

Machine learning

Machine learning is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

The era of data

Machine learning

The era of data

*“A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”*

The ability of learning is one of the most important aspects of intelligence. Translating that power to machines, sounds like a huge step towards making them more intelligent. And in fact, Machine Learning is the area that is making most of the progress in Artificial Intelligence today; being a trendy topic right now and pushing the possibility to have more intelligent machines.

**The following are some of the most common categories of practical Machine Learning applications:**

### Image Processing

### Text Analysis

### Data Mining

### Video Games & Robotics

### And much more

Among the different types of ML tasks, a crucial distinction is drawn between supervised and unsupervised learning. The main difference between both approaches resides in the way we feed training examples to our algorithm, how the algorithm uses them and the type of problems they solve.

* **Supervised machine learning:** The program is “trained” on a pre-defined set of “training examples”, which then facilitate its ability to reach an accurate conclusion when given new data.
* **Unsupervised machine learning:** The program is given a bunch of data and must find patterns and relationships therein.

# The goal of ML is never to make “perfect” guesses, because ML deals in domains where there is no such thing. The goal is to make guesses that are good enough to be useful.

The future holds great prospect to machine learning as it further develops day by day, an example being the development of deep learning from machine learning to achieve more sophisticated tasks.

Examples of the future application include:

* Deeper personalization.
* Automatic cars
* Mobile experience automation.
* Real-time speech translation
* Health and fitness
* And much-much more

Top companies have shown great interest in machine learning by setting up their own dev teams such as google with google brain and tenser flow, or by making accusations as shown below by recent news:

* [Apple](http://fortune.com/fortune500/apple-3/) has bought a machine-learning startup called Turi in a deal reported to be worth around $200 million. Turi provides a machine-learning platform with the tools to use it.
* [Intel](http://fortune.com/fortune500/intel-51/) is bolstering its artificial intelligence efforts by acquiring Nervana Systems, a startup considered among the leaders in developing machine learning technology for $350 million.

# **The use of machine learning in the airline industry**

**Maintenance**

Nowadays, airlines are paying for engines based on a time on wing metric, which identifies the operational reliability of an aircraft engine. Due to this payment model, companies manufacturing such machines are obligated to make these engines more dependable. This is where the pattern recognition capabilities of machine learning come into play. These algorithms can isolate vulnerabilities within these implementations and accurately identify what kind of maintenance is required to repair them. The assessment doesn't stop there, multiple operational components caused by frequent use and external factors can be analyzed through the intelligent eye of a computer.

**Customer Engagement**

Instead of having users self-select an issue and fill out endless form fields, machine learning is helping companies look at the substance of a request and route it to the right place. Ticket tagging and routing can be a massive expense for big businesses. Machine learning, by helping automate the process is helping customer service save significant time and money, all while making sure issues get prioritized and solved as fast as possible.

**Understanding customer behavior**

Role of sentiment, mood, etc. in consumer decision making behavior is increasingly being recognized as a key factor that drives a lot of big decisions. By understanding what the users are saying about a certain incident by using machine learning to read what they post on social media, an inference can be made weather the incident caused was good or bad for the company.

****

Artificial intelligence and data science isn’t new in the airline industry. However, its development has only recently begun to make a leap forward for two reasons:

* Machines started learning and filling gaps for themselves based on their experience, improving precision from what used to be exclusively rational critical data analysis and interpretation.
* Airlines began looking beyond historical trends analysis and envisioned projective data and behavioral analytics.

In October 2015, the University of Oxford opened a new Data Science Lab in collaboration with Emirates, to use cutting-edge analysis to help the airline “make its services more efficient and customer-focused”. From the University, a select group of mathematicians, scientists, engineers and social scientists joined carrier’s experts in the industry ad data, to research how to apply optimization techniques to Emirates' data, and develop machine learning to interrogate complex data sets.

The industry has been behind in using these pools of big data analysis tools, despite the latest trends in sectors such as the retail business, who has been able to personalize their value proposal with demand segmentation in e-commerce channels.  Some airlines have taken this trend a step forward, as they realized that, since data science expertise was transversal to all units of activity, it needed to be at the top of the management funnel rather than just at a service-level position.

In 2015, EasyJet appointed his first head of data science, Alberto Rey-Villaverde, to accelerate the use of artificial intelligence (AI) to improve efficiency, reduce cost and increase revenue and customer satisfaction. His profile sums up greatly how has this discipline also risen up the ranks in the airline industry.

An M.Sc. in Data Mining from the University of Westminster – a program  that will later be upgraded to Business Intelligence and Analytics, Mr. Rey-Villaverde began his career in 2006 at EasyJet as a Yield Developer, and was subsequently promoted to Network Pricing Manager and Yield Strategy and Development Manager, before becoming head of data science.

