BOSTON UNIVERSITY

Introduction to Machine Learning

Boston University
AD 688 Web Analytics
Michael Golden

What is Machine Learning?

"Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel, IBM 1959

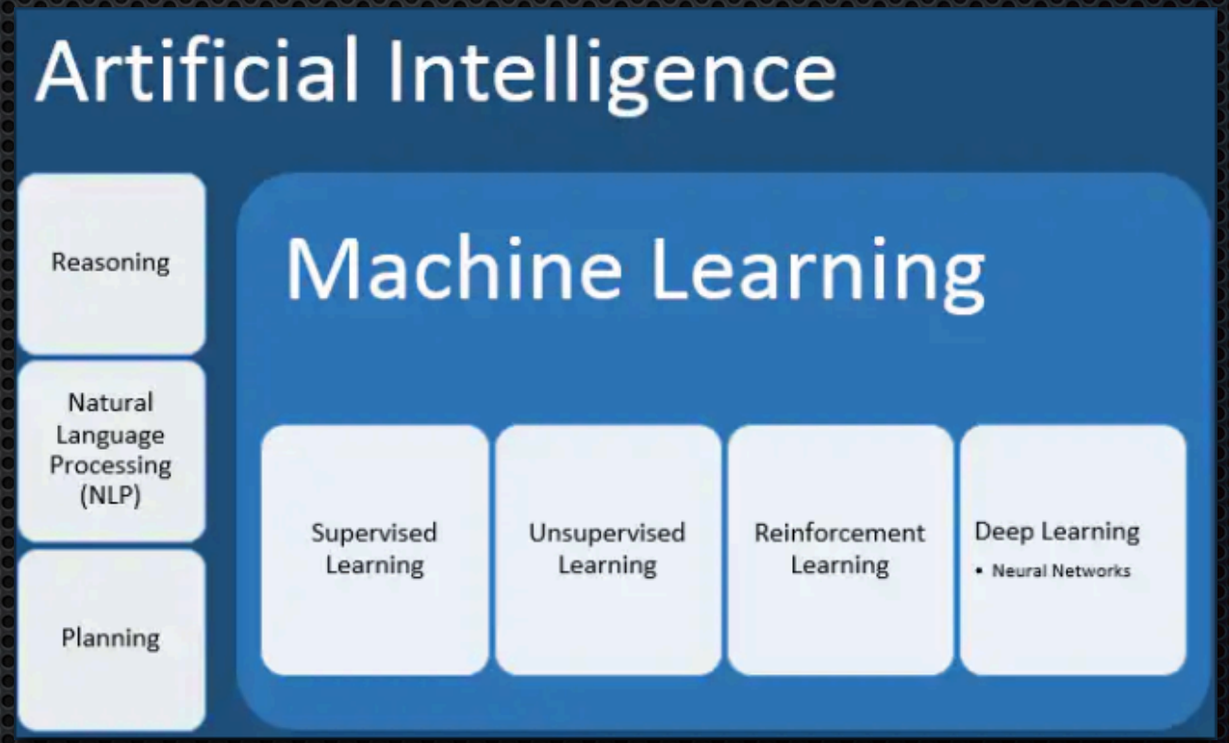
BOSTON UNIVERSITY

The Machine Learning Process

- Clean your data
- Select a proper algorithm for building a prediction model
- Train your model to understand your data
- Evaluate your model
- Parameter tuning
- Deploy model

Approaches to Machine Learning [1]

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Deep Learning



BOSTON UNIVERSITY

Supervised Learning Techniques

- Classification - the process of predicting discrete class labels or categories; using labeled data
- Regression - the process of predicting continuous values
- Anomaly Detection - the process of discovering abnormal and unusual cases

Unsupervised Learning Techniques

- Clustering - finds patterns and groupings from unlabeled data
- Dimension Reduction - reduce redundant data
- Density Estimation - mostly used to explore the data to find some structure within it
- Market Basket Analysis - based on the theory that customers of certain items are more likely to purchase another group of items

BOSTON UNIVERSITY

Major Machine Learning Techniques

- Regression/Estimation — predicting continuous values
- Classification — predicting the item class/category of a case
- Clustering — grouping of data points or objects that are similar
- Associations — used for finding frequent co-occurring items/events
- Anomaly Detection — discovering abnormal and unusual cases
- Sequence Mining — predicting next events
- Dimension Reduction — reducing the size of data
- Recommendation Systems — associating people's preferences and recommending items

Real-World Applications

- Infectious disease researchers searching for the cause of a mysterious polio-like illness (AFM -Acute Flaccid Myelitis) that is paralyzing children [2]
- Converting complex media and raw data into useful information, such as automatic speech and handwriting transcription, and automatic face recognition [3]
- Serve as a cost and time-saving tool when compared with manual classification of public health expenditures. [4]
- Brain tumors are often classified by visual assessment of tumor cells, yet such diagnosis can vary depending on the observer. Machine-based learning approaches are being developed to aid the diagnosis of clinical samples and classifying brain tumors [5]

BOSTON UNIVERSITY

References

1. IBM. (n.d.) Retrieved from https://www.ibm.com/analytics/machine-learning

2. Reardon, S. (2018). Machine learning hunts for cause of paralyzing illness. *Nature*, 564(7735), 170.

3. Camastra, F., & Vinciarelli, Alessandro. (2008). Machine learning for audio, image and video analysis theory and applications (Springer computer science). London: Springer.

4. Brady, E., Leider, J., Resnick, B., Alfonso, Y., & Bishai, D. (2017). Machine-Learning Algorithms to Code Public Health Spending Accounts. *Public Health Reports*, 132(3), 350-356.

5. Wong, D., & Yip, S. (2018). Machine learning classifies cancer. *Nature*, 555(7697), 446-447.

BOSTON UNIVERSITY