IS883: Synthesizing Digital Efforts

Mohannad Elhamod

BOSTON UNIVERSITY

Boston University Questrom School of Business

Welcome!

What is this class about?

BOSTON UNIVERSITY

ChatGPT... Absolute Terror!

The Washington Post
Democracy Dies in Darkness

TECH Artificial Intelligence Help Desk Internet Culture Space Tech Policy

INNOVATIONS

ChatGPT took their jobs. Now they w dogs and fix air conditioners.

Technology used to automate dirty and repetitive jobs. Now, artificial intelligence chatbots are coming aft ones.

By Pranshu Verma and Gerrit De Vynck
June 2, 2023 at 6:00 a.m. EDT

Forbes

FORBES > LEADERSHIP > EDUCATION

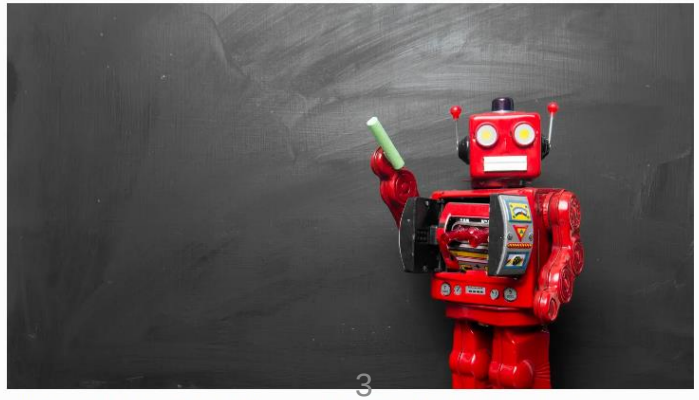
ChatGPT Will Not Replace Teachers But Could Act As A Teaching Assistant

Ray Ravaglia Contributor
I write about education at the intersection of tech and innovation.

Follow

0 Mar 31, 2023, 01:15pm EDT

Listen to article 4 minutes



While Not Replacing Teachers Anytime Soon, GPTs Can Still Be Useful in the Classroom GETTY

The Guardian

Sport Culture Lifestyle More

tics Business Tech Science Newsletters Fight for democracy

'ChatGPT said I did not exist': how rtists and writers are fighting back gainst AI

Forbes

> INNOVATION > ENTERPRISE TECH

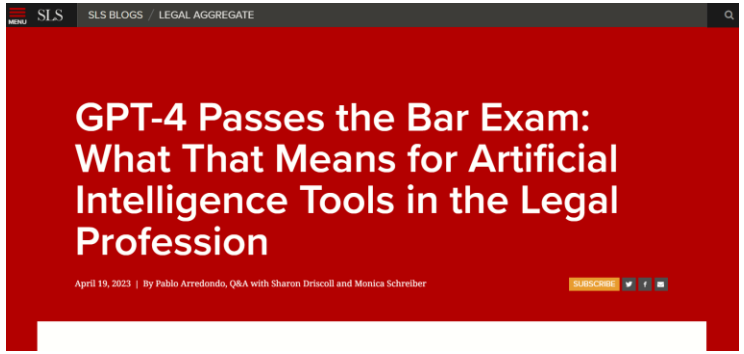
ill ChatGPT Put Data Analysts ıt Of Work?

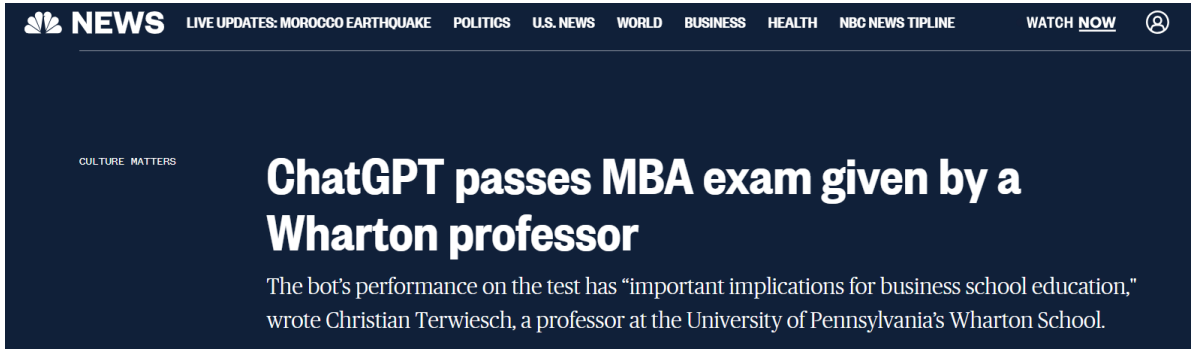
rd Marr Contributor

Follow

3 Feb 7, 2023, 01:40am EST

ChatGPT can do anything!



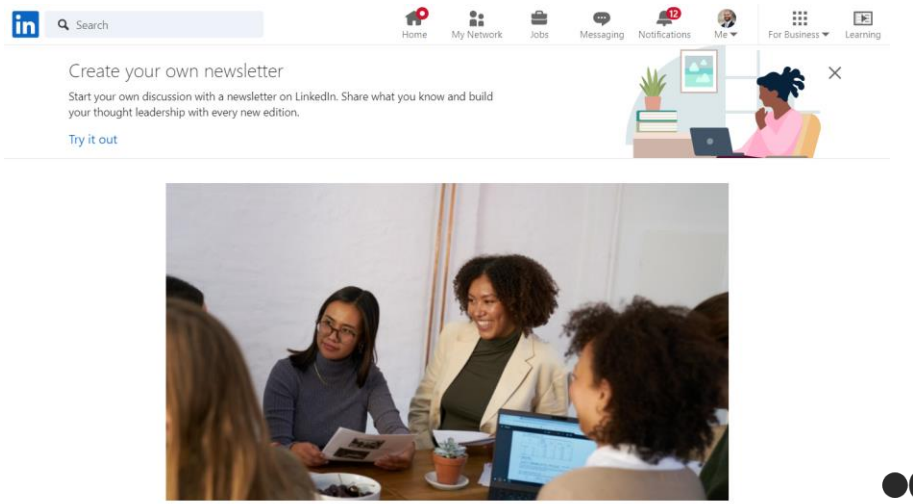


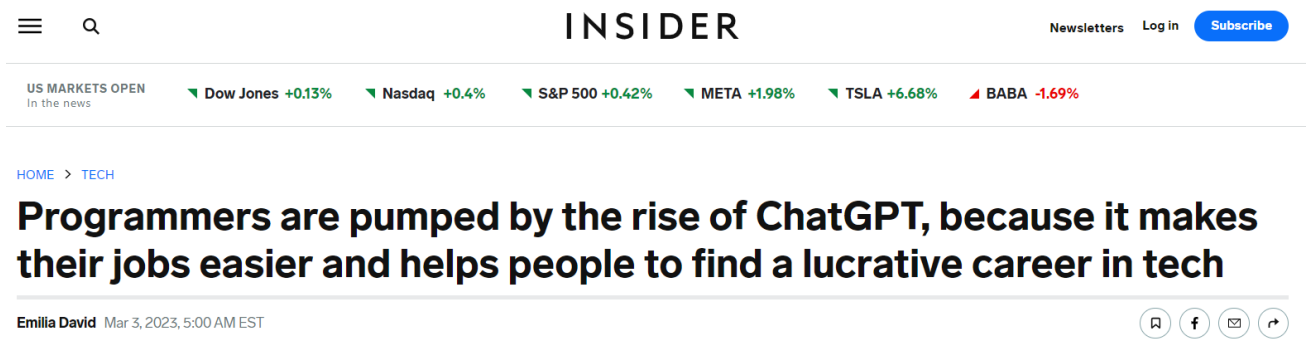
The New York Times

When Doctors Use a Chatbot to Improve Their Bedside Manner

Despite the drawbacks of turning to artificial intelligence in medicine, some physicians find that ChatGPT improves their ability to communicate empathetically with patients.

ChatGPT is amazing!



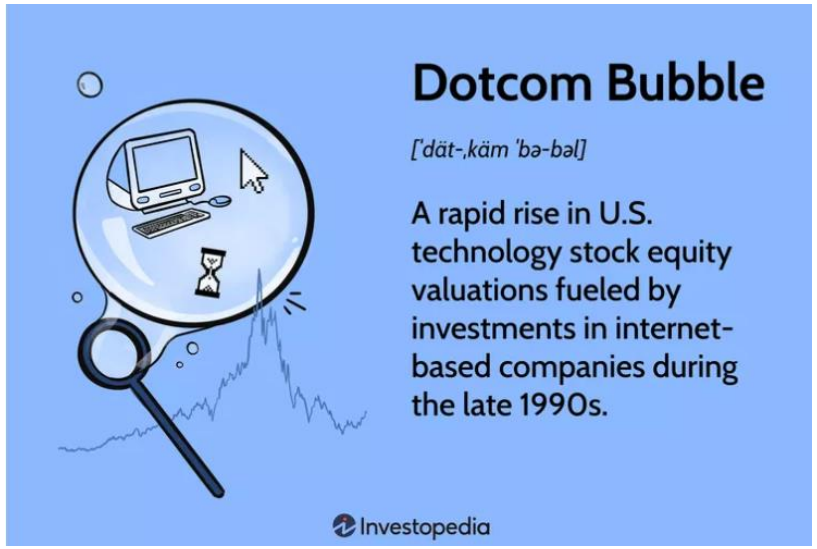


ChatGPT is horrible!

