The Present and Future of Machine Learning

Jacob Staples and Matthew Caird

†Virginia Polytechnic Institute and State University

What Machine Learning Is

Machine Learning is the process of how AI would interpret data. It uses algorithms to process vast amounts of data and separates them into groups based on what they already know [10]. The three different main types of machine learning [9]:

- **Supervised:**
  - Humans give machine an algorithm
  - Machine displays output, humans change algorithm to achieve desired output and minimal bias
  - Machine is expected to learn pattern
- **Unsupervised:**
  - Human gives machine input data
  - Machine learns as much as possible and attempts to solve problem
  - Receives errors and is expected to learn from them
- **Reinforcement:**
  - Algorithm learns as much about the environment as possible
  - Attempts to get ideal output with minimal risk
  - Humans provide feedback to allow algorithm to learn

Data Mining and its Relation to Machine Learning

- **Data Mining** is a process of finding and extracting useful data from vast, raw data. This is often done by using software or programs, which is often machine learning, to find patterns in data and to then extract and make use of the data.
- **Differences**:
  - Machine learning is designed to grow and learn from itself, while data mining can't do that as it is given a pre-determined set of rules and parameters to make use of a data set.
  - Data mining's purpose is to gather data and determine a particular outcome or trend, while machine learning's is to complete tasks, harvest data, gain experience, and learn by using algorithms to do so.
  - Data mining relies on human intervention, while machine learning is designed so that it can teach and run by itself.

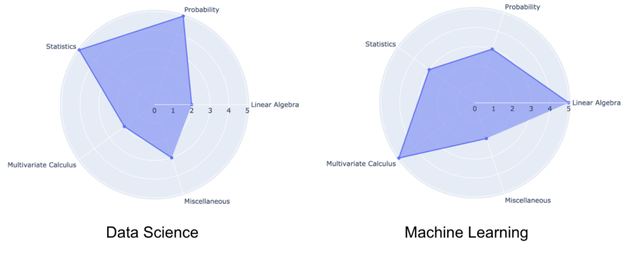


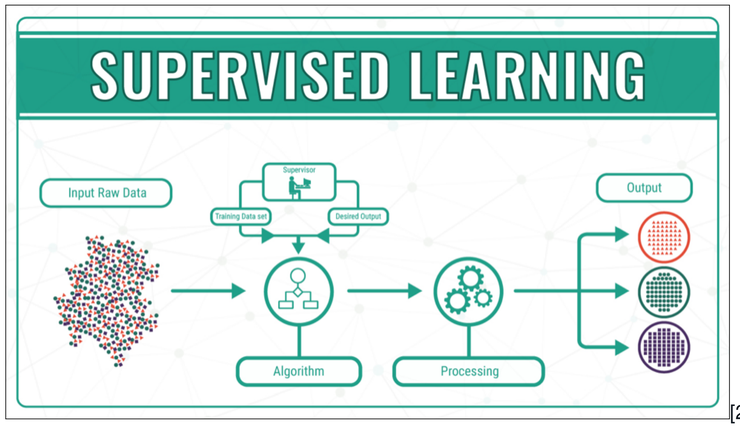
Fig. 1: Concept comparison of Data Mining and Machine Learning [5]

- **Similarities**: Both processes are used to solve complex problems, and the algorithms they use to find patterns in data are typically quite similar. [1]

The Math in Machine Learning and Specifically In Supervised Learning Algorithms

There are many concepts that can be useful in machine learning, but the four main ones that come up most often are Linear Algebra, Multivariate Calculus, Probability, and Statistics. [5]

- **Linear Algebra**: This acts as a stage or platform in which machine learning algorithms produce their results. Machines can solve systems of linear equations much quicker than humans can, which is useful as data is often large and require these types of calculations.
- **Multivariate Calculus**: This is typically used when dealing with numerical optimisation. Gradient decent, which consists of moving in the direction of steepest descent in iterations, is used to minimize a function and update the parameters of the model.
- **Probability**: This is used in machine learning to carry out hypothesis testing. One important theorem for this is Bayes' Theorem, which is $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. This is used to find a new assumption, or posterior, based on previous assumptions and constraints. In machine learning, this is done many times to produce a more accurate parameter result.
- **Statistics**: Some topics show up here in measures of central tendency, spread of data, distributions, and hypothesis testing.



