Overview

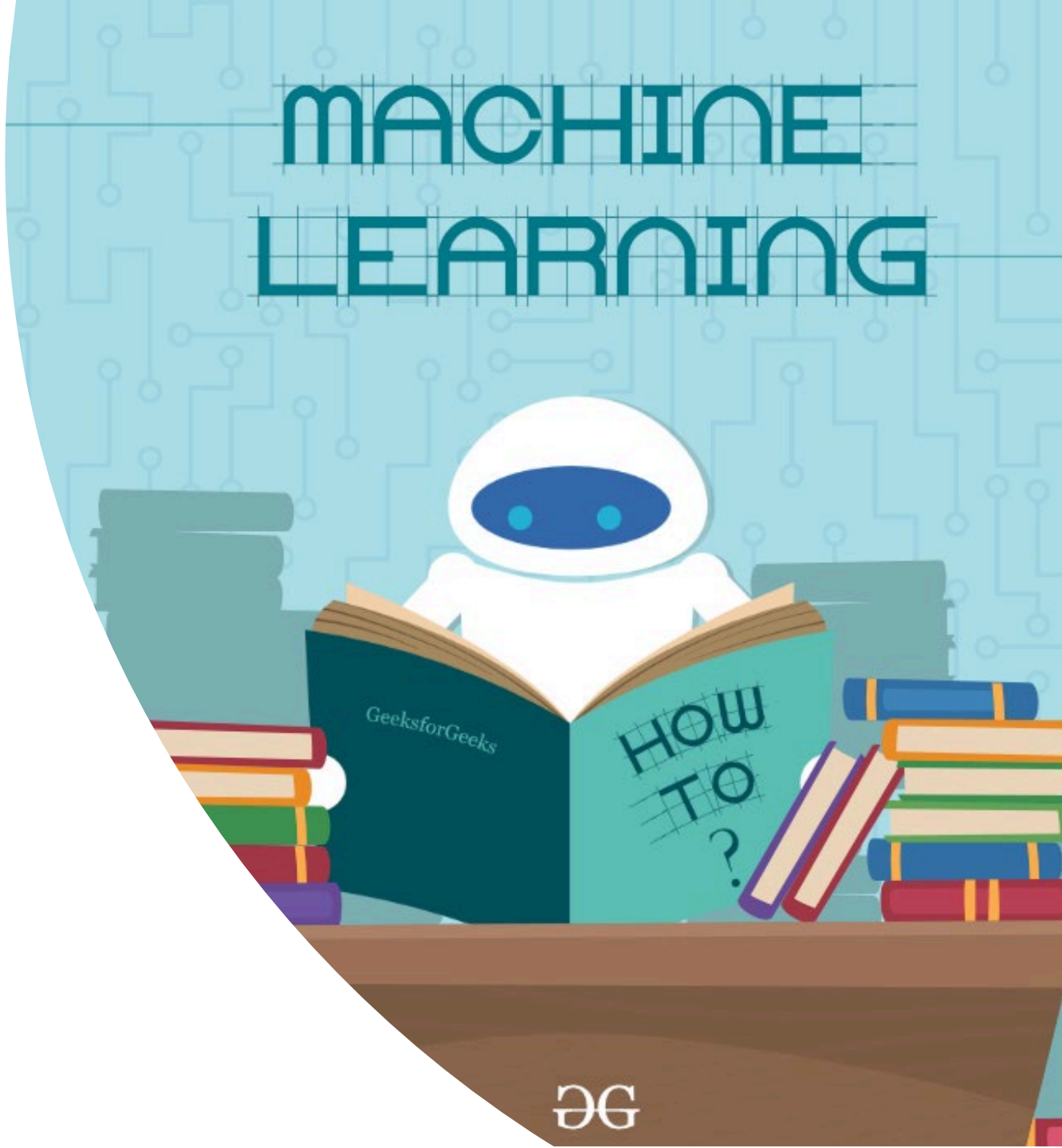
CE 5331 Machine Learning for Civil Engineers

Venki Uddameri, Ph.D. , P.E.

Goals

- Class overview and Objectives

- What is Machine Learning and
- Why should Civil Engineers care?
- Next Steps

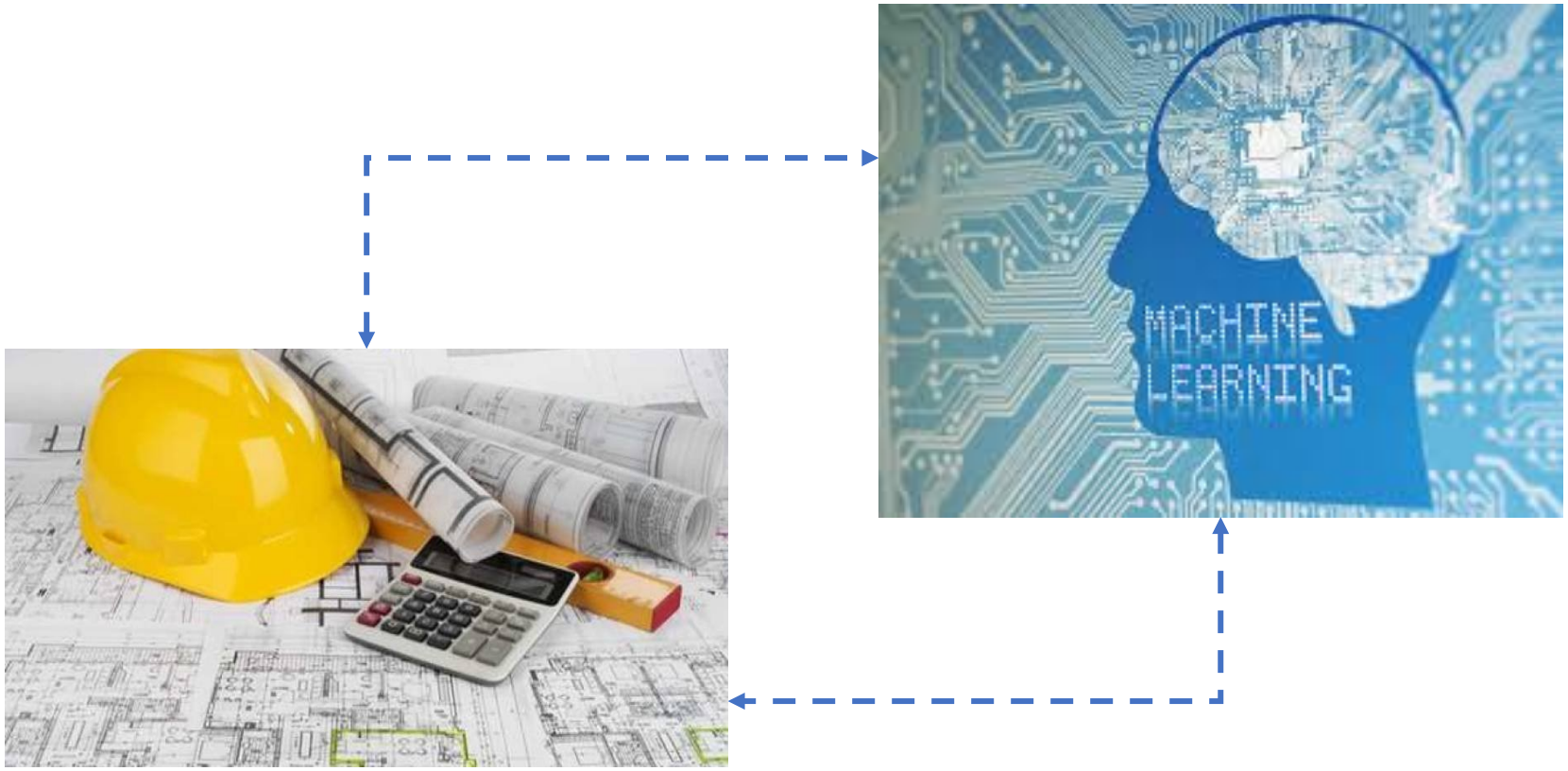


What is Machine Learning?

- Study of computer algorithms that exhibit certain learning characteristics exhibited by living beings
  - Coined in 1959 by Arthur Samuel of IBM
  - Popularized in recent times by their prolific use

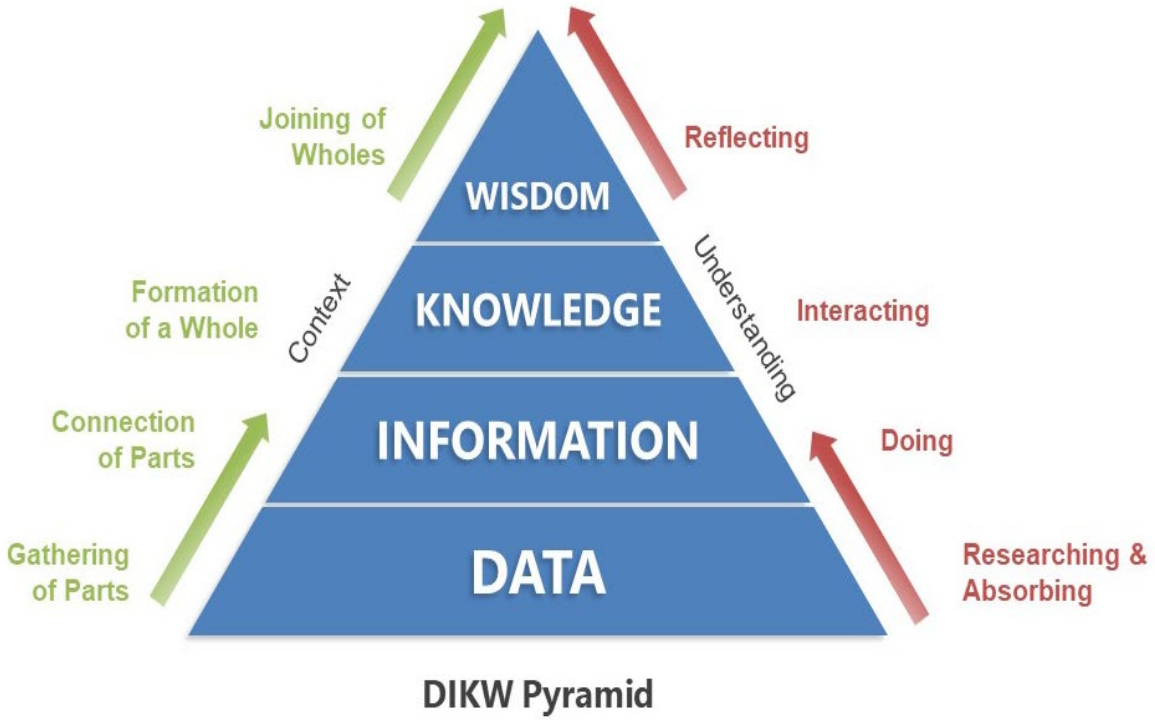
- It is a sub-discipline of Artificial Intelligence whose focus is making machines have cognition
  - Create a device that senses the environment and takes necessary adaptive action

Advances in computing (both processing and storage) has helped machines perform tasks that are similar to human cognition



Why Should Civil Engineers Learn Machine Learning?

