Its job is to fool the discriminator
- The discriminator learns to distinguish the generator's data from real data
  - Its job is to detect data from the generator



Google

Generative Adversarial Networks

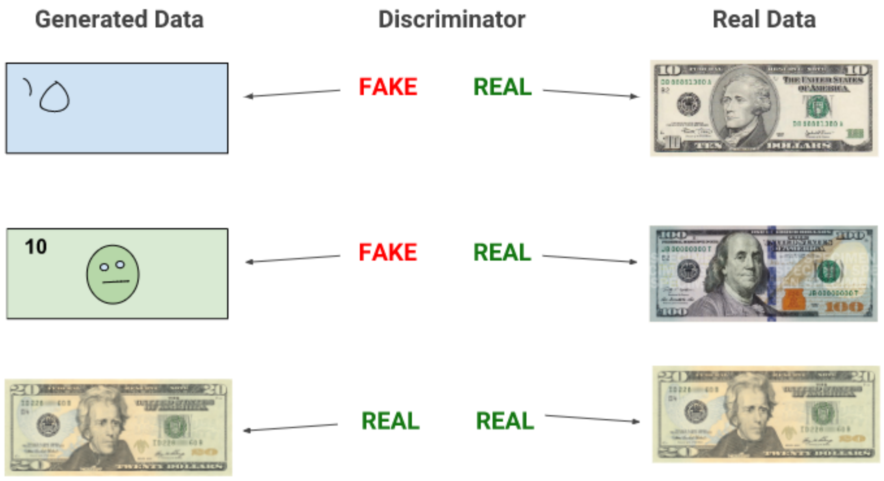
- The generator learns to generate plausible data
  - Its job is to fool the discriminator
- The discriminator learns to distinguish the generator's data from real data
  - Its job is to detect data from the generator



Google

Generative Adversarial Networks

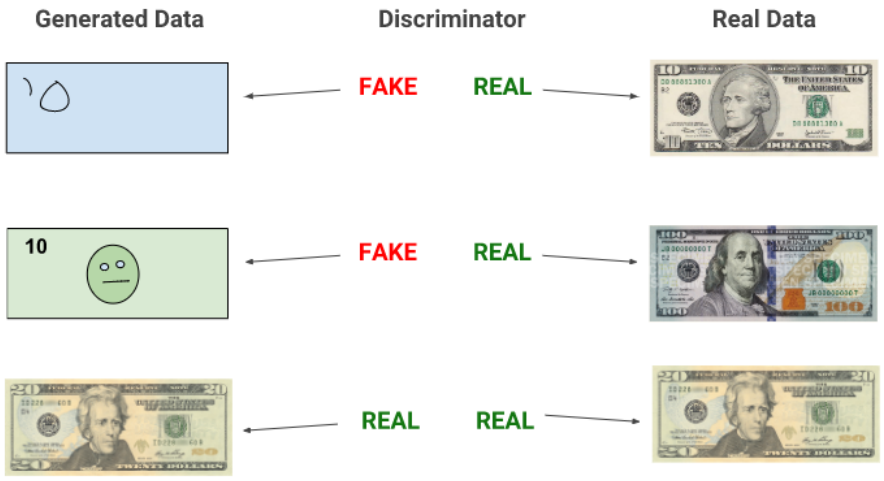
- The generator learns to generate plausible data
  - Its job is to fool the discriminator
- The discriminator learns to distinguish the generator’s data from real data
  - Its job is to detect data from the generator



Google

Generative Adversarial Networks

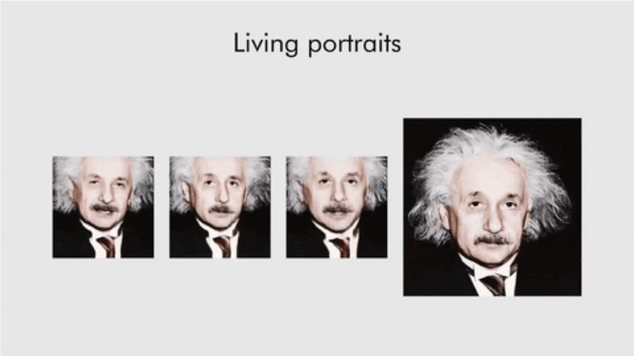
- The generator learns to generate plausible data
  - Its job is to fool the discriminator
- The discriminator learns to distinguish the generator's data from real data
  - Its job is to detect data from the generator