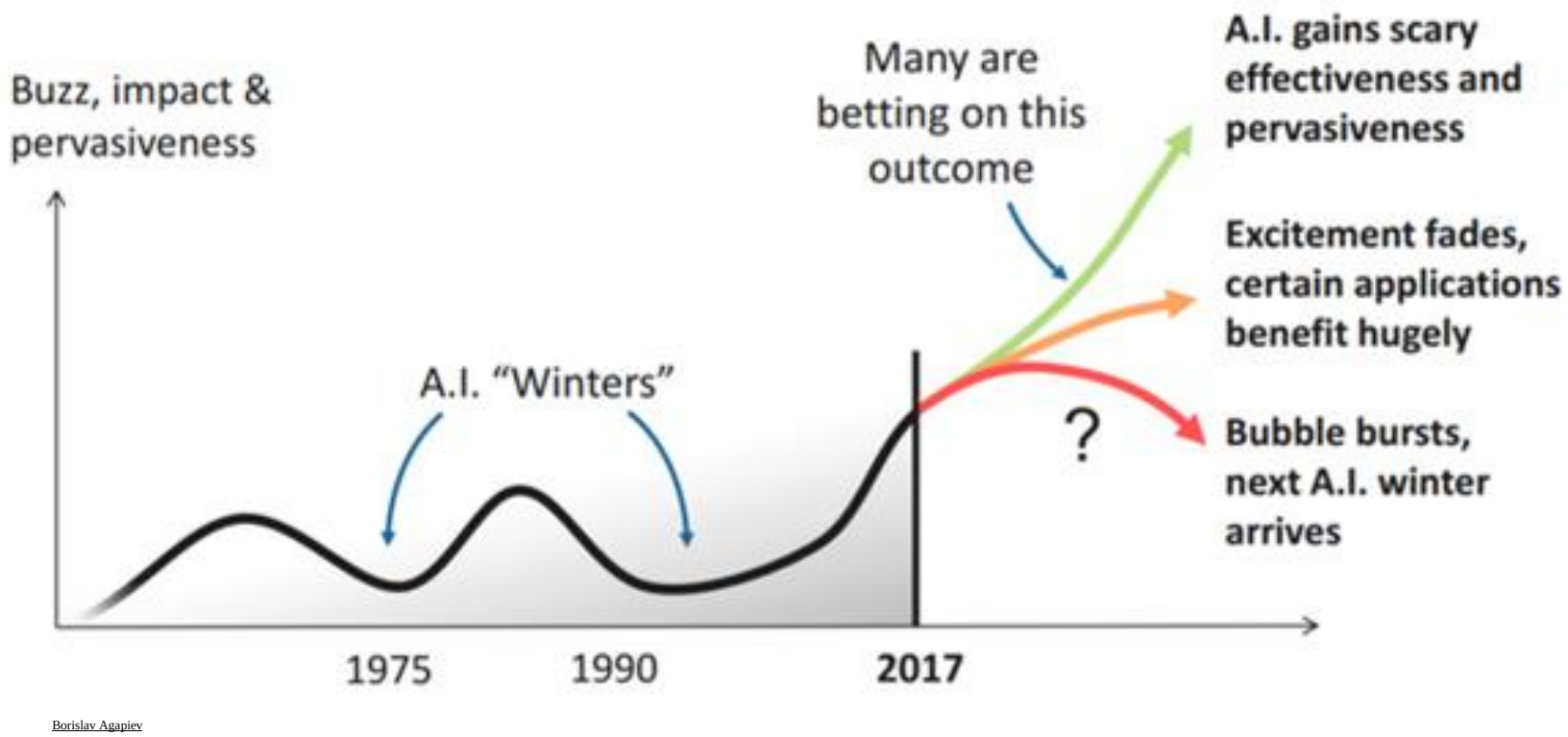
ChatGPT.. What is it really about though?

Let’s play a bit!

Is this a hype?

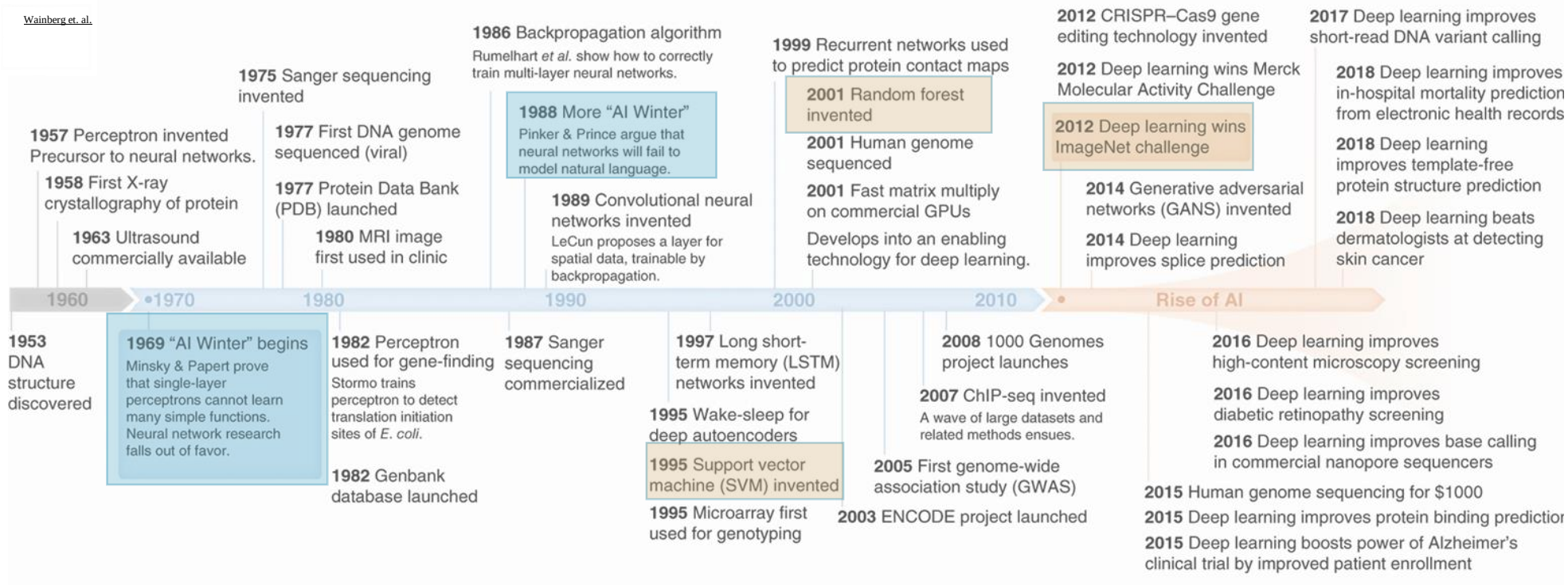


Is this a hype?



Borislav Agapiev

Is this a hype?



General Guidelines

BOSTON UNIVERSITY

Why are you here?

- I am sure it eventually leads to money….
- Maybe it is just a required class?

- Nonetheless, understanding how things work is essential for decision making and innovation.

> Someone who had begun to [study] geometry asked Euclid, 'What shall I get by learning these things?' Euclid called his slave and said, 'Give him [some money], since he must make gain out of what he learns'.
>
> (Heath, 1981, loc. 8625)

Euclid

Fundamentals are important!

- The more foundational knowledge you skip, the more fundamental errors you will make.
- Work hard. **Be patient!**

Zen Speaks: Shouts of Nothingness

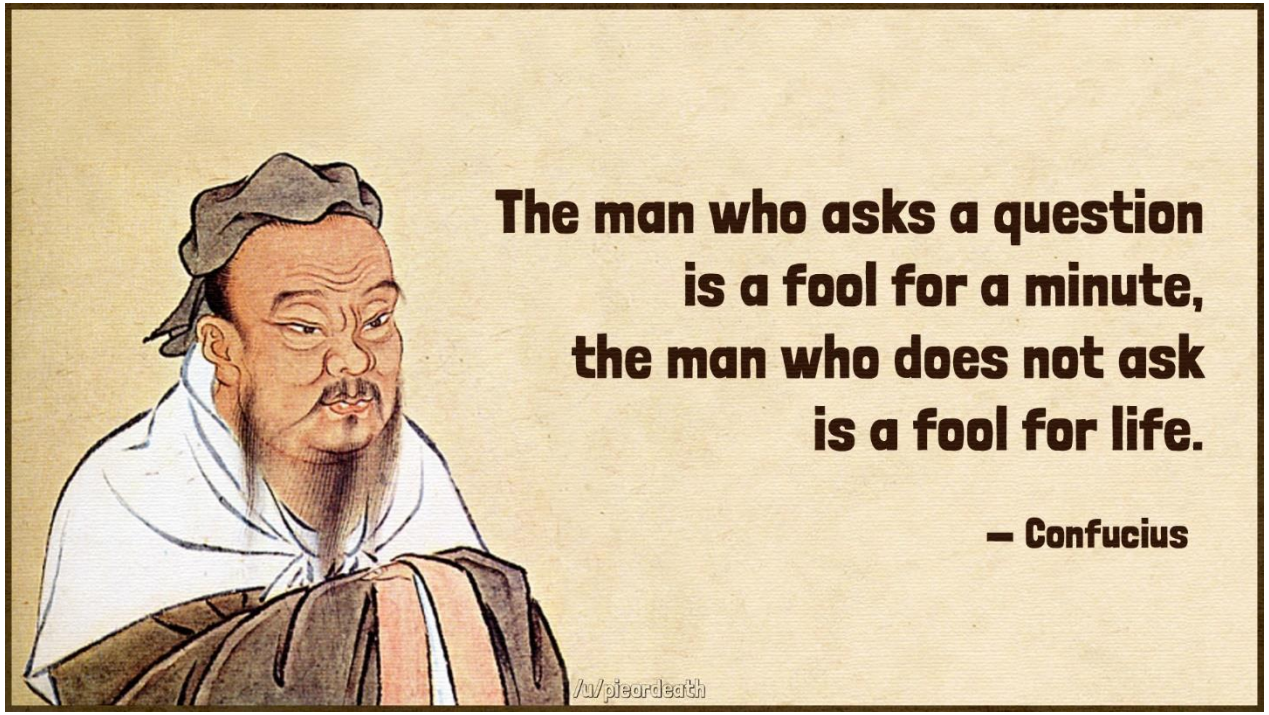
Your professor is not a God

- Machine learning is a fast-growing domain.
- The internet nowadays has all kinds of learning material.

- Your professor is **NOT** here as a walking encyclopedia. He is here to guide your learning experience and build you a solid foundation, so you could continue learning on your own later.

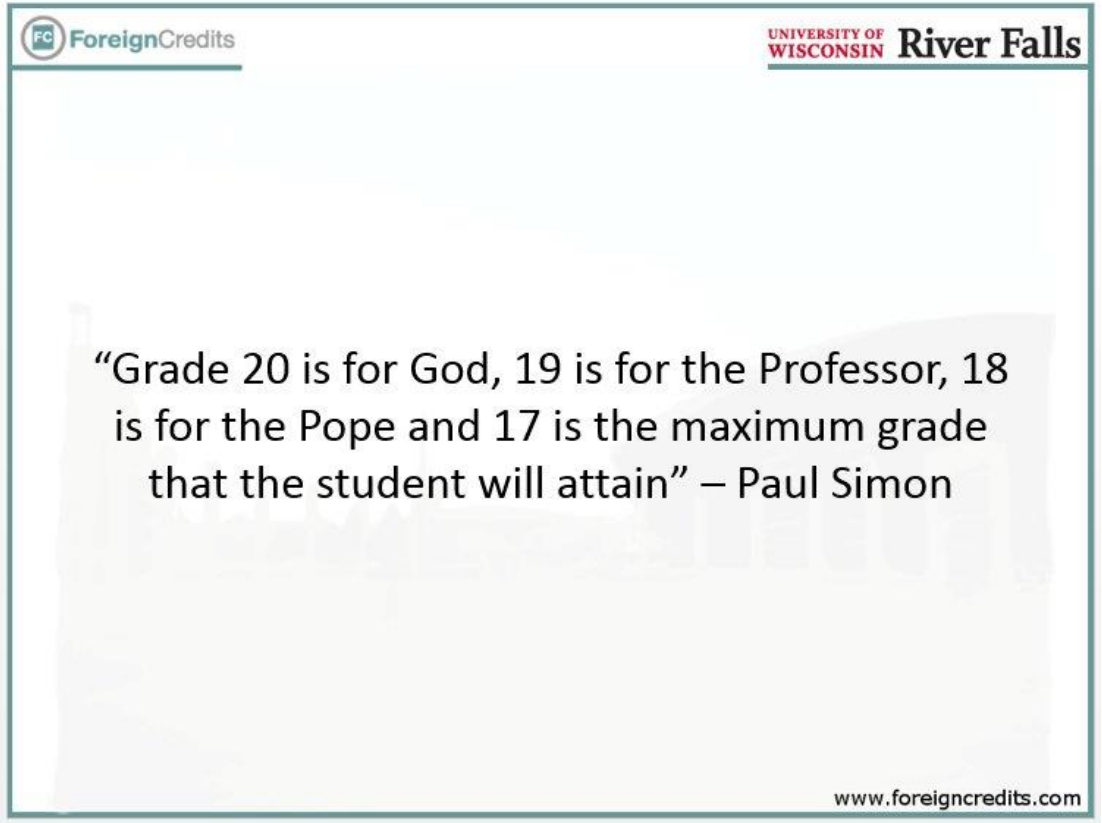
No Question is Foolish

- Other might have the same question.
- Knowledge is hierarchical.