Why should Civil Engineers Learn it?



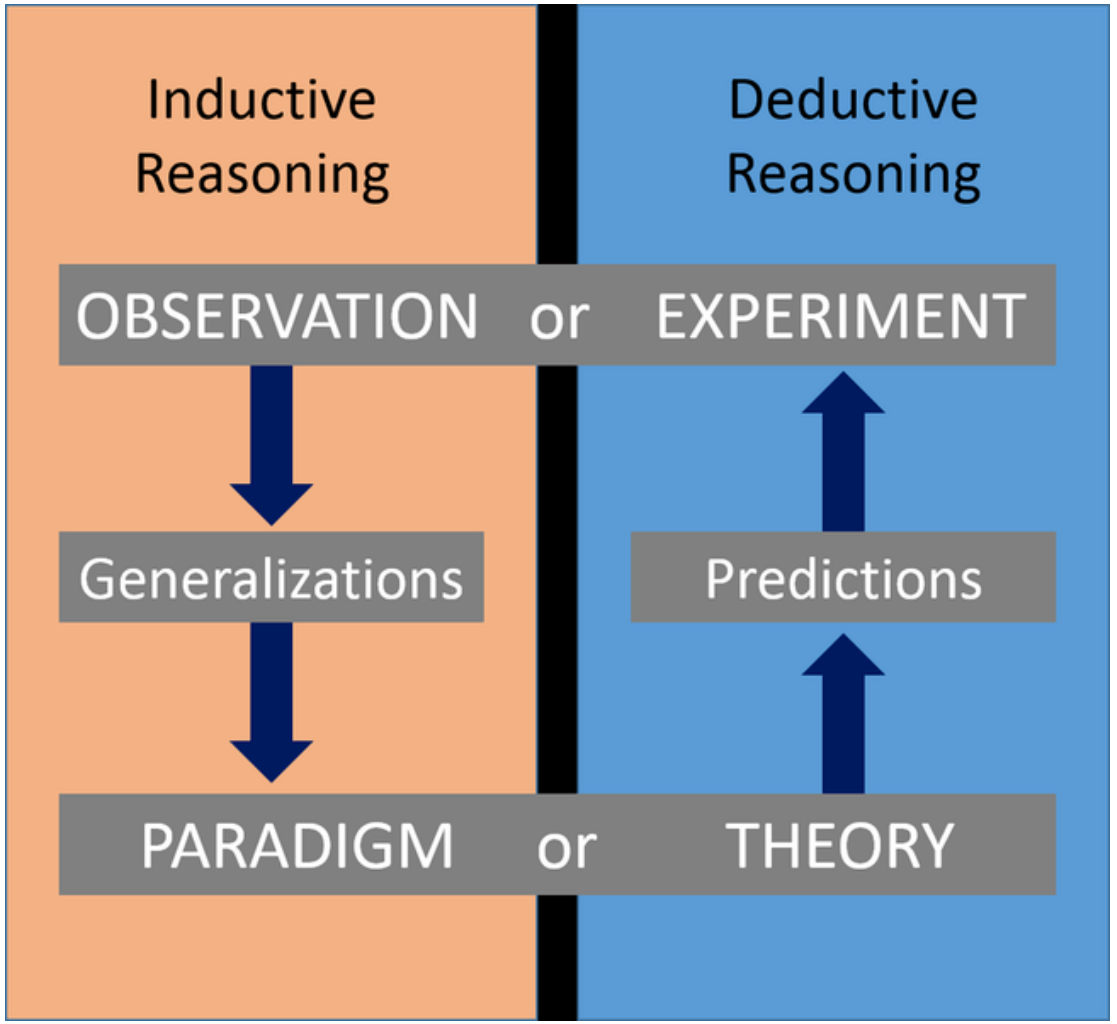
DIKW Pyramid

- Data: Fragmented Pieces of Information
  - Data by themselves are of limited value
- Information: process data to make sense
  - Focused on a specific problem at hand
    - Localized
  - Useful to solve the specific problem
- Knowledge: Synthesis of information
  - Formation of a whole
  - Knowledge can be transferred
    - Not localized (generalizable)
- Wisdom: Synthesis of knowledge
  - Putting knowledge to good use

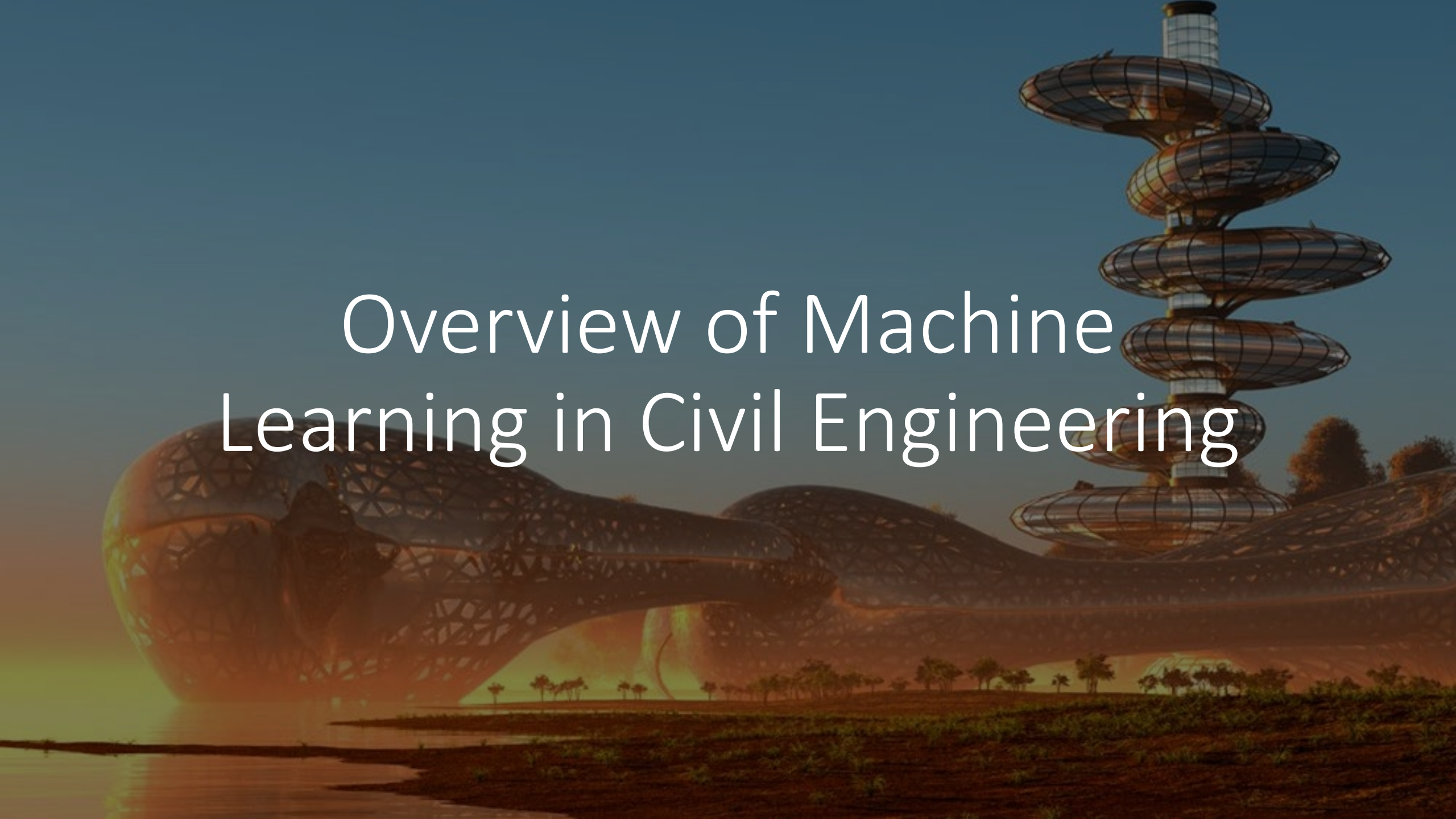
Machine learning and Civil Engineering Share Similar Goals

CE - Knowledge

- Generalization is a key element of Civil Engineering practice
  - Body of knowledge (BOK)
- Generalization helps codify our knowledge and apply it
  - Proper reconciliation of uncertainty
- We establish our body of knowledge in several ways
  - Use of theory
  - Experimentation
  - Observations of natural phenomena



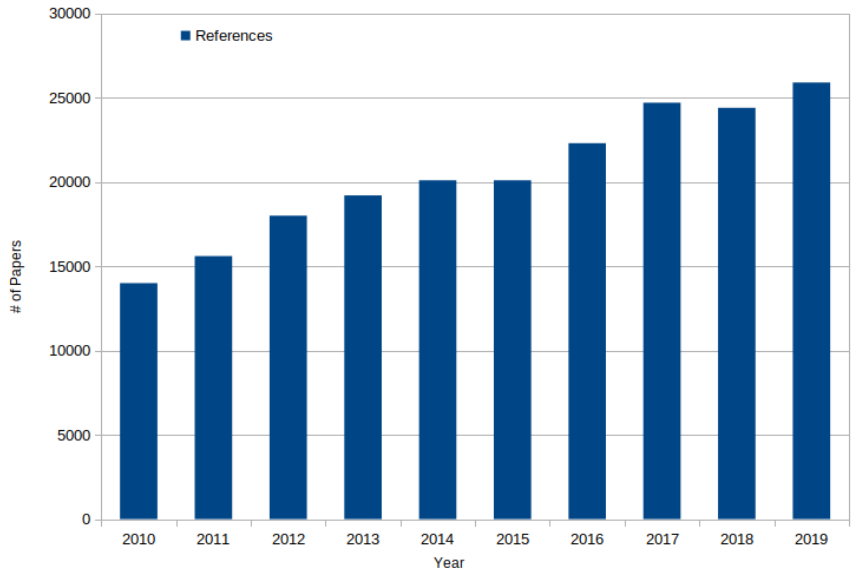
Both Inductive and Deductive Reasoning are combined in Civil Engineering Practice



Overview of Machine Learning in Civil Engineering

Where is Machine Learning being used in Engineering?