“Technology and data is not something that falls either on the customer’s side or on the business side. We see it as something integrated that goes across the whole business”, he stated at the [BTiQ](http://www.businesstravel-iq.com/) 2015 conference. “The airline seat is a perishable asset. Once the plane departs, if we have a flight taking off from Luton to Malaga at 6 PM, if my team didn’t manage to sell that seat, the seat is lost. All the effort that was put in place to bring it to market is wasted”. “What we do is look at the past performance, we learn a lot. After 20 years flying we earn a lot of knowledge on how the customers behave, what are the patterns, how much people is willing to pay at different times of day”, he also said. “We need a lot of data. Not only historical data but that of what is going on, on a day-to-day level. Those 95% of price adjustments that we make every day, they’re coming from what is happening every time every minute”.

 The data scientist speaks about the applications of this for airline business intelligence and yield management: “Data supports changes in the current market conditions, and we started to use machine learning as traders weren’t making smart decisions and the algorithms weren’t picking up everything”, he revealed. “To date, analytics has been about diagnostic capability and looking backward. Now advanced AI is more focused on predictive capability so we can better understand the future and plan for it”.

 EasyJet has applied data science to revenue management, setting fares by responding to passenger demand, helping increase revenue per seat in nearly 20% between 2010 and 2014, according to the Financial Times.  According to the newspaper, EasyJet plans to use AI to predict demand and analyze over 1.3 Billion searches made on its website each year, to optimize destinations and flight times.

Despite these trends, Legacy airlines are still some steps behind, and need further development, as their yield management is away from EasyJet's simpler logistical and revenue structure: its entire booking procedure is online, their fleet has only one type of plane and only operates short haul routes. “This combination of simplicity and scale generates an enormous quantity of data”, Mr. Rey- Villaverde admits at EasyJet’s Spanish blog. He also addressed a growing internal fear at the [BTiQ](http://www.businesstravel-iq.com/) 2015 conference: machines are not replacing humans. “We are using machine learning on our team. We did not displace anybody. We didn’t have to reduce the number of positions that we have in there. On the contrary, we have more people that are capable now, just because we are using machine learning.  "Machine learning is used to automatize what we are doing”. He points out: “I think we are far away from machines that think for themselves, but they are more intelligent. Machines pick up the data, learn from it, spot patterns and get a program from that”.

What will happen to Data Science, Predictive Analytics and Artificial Intelligence for medium to long-haul flights and legacy sales? The escalation of AI in the entire industry is an upcoming challenge, one that benefits all. Rey Villaverde will discuss these topics as a Keynote address “How to apply data science to revenue management to maximize sales through supply, demand and competitor pricing” at the Aviation IT Show Americas, part of the Aviation Festival Americas, held on May 24th-25th 2016 in Miami. He’ll discuss the identifying of prescriptive, predictive, diagnostic, and descriptive analytic capabilities, the evolving perception of data sharing and automation, as well as talk about segmenting the market and daily price adjustments through historical and real-time data, among other issues. Airnguru will be present at the event to engage with industry representatives and continue to develop airline pricing intelligence.

Also, currently available forecasting models cannot correctly forecast behaviors that they haven’t “seen” in history, unless the users manually recalibrate them to do so. This produces high supervision costs from revenue management teams. We expect that next generation of forecasters, based on machine learning, may be able to automatically introduce controlled biases outside the boundaries of historic data in order to discover unseen behaviors and recalibrate the forecasts accordingly.

Asset-intensive organizations of all sizes and levels of manufacturing maturity are unleashing the power of predictive analytics to gain significant improvements in asset reliability.

We can also take a moment to break down the very complex problem of airport security and how Machine Learning could help, specifically with “random” search selection to make it…. well… not so random.

Let’s start off with a basic assumption. Airport security is an “unsupervised machine-learning problem”. Why is that? The TSA screens millions of people every day and rarely if ever do they find something noteworthy. Sure, they find an overly large container of body wash and the odd pocketknife, but do they routinely find something really “big”? Nope. It would be all over the headlines if they did, the story would help boost confidence that the TSA is actually doing something and catching the bad guys.

So from a machine learning perspective, if each data point is an individual taking a flight, most if not all data points are non-malicious. That is, there is nothing going on but a normal trip from point A to B. So that’s important because we don’t have labeled data. We don’t have good examples of malicious “trips” or activity. This is an unsupervised learning problem, because we are looking for outliers, without knowing in advance what those outliers look like. This is sometimes referred to as anomaly detection in security operations terms. We could use something called Gaussian distribution, which is sometimes also called Normal Distribution, to model our data and then generate an ML model that learns complex non-linear relationships between our data dimensions or features. The mathematic details aren’t important, there are plenty of good toolkits to assist here. Let’s just assume the math exists and is well tested.

So now that we’ve identified the problem and the potential algorithm, let’s describe what some potential feature of a “trip” might be. The basic features are attributes of the trip itself. The much more interesting features come from the individual taking the trips. Here we need to “de-normalize” the data so that everything about the trip and its associated individual is represented as 1 “row” of data which becomes, as we will see later, 1 “vector” of data.

