A PRIMER ON MACHINE LEARNING

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

Machine learning algorithms iteratively learn from data by generalizing their experience into models. These models allow computers to find insights that might be difficult or impossible for humans to find.

The iterative aspect of machine learning is important because as models are exposed to new data, they can take into account and adapt to new observations. They learn from previous computations to produce reliable decisions and results.

Why do we hear so much about machine learning?

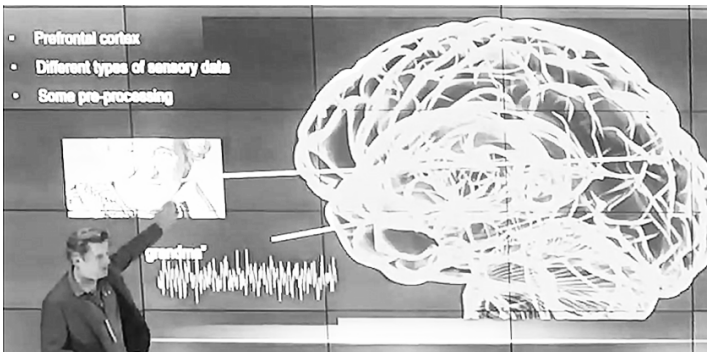
Interest in machine learning has grown exponentially over the past two decades, mostly due to a couple of underlying factors. First, the expansion of computers, the internet, and the information economy have generated increasing volumes and varieties of data, many of which are unstructured (meaning they cannot be processed by computers without first requiring human effort to structure them into machine-readable form). At the same time, computational processing has become cheaper and more powerful, enabling faster and more complex mathematical calculations and increasingly affordable data storage.

It's now possible to quickly and automatically produce models that can analyze bigger, more complex data, and deliver faster, more accurate results – even on a very large scale. The result? High-value predictions that enable people like you to create a new generation of intelligent services and devices, and to guide better decisions that produce smarter actions.

Some examples of machine learning applications you may be familiar with:

- Self-driving cars
- Online recommendation, like you get from Amazon and Netflix
- Automatic suggestions of the names of your friends in the pictures you post on Facebook, Flickr or Google
- Email spam filtering
- Contextual ads on web pages and mobile devices

Loop Q, the only unsupervised learning machine in the industry



What are some popular machine learning methods?

Machine learning is very broad, but the three main approaches are supervised learning, semi-supervised learning and unsupervised learning. Most applications of machine learning today use supervised learning. Semi-supervised learning accounts for a minimal part and unsupervised learning is the most difficult, and hence the least common.

Supervised learning is the machine learning task of inferring a function from labeled training data. Labeled data for a learning problem is usually provided by a skilled human agent and is therefore costly. Training data consist of a set of training examples.

Intuitively, we can think of the learning problem as an exam, and labeled data as the few example problems that the teacher solved in class. The teacher also provides a set of unsolved problems. In supervised learning, each example is a pair consisting of some input object and the desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a reasonable way.

Semi-supervised learning is a class of supervised learning techniques that also use unlabeled data for training - typically a small amount of labeled data with a large amount of unlabeled data (because unlabeled data is less expensive, and takes less effort to acquire). Many machine learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. While the cost associated with the labeling process may render a fully labeled training set infeasible, acquisition of unlabeled data is relatively inexpensive. However, semi-supervised learning still relies heavily on the availability of human-labeled data and therefore requires the expertise of a very scarce resource, data scientists. This obstacle makes this a more cumbersome, time-consuming, and expensive learning method than the unsupervised learning.

Unsupervised learning involves algorithms that identify patterns in unlabeled data with very little or no human guidance. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. The system is not told the “right answer” for any of the data, and the algorithm must figure out the underlying structure of the data all on its own. This human-capacity ability to learn and reason independently allows for the machine to continuously learn as new data becomes available, deepening its understanding in real-time. The lack of human interaction and pre-programming means that the machine learns in any language (human or machine) like a native speaker, exclusively from the domain. The learning speed with this learning method is substantially faster, the difference is between using days or even months of manual work compared to a matter of minutes with the unsupervised technology.

What's the difference between statistics, data mining, machine learning, and deep learning?

Let's try to briefly define each:

Statistics is a sub-topic of mathematics that focuses on techniques for summarizing and making inferences from data. These inferences are usually made by drawing a sample from a set of data that is subject to random variation and applying probabilistic principles to draw conclusions about the data as a whole. The most common statistical methods used today have been developed over several centuries, and provide the mathematical foundation for modern data science and machine learning.

Data mining is a sub-field of machine learning that focuses on extracting useful knowledge from structured or unstructured data sets. Data mining techniques seek to discover patterns in data, and are often painstakingly implemented to satisfy some particular knowledge need in a particular situation with a particular data set. The goal is generally not to develop a more sophisticated understanding of the underlying processes that generated the data.

Machine learning involves inducing new knowledge from experience and storing it in the form of mathematical models that can be used later to make inferences or detect patterns. The simplest implementations of machine learning have been used for over a century, in the form of regression and the generalized linear model. A newer variety of machine learning uses artificial neural networks (ANN), which are a group of algorithms loosely based on our understanding of the brain. ANNs can – in theory – model any kind of relationship within a data set. But in practice, getting reliable results from neural networks can be very tricky. Artificial intelligence research dating back to the 1950s has been punctuated by the successes and failures of artificial neural networks.

Deep learning is based on endowing a neural network with many hidden layers, enabling a computer to learn tasks, organize information, and find patterns on its own. Deep Learning is part of a broader family of machine learning methods based on learning hierarchical representations of data where an observation can be represented at many levels simultaneously. For example, a picture may be represented as a set of pixel intensity values, or at a more abstract level as a set of edges, or regions of a particular shape. Research in this area attempts to make better representations and create models to learn these representations from large-scale, unlabeled data using different types of deep neural networks. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in the nervous system.

Loop AI Labs machine learning experience and expertise

At Loop AI Labs, we started to work with deep learning and unsupervised learning in 2012, and after over three years of intense R&D and intensive trials with selected clients, we launched our commercial grade Loop AI Labs cognitive computing platform Q in 2015. We continue to search for and evaluate new techniques.

While the company is young, our team collectively represents a long history of experience implementing artificial intelligence in prestigious organizations as the Artificial Intelligence Center at SRI International (founded as Stanford Research Institute), one of the largest private research facilities in the country, and Stanford's Computational Semantics Lab, including participation in the largest government-funded artificial intelligence project in history (DARPA CALO).

We combine our rich, sophisticated artificial intelligence heritage with new architectural advances to ensure that Loop Q runs autonomously and as fast as possible – even in huge enterprise environments that span the globe.