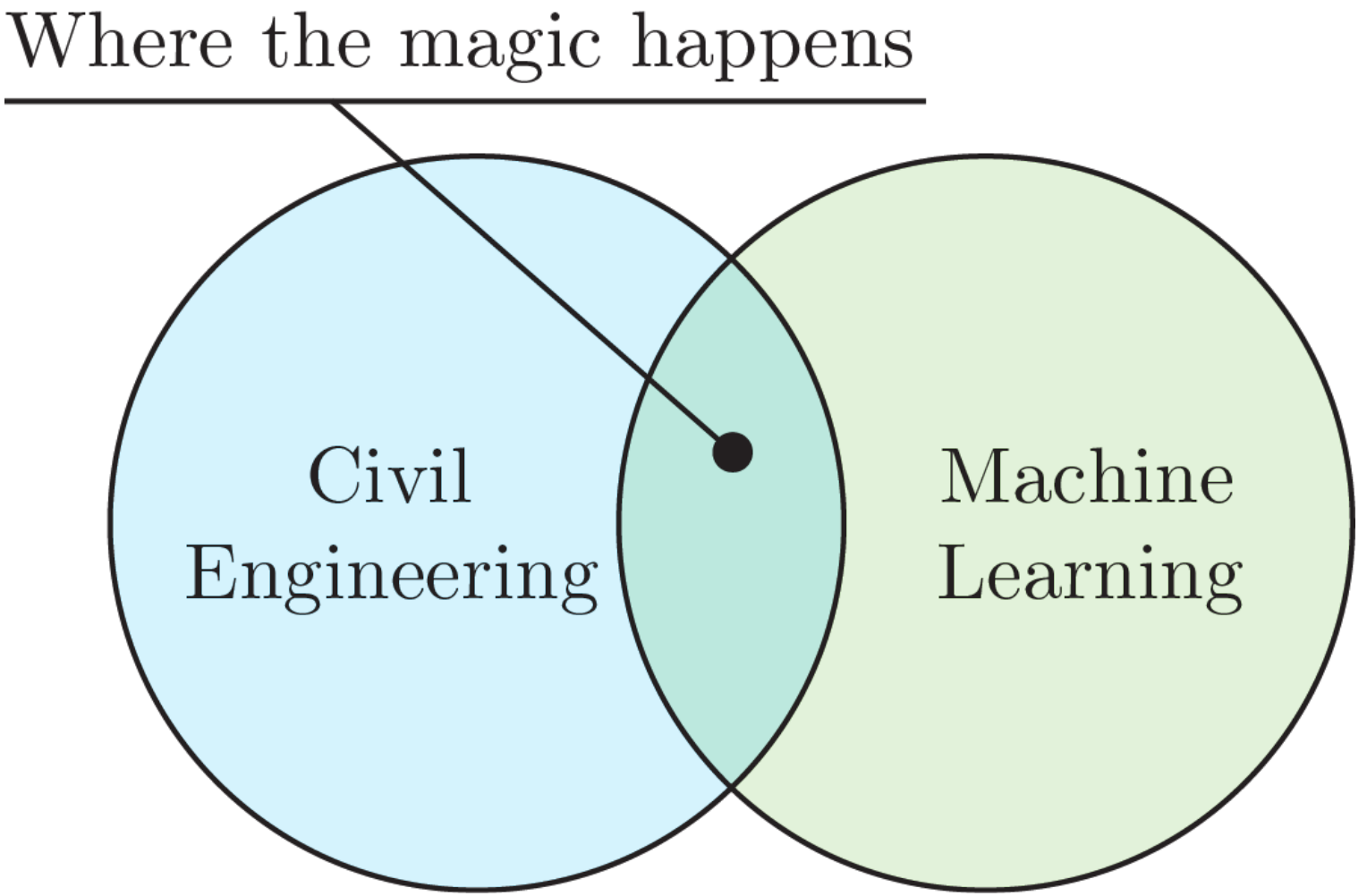
- The use of machine learning is not new to civil engineering
  - Early research applications were presented in 1980s or even earlier
- The use of these techniques has however grown exponentially in recent years
  - Availability of data
  - Increased computational power
  - Availability of software

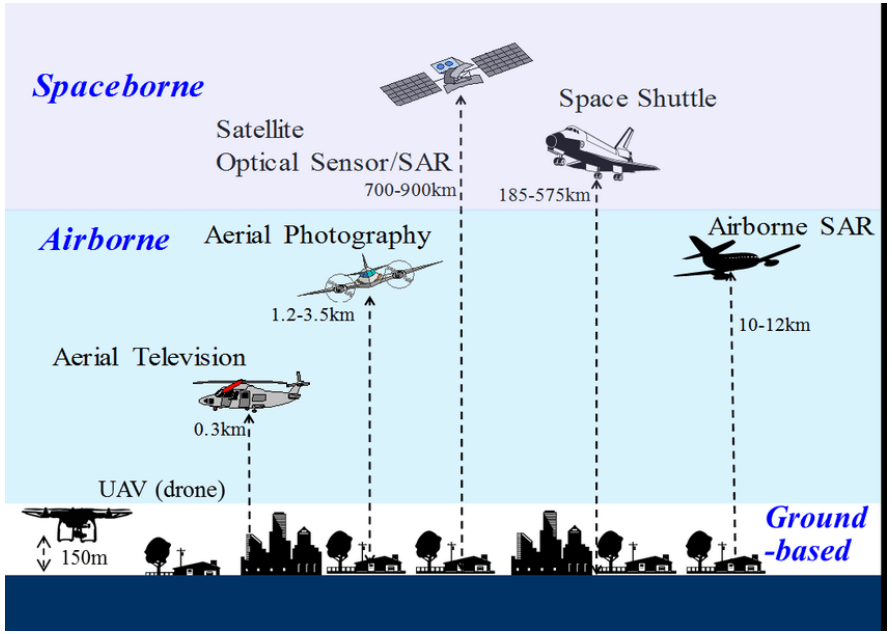


Google Scholar Cites with key-words:
"Machine Learning and Civil Engineering"

Machine-Learning methods are slowly transitioning from research to practice domain in CE

Overview of Machine Learning in Civil Engineering

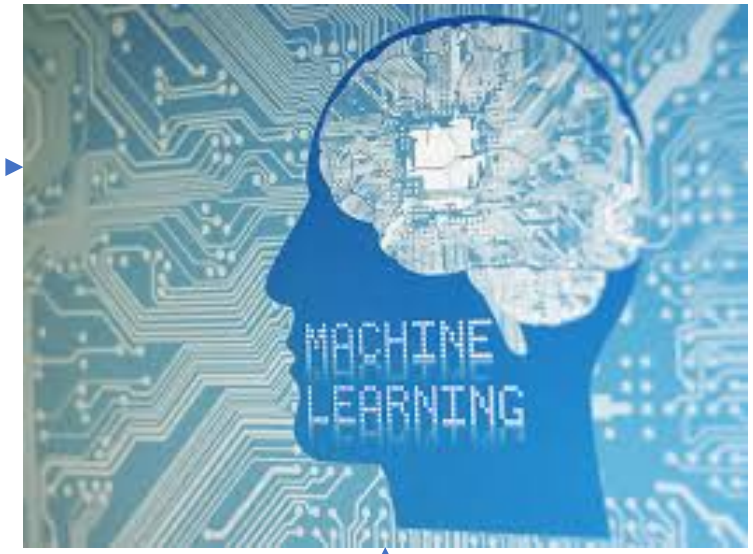




There are wide range of engineering problems that can be solved using Machine Learning Methods

What can Machine Learning be used for?

- Assessment of Risks of natural and manmade hazards
  - Earthquakes, floods, droughts
  - Cracks in Concrete structures
- Analysis and Prediction of external stressors
  - Wind loads, traffic flows, rainfall extremes
- Design of Civil Structures
  - Transportation networks, structural health monitors
- Planning and Management
  - Construction scheduling, cost optimization
- Geotechnical and Geo-environmental
  - Mapping soil properties and pollutant transport



What is the purpose of this class?

Purpose of this class

- While Machine Learning is gaining traction there are very few universities that offer courses in ML within Civil Engineering curricula
  - Students have to take classes from other departments
    - Computer Science, Electrical Engineering
  - Too theoretical or examples are not relevant to civil engineering
  - Online courses focus on software and do not go in required theory
    - Learn software without knowing limitation/assumptions or their proper usage

A balanced approach with a focus on Civil Engineering applications

What does the course cover?

