

Machine Learning for Business Leaders

Demystifying Artificial Intelligence

Eszter Windhager-Pokol
Head of Data Science

Contents

| | |
|--|---|
| Introduction..... | 2 |
| What is Machine Learning? | 2 |
| How ML differs from traditional problem solving..... | 2 |
| Features and Labels, learning data science jargon..... | 3 |
| Types of machine learning: | 4 |
| Supervised | 4 |
| Common algorithms used in supervised learning: | 5 |
| Applications of supervised learning include | 6 |
| Unsupervised learning | 6 |
| Reinforcement learning | 6 |
| Conclusion..... | 7 |
| About the Author | 7 |

Introduction

If you have ever interacted with the chat function on a website, there is a good chance you experienced the power of Machine Learning (ML) first hand. From recommender systems on streaming services to personal assistants like Siri or Alexa, ML is nearly everywhere today. Applications such as predictive maintenance in manufacturing, recommender systems for streaming services, fraud detection in banking and many others demonstrate the potential for ML to revolutionize business functions.

Despite the seeming ubiquity of ML, a 2018 research report by McKinsey states that only 39% of surveyed firms embedding ML in daily business functions and a further 31% having completed ML pilot projects.ⁱ Another key finding of the survey was the 43% of respondents said a lack of clear strategy for AI was the most significant barrier in adoption.

This white paper demystifies Machine Learning for non-technical business stakeholders to better collaborate in, and derive more value from, data science initiatives.

What is Machine Learning?

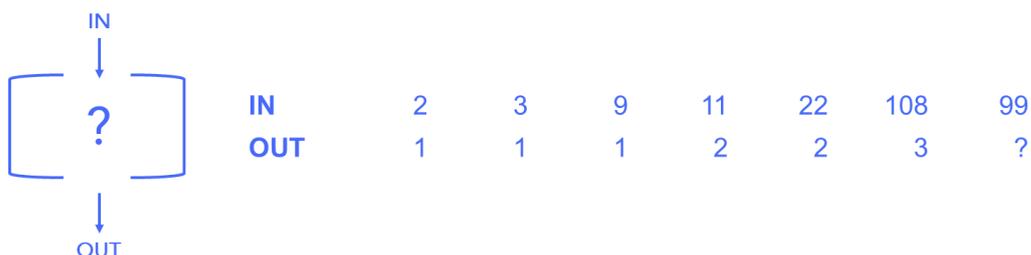
Many people are surprised to learn that the term Machine Learning was first introduced as early as 1959 by Arthur Samuel when he attempted to teach computers to play checkers. The most cited definition of ML comes from Tom M. Mitchell:

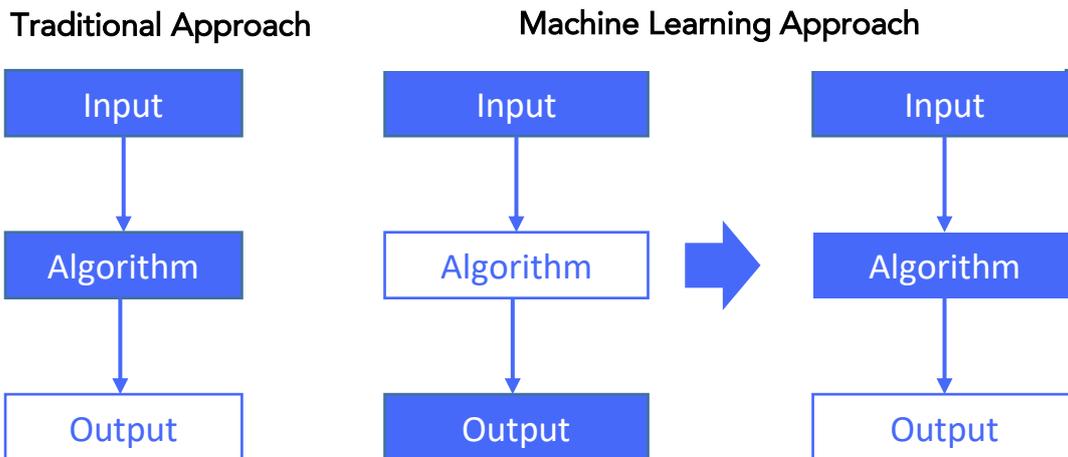
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."

For the non-technical person, it's easier to think of ML a series of algorithms – sets of instructions that a computer performs to solve a task – that enable computers to learn how to solve tasks and improve their performance as they process data. In the context of the explosion of data and computing power in the last several decades, it's easy to see how ML has emerged as a powerful tool to solve a wide variety of difficult problems.

