AI Uses in Education Research

AERA Virtual Research Learning Series
June 11, 2024

Min Sun, Lovenoor (Lavi) Aulck,
Jing Liu, Shawon Sarkar

Institute of Education Sciences

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

AmplifyLearn.AI

UNIVERSITY OF OREGON

Outline

- Intro
- What is ML? What is AI?
- Use case examples part 1
- Use case examples part 2
- Summary

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

ISEA

- Innovation Science for Education Analytics (ISEA)
  - Funded by IES and the Department of Education
  - www.amplifylearn.ai/isea/

With the rise of artificial intelligence (AI) in all aspects of society, there has been an increasing talent gap in AI and machine learning, especially in applying these tools in education. Through a collaboration between the UW College of Education, the UW eScience Institute, and faculty from other higher education institutions including the University of Oregon, the University of Maryland, and Vanderbilt University, a training program called Innovation Science for Education Analytics (ISEA) will launch in January 2024 thanks to a 3-year grant from the Institute of Education Sciences (IES).

ISEA seeks to advance computational and analytic capacity in K-12 education. The program will recruit a cohort of 15-20 fellows each year with the goal of developing a new pool of talented individuals who have integrated expertise across engineering, statistics, and K-12 education. ISEA is designed as a targeted training program that emphasizes state-of-the-art data analytics informed by education domain knowledge, computational workflows, immediate applications, and career advising.

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

Multiple Paradigms of Regression

- Statistics/economics: explain a relationship
  - Ties in with the idea of inference

Multiple Paradigms of Regression

- Statistics/economics: explain a relationship
  - Ties in with the idea of inference
- Machine learning: predict the future
  - Ties in with the idea of prediction

Example: Earnings

- Inference
  - What are the determinants of income?
  - Do people with children earn more?
  - On average, how much more will a person earn for each additional year of schooling?

Example: Earnings

- Inference
  - What are the determinants of income?
  - Do people with children earn more?
  - On average, how much more will a person earn for each additional year of schooling?

- Prediction
  - What is the predicted income for person X?
  - What are the descriptors of a person with income Y?

Example: Earnings

- Linear regression offers a concise summary of the mean of one variable as a function of the other variable(s) through two parameters: the slope and intercept of the line

Example: Earnings

- Linear regression offers a concise summary of the mean of one variable as a function of the other variable(s) through two parameters: the slope and intercept of the line

- Model:

$$wages_i = \beta_0 + \beta_1 * education_i + \epsilon_i \qquad wages_i = 12409 + 2518 * education_i + \epsilon_i$$

Example: Earnings

- Linear regression offers a concise summary of the mean of one variable as a function of the other variable(s) through two parameters: the slope and intercept of the line

- Model:

$$wages_i = \beta_0 + \beta_1 * education_i + \epsilon_i \qquad wages_i = 12409 + 2518 * education_i + \epsilon_i$$

- Have we measured the causal effect of an additional year of education on wages?

Example: Earnings

- Linear regression offers a concise summary of the mean of one variable as a function of the other variable(s) through two parameters: the slope and intercept of the line

- Model:

$$wages_i = \beta_0 + \beta_1 * education_i + \epsilon_i \qquad wages_i = 12409 + 2518 * education_i + \epsilon_i$$

- Have we measured the causal effect of an additional year of education on wages?
  - With another year of education, will I definitely earn $2,518 more?

The Model Is Right…?

> “All models are wrong but some are useful”
>
> \- George Box

> “… all models are limited by the validity of the assumptions on which they ride”
>
> \- Collier, Sekhon, and Stark

> “Assumptions behind models are rarely articulated, let alone defended.”
>
> \- David Freedman

Models for Predictions

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

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

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?

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

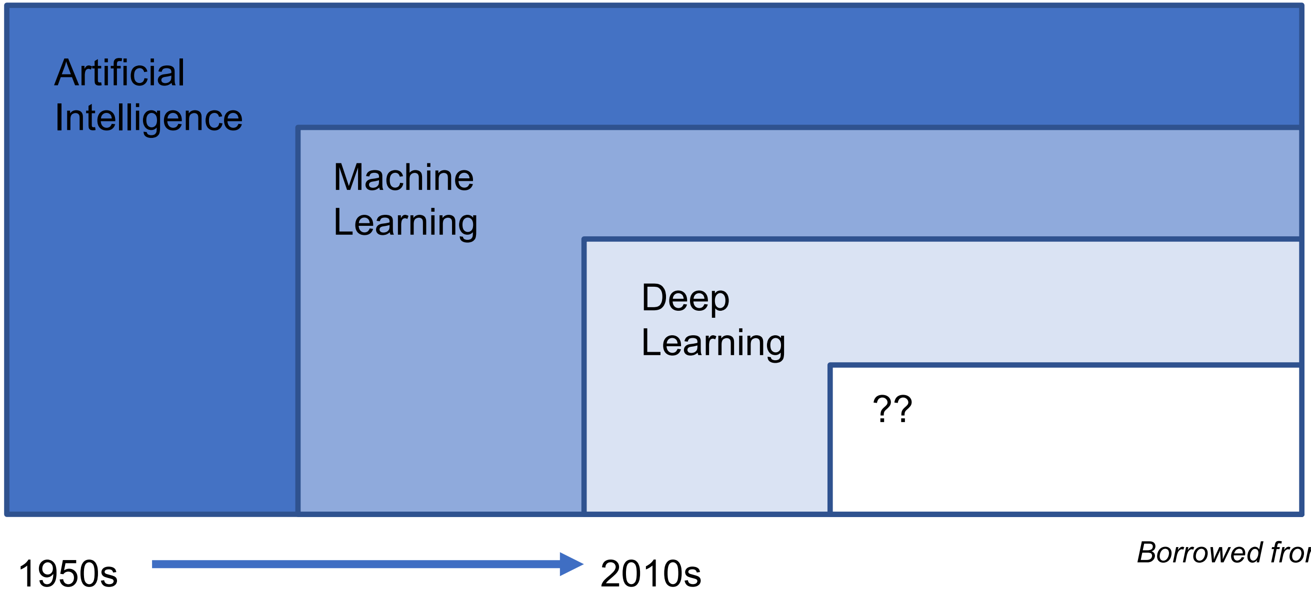
What is Machine Learning?