- We will cover a suite of machine learning methods
  - Supervised Learning
    - Statistical, tree-based, ensemble (boosting and bagging), neural networks and variants
  - Unsupervised Learning
    - Clustering algorithms, Self-organizing Maps, Hidden Markov Models
  - Introduction to semi-supervised learning and reinforcement learning

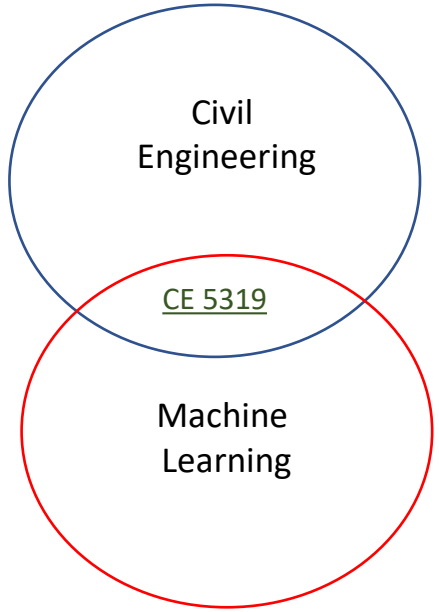
- We shall cover basic theory related to algorithms and mathematical methods
  - Stochastic Gradient Descent, automatic differentiation

- We will cover practical model building concepts
  - Feature selection, model evaluation

- We will use industry-standard software
  - R and Python – scikit-learn, Keras, Tensorflow, PyTorch and other libraries

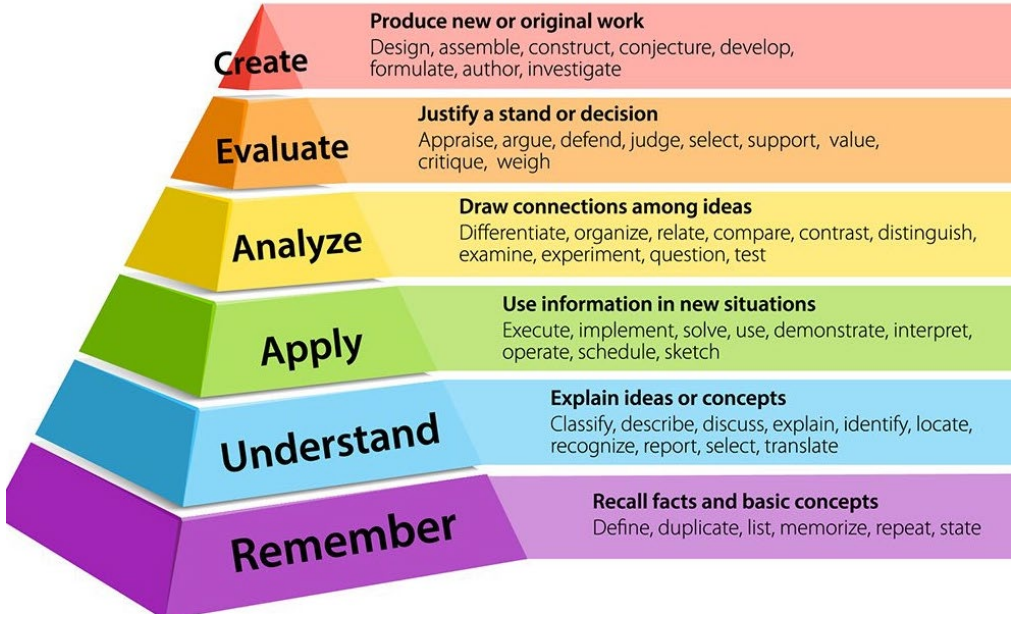
- We will use 'real-world' engineering datasets
  - Outlier detection, handling missing data

- We will use examples from various sub-disciplines of civil engineering
  - Systems-focused approach



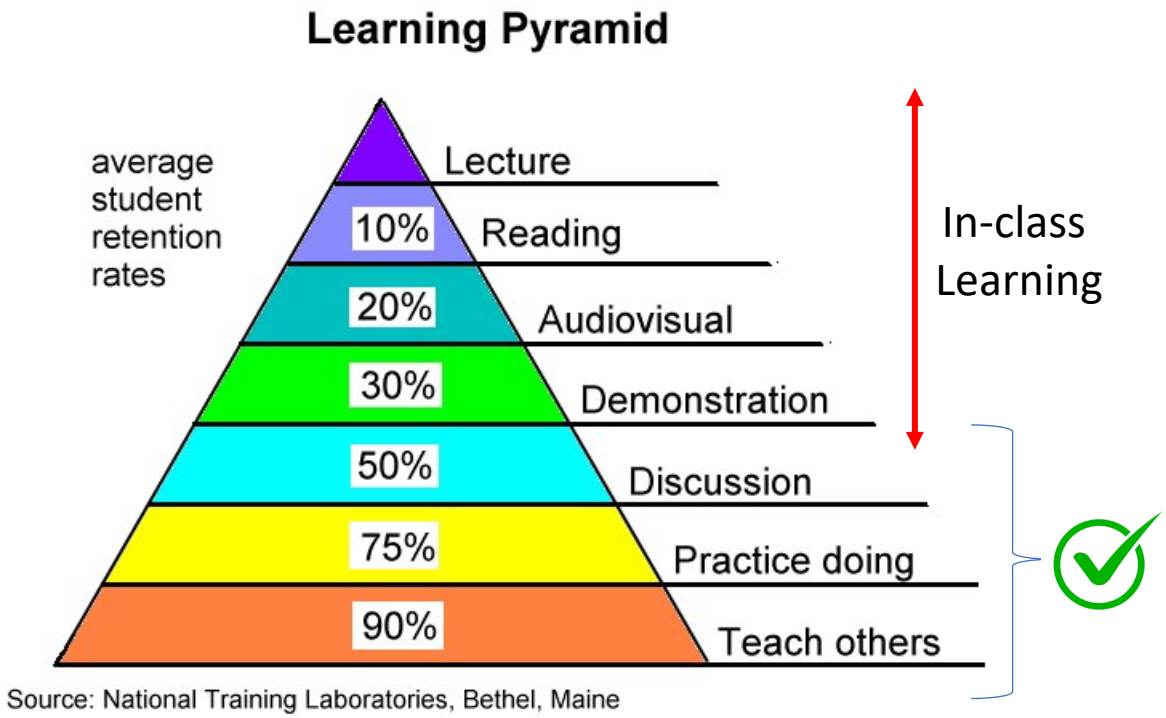
Class Expectations

Bloom's Taxonomy



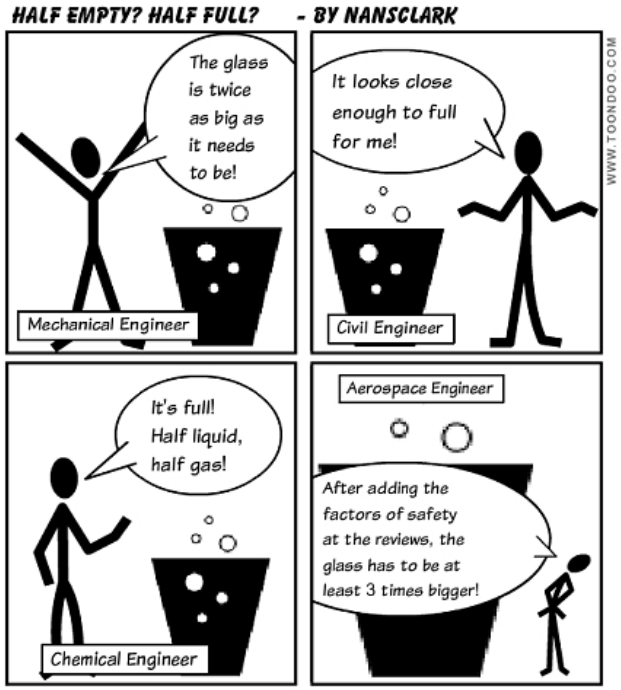
Graduate

Undergraduate



How much you get out of this class is a function of how much you put in

9GAG.COM/GAG/4022820



"Whenever something goes wrong, I just push this little button and restart. I wish my whole life was like that!"

Let's get started.

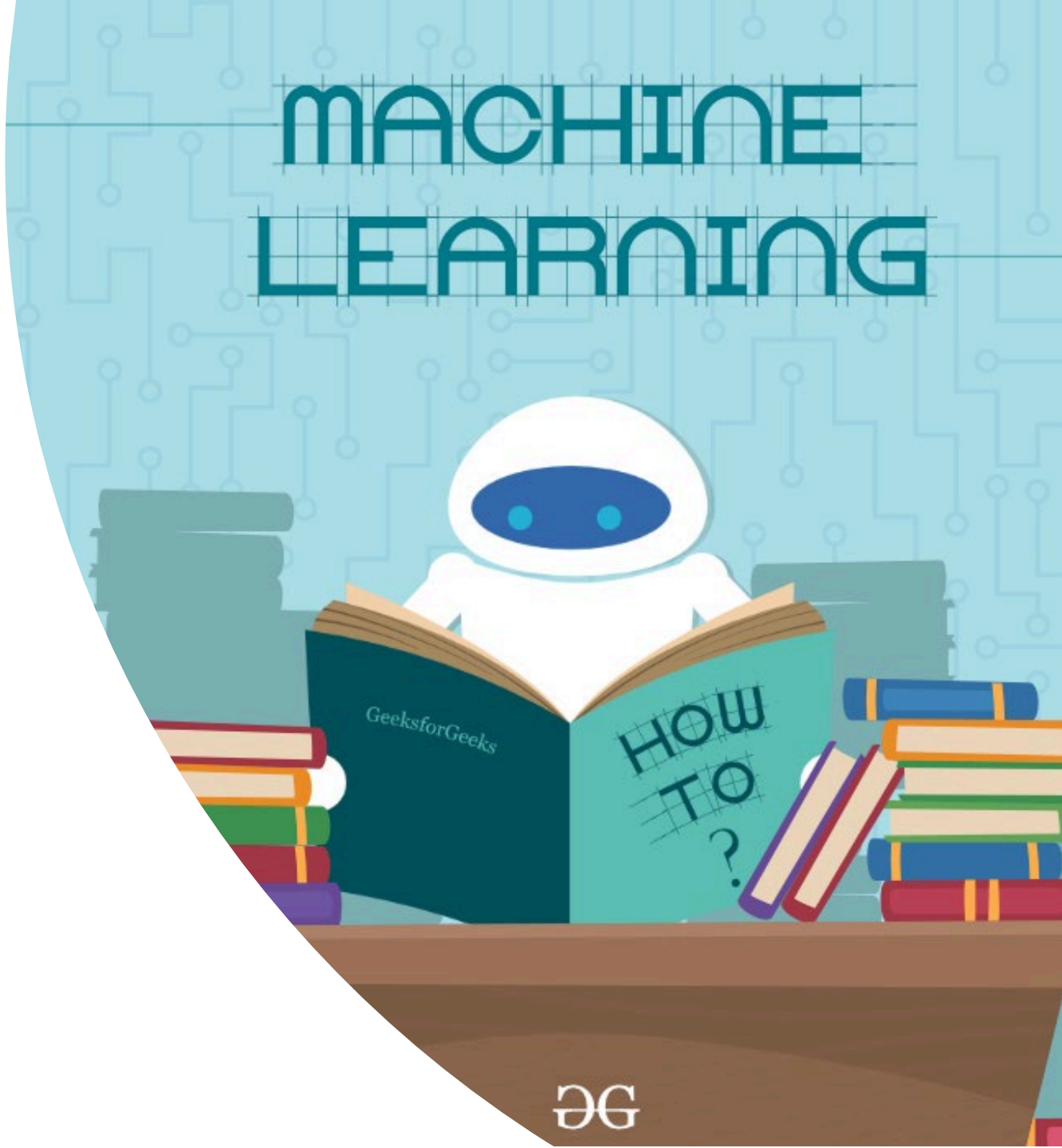
Introduction

CE 5331 Machine Learning for Civil Engineers

Venki Uddameri, Ph.D. , P.E.