Can I get an A?

- Questrom policy
- But you will get your fair chance.



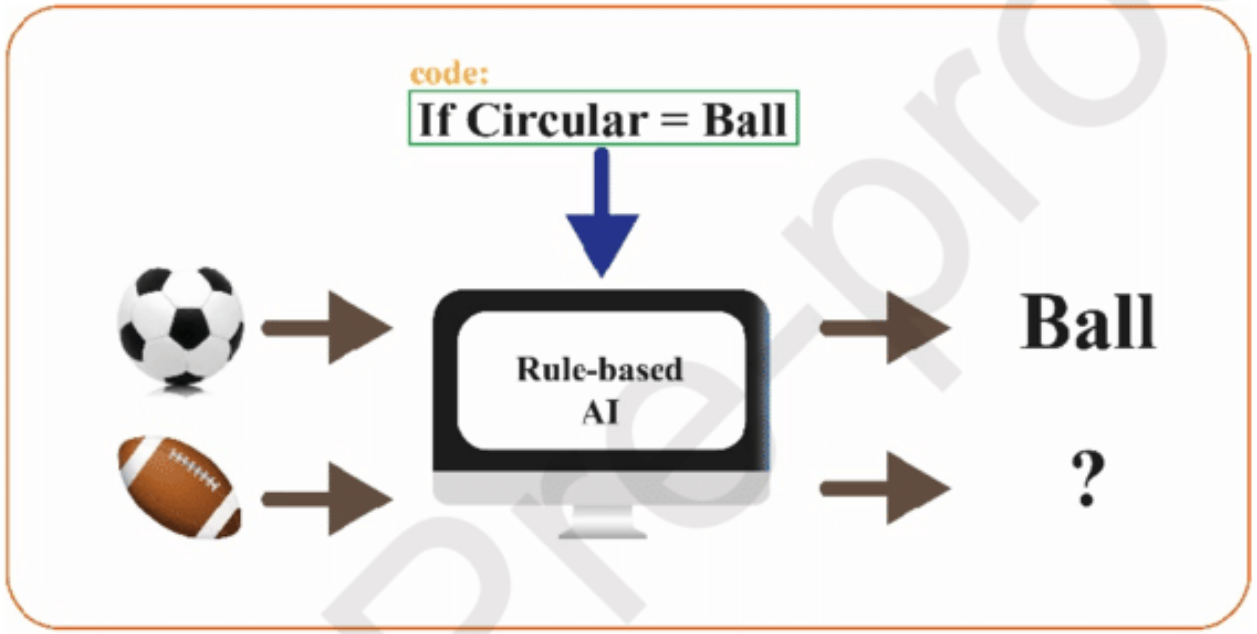
Intro to ML

BOSTON UNIVERSITY

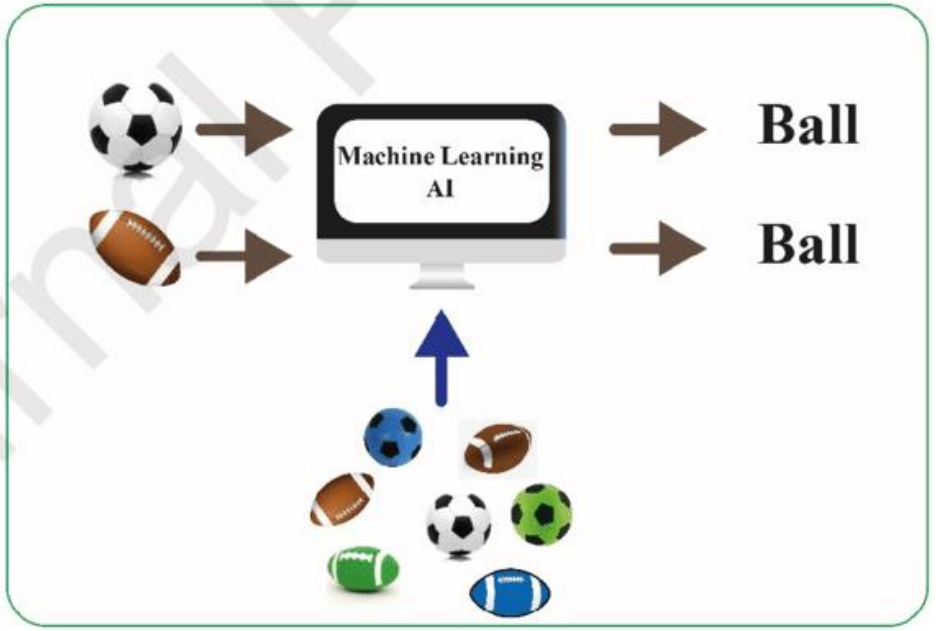
Boston University Questrom School of Business

What is Machine Learning?

Elbadawi, Moe & McCoubrey, Laura & Gavins, Francesca & Ong, Jun Jie & Goyanes, Alvaro & Gaisford, Simon & Basit, Abdul. (2021). Harnessing Artificial Intelligence for the Next Generation of 3D Printed Medicines. Advanced Drug Delivery Reviews. 175. 10.1016/j.addr.2021.05.015.



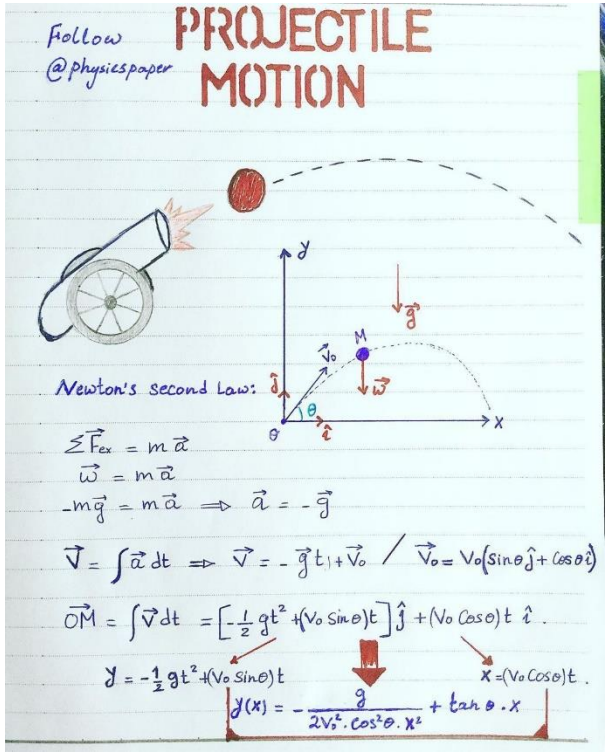
Knowledge-based modeling

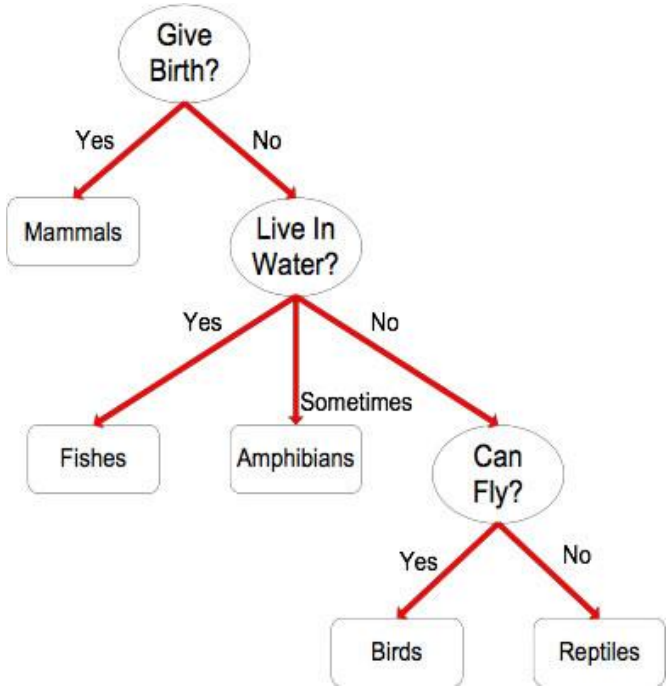


Data-driven modeling

Knowledge-based Modeling

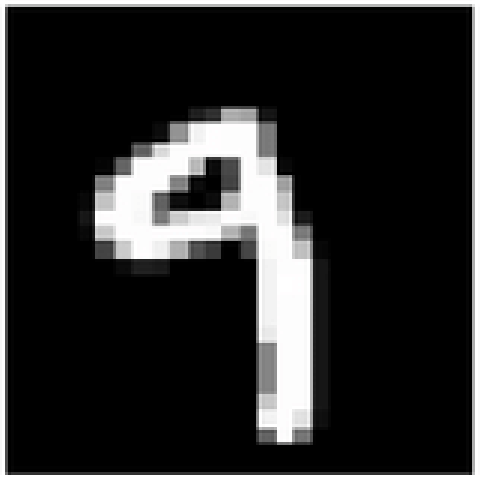
- What challenges you find with this approach?



