Machine Learning: Basic Analysis

Dr. SheelGhule[1], Mrs. RupaliGhule (Chikhale)[2]

[1]*Persistent Systems Limited, Nagpur,*

[2]*G. H. Raisoni Institute of Information Technology, Nagpur*

Abstract: Machine learning is a subfield of AI (artificial Intelligence). Machine learning aims at investigating how computers can learn based on data. Main aim of machine learning is to understand the structure of data and fit that data into some models that can be easily understood and utilized by people. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

Here we'll look into the common machine learning methods of supervised and unsupervised learning, and common algorithmic approaches in machine learning. Additionally, we'll discuss biases that are perpetuated by machine learning algorithms, and consider what can be kept in mind to prevent these biases when building algorithms. Algorithms is a major part of any aspect of business.

Keywords: Machine Learning (ML), Algorithm, Artificial Intelligence, supervised, unsupervised

I. INTRODUCTION

Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

Machine Learning refers to the techniques involved in dealing with vast data in the most intelligent fashion (by developing algorithms) to derive actionable insights.

The reason stems from the seemingly unlimited use cases in which machine learning can play a role, from fraud detection to self-driving cars to identifying your "gold card" customers to price prediction.

As regards machines, we might say, very broadly, that a machine learnswhenever it changes its structure, program, or data (based on its inputs or inresponse to external information) in such a manner that it's expected futureperformance improves. Some of these changes, such as the addition of a recordto a data base, fall comfortably within the province of other disciplines and arenot necessarily better understood for being called learning. But, for example,when the performance of a speech-recognition machine improves after hearingseveral samples of a person's speech, we feel quite justified in that case to saythat the machine has learned.

Machine learning usually refers to the changes in systems that perform tasksassociated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc.

Use of Artificial Intelligence in Machine Learning

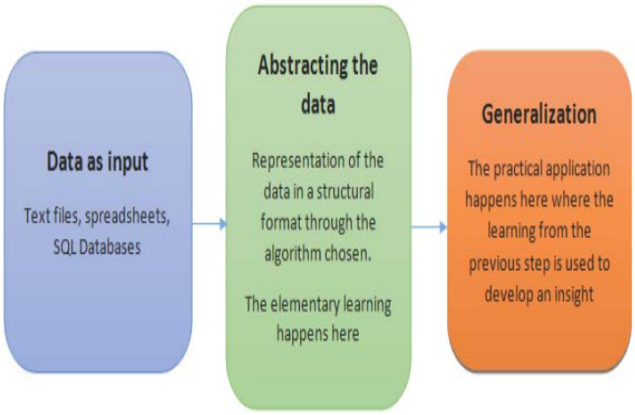
Machine Learning is a subset of AI where the machine is trained to learn from its past experience. The past experience is developed through the data collected. Then it combines with algorithms such as Naïve Bayes, Support Vector Machine (SVM) to deliver the final results.

Use of Statistics in Machine Learning

Let's understand this. Suppose, I need to separate the mails in my inbox into two categories: 'spam' and 'important'. For identifying the spam mails, I can use a machine learning algorithm known as Naïve Bayes which will check the frequency of the past spam mails to identify the new email as spam. Naïve Bayes uses the statistical technique Baye'stheorem (commonly known as conditional probability). Hence, we can say machine learning algorithms uses statistical concepts to execute machine learning.

How exactly do we teach machines?

Teaching the machines involve a structural process where every stage builds a better version of the machine. For simplification purpose, the process of teaching machines can broken down into 3 parts:



Machine Learning Methods

In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed.

Two of the most widely adopted machine learning methods are **supervised learning** which trains algorithms based on example input and output data that is labeled by humans, and **unsupervised learning** which provides the algorithm with no labeled data in order to allow it to find structure within its input data. Let's explore these methods in more detail.

Supervised learning:

- Suppose you had a basket and it is fulled with some fresh fruits your task is to arrange the same type fruits at one place.
- Suppose the fruits are **apple,banana,cherry,grape.**
- So you already know from your previous work that, the shape of each and every fruit so it is easy to arrange the same type of fruits at one place.
- Here your previous work is called as **train data** in data mining.
- So you already learn the things from your train data, This is because of you have a **response variable** which says you that if some fruit have so and so features it is grape, like that for each and every fruit.
- This type of data you will get from the **train data.**
- This type of learning is called as **supervised learning**.
- This type solving problem come under **Classification.**
- So you already learn the things so you can do your job confidently.

Unsupervised learning:

- Suppose you had a basket and it is fulled with some fresh fruits your task is to arrange the same type fruits at one place.
- This time you don't knowanything about that fruits, you are first time seeing these fruits so how will you arrange the same type of fruits.
- What you will do first you take on fruit and you will select any physical character of that particular fruit. Suppose you taken color.
- Then you will arrange them base on the color, then the groups will be something like this.
- **RED COLOR GROUP: apples & cherry fruits.**
- **GREEN COLOR GROUP: bananas & grapes.**
- So now you will take another physical character as size, so now the groups will be something like this.
- **RED COLOR AND BIG SIZE:** apple.
- **RED COLOR AND SMALL SIZE:** cherry fruits.
- **GREEN COLOR AND BIG SIZE:** bananas.
- **GREEN COLOR AND SMALL SIZE:** grapes.
- Job done happy ending.
- Here you didn't know learn anything before means **no train data and no response variable.**
- This type of learning is knownunsupervised **learning.**

Semi-supervised learning:

Semi-supervised learningis used for the same applications as supervised learning. But it uses both labeled and 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). This type of learning can be used with methods such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with labeling is too high to allow for a fully labeled training process. Early examples of this include identifying a person's face on a web cam.