Goals

- What are Data?
- Data Archival Models
  - Flat Files
  - Relational Databases
  - NO SQL
- What is Machine Learning
  - Types of learning
- Next Steps



Data

- The knowledge hierarchy represents a generic pathway in increasing the value of engineering endeavors
  - However the path is not linear
- Data are the fundamental building blocks of information and knowledge
- There are several important questions when handling data
  - How to acquire required data?
    - Sensing technologies
  - How to sift ‘bad’ data from ‘good data’?
    - Data QA/QC, validation
  - How to store acquired data?
    - Databases
  - How to easily retrieve data?
    - Query databases

- Civil Engineering Data can be categorized in many ways
  - Generally they are collected in space and/or time
  - They can be univariate or multivariate
  - They can be discrete, continuous
  - They can be on nominal, ordinal, interval of ratio scales
  - They can be structured or unstructured

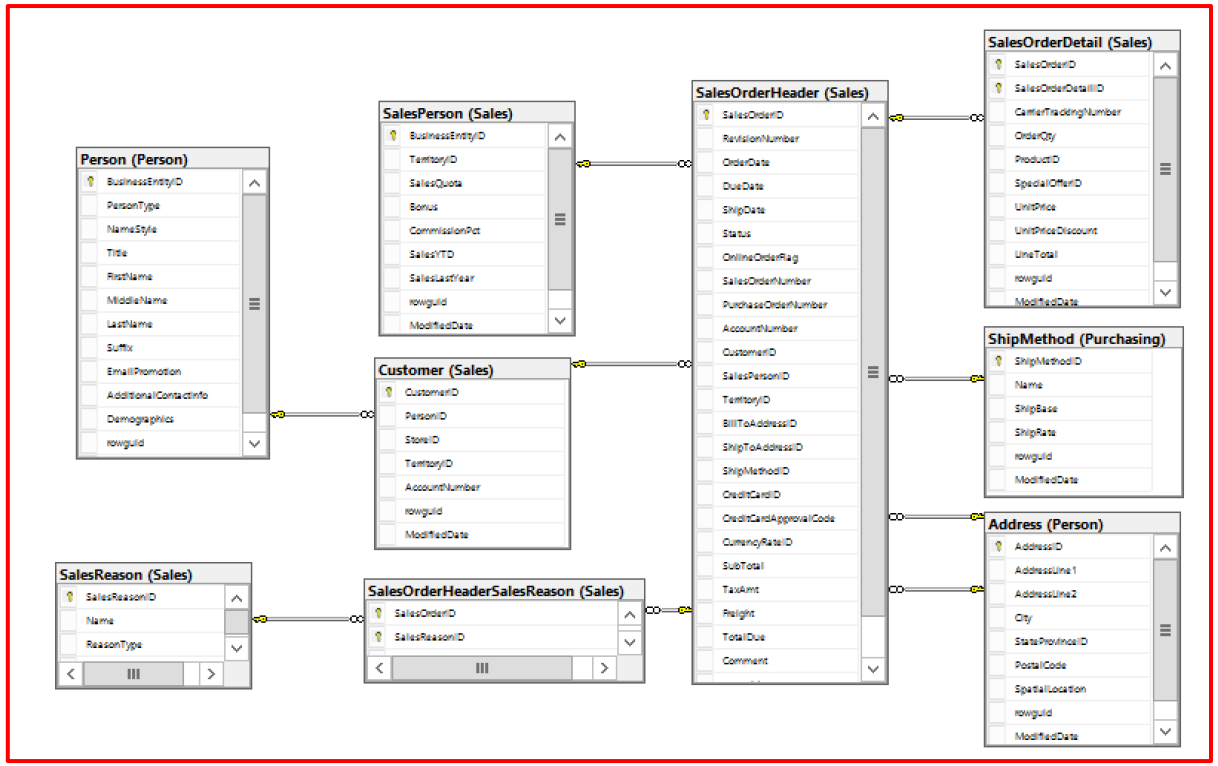
https://www.forbes.com/sites/adrianbridgwater/2018/07/05/the-13-types-of-data/#f2ab12e33624

Databases

- A database is a collection of data usually on a computer
  - A database typically stores structured data
- A ‘flat file’ is the simplest database
  - Single file
  - Most data are stored in this format (EXCEL, CSV)
  - Not efficient for large data
- A ‘relational database’ stores data in multiple tables
  - A table is a 2D (row x column)
  - Columns contain attributes or features
  - Rows contain values for each attribute
  - A relational database has multiple tables
    - Each table has a unique identifier attribute (Primary Key)
  - Primary Key of one table is used to connect values to another table

Relationships between table are used to join and query relational databases

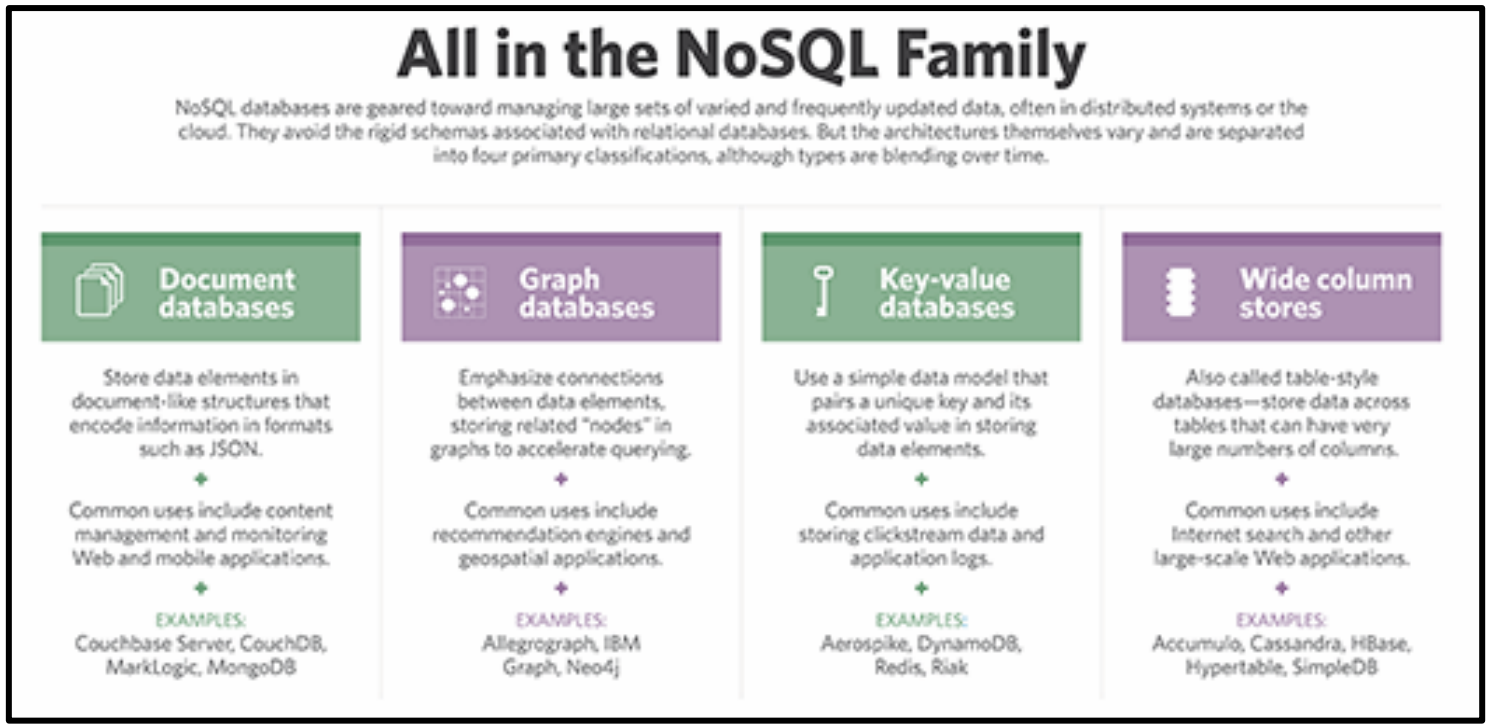
Data in a relational database is queried using the structured query language (SQL) Command



Databases these days store not just numeric and text data but also images, maps, audio, video files

Databases – Recent Trends

- A recent trend has to been to mine ‘unstructured data’
  - Text data that is unstructured
  - Data in documents, tweets, reports, posts
- Relational databases are not good with unstructured data



NoSQL which stands for ‘not only SQL’ is becoming popular in its use

Data and Databases – Cloud Services

- Most of the data these days are being stored on the ‘cloud’
- Data storage is distributed
- Data needs to be downloaded from the cloud
  - A website shows the data which needs to be downloaded
- Web scraping is the process of extracting data from a website and storing locally on your computer for data analysis