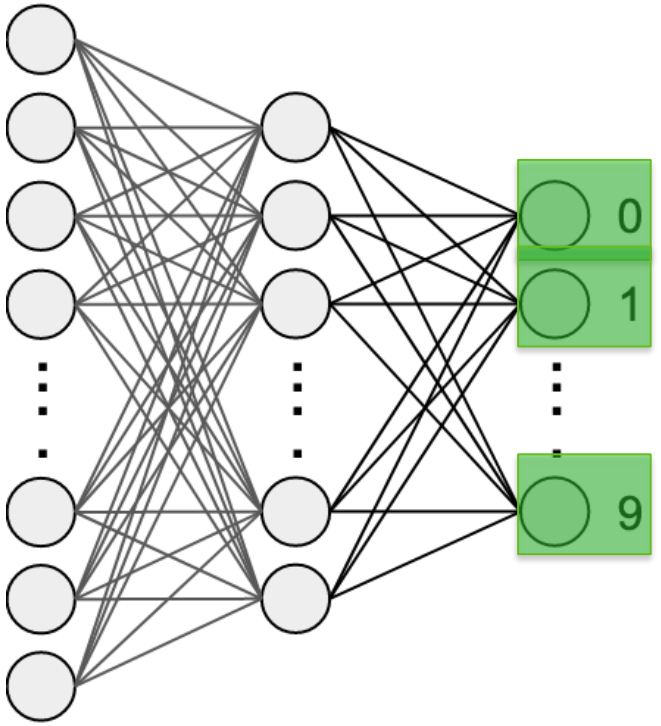


http://jcsites.juniata.edu/faculty/rhodes/ml/rulebasedClass.htm
https://www.pinterest.com/pin/598626975454754006/

