Machine Learning Algorithms

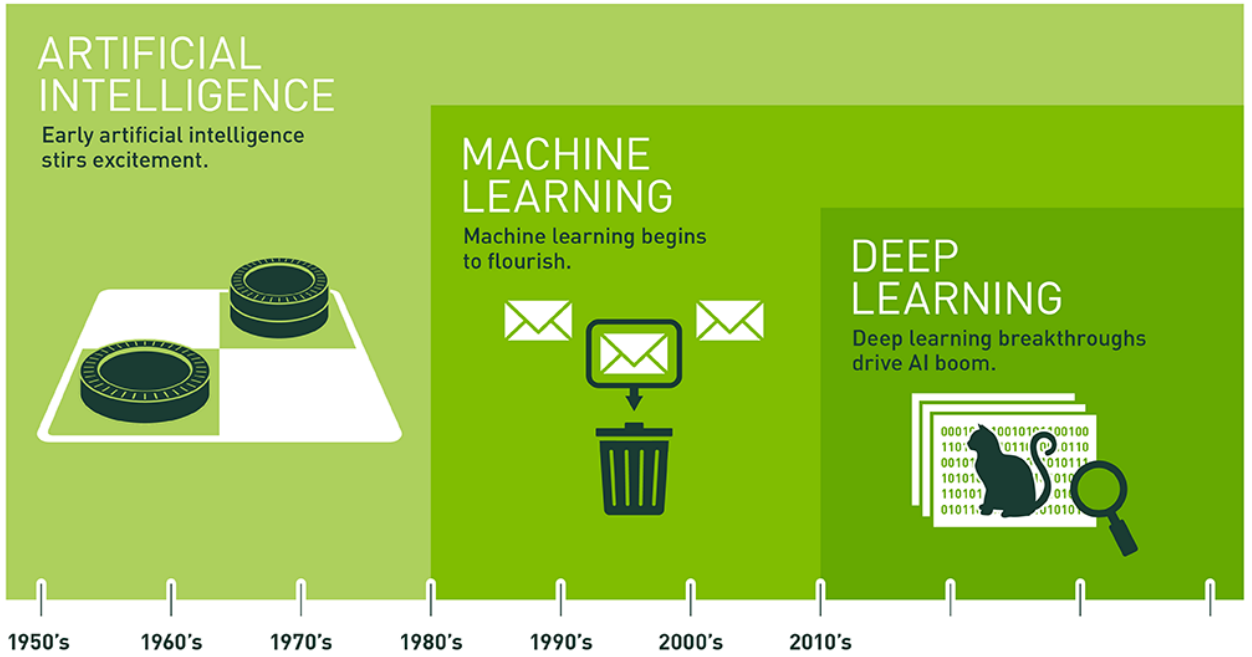
PRESENTED BY: DAVID SWANSON

What is Machine Learning?

- Learning without explicit programming
- Sub-set of artificial intelligence
  - Predictions based on data
- Tom M. Mitchell's Definition
  - "A computer program that learns from experience E with respect to some class of tasks T and performance measure P. If it's performance at tasks in T, measured by P, improves with experience E."