Data you need is seldom stored in the format you want it. It may have

Data Munging or Data Wrangling is the process by which data is ready compatible for further analysis

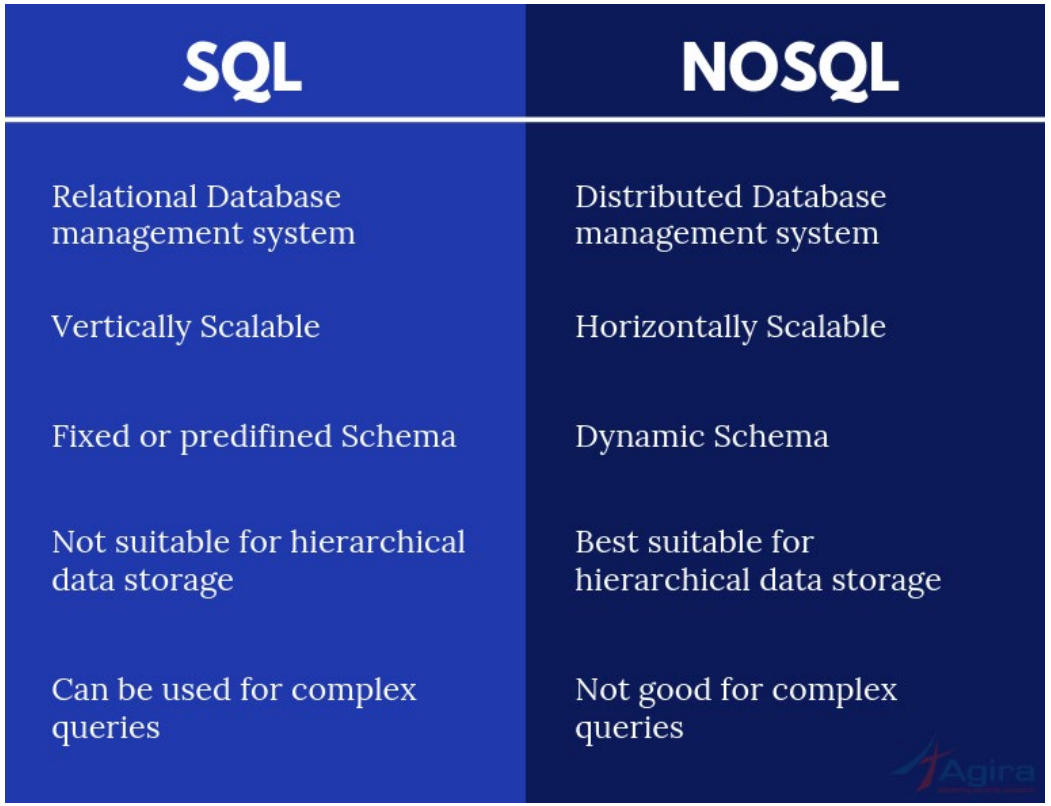
Data Munging typically refers to manual methods

Data Wrangling typically refers to use of automation for data cleaning

Both words are also used interchangeably

Data and Databases

- A good amount of time is spent munging and wrangling with data in ‘real-world’ data science projects
- Our focus is not to dwell on this issue in the interest of time
- Students are encouraged to learn more about SQL and relational databases
  - as well as ways to archive and query unstructured data



Several books, tutorials are available online (My SQL and POSTGRE SQL are two popular open-source RDBMS)

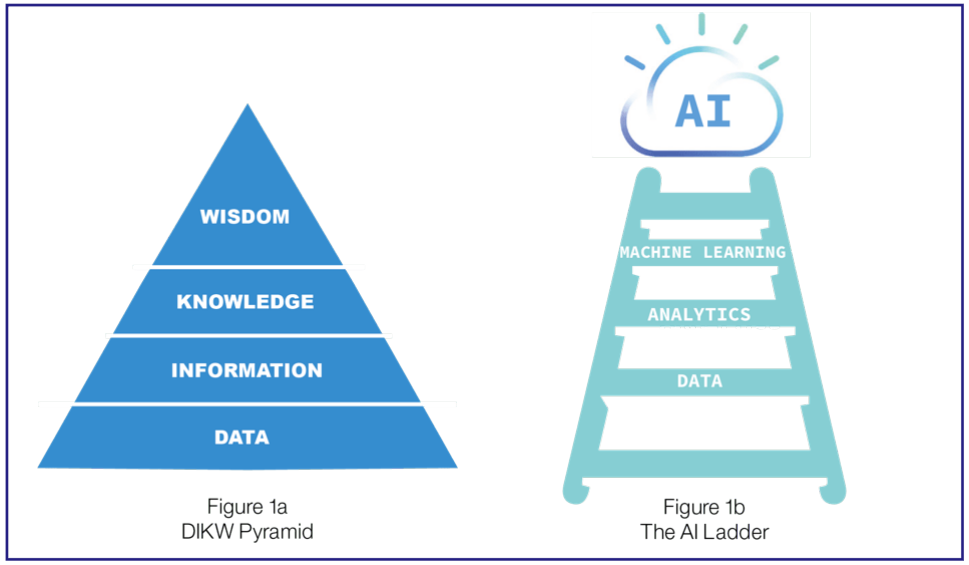


Figure 1a
DIKW Pyramid

Figure 1b
The AI Ladder

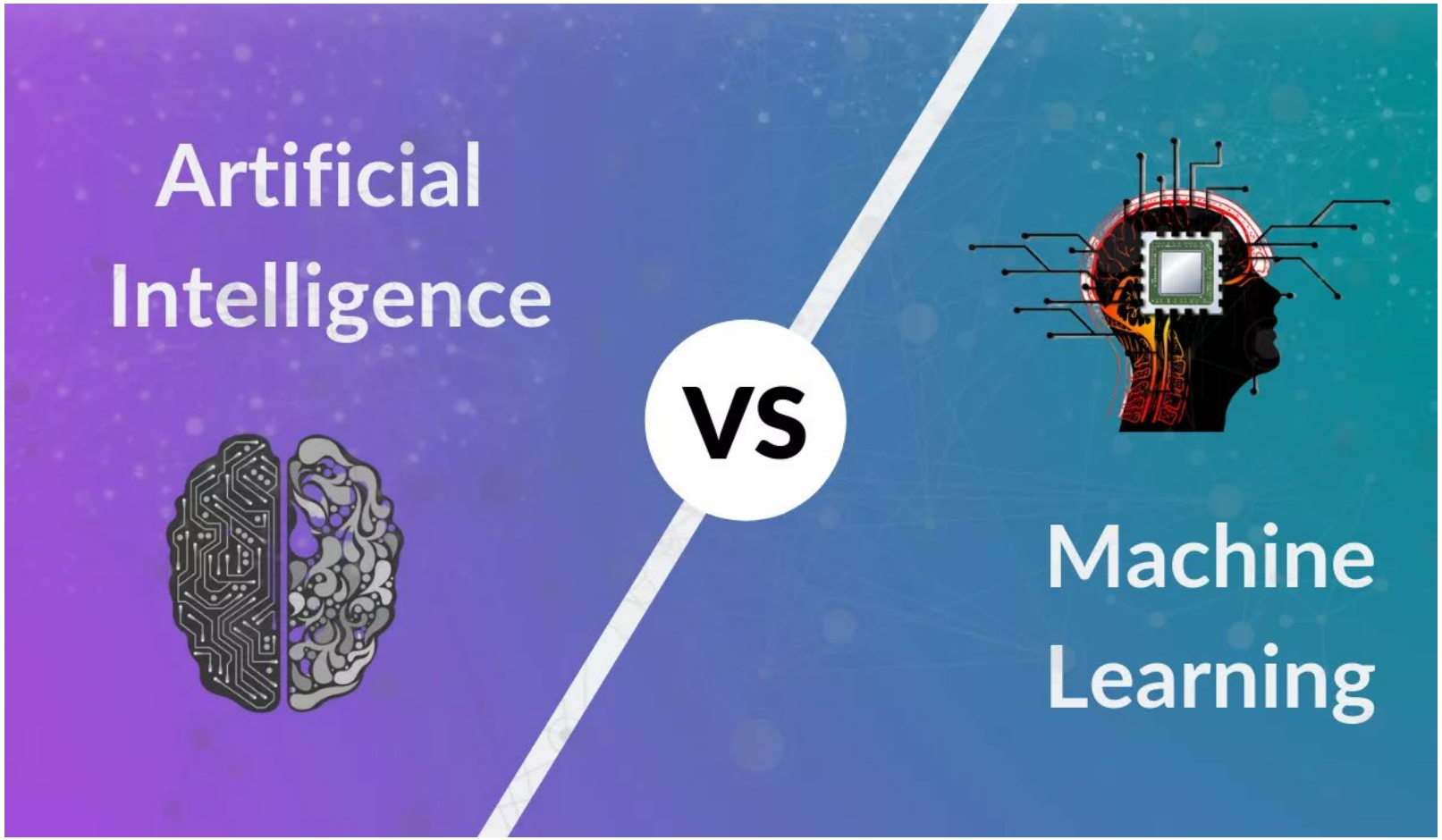
https://www.ibm.com/cloud/garage/architectures/dataAnalyticsArchitecture

Machine Learning

Study of algorithms and models that computers use to perform specific tasks

Machine Learning is a subset of Artificial Intelligence

AI versus Machine Learning



Artificial Intelligence (AI) – Human Cognition

- AI is a discipline of Computer Science that focuses on incorporating cognition into computers
- Create computers or computer systems that can act like humans
- Cognition is fundamental to human intelligence
  - Ability to learn
  - Ability to think

The end goal of AI is to create computers with wisdom

Weak AI → Moderate AI → Strong AI

Elements of Human Cognition

Human cognition is:
conscious and unconscious,
concrete or abstract,
intuitive and conceptual
It encompasses processes such as memory, association, concept formation, pattern recognition, language, attention, perception, action, problem solving and mental imagery