[2]

Fig. 2: Simple diagram of how Supervised Learning works

How a supervised learning algorithm works:
Given a set of N training examples, denoted as pairs $(x, y) = (\text{feature vector}, \text{label/class})$, the algorithm seeks some function $g : X \to Y$, where X is the input space and Y is the output space. This function g is an element of the space of possible functions G, called the hypothesis space, and g can be represented using a scoring function $f : X \times Y \to \mathbb{R}$ so that g is defined by giving the highest score y value represented in $g(x) = \underset{y}{\operatorname{argmax}} f(x, y)$.

Most learning algorithms are probabilistic models where g takes the form of a conditional probability model $g(x) = P(y|x)$ or f takes the from of a joint probability model $f(x, y) = P(x, y)$. f or g can be chosen using empirical risk minimization, which finds the function to best fit the training data, or structural risk minimization, which imposes a penalty function to control bias and variance tradeoff. A loss function $L : Y \times Y \to \mathbb{R}^{\geq 0}$ is defined to measure how well the function fits the training data and then the risk $R(g)$ of function g is defined as $R_{emp}(g) = \frac{1}{N} \sum_i L(y_i, g(x_i))$ [7]

Machine Learning Used Today

Machine learning is used across many fields in today's world. Some examples of machine learning today include:

- **Law:** A study found that with a trained algorithm performed much better than experienced judges at determining whether the person could wait for trial at home or in jail [6]
- **Translation:** Machine learning is used in the translation of foreign languages to one's native language [6]
- **Finance:** Machine learning is used today in the prevention of money laundering in which PayPal uses deep learning to filter through transactions and determine legitimacy [6]
- **Entertainment:** Streaming services use machine learning to help the user find more shows they may be interested in [4]

Future Questions about Machine Learning

With Machine learning being used in a wide variety of fields today one question towards the future is where could machine learning be seen more in than it is today:

- **Healthcare:** In healthcare there has been an increase in the usage of Machine learning to help with:
  - Diagnosis of diabetic retinopathy [8]
  - Prognosis of cancer [3]

References

[1] *Data Mining Vs. Machine Learning: What Is the Difference?* Apr. 2020. URL: https://www.simplilearn.com/data-mining-vs-machine-learning-article.

[2] *Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning*. Apr. 2020. URL: https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/.

[3] *Machine Learning Is The Future Of Cancer Prediction*. Apr. 2020. URL: https://towardsdatascience.com/machine-learning-is-the-future-of-cancer-prediction-e4d28e7e6dfa.

[4] Bernard Marr. "27 Incredible Examples Of AI And Machine Learning In Practice". In: *Forbes* (Apr. 2018). URL: https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice.

[5] *Mathematics behind Machine Learning – The Core Concepts you Need to Know*. Apr. 2020. URL: https://www.analyticsvidhya.com/blog/2019/10/mathematics-behind-machine-learning/.

[6] R. Ramzipoor. "10 Surprising Ways Machine Learning is Being Used Today". In: *InformationWeek* (Dec. 2017). URL: https://www.informationweek.com/big-data/ai-machine-learning/10-surprising-ways-machine-learning-is-being-used-today/a/d-id/1330613.

[7] *Supervised Learning*. Apr. 2020. URL: https://en.wikipedia.org/wiki/Supervised_learning.

[8] *The Future of Machine Learning*. Apr. 2020. URL: https://towardsdatascience.com/the-future-of-machine-learning-ce0a9dc18cb8.

[9] *Types of Machine Learning Algorithms You Should Know*. Apr. 2020. URL: https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861.

[10] *What is machine learning?* Apr. 2020. URL: https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/.