> Machine learning is the science of getting computers to act without being explicitly programmed

> Machine learning is a scientific discipline that explores the construction and study of algorithms that learn from data

> Machine learning is a natural outgrowth at the intersection of computer science and statistics

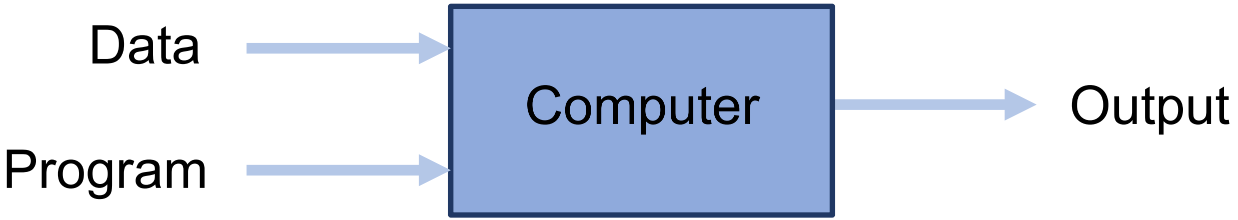
What is Machine Learning?



Borrowed from Nvidia

What is Machine Learning?

Traditional Programming

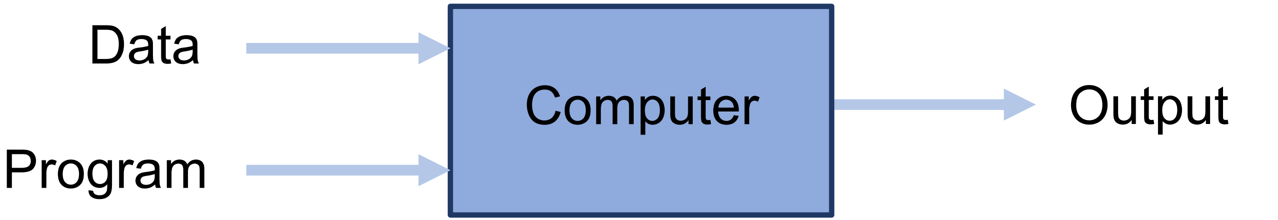


Machine Learning

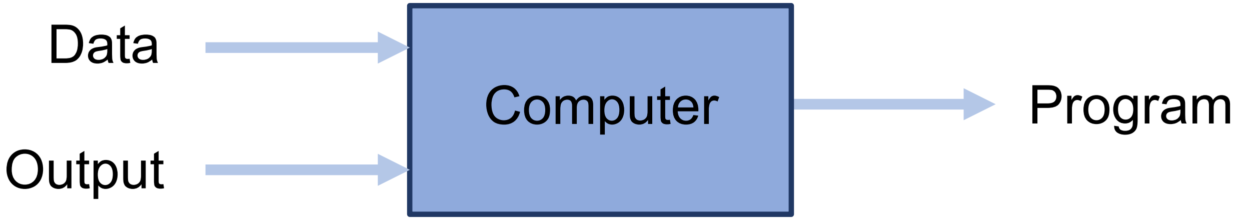
?

What is Machine Learning?

Traditional Programming



Machine Learning



“Flavors” of Machine Learning

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

Supervised Learning

- We know the "correct" answer for values
  - We're trying to model the input (independent) variables as they relate to the output (dependent) variables
- Examples
  - Predicting whether a student will graduate from a university
  - Predicting salary upon graduation
  - Other examples?

Unsupervised Learning

- We don't know the "correct" answer for values
  - We're trying to discover underlying structure
- Examples
  - Extracting topics/themes from textual surveys of students
  - Understanding the relationships between different departments/programs
  - Finding bottlenecks in student course offerings

"Pop" Quiz

Would you address each of the below with a supervised or unsupervised learning algorithm?

- Given email labelled spam or not, build a spam detector
- Given news articles on the web, group them into sets based on topic
- Given a database of customer data, find market segments/customer groups
- Given patients with a disease, determine if new patients have the disease
- Given phone records of individuals, determine which are the wealthiest
- Given phone records of individuals and survey data about their income, predict the incomes of new subscribers

Semi-supervised Learning

- We know the “correct” answer for some values
  - The input data contains both labelled and unlabeled instances
- Examples?

Reinforcement Learning

- We develop “agents” that seek to maximize “reward”
  - We’re trying to build models that will seek to maximize some gain
- Examples
  - Creating personalized curricula for students
  - Creating educational content for students

ML Basics

- Thousands of ML algorithms
  - No one knows them all
  - Often tweaks or small improvements to existing approaches

ML Basics

- Thousands of ML algorithms
  - No one knows them all
  - Often tweaks or small improvements to existing approaches
- Every ML algorithm has three components
  - The representation (the model)
  - The evaluation (the cost/objective function)
  - The optimization (the search)

Designing for Prediction: Key Ideas

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

Determining Model Fit

- Adding more features will generally improve the “fit” of your model

Determining Model Fit

- Adding more features will generally improve the “fit” of your model

- As the number of features we have approaches or exceeds the number of observations, the model can fit really, really well

Determining Model Fit

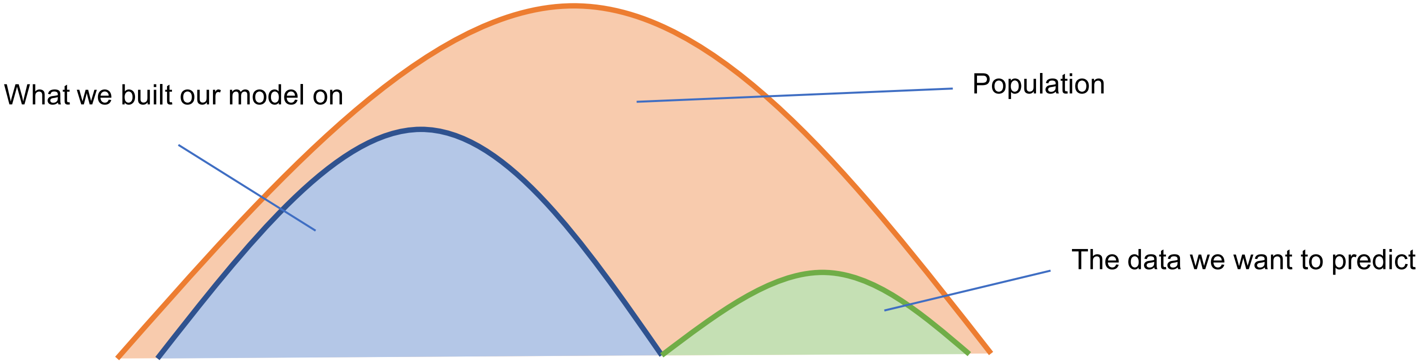
> Adding more features will generally improve the “fit” of your model

> As the number of features we have approaches or exceeds the number of observations, the model can fit really, really well

> Should we keep adding features? Is this what we want?