How ML differs from traditional problem solving

The traditional approach to computational problem solving involves a known input and an unknown output. Data scientists use machine learning algorithms to solve the problem, i.e. to find the functions and patterns in the data to reproduce outputs for different inputs.





Two competitions between man and machine best demonstrate the difference between the two approaches. In 1996, IBM's Deep Blue computer programmed to compete in chess took on world champion Gary Kasparov. Deep Blue used a fast, computational rule engine that systematically enumerated all possible movements. Gary Kasparov won three of five chess matches, drawing the other two.

Fast forward to 2015, when a computer program named AlphaGo, designed by DeepMind technologies, later acquired by Google, competed and won in the board game Go. Originating in Asia, many consider Go to be significantly more difficult for computers to win than other games such as chess because it contains far more possible moves than chess.

Instead of employing a rule engine like Deep Blue, AlphaGo used machine learning to analyze previous games and games it played against itself, a process known as training, to develop new strategies. It then applied the strategies learned to new input. Where Deep Blue failed against the top-rated chess player, AlphaGo beat Ke Jie, the number one ranked player at the time.

Features and Labels, learning data science jargon

When talking with data scientists about ML projects, you'll often hear the following terms: features, labels, training data, supervised and unsupervised learning. These definitions are critically important to understanding ML and effectively collaborating with your data science team.

A feature is an individual measurable property of a phenomenon being observed.ⁱⁱ Simply put, a feature is the input and a label is the output. If we are trying to predict the credit worthiness of a loan applicant, we can take many factors, i.e. features, into consideration such as applicant age, geography, employment history and more. The label in this example would be a credit score or the risk of default.

Types of machine learning:

There are three types of ML models: supervised, unsupervised, and reinforcement learning. Each type of ML uses a variety of algorithms.

Supervised

An easy way to think of supervised learning is learning by example. In supervised learning a data set containing training examples with associated correct labels exists. A classic example of supervised ML is distinguishing between cats and dogs.

How humans see dogs and cats



Data scientists can create a model, for example consisting of artificial neural networks, and feed it with pictures of cats and dogs. This process is called training the model. If enough data is used, the model learns what distinguishes a dog from a cat.

How computers see dogs and cats

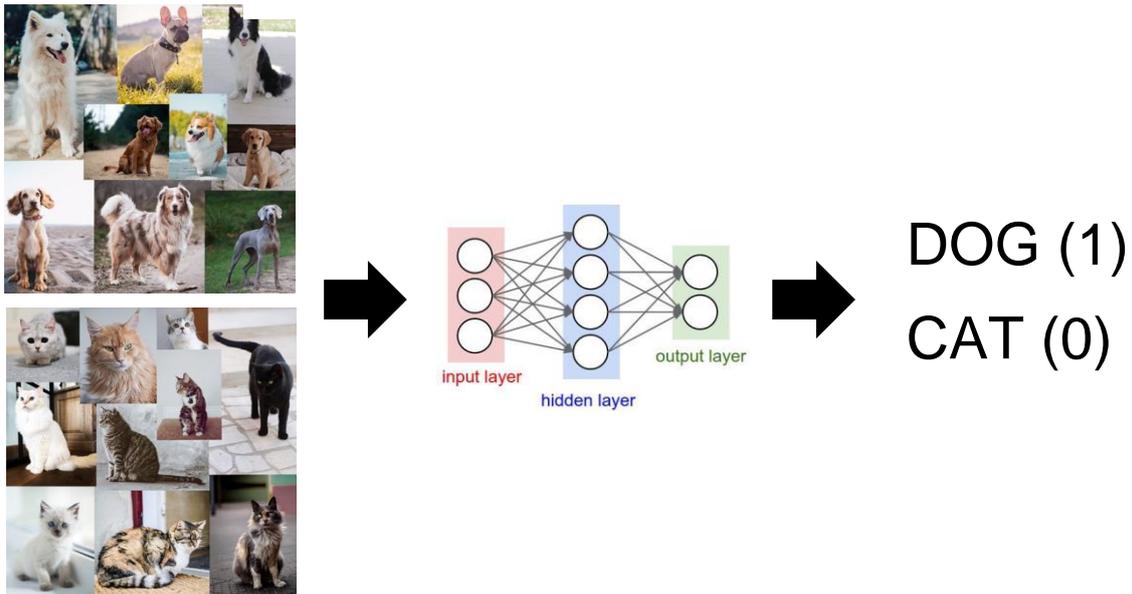
```

11010110101101001101100011
01101011110111111010000101
01101111011110111101110110
10110101101101011111011110
11110001101101111010110100
10101000010010001010001101
01100100101010011101011010
11010011011000110110101111
01111110100001010110111101
11101111011101101011010110
11010111110111101111000110
11011110101101001010100001
00100010100011010110010010
  