Piaget model Cognitive Development in Children

Stage	Years	Description
Sensorimotor	0 – 2	Causality, time and space
Pre-operational	2 – 7	Memory, imagination, basic relationships, irreversible thinking
Concrete operational	7 – 12	Reversible thinking, non-ecocentric thinking,
Formal operational	12+	Flexible and abstract thinking, hypothesis formulation and testing

Machine Learning (ML)

- Machine Learning is a subset of Artificial Intelligence
- Study of algorithms and models used to perform specific tasks
- ML focuses on creation of computer programs that can access data and learn from it
- Word was coined by Arthur Samuel in the year 1959

- Machine Learning is closely related to several other fields
  - Statistics
  - Optimization and operations research
  - Data mining and Information theory

The end-goal of machine learning is to create computers that can transform data to information and knowledge

Machine Learning methods and algorithms seek to mimic the learning processes and styles employed by humans

What is Machine Learning

- Machine learning broadly is making machines (computers) mimic human behavior
  - Learn and adapt
  - A Subset of Artificial Intelligence
- Some tasks where we seek to machines to learn include
  - Classification tasks
    - Divide and group objects
  - Regression (prediction) tasks
    - Predict the response of a variable using other variables
    - Forecast values into future times
  - Create Association Rules
    - IF-THEN rules

Machine Learning Methods typically Convert Data to Information or Knowledge

Machine Learning

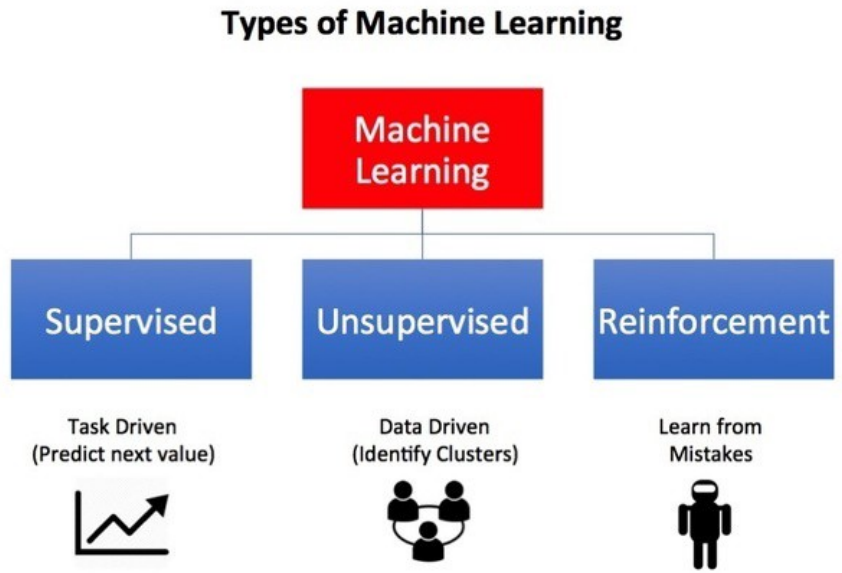
- Machine learning is gaining popularity because:
  - Large volumes of data are becoming accessible worldwide
  - Computational sources have become cheap
  - Computation can be carried out in parallel
  - It is easier to implement machine learning algorithms
    - Software availability
- Machine learning can map highly nonlinear phenomena
  - Therefore, it can capture the dynamics associated with natural systems
    - Very useful in all civil engineering fields

Coupling Physics-Based Models with Machine Learning Methods is a hot topic in Civil Engineering

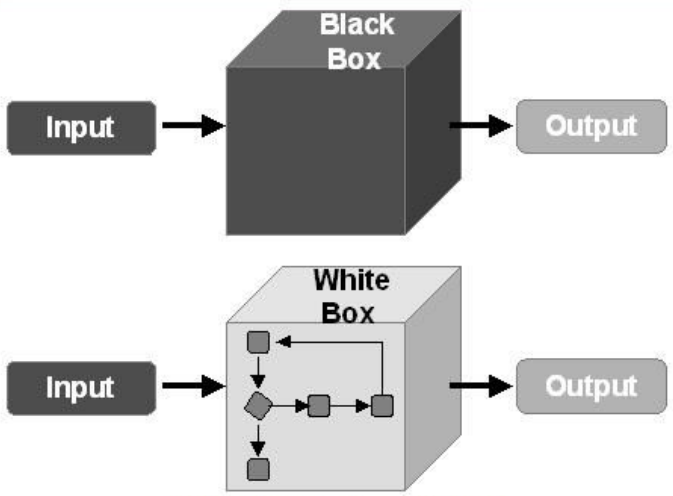
Traditional Physics-based Models often linearize Civil Engineering Systems and therefore cannot fully capture the system's dynamics

Machine Learning

- ML models can be categorized in several ways
- Broadly they can be classified as
  - Supervised learners
    - Need both input and output data
  - Unsupervised learners
    - No need for output data
  - Semi-supervised learners
    - Falls between the two extremes of supervised and unsupervised classificati…
  - Reinforcement learners
    - Interact with the environment and adapt (optimize) system behavior in time



- They can also be classified as “black-box” or While(Grey)-box models
  - Blackbox models
    - The learning/mechanisms by which the ML learns and processes data is unknown
  - While(grey)box models
    - The learning is more transparent



Supervised Learning

- The goal is for the algorithm to seek to map the relationship between Y (output) and X (input)
- Learning is from a training dataset
  - You show examples to the algorithm to facilitate the learning process
- Supervised learning is commonly used for 2 classes of problems
  - Classification
    - Used with discrete variables
  - Regression
    - Used with continuous variables

Supervised learning is used when the data are labeled (i.e., have an output) – Primary Goal is to make Predictions

Unsupervised Learning

- The goal of unsupervised learning is primarily to cluster the data into a set of groups or seek associations within the data
- Clustering is often carried out on a multivariate dataset
- The number of groups (clusters) need to be known a-priori
  - In most techniques this is specified
    - K-means clustering
  - We can also optimize (estimate) it
    - Becomes supervised learning at that point
- Creation of association rules is another unsupervised learning method

The primary goal is to understand (clarify) the structure or distribution of the data

Semi-Supervised learning

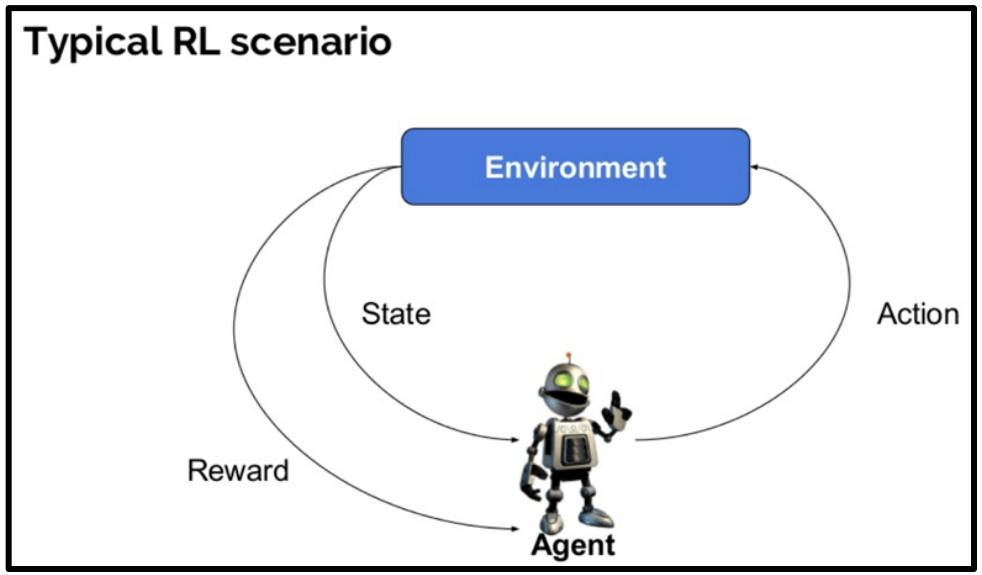
- Useful when some data are labeled but a large amount of data is not labeled
- Easy to collect inputs but hard to collect outputs
- Use unsupervised learning to cluster input data (step 1)
- Make predictions for each cluster (step 2)
- Use the inputs and outputs from step 1 and 2 to create a supervised learning model

Combining supervised and unsupervised learning methods create enormous possibilities to draw insights from available data