The ML Dilemma

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



Institute of Education Sciences

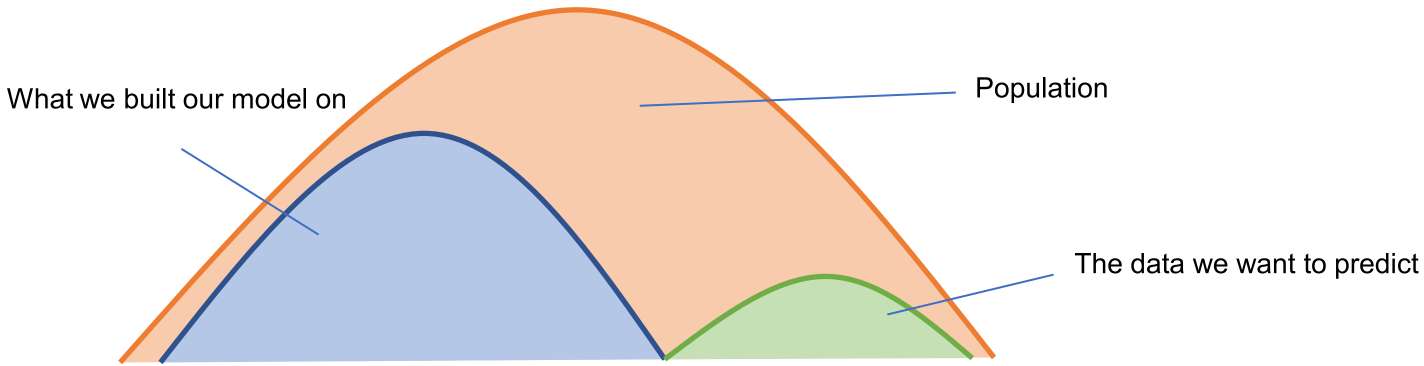
AmplifyLearn.AI

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

UNIVERSITY OF OREGON

The ML Dilemma

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

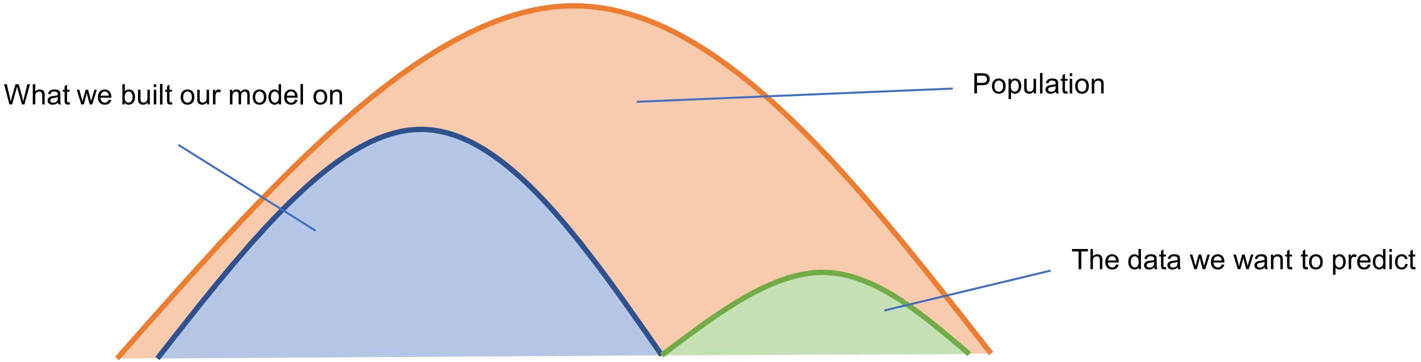


The ML Dilemma

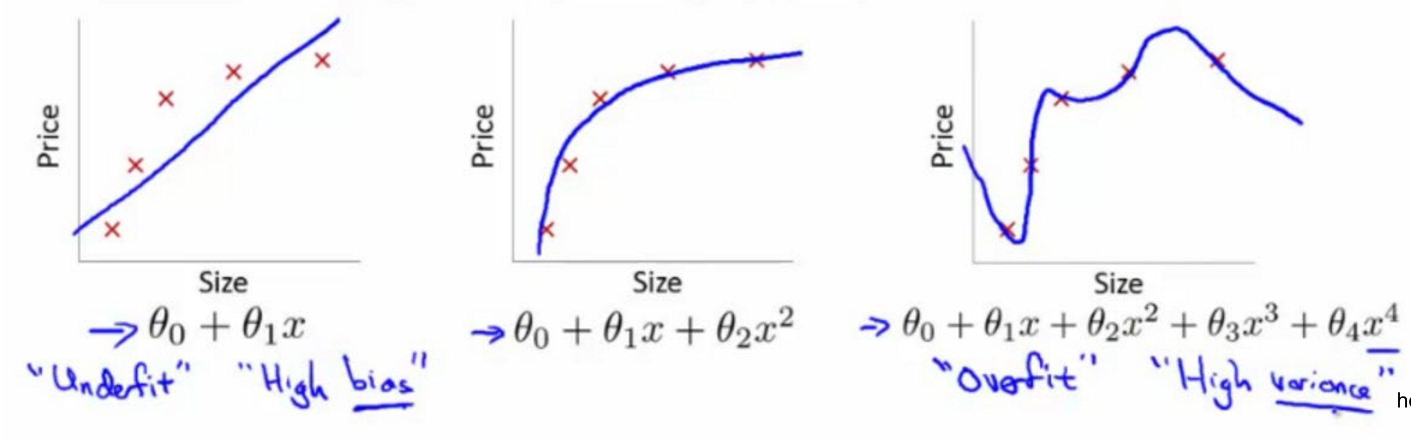
variance

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

bias



Bias - Variance Tradeoff



holehouse.org

Bias-Variance Tradeoff: the problem of simultaneously minimizing two sources of error that prevent generalization

Overfitting and Underfitting

- More formally: <u>overfitting</u> is when we have too high of variance in our model with respect to bias

- More formally: <u>underfitting</u> is when we have too high of bias in our model with respect to variance

Overfitting and Underfitting

- More formally: overfitting is when we have too high of variance in our model with respect to bias

- More formally: underfitting is when we have too high of bias in our model with respect to variance

- How do we know?