Google

Generative Adversarial Networks

Derpfakes



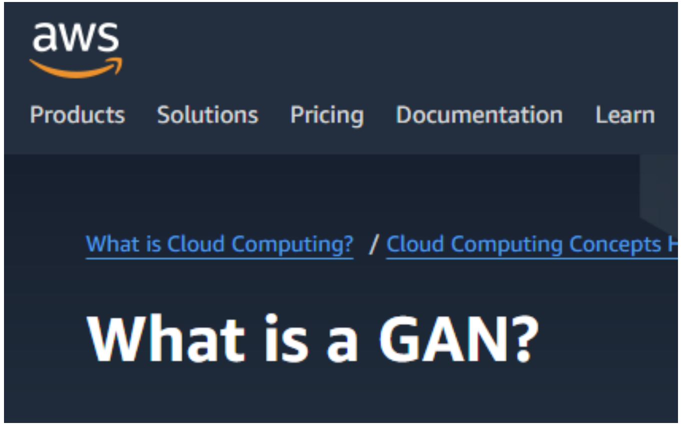
Samsung

Kim et al

Generative Adversarial Networks

Generative Adversarial Networks

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio



Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Transformers

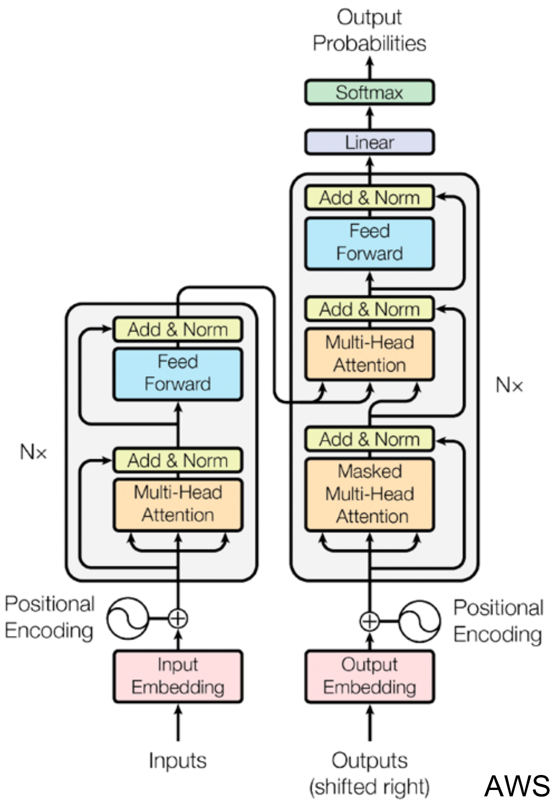
> Basis of much of modern “AI” and LLMs

Transformers

- Basis of much of modern “AI” and LLMs

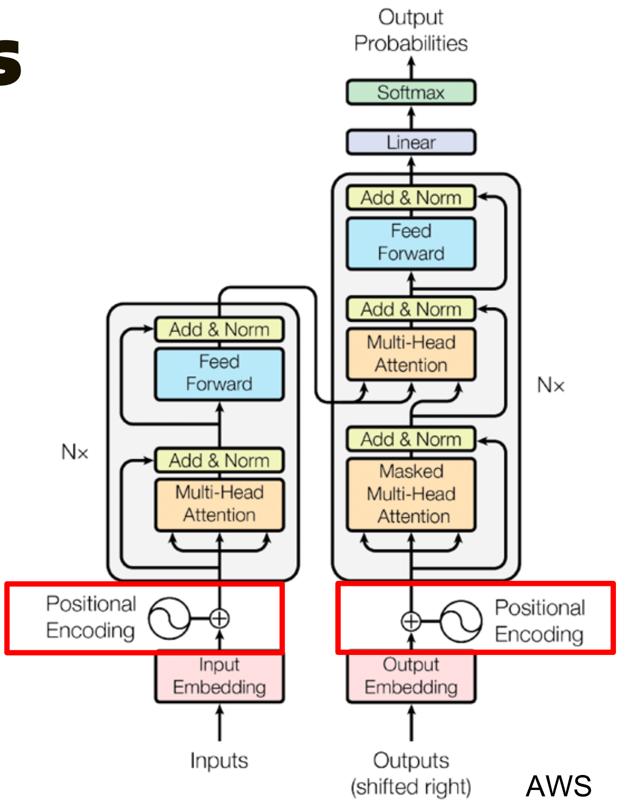
- General idea:
  - A neural network that learns context by tracking relationships in sequential data
  - Traditional AI frameworks often rely on encoder/decoder relationships
  - Attention: allows a model to pay “attention” to different parts of a sequence at once

Transformers



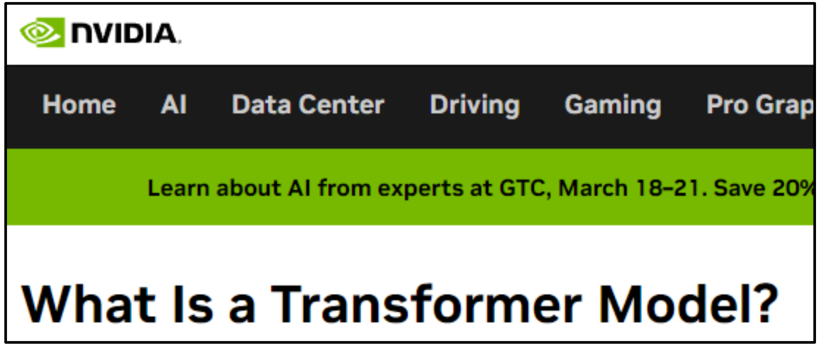
AWS

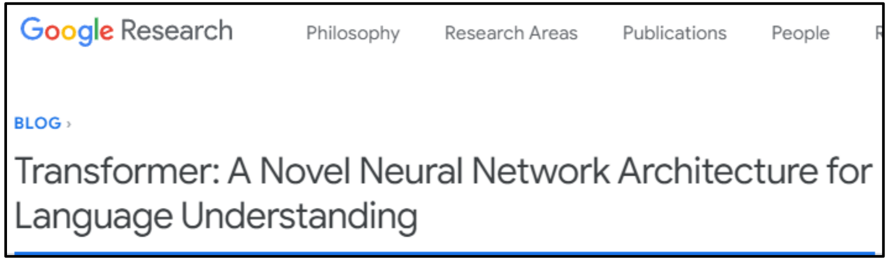
Transformers



AWS

Transformers





Notebook

Outline

> Recap
> Continuing Modeling Exercise
> Break/Discussion
> Neural Networks
> Break
> Neural Network Flavors
> Programming Exercise (time permitting)