Brief History of Machine Learning

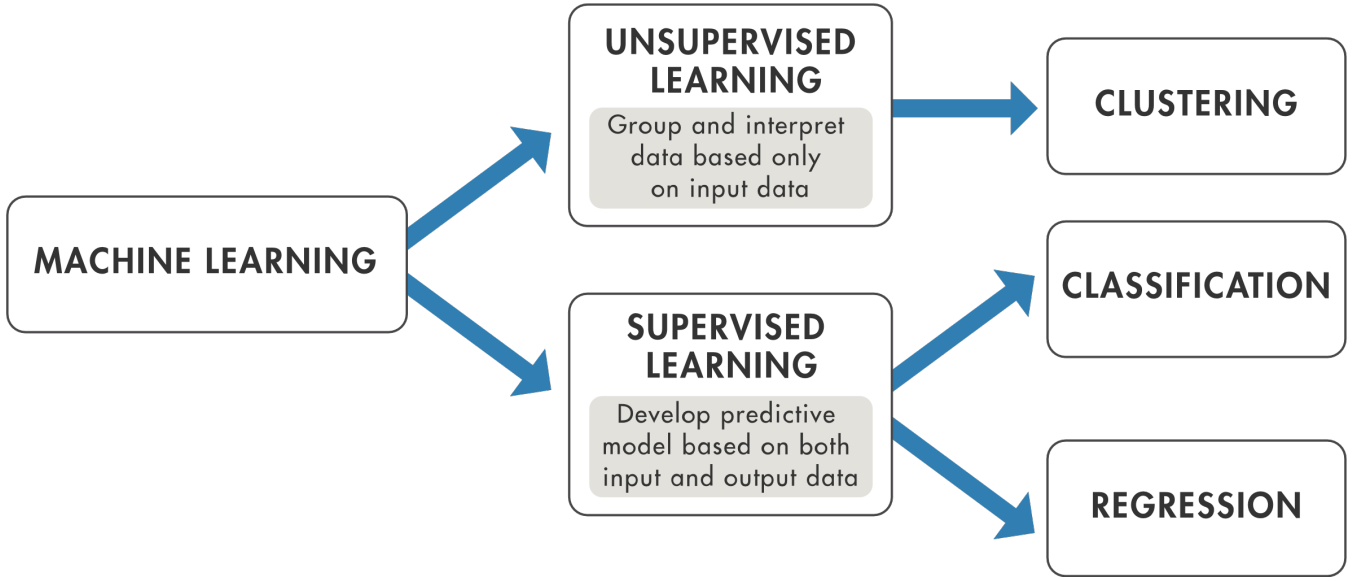
- "Machine Learning" – Coined in 1959
- Evolved from topics in artificial intelligence
- Emphasis on logical, knowledge-based approach caused division
- Reemerged as a separate field in 1990s
- New goal of solving practical problems.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Types of Machine Learning

- Classification based on existence of "feedback".
- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning



Supervised Learning

- Input X mapped to output Y by function Y = f(X)
- Called supervised because presence of "teacher"
- Learning continues based on performance
- Further groupings: Regression and Classification Problems
  - Classification:
    - Output is a category
  - Regression:
    - Output is a variable with real value

Unsupervised Learning

- Input data (X), No corresponding output
- Goal: model data distribution in order to learn about data
- Called unsupervised because of the lack of a “teacher”.
- Further Grouped into:
  - Clustering:
    - Discover inherent groupings in the data
  - Association:
    - Discover rules that describe large portions of data

Semi-Supervised Learning

- Large amounts of input data (X), only some data labeled as (Y)
- In between other options
- Only some data is labeled because of time constraints
- Used to discover structure of input and make best case predictions
- Many real-world machine learning problems fall into this area
- Example:
  - Photo archive where only some images are labeled (e.g. dog, cat, person)

Most Used Algorithms

- Supervised Learning:
  - Decision Trees – Tree-like graph of decisions and their possible outcomes
  - Logistic Regression – Models binomial outcomes with one or more variables
  - Support Vector Machines – Binary classification algorithm
  - Ensemble Methods – Construct a set of classifiers then classify by predictions
- Unsupervised Learning:
  - Clustering Algorithms – Grouping sets of objects in similar clusters
  - Principal Component Analysis – Convert observations to principal components
  - Independent Component Analysis – Reveals hidden factors in sets

Parallel Computing Application

- Main slow-down of machine learning algorithms:
  - Large data sets
- Algorithms consist of mostly linear algebra
- Operations on signal data points with little dependencies
- Ideal conditions for embarrassingly parallel algorithm application
- Ongoing research for parallelizing more complex problems (PML)

References and Literature

- SAS article on Machine Learning
  - https://www.sas.com/en_us/insights/analytics/machine-learning.html
- Supervised vs. Unsupervised Learning
  - https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
- Most common Machine Learning Algorithms
  - https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html
- Parallel Computing for Machine Learning Article
  - http://people.eecs.berkeley.edu/~mme/cs267-2016/hw0/results/pan.pdf