Data-driven Modeling

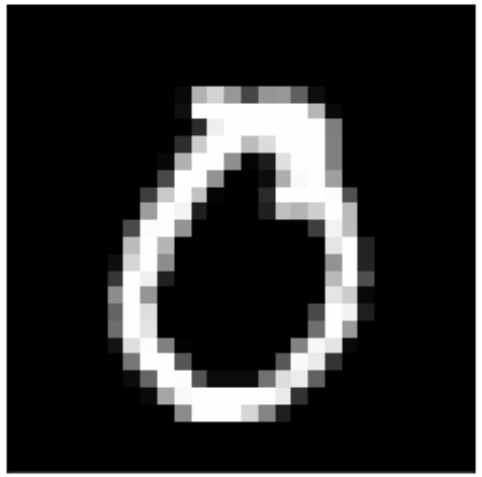


MNIST Dataset

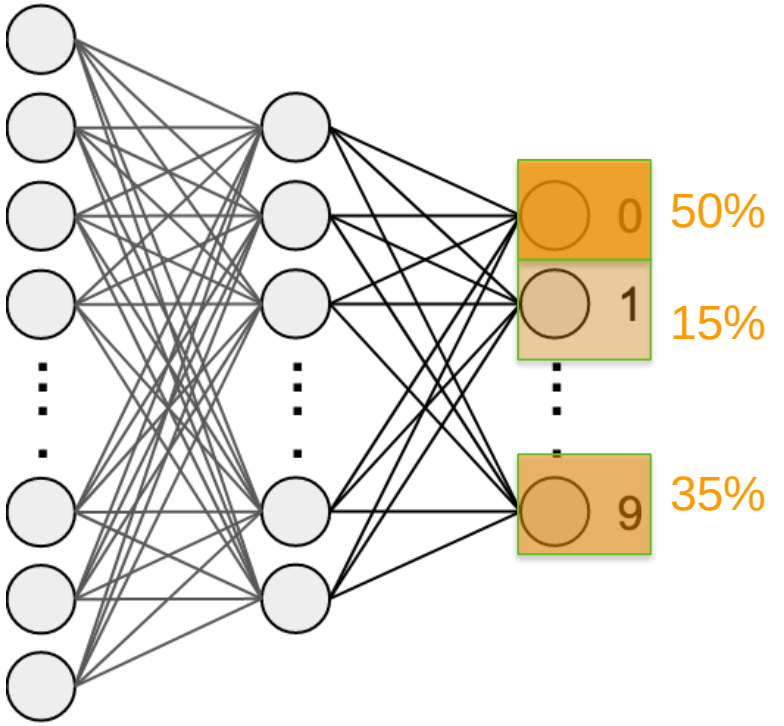


Phase 1: Training

Data-driven Modeling



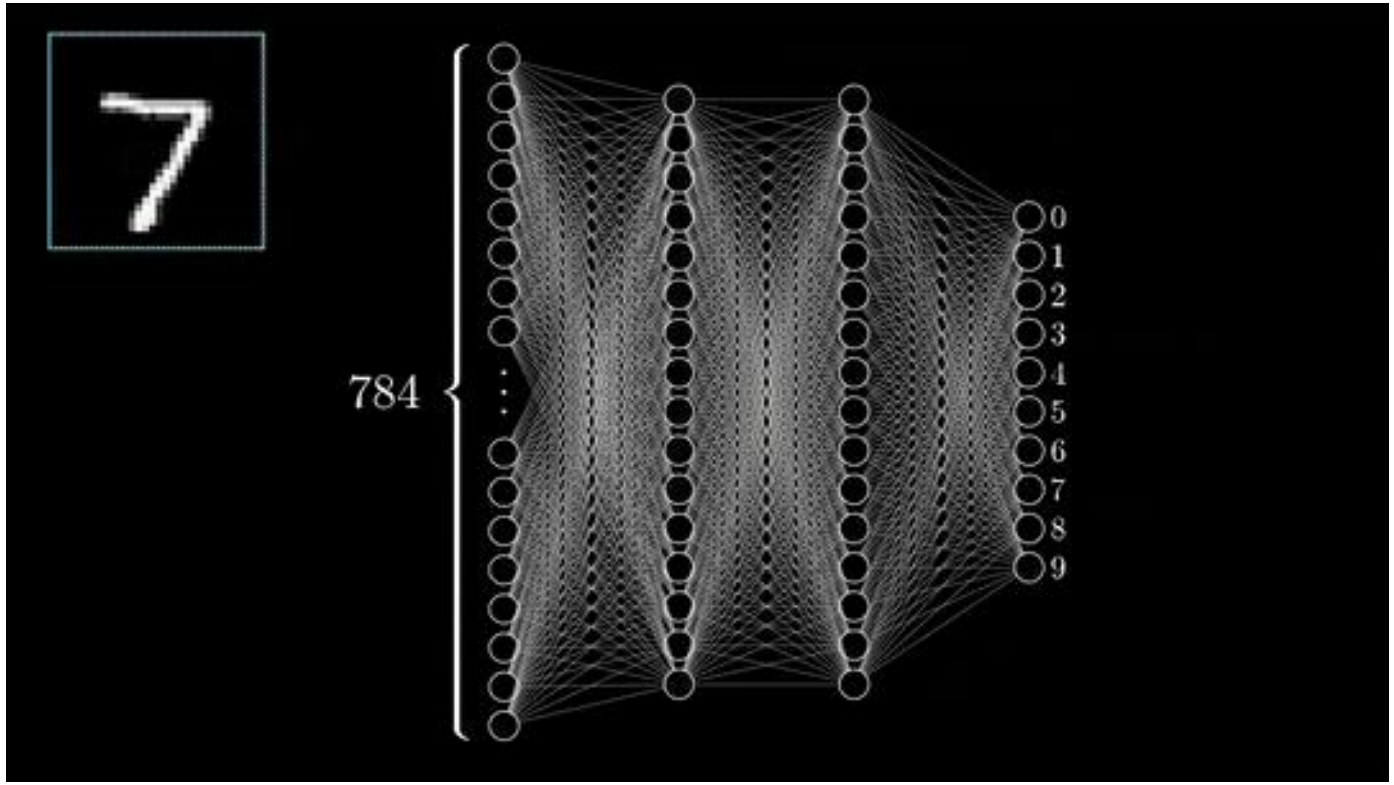
MNIST Dataset



Phase 2: Validation/Testing

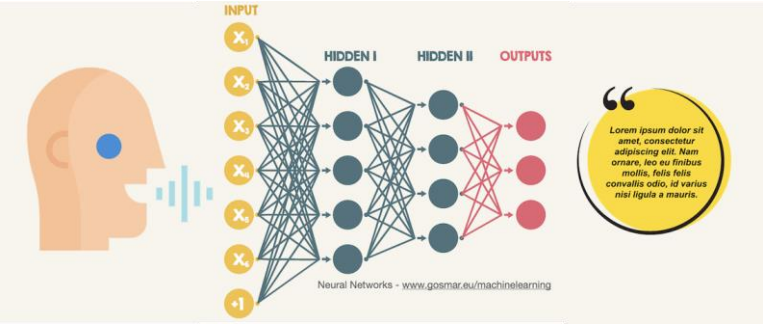
Data-driven Modeling

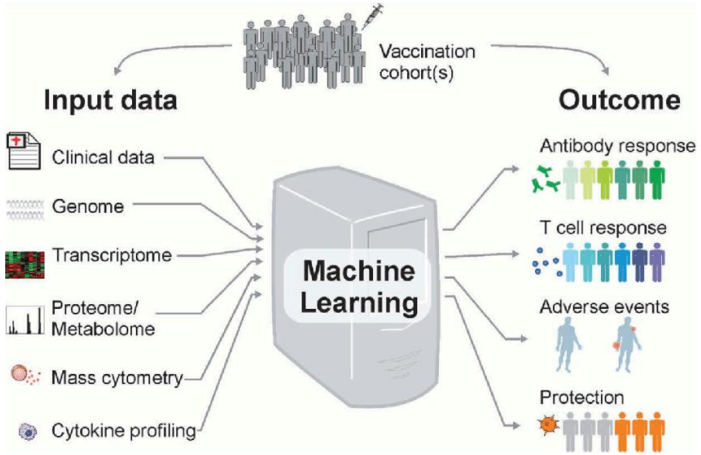
- Demo



https://gfycat.com/gifs/tag/3b1b

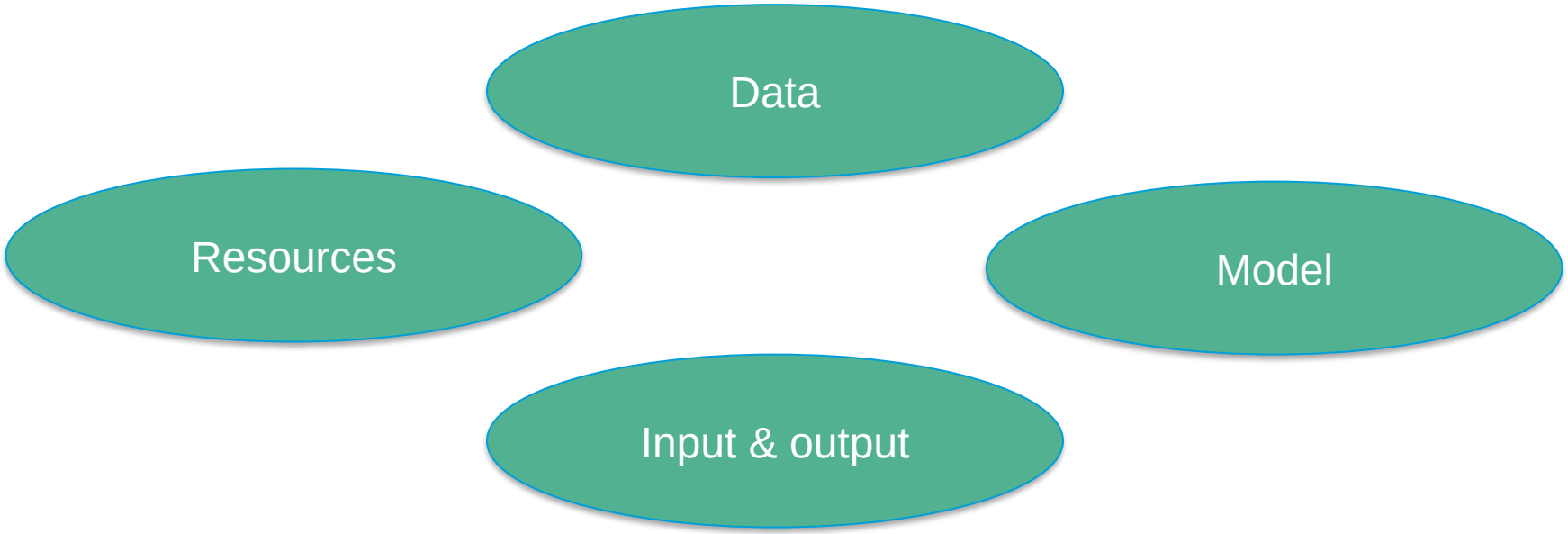
Applications





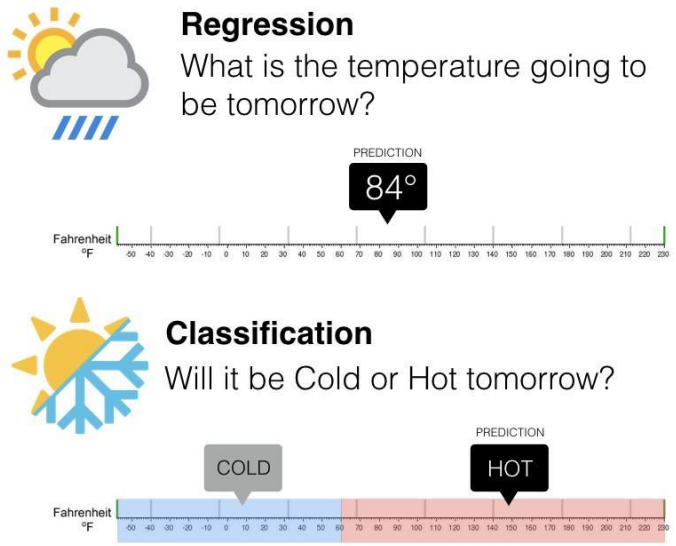
•https://www.gosmar.eu/machinelearning/2020/05/25/neural-networks-and-speech-recognition/
•https://www.ufrjnautilus.com/post/vis%C3%A3o-computacional-e-carros-aut%C3%B4nomos
•https://www.nature.com/articles/d41586-019-03298-6
•https://www.forbes.com/sites/johnkoetsier/2023/04/14/generative-ai-music-platform-creates-forever-songs-with-artists-unique-sounds-melodies-and-beats/?sh=75a691eada47
•Gonzalez-Dias, Patrícia & Lee, Eva & Sorgi, Sara & Lima, Diógenes & Urbanski, Alysson & Silveira, Eduardo & Nakaya, Helder. (2019). Methods for predicting vaccine immunogenicity and reactogenicity. Human Vaccines & Immunotherapeutics. 16. 1-8. 10.1080/21645515.2019.1697110.

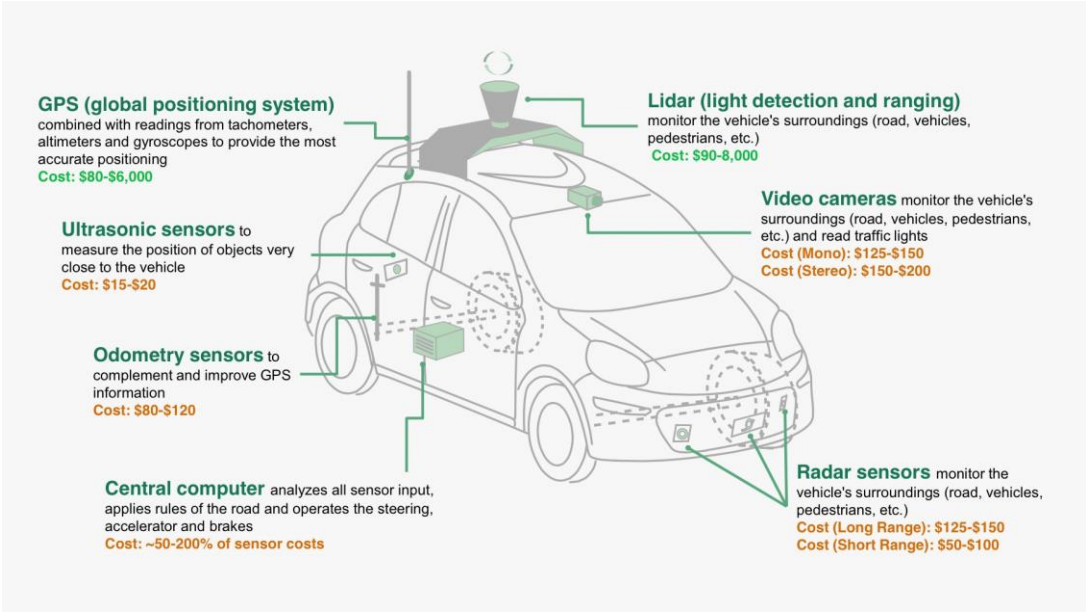
"Pillars" of Using ML



Input & Output

- What exactly are you trying to model?





https://www.wired.com/2015/04/cost-of-sensors-autonomous-cars
https://medium.com/@ali_88273/regression-vs-classification-87c224350d69

Data

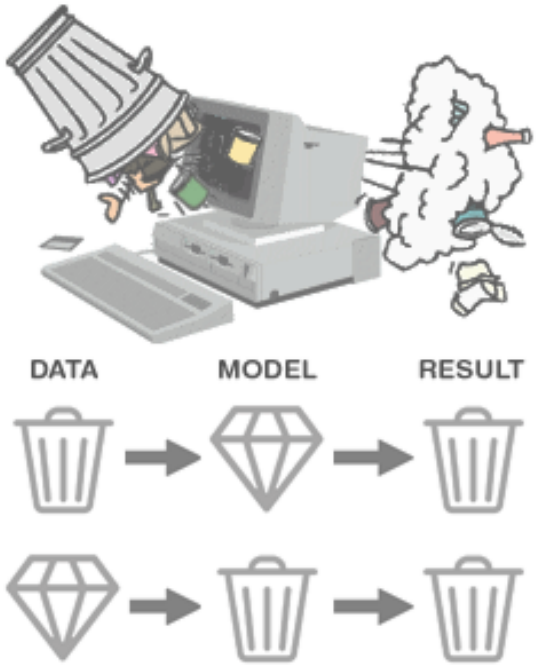
- Is there enough of it?
- Does it need clean-up?

Damaged specimen

Missing Features

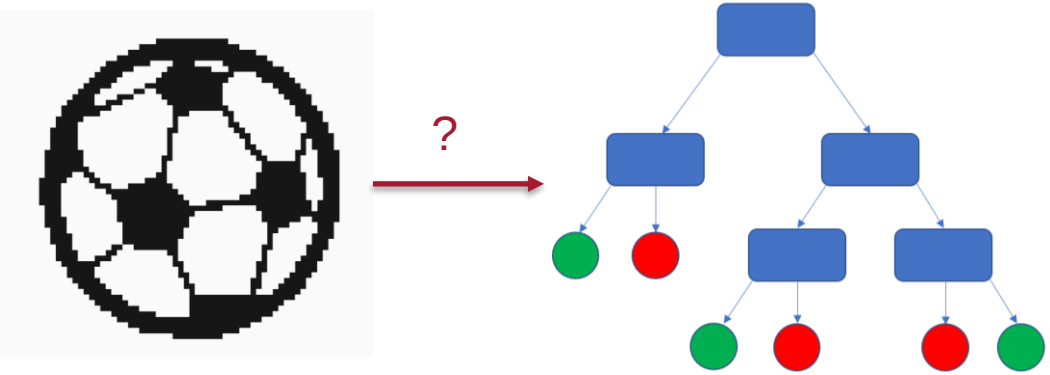
Occluded Features

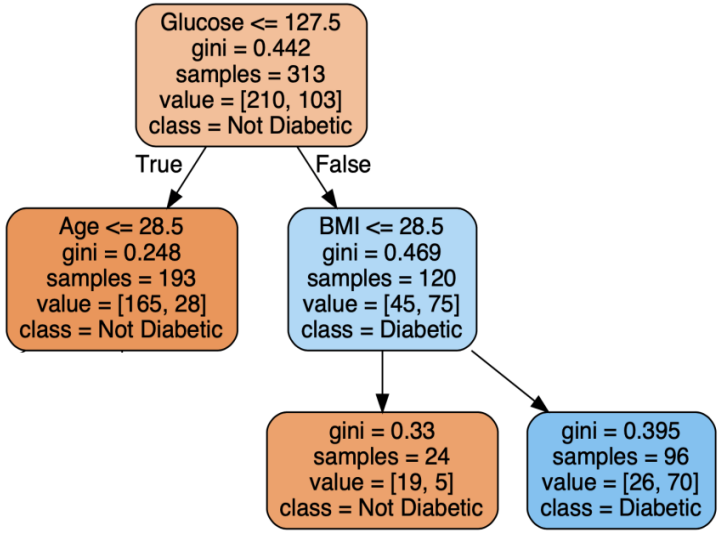
Elhamod, M., Diamond, K. M., Murat Maga, A., Bakis, Y., Bart, H. L., Mabee, P., Dahdul, W., Leipzig, J., Greenberg, J., Avants, B., & Karpatne, A. (2022). Hierarchy-guided neural network for species classification. *Methods in Ecology and Evolution*, 13, 642–652. https://doi.org/10.1111/2041-210X.13768



Model

- What kind of model would be sufficient/suitable for modeling your data?

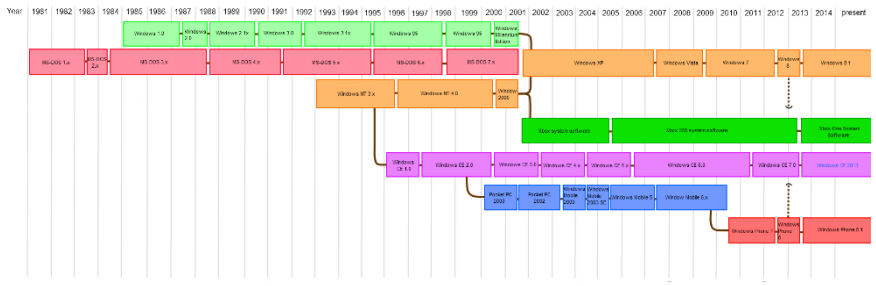




https://statisticallyrelevant.com/decision-trees-in-python-predicting-diabetes/

Resources

- What compute resources are available?
- What is the project's timeline?