Overfitting and Underfitting

- More formally: overfitting is when we have too high of variance in our model with respect to bias

- More formally: underfitting is when we have too high of bias in our model with respect to variance

- How do we know? Validation!

Training, Validation, and Test Data

- **Training set**: a set of examples used for learning, to which we fit parameters of a model

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

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)

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!!

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!!

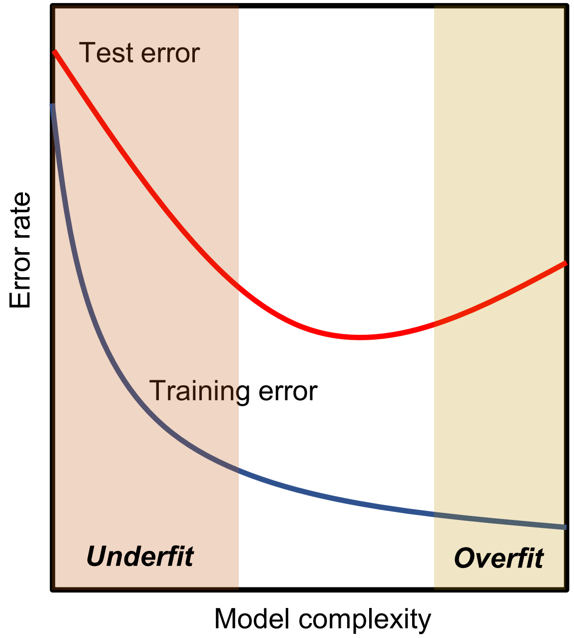
Test Data

Until you're ready to report final results…

Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!
Do not peek at your test data!! Do not peek at your test data!!

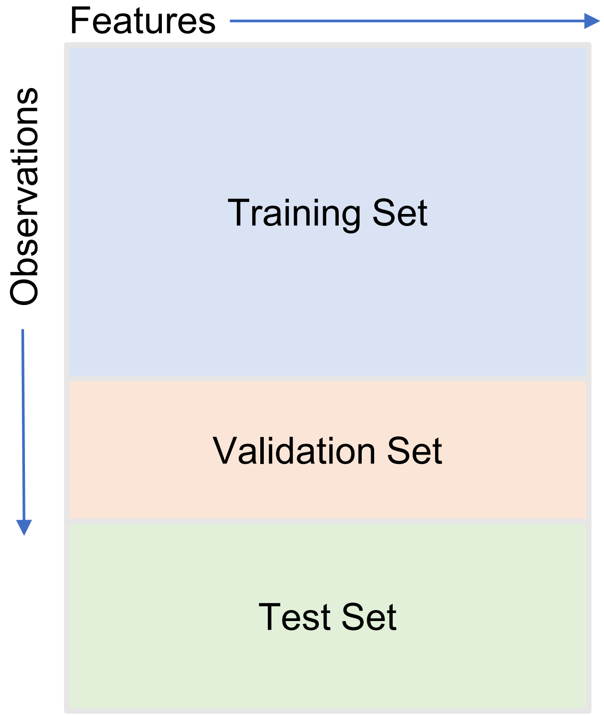
How Prediction Experiments Work

- ML experiments typically separate data into a training/test set
- Model is fit on training set
  - Validation set is often used to fine tune
- Performance is measured on the test set
  - This gives something of a “real-world” approximation of how well the model performs



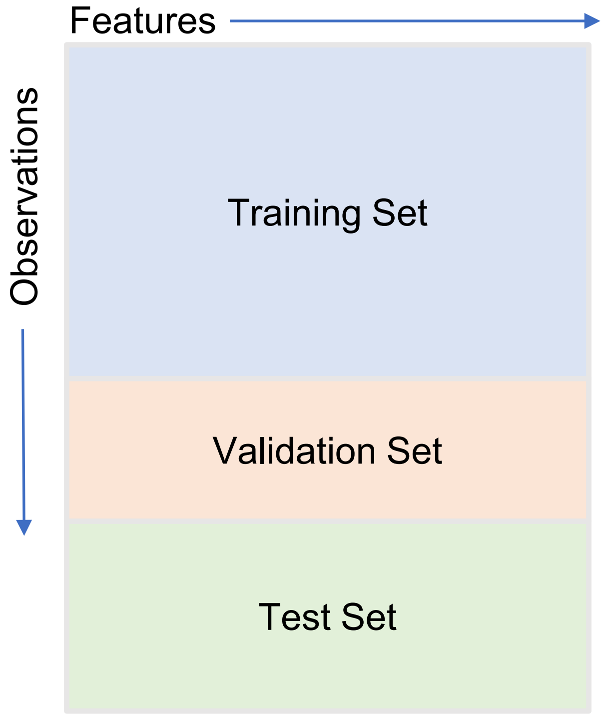
A Typical (Supervised) ML Experiment

> Data with labelled instances
  - Split into training, validation, and test sets



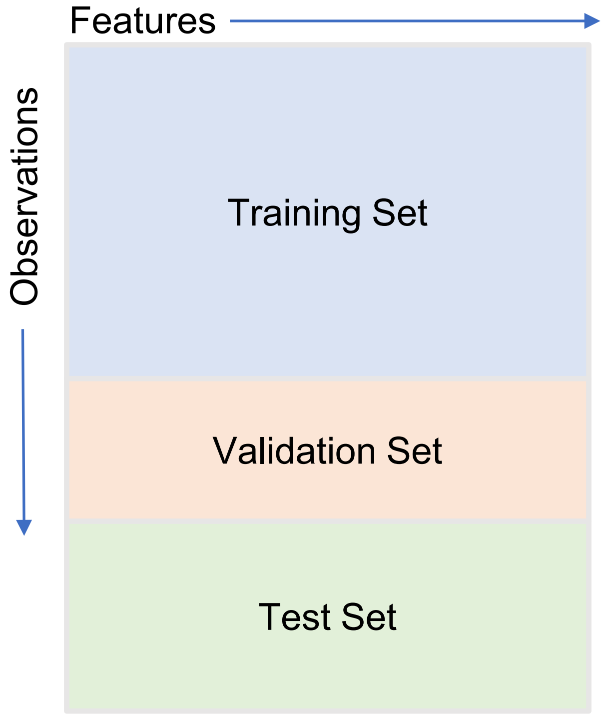
A Typical (Supervised) ML Experiment

- Data with labelled instances
  - Split into training, validation, and test sets
- Training your model
  - Estimate parameters on training set
  - Tune hyperparameters on the validation set
  - Report results on the test set