```

```

11010110101101001101100011
01101011110111111010000101
01101111011110111101110110
10110101101101011111011110
11110001101101111010110100
10101000010010001010001101
01100100101010011101011010
11010011011000110110101111
01111110100001010110111101
11101111011101101011010110
11010111110111101111000110
11011110101101001010100001
00100010100011010110010010
  
```

Training Models with Data



Common algorithms used in supervised learning:

Linear and non-linear regression – Regression algorithms help analysts model relationships between data points. Regression algorithms can quantify the strength of correlation between variables in a data set. In addition, regression analysis can be useful for predicting the future values of data based on historical values.

Decision Tree, and other tree-based methods – These include Random Forest, Gradient Boosting Machines, and LightGBM. Decision tree algorithms use a branching structure to illustrate the results of a decision. Decision trees can be used to map the possible outcomes of a decision. Each node of a decision tree represents a possible outcome. Percentages are assigned to nodes based on the likelihood of the outcome occurring.

Bayesian models - Bayesian algorithms allow data scientists to encode prior beliefs about what models should look like, independent of what the data states. These algorithms are especially useful when there is a lack of sufficient data to confidently train a model.

Neural networks and deep learning - A neural network attempts to mimic the way a human brain approaches problems and uses layers of interconnected units to learn and infer relationships based on observed data. A neural network can have several connected layers. Neural network models are able to adjust and learn as data changes. Neural networks are often used when data is unlabeled or unstructured. Deep learning is when a neural network has a large number of layers. It is being used in a variety of applications, including self-driving cars and in applications that radiologists use to help interpret medical images.

Applications of supervised learning include:

- Credit risk modelling
- Churn predictions
- Cross-sell models
- Customer Lifetime Value
- Fraud detection
- Predictive Maintenance
- Demand forecasting
- Customer segmentation

Unsupervised learning

Unsupervised ML, where the outputs, or labels, are unknown, is commonly used to detect anomalies, find groups of similar subjects such as customers or decrease the noise in data through dimensionality reduction. It is most appropriate when the problem involves a large amount of data that is unlabeled, such as social media data.

Clustering, or finding groups of similar customers, products, machines, stocks etc. is a very common unsupervised ML application. All objects in a cluster are more similar to each other than objects in other clusters. Clustering is a type of unsupervised learning because the data is not labeled. K-means or Hierarchical clustering algorithms interpret the parameters that make up each item and then group them accordingly.

It's important to classify the unlabeled data by finding patterns or clusters. For example, unsupervised learning is used for email spam detection since there are far too many variables in legitimate and spam emails for a person to flag. Instead, machine learning classifiers using clustering are applied in order to identify spam.

Dimensionality reduction refers to decreasing the noise in data and can help algorithms process data faster. Storing and analyzing data that isn't useful adds overhead to the computational power of a data platform.

Reinforcement learning

Similar to supervised learning, reinforcement learning uses data to improve the accuracy of algorithms. However, reinforcement learning differs from other types of supervised learning because the data being used to train the model isn't a sample data set but rather production data so the model is "learning on the job". In reinforcement learning, the model learns appropriate strategies based on what has and has not worked in the past.

Conclusion

When deploying Machine Learning, organizations often struggle to overcome two major challenges: data and culture. Without enough reliable data in the right format, ML projects are doomed to failure. Understanding the data underpinning the ML models is a critical success factor.

The other area in which organizations fall short in their attempts to leverage ML is culture. The aforementioned McKinsey research suggests that a lack of strategy is often an obstacle. To leverage ML, business leaders need to identify a business problem worth solving, whether that is high customer churn or the high cost of maintenance. ML can only add value when used to solve a well-defined business problem.

To ensure a successful Machine Learning initiative, management needs to ensure collaboration between a variety of stakeholders, of whom data scientists are but one. Finally, the organization must have the ability and willingness to change.

About the Author

Eszter Windhager-Pokol, the Head of Data Science at Starschema, holds a MSc degree in Applied Mathematics from Eötvös Loránd University. She has eight years' experience in supporting data driven decision making as a consultant, four years research experience in the field of collaborative filtering and three years background in developing user behavior analytics product for IT security purpose.

Eszter holds data science trainings regularly for business users from the basic level to the advanced analytics courses and teaches Mastering the Process of Data Science at CEU as a visiting faculty instructor. She is an organizer of the R-Ladies Budapest meetup group and participates in program committees of several international data science conferences.



ⁱ Notes from the AI frontier: AI adoption advances, but foundational barriers remain. McKinsey&Company 2018

ⁱⁱ Bishop, Christopher. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006.