Demo

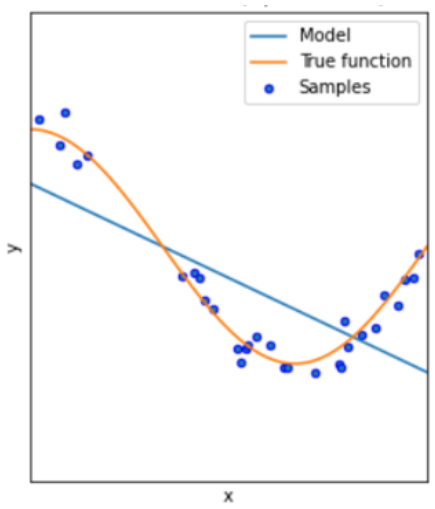
BOSTON UNIVERSITY

Intro to ML

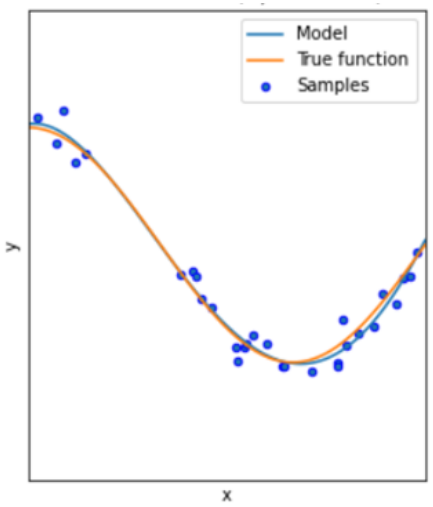
Continued…

BOSTON UNIVERSITY

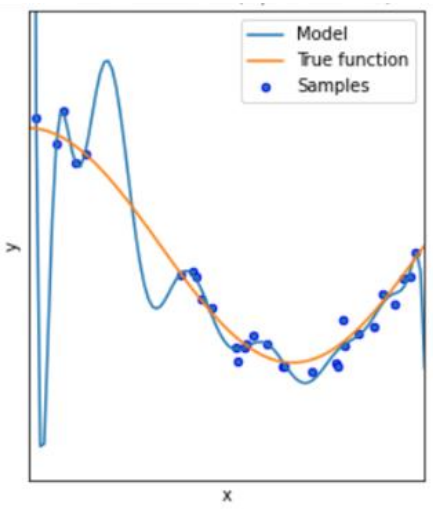
Model Generalization



Underfitting



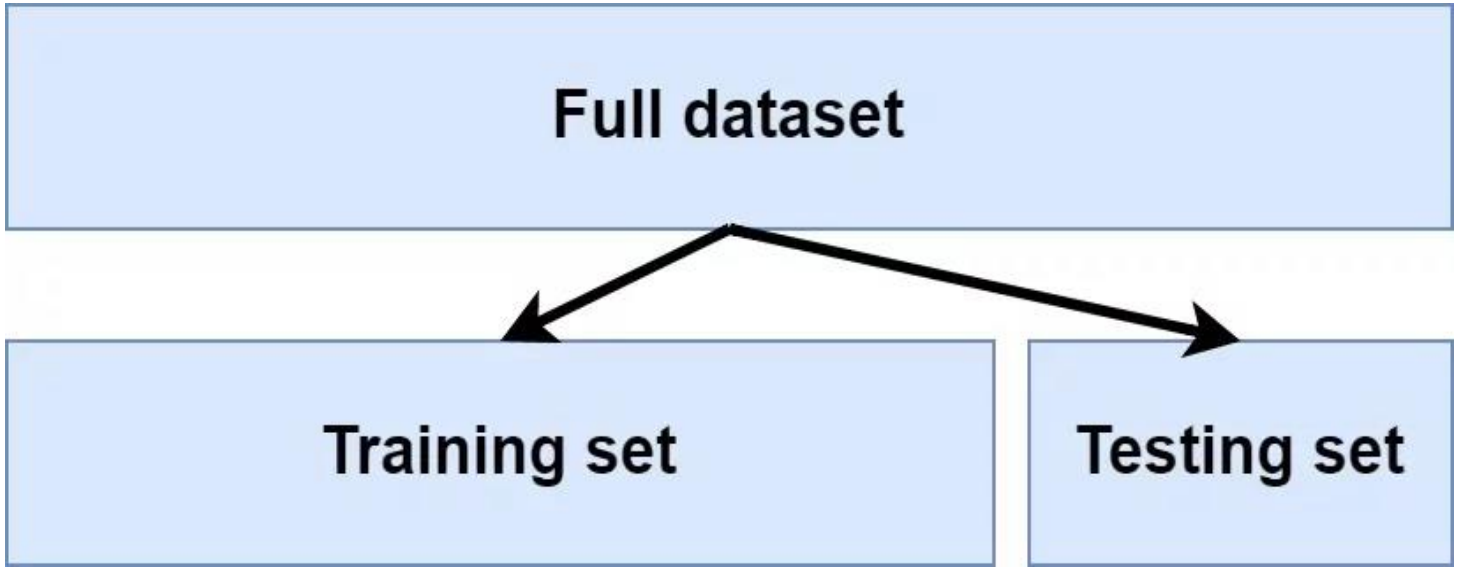
Well-fitted



Overfitting

https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning

Model Generalization



https://www.machinecurve.com/index.php/2020/11/16/how-to-easily-create-a-train-test-split-for-your-machine-learning-model/

Model Generalization

https://medium.com/trusted-ai/explaining-ai-model-behaviour-with-ibm-watson-openscale-86515702c177

Neural Networks

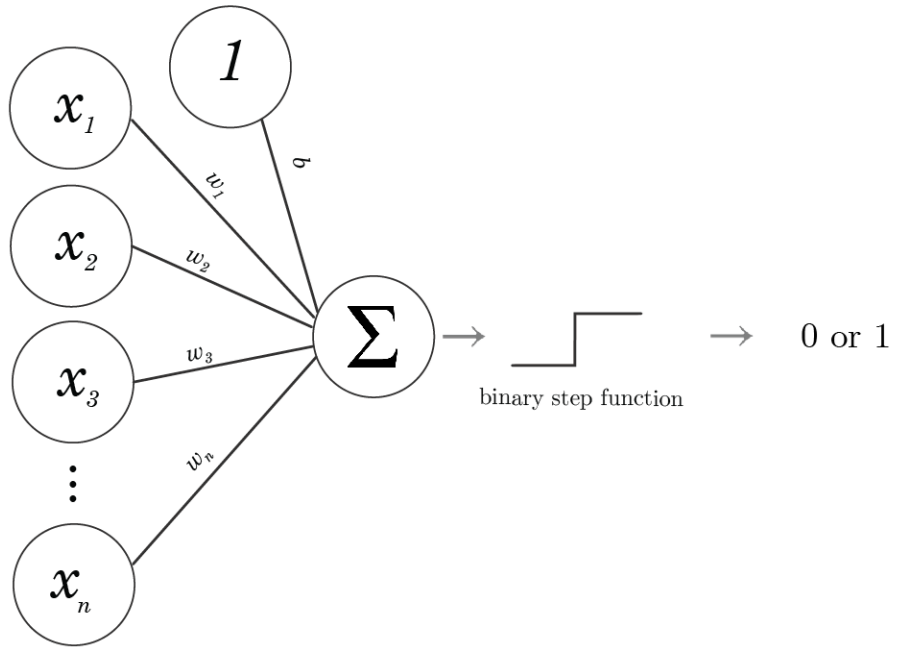
BOSTON UNIVERSITY

Boston University Questrom School of Business

The building block: The Perceptron

- Demo

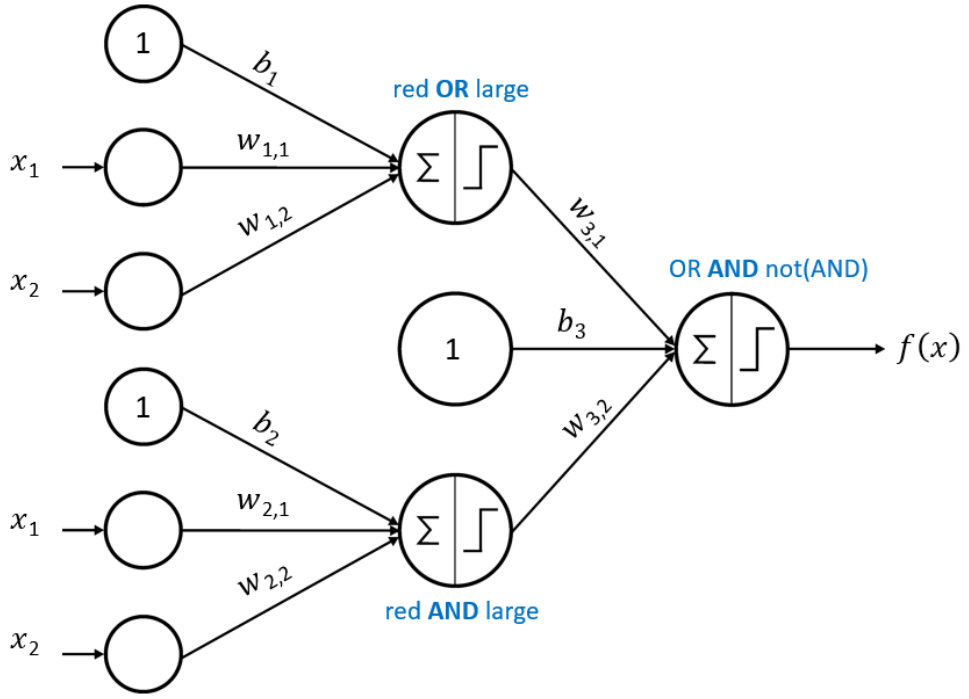
- Can we work around the “linear separability” issue”?