A Typical (Supervised) ML Experiment

- Data with labelled instances
  - Split into training, validation, and test sets
- Training your model
  - Estimate parameters on training set
  - Tune hyperparameters on the validation set
  - Report results on the test set
- Evaluation
  - Many metrics and context-dependent
  - Ideally, what we use to train should be our final evaluation criteria



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

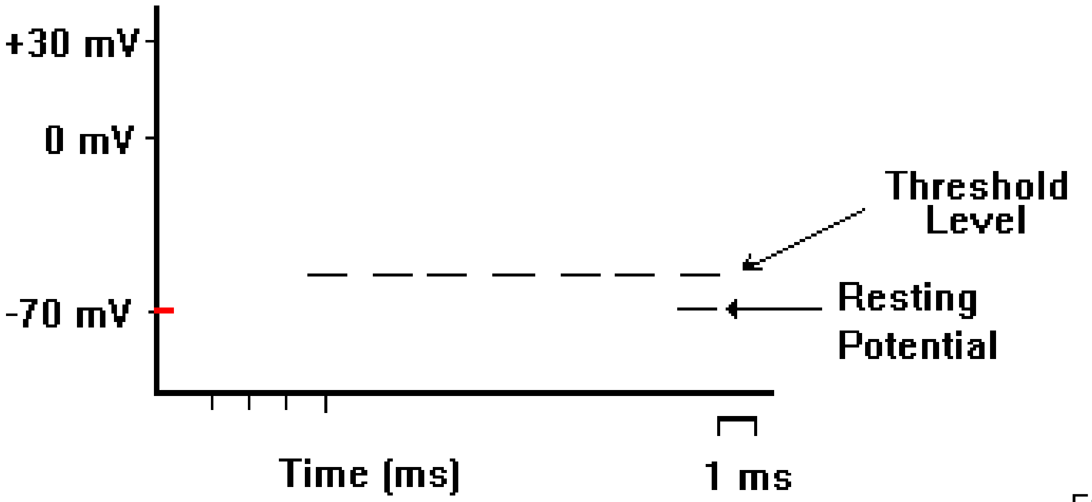
Neurons

- 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

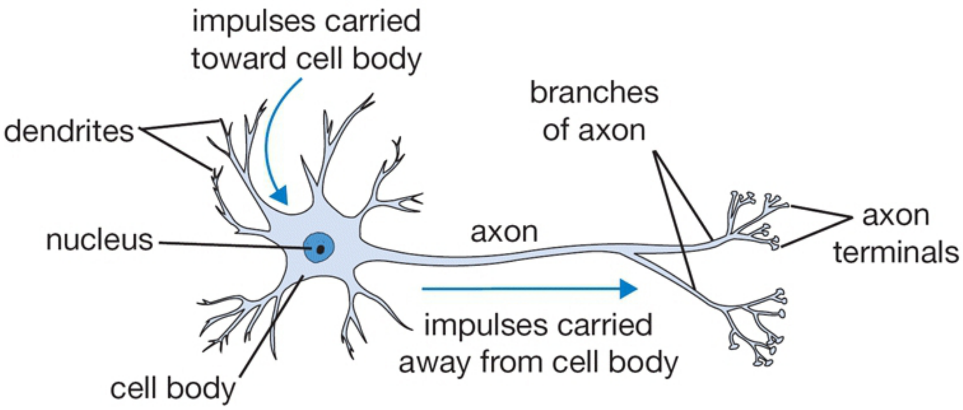
- Rosenblatt's (a psychologist's) attempt at trying to mathematically model a neuron

- 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!

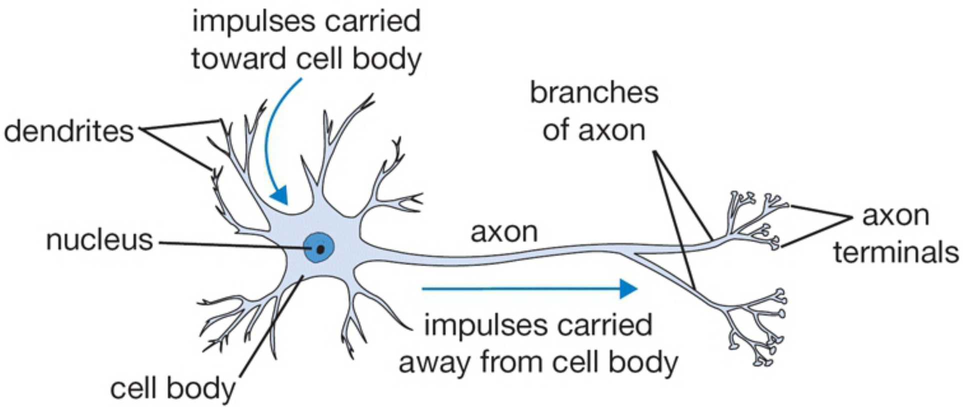
The Perceptron

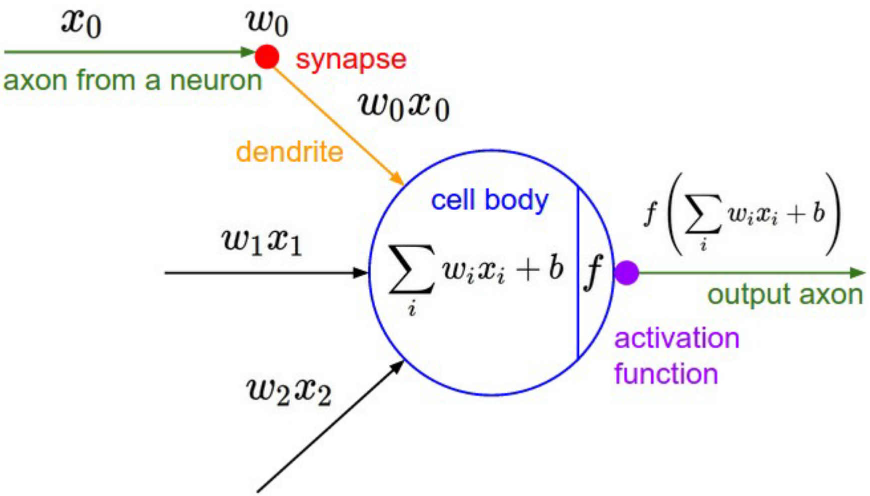
- Rosenblatt's (a psychologist's) attempt at trying to mathematically model a neuron
- The perceptron takes in a set of inputs, multiplies them by a weight, and then aggregates
  - Keep this process in mind! It is central to how neural networks work!
- Perceptron "fires" if the aggregate reaches some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