Reinforcement learning:

Reinforcement learning is often used for robotics, gaming and navigation. With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards. This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do). The objective is for the agent to choose actions that maximize the expected reward over a

given amount of time. The agent will reach the goal much faster by following a good policy. So the goal in reinforcement learning is to learn the best policy.

Types of Learning

These are the main machine learning problems:

Classification:Learn to put instances into pre-definedclasses.

A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.

Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Association:Learn relationships between attributes.

An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Numeric prediction:Learn to predict a numeric quantityinstead of a class.

Clustering:Discover classes of instances that belong together

A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

Clustering algorithms divide a data set intonatural groups(clusters). Instances in the same cluster aresimilarto eachother, they share certain properties

II. APPLICATIONS

Most readers will be familiar with the concept of web page ranking. Thatis, the process of submitting a query to a search engine, which then _ndswebpages relevant to the query and which returns them in their order ofrelevance. See e.g. Figure A. for an example of the query results for \machinelearning". That is, the search engine returns a sorted list of webpagesgiven a query. To achieve this goal, a search engine needs to `know' which

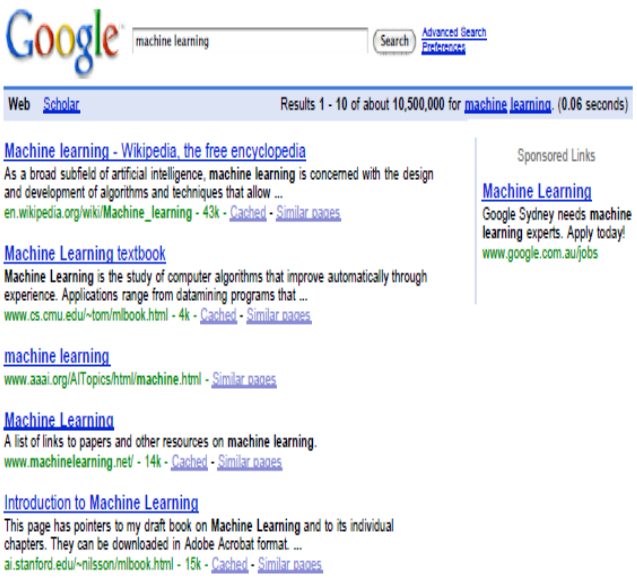


Fig A. The 5 top scoring webpages for the query \machine learning"

are relevant and which pages match the query. Such knowledge can be gained from several sources: the link structure of webpages, their content, the frequency with which users will follow the suggested links in a query, or from examples of queries in combination with manually ranked webpages. Increasingly machine learning rather than guesswork and clever engineering

is used to automate the process of designing a good search engine

What are the applications of Machine Learning?

It is very interesting to know the applications of machine learning. Google and Facebook uses ML extensively to push their respective ads to the relevant users. Here are a few applications that you should know:

- **Banking & Financial services:**ML can be used to predict the customers who are likely to default from paying loans or credit card bills. This is of paramount importance as machine learning would help the banks to identify the customers who can be granted loans and credit cards.

- **Healthcare**: It is used to diagnose deadly diseases (e.g. cancer) based on the symptoms of patients and tallying them with the past data of similar kind of patients.

- **Retail**: It is used to identify products which sell more frequently (fast moving) and the slow moving products which help the retailers to decide what kind of products to introduce or remove from the shelf. Also, machine learning algorithms can be used to find which two / three or more products sell together. This is done to design customer loyalty initiatives which in turn helps the retailers to develop and maintain loyal customers.

These examples are just the tip of the iceberg. Machine learning has extensive applications practically in every domain. You can check out a few Kaggle problems to get further flavor. The examples included above are easy to understand and at least give a taste of the omnipotence of machine learning.

Other Advantages of Machine Learning

Machine learning is proactive and specifically designed for "action and reaction" industries. In fact, systems are able to quickly act upon the outputs of machine learning - making your marketing message more effective across the board. For example, newly obtained data may propel businesses to present new offers for specific or geo-based customers. However, data can also signify cutting back on unnecessary offers if these customers do not require them for conversion purposes.

The latter may even be a form of learning from past behaviors. Machine learning models are able to learn from past predictions, outcomes and even mistakes. This enables them to continuously improve predictions based on new incoming and different data.

III. CONCLUSION

This article summarized some of the most salient items. This tutorial reviewed some basic concept of machine learning, common methods and popular approaches used in the field.

Because machine learning is a field that is continuously being innovated, it is important to keep in mind that algorithms, methods, and approaches will continue to change. Hope this article helped you to get acquainted with basics of machine learning.

IV. REFERENCES

[1]. https://datafloq.com/read/12-algorithms-every-data-scientist-should-know/2024#datascience
[2]. http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/fksummary.pdf
[3]. https://www.sas.com/it_it/insights/analytics/machine-learning.html
[4]. http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/fksummary.pdf
[5]. http://robotics.stanford.edu/people/nilsson/MLBOOK.pdf
[6]. https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/
[7]. http://alex.smola.org/drafts/thebook.pdf
[8]. https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning
[9]. https://www.lumagate.com/news/the-advantages-of-machine-learning
[10]. https://www.quora.com/What-is-the-difference-between-supervised-and-unsupervised-learning-algorithms
[11]. https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms
[12]. https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning
[13]. https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/
[14]. https://www.lumagate.com/news/the-advantages-of-machine-learning
[15]. http://alex.smola.org/drafts/thebook.pdf
[16]. https://dzone.com/articles/5-predictions-about-the-future-of-machine-learning