Adam Dhala

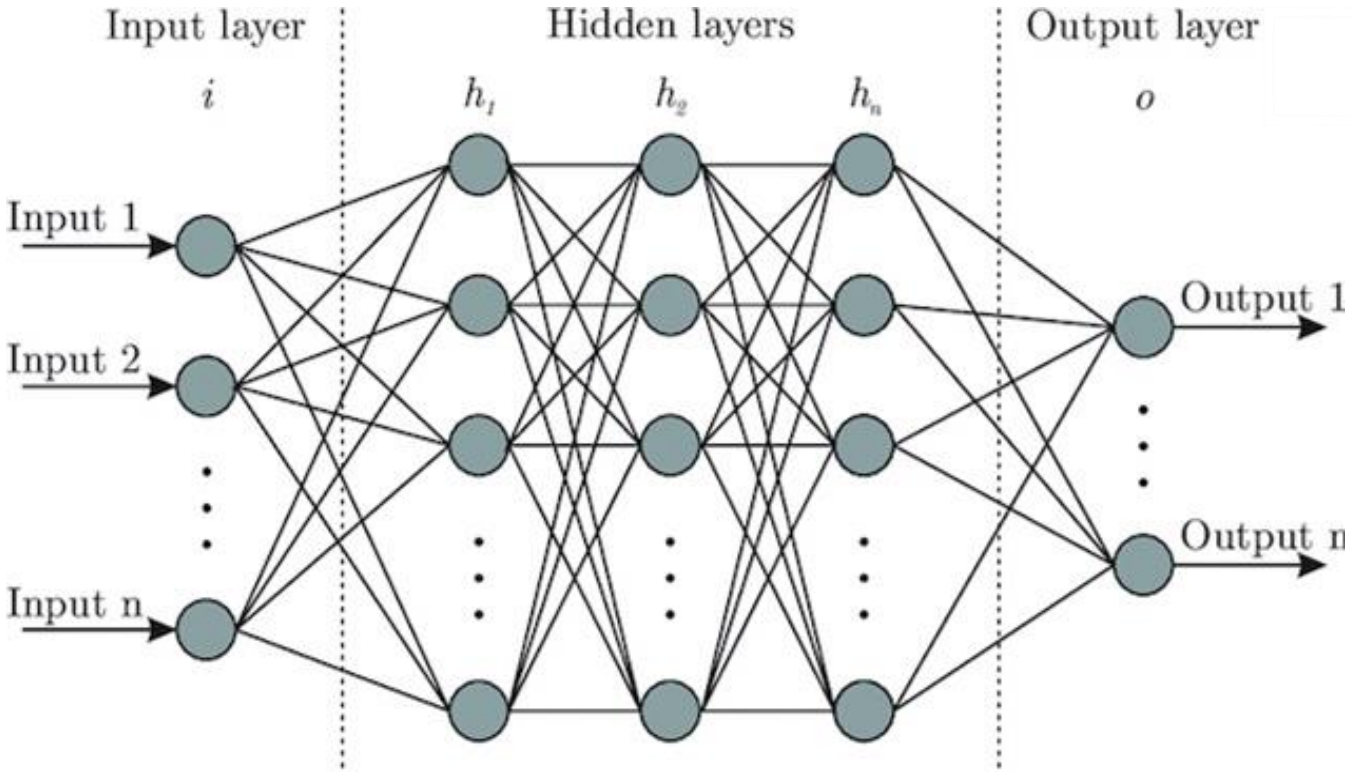
Power in Numbers: Multiple Perceptrons

- Demo



western-neuralnets.ca

Neural Newtorks

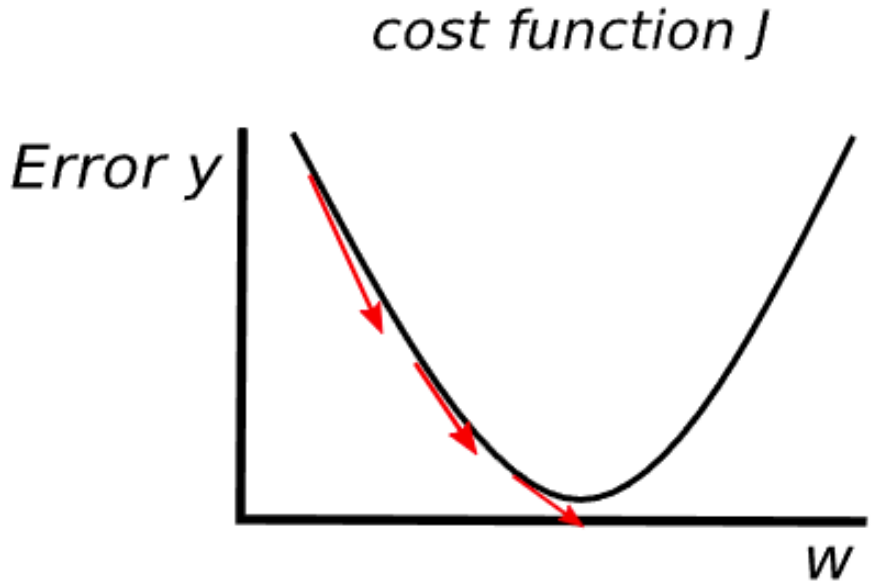


Neural Newtorks

- In theory, a single layer with an “infinite” number of neurons can approximate ANY function
  - In practice though…

Optimization

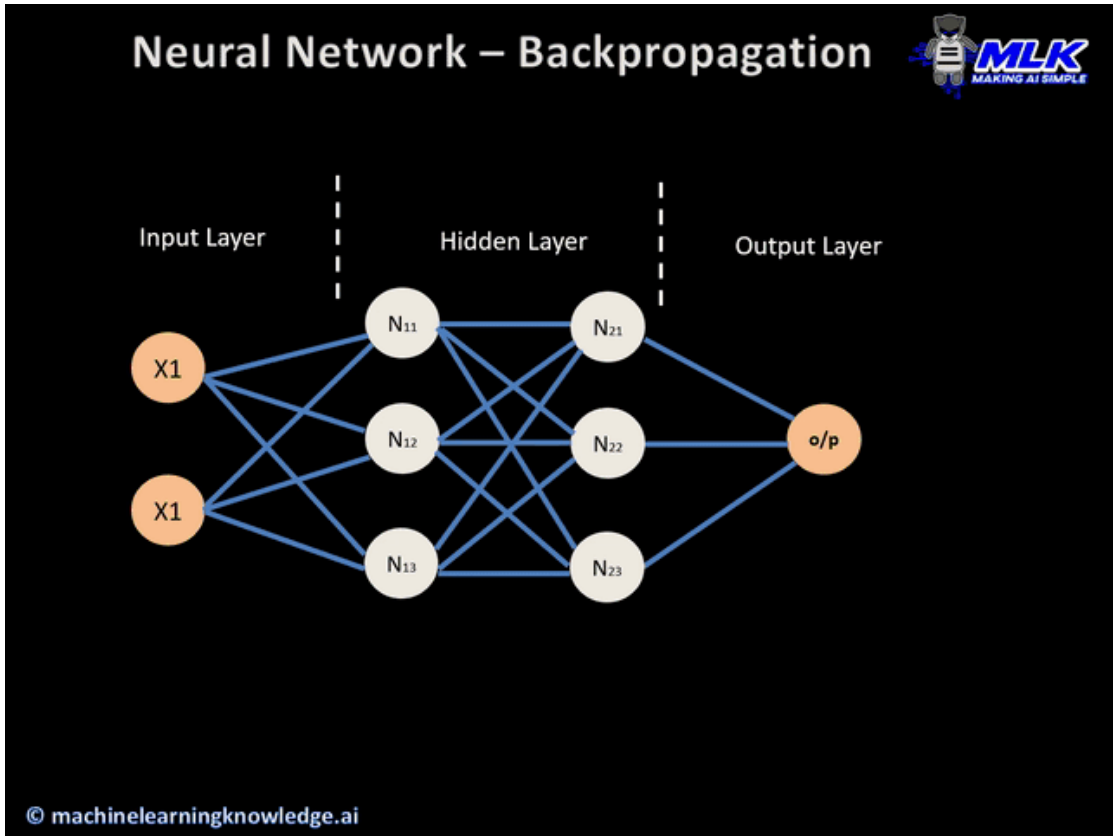
- The method that finds the best weights (i.e., weights that lead to lowest error).
  - Error ≡ loss ≡ cost function
  - Generally using gradient descent with backpropagation.



Elvira Siegel

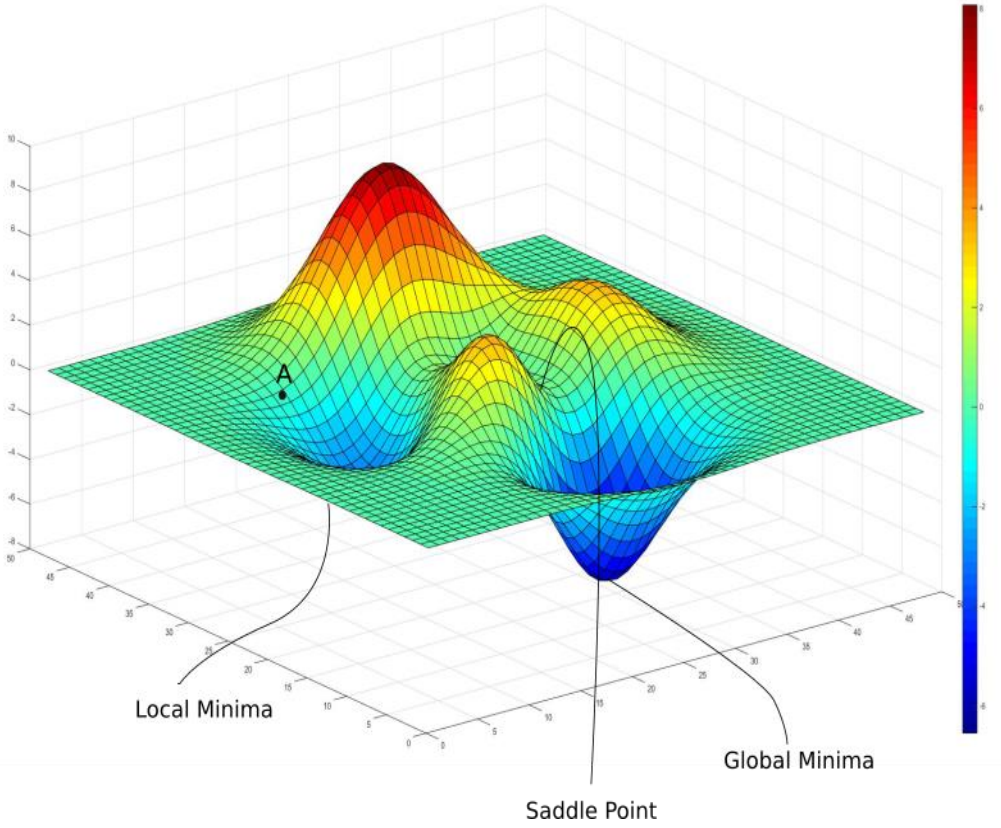
Optimization

- The method that finds the best weights (i.e., weights that lead to lowest error).
  - Error ≡ loss ≡ cost function
  - The method generally used is gradient descent with backpropagation.