The Perceptron



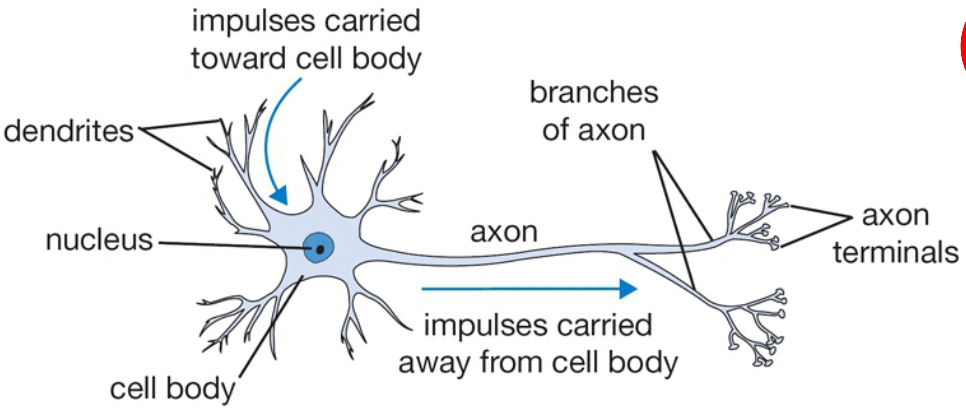
The Perceptron

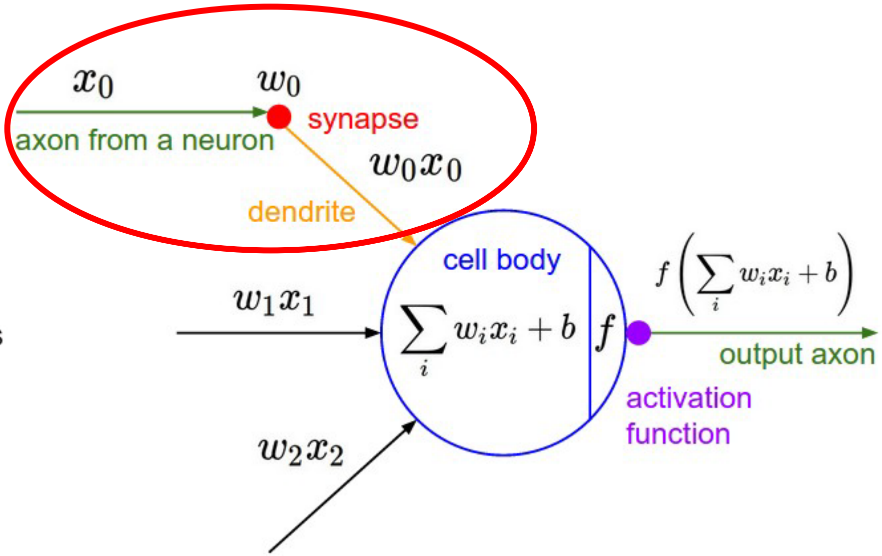




Stanford CS231n

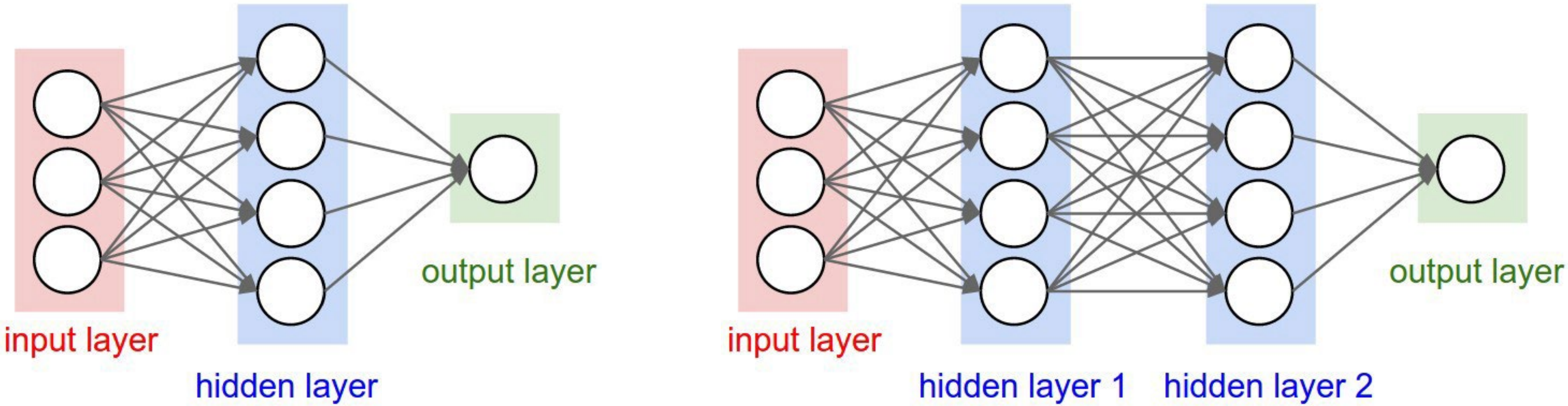
The Perceptron





Stanford CS231n

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 (ŷ)
  - Basically “make predictions”

Training Multi-Layer Networks

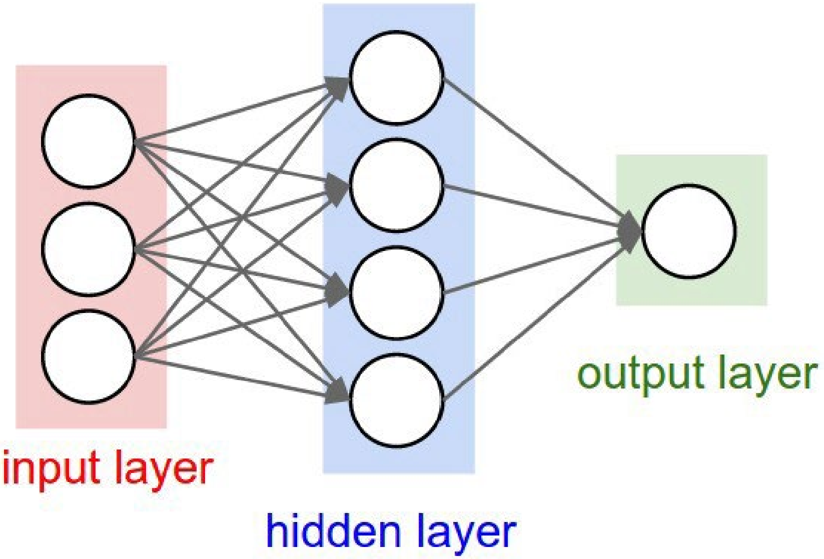
- Forward propagation
  - Pass the input data through the network and calculate outputs (ŷ)
  - Basically “make predictions”