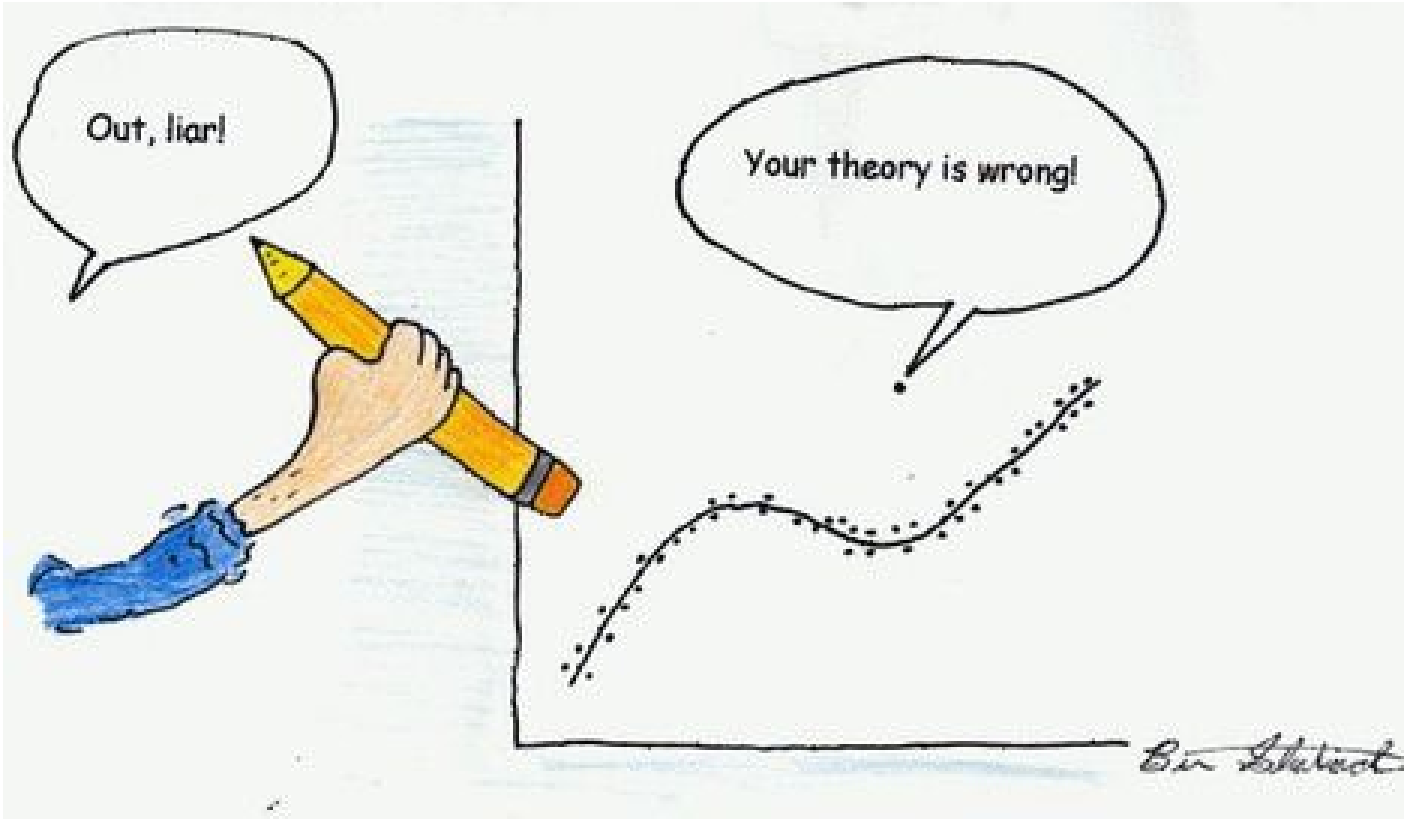
Reinforcement Learning

- Reiforcement learning is an adaptive learning paradigm
- The Agent identifies an optimal state by interacting with the environment
  - A reward mechanism is used in this identification
- Learning to progress towards the optimal goal is guided by rewards that are obtained by taking certain actions

Reinforcement Leanring has wide range of Civil Applications ranging from Intelligent Transportation to Water Resources Management



Dynamic Programming, Markov Methods are some common traditional RL Methods

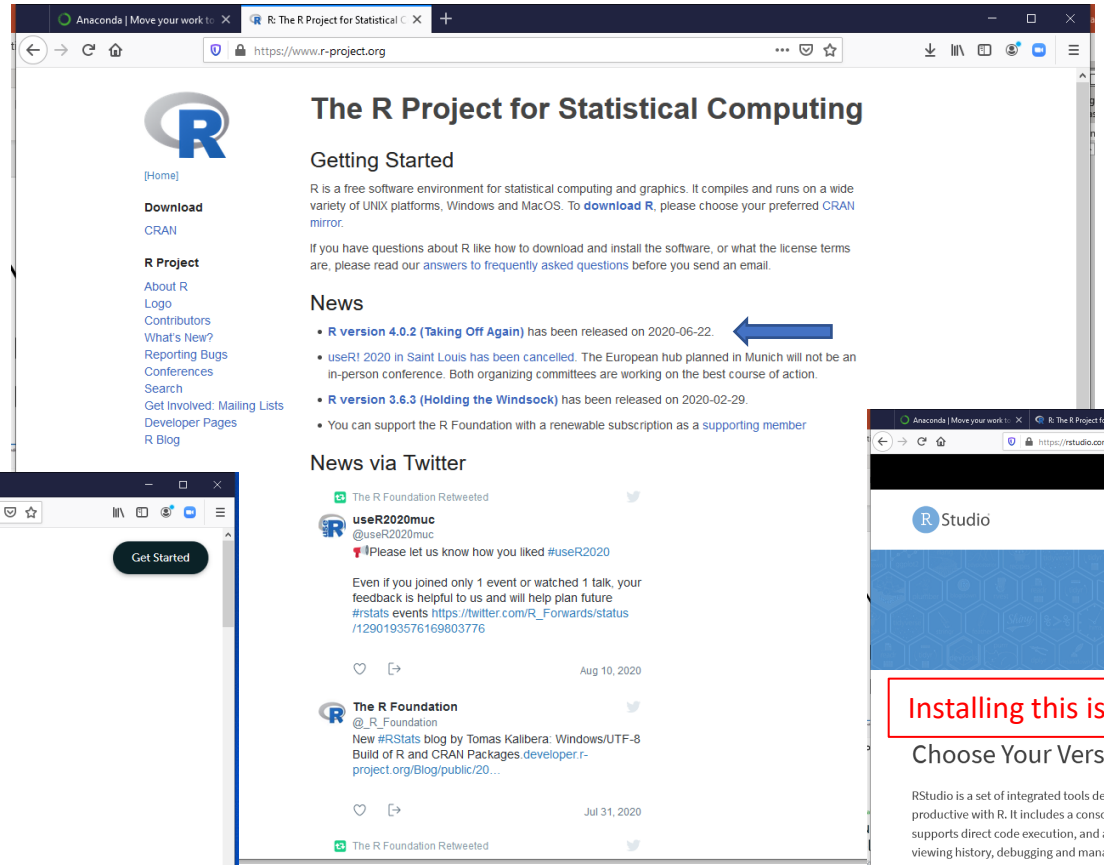


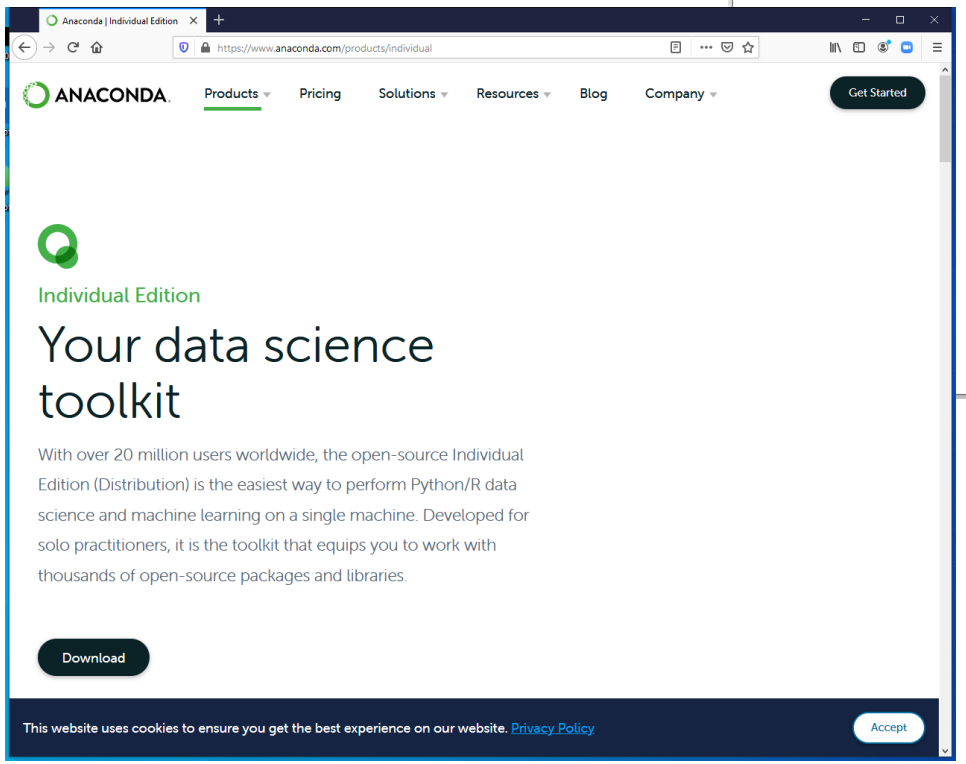
Challenges of Machine Learning

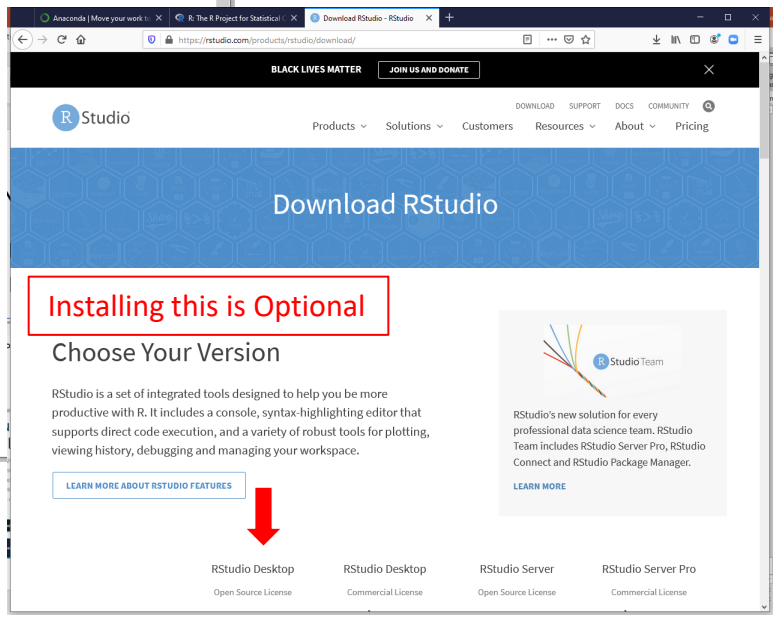
- The relationship between the input and (output) is not fully known or understood
  - How do we select which inputs to use in the model?
  - How do we justify our selection?
- Which algorithm/technique to choose?
  - Why are we choosing this algorithm?
- The architecture of the model is not known a priori even for a selected algorithm?
  - What is the functional relationship between a given input and output?
  - Why are we selecting (or how do we select) an appropriate function?
- How do we train the model?
  - This is essentially a parameter estimation or optimization model
  - There are many choices
- How do we know that the model we developed is good?
  - Test with data the model has not seen – How much?
  - What if there is not enough data to build and independently test the model?

Next Steps

- Install R
- Install Anaconda







Please have these programs installed prior to the next class

Let me know if there are any questions!!