Optimization

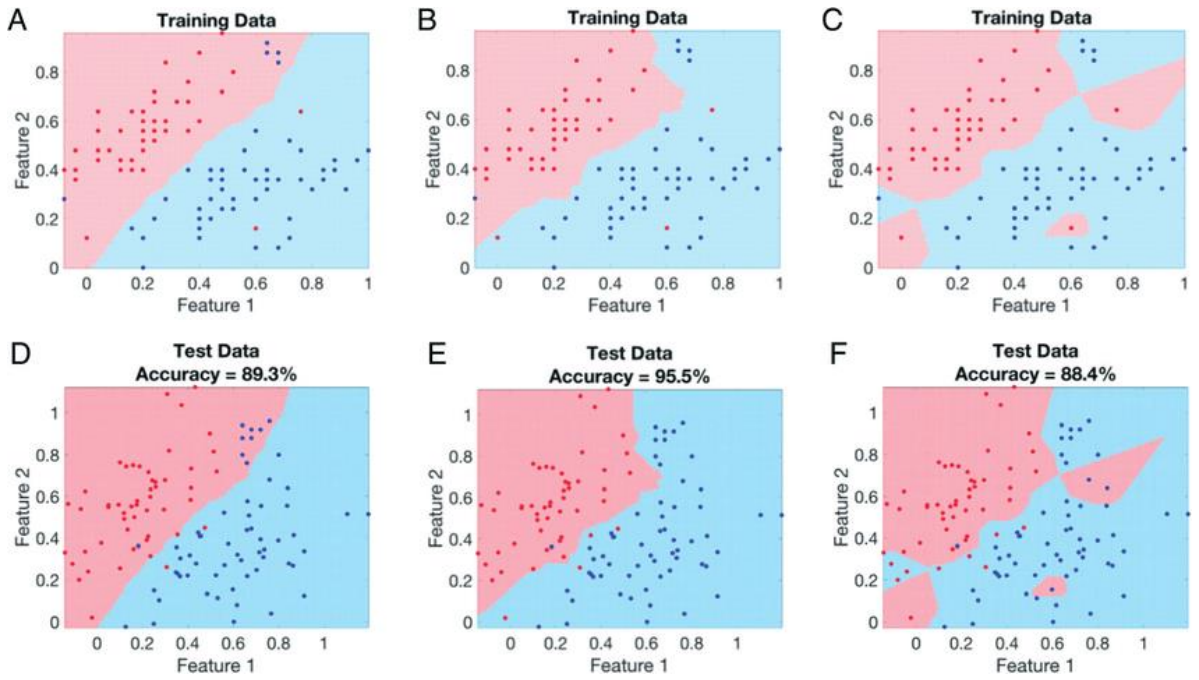
- Can we always achieve lowest error?
- Demo



TechTalks

Overfitting and Underfitting

Better fit on the training data

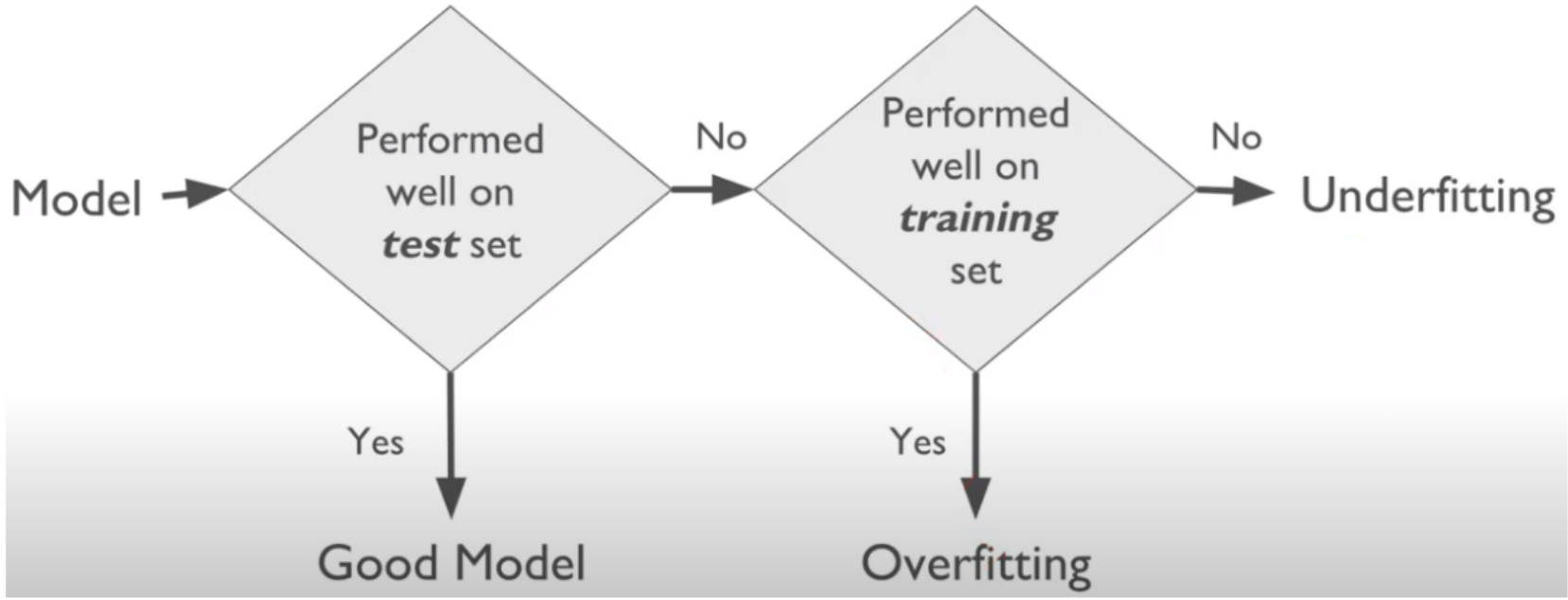


underfitting

overfitting

Jina Ko

Overfitting and Underfitting



http://jcsites.juniata.edu/faculty/rhodes/ml/clusterAn.htm

Results are bad?

- Check against a benchmark!
  - paperswithcode.com
  - kaggle.com
- Are you overfitting or underfitting?

How do I improve my results?

- Best way: Get more **GOOD** data
  - If not, clean-up existing data.
- Are you overfitting or underfitting?
  - Overfitting: get more data or use a less complex model.
  - Underfitting: get a more complex model.
- Keep it simple!
  - Start with a simple model, simple data, simple code.
  - Test by component
  - Test by example

Class Logistics

BOSTON UNIVERSITY

Survey Results

- Office hours
- Results

Pass through the Syllabus

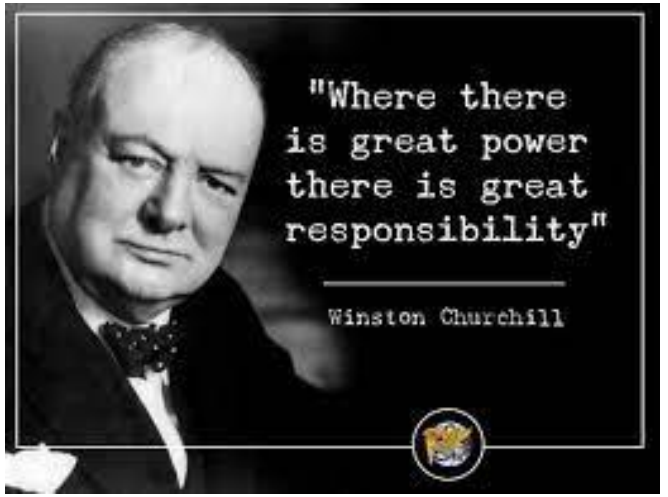
- Did you actually read it?

General Objectives

Assignments are meant to simulate challenges in real-life.

- At the workplace, you can use the internet, code snippets, and even ChatGPT.
- You can also discuss with others.
- But you must cite your resources!

- You cannot have someone do the work for you though.
  - No copy-paste of others' solutions.
  - You must own your work!
  - Zero-tolerance for cheating.



Projects

- You will form 6 groups, 4-5 students each.
- You will pick a topic for your project such that it uses language models and targets a real-world application.
  - There should be some form of originality.
  - If in doubt or having trouble, talk to your professor.
  - Be creative and ambitious!

Projects

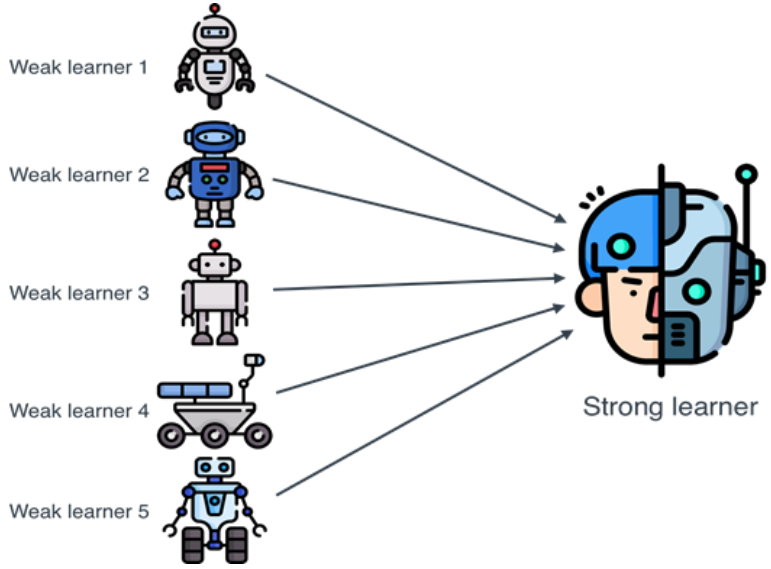
- Your grade will depend on
  - A balanced workload among the team members.
  - The ability to present and document your problem, proposed solution, and results clearly.
  - The ability to achieve good (or promising) results, adequately justify any limitations, and propose solutions to those limitations.
  - Others' ability to interact with your solution and their feedback.
  - The ability to iteratively improve your results over time according to plan.

Projects

- Generally, all team members will get the same grade **given they have contributed equally to the project.**
- Team members should also contribute, more or less, across the different component of the project (i.e., coding, writing, presenting, etc.)
- Equal contribution:
  - For each submission, you will include a schedule of current and expected contributions per team member.
  - After the 1st milestone, you will fill a survey grading yourself and fellow teammates and reflecting on the health of your team. This is a chance to communicate any concerns you have regarding group dynamic or unfair work distribution.
  - If it becomes evident that some team members have significantly under-contributed to the project, they will be scored individually.

Dicussions

- Being able to communicate your thoughts and defend them is of utmost importance in the workplace.
- Your active presence is also of utmost importance in the workplace.
- You do not have to (and probably cannot) always be right. But, you will most certainly contribute to the learning experience.
- Discussions can also take place on Blackboard.
- They also include constructively critiquing other classmate's projects.
- Pre-class readings will be brought up in class and will count towards your grade.



https://livebook.manning.com/concept/machine-learning/ensemble-method

Individual Work

- You will have 3-4 different individual assignments. You will be working on these on your own.
- Some of these will be completely independent from the work in class, others will be a follow-up on a class demo.
- Before reaching out to professor/TA with questions, make sure you have done your homework, given your best effort, and researched well on your own first.

Coding Frameworks and Platforms

Hugging Face

Shared Computing Cluster (SCC)

- BU's computing cluster with access to CPUs and GPUs.
- Each of you will get a limited number of compute hours. Use your allocations wisely!
- When facing issues, ask the professor or TA for help.
- **Don't leave things till last minute!**

BU TechWeb

Assignment

BOSTON UNIVERSITY

Take a look at this other example

- SCC Guide

Homework

- Finish the Jupyter notebook at home and submit the notebook on Blackboard.

BU TechWeb