- Calculate error
  - Determine how far off from y we are with ŷ

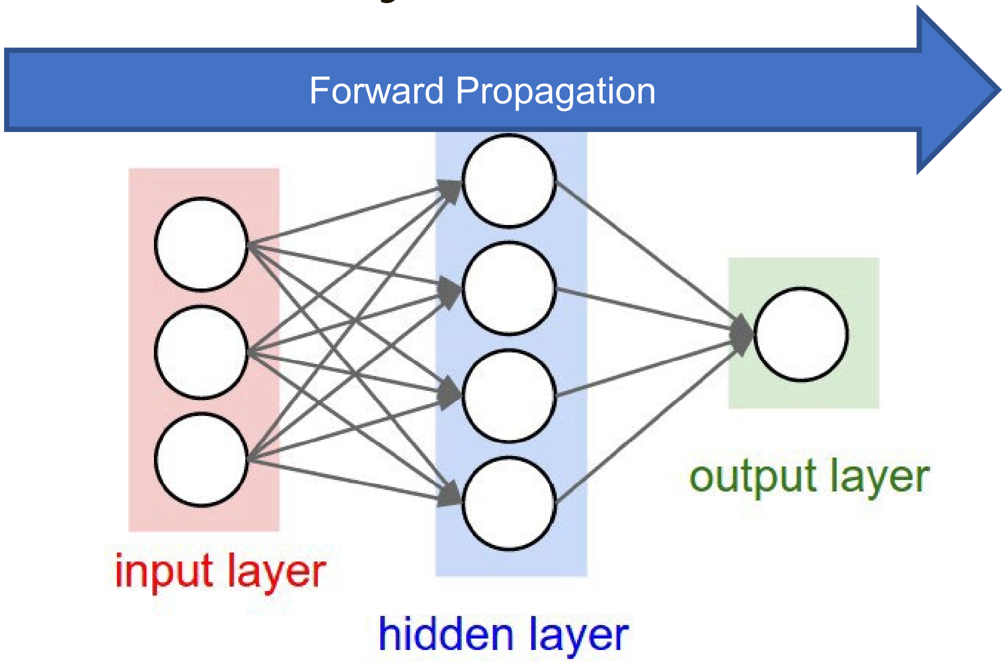
Training Multi-Layer Networks

- 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

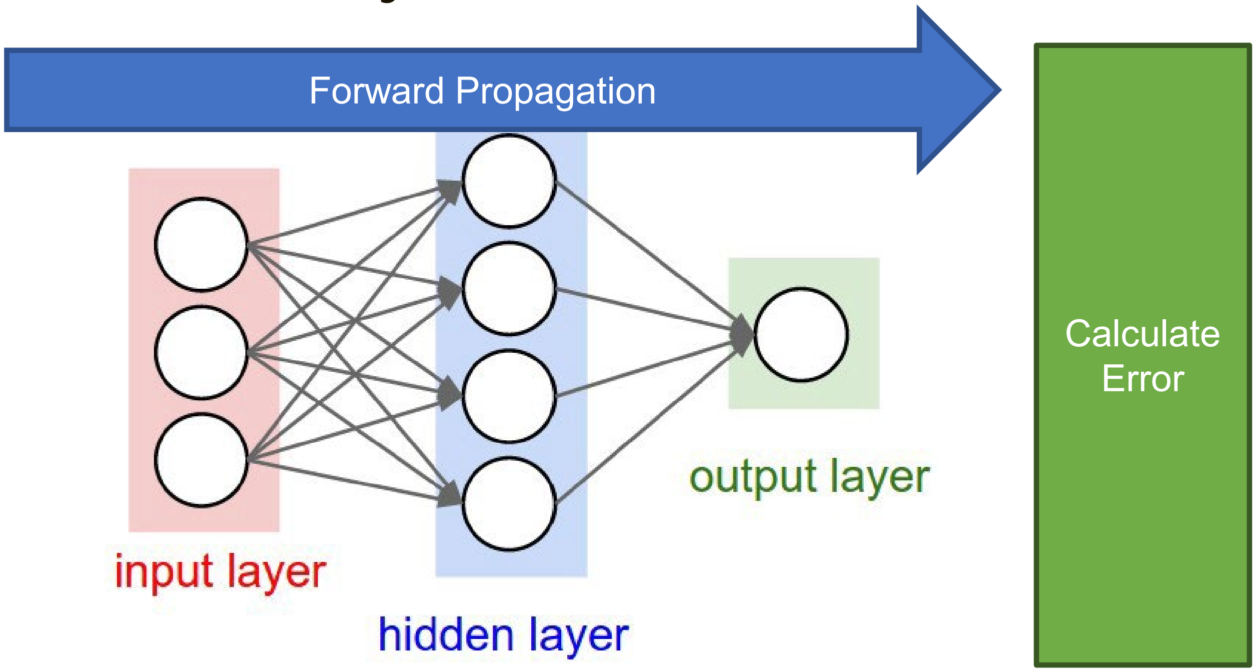


Training Multi-Layer Networks

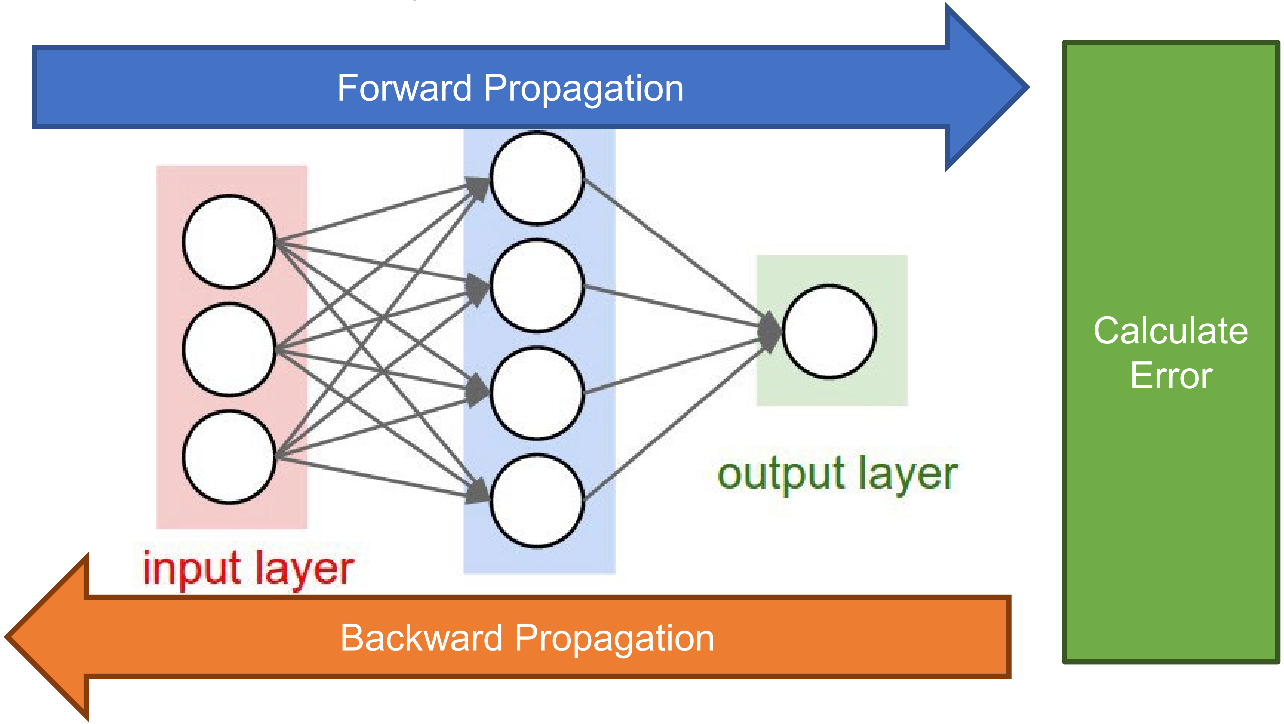


Institute of Education Sciences · AmplifyLearn.AI · University of Washington eScience Institute · University of Oregon

Training Multi-Layer Networks



Training Multi-Layer Networks



Transformers

- Basis of much of modern “AI” and LLMs

Transformers

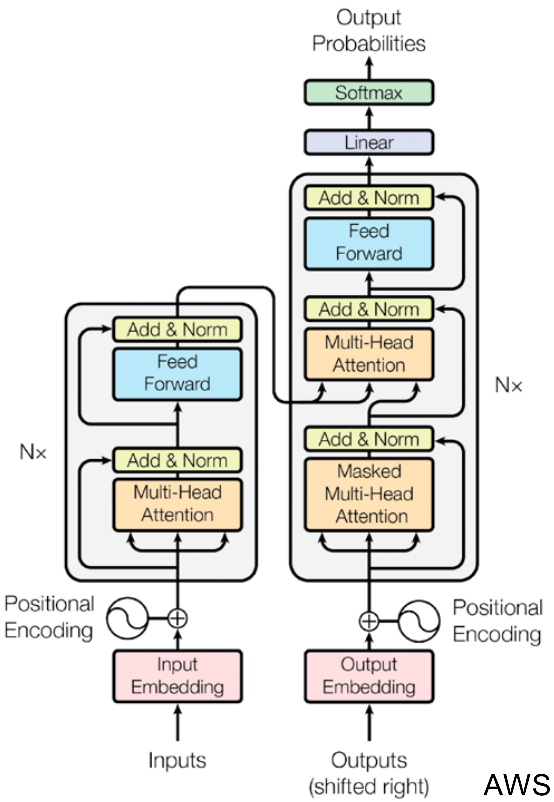
- Basis of much of modern “AI” and LLMs

- 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

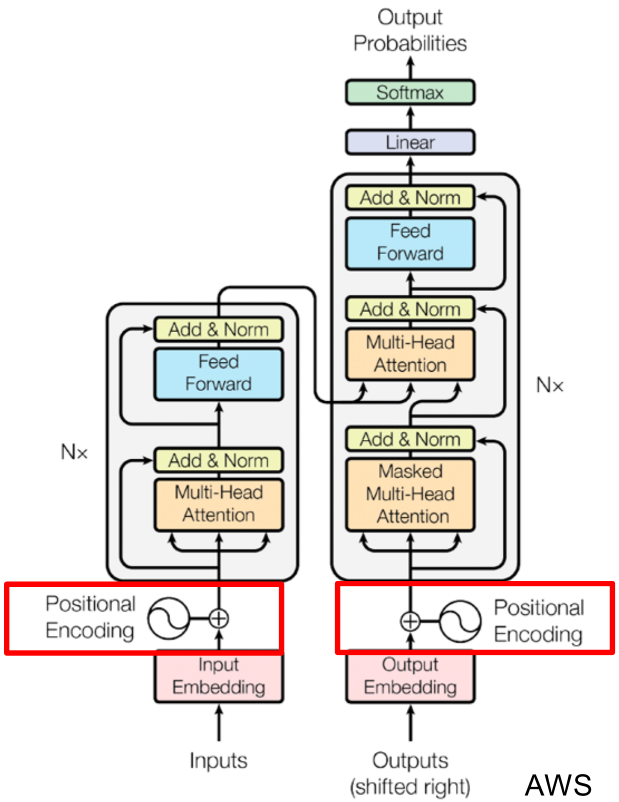
- Basis of much of modern “AI” and LLMs
- General idea:
  - A neural network that learns context by tracking relationships in sequential data
  - Traditional AI frameworks often rely on encoder/decoder relationships
  - Attention: allows a model to pay “attention” to different parts of a sequence at once
- What most of the examples will leverage
  - Keep in mind the basics of ML and deep learning!

Transformers



AWS

Transformers



AWS

Institute of Education Sciences · AmplifyLearn.AI · University of Washington eScience Institute · University of Oregon

Transformers

NVIDIA

Home AI Data Center Driving Gaming Pro Grap

Learn about AI from experts at GTC, March 18–21. Save 20%

What Is a Transformer Model?

Google Research Philosophy Research Areas Publications People

BLOG ›

Transformer: A Novel Neural Network Architecture for Language Understanding

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