* From – To airports / countries
* Flight time
* Check-in time
* Days booked in advance
* Method of payment
* 1 way or 2 way ticket
* Destination is “Home”?
* Individual details (which the airlines has, so by default the TSA has these)
  + Passport country
  + Place of Birth
  + Citizenship
  + Number of average trips per month, year
  + Trip rate deviation
  + Etc., and many more

You can see that the individual details are pretty ordinary, and honestly fairly boring. You can still get a tremendous amount of insight by modeling these “ordinary” features.

The data now needs to be represented as an n by m (n\*m) matrix or tensor where n is the number of features we’ve identified, and m is the total number of trips to analyze. Needless to say, n is going to be in the hundreds or thousands and m will be in the hundreds of millions. This is a bigdata problem. We can dive deeper into possible solutions and how machine learning can be achieved in practical terms if there is interest.

Finalizing the solution, all we need to do is generate a “probability” classifier that flags a particular trip as out of the ordinary. If the answer is yes, a real-time indicator can be sent to the TSA that extra care is to be taken with the individual. It doesn’t mean the trip is abnormal, or the person is malicious. It just means, there is a higher risk, and a search is warranted. Sure beats a random search!!

Can we do anything more interesting than the above approach? As you probably suspect the answer is YES**.**

What if we could collect fuzzy information about an individual, data that is all available in the public domain? People openly post of Face Book, Twitter, bogs, etc. Using another Machine Learning technique, SentimentAnalysis**,** which is a byproduct of Natural Language Processing, we could gather insight into the mood or level of an individual’s disgruntlement.

# The Question

To make an application by only using machine learning, where we can input 3 numbers, say the length, height and breadth of a cuboid,

Such that we get an output of its volume, without using the volume formula.

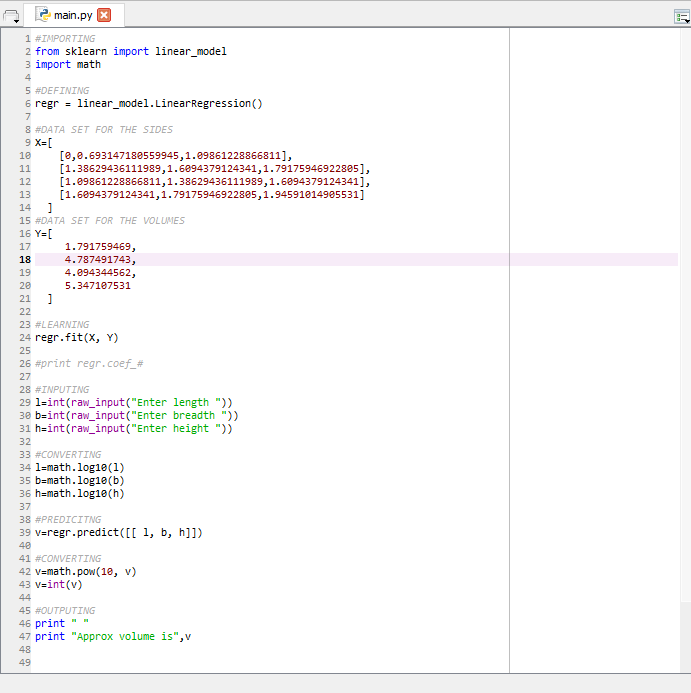
 i.e. L\*B\*H.

# The Answer

If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of *y* and *X* values.

After developing such a model, if an additional value of *X* is then given without its accompanying value of *y*, the fitted model can be used to make a prediction of the value of *y*.

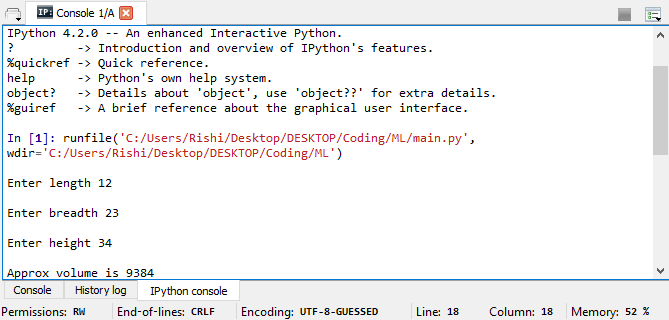
# The Code



# The Explanation

1. To use Linear Regression, we use the scikit-learn package along with python to code it.
2. We define data sets of the sides and respective volumes from which the machine can learn from.
3. The machine learns the data sets using **.fit** function.
4. The program inputs the length, breadth and height.
5. The numbers get converted into logarithms of base 10.
6. The program predicts a new volume from the entered values.
7. The volume is anti-logged and then displayed.

# The Output



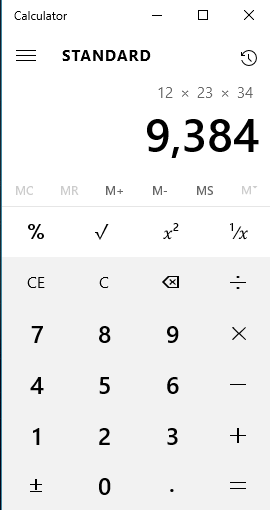
Length = 12

Breadth = 23

Height = 34

Volume = 9384

# The Conclusion



Therefore, by Linear Regression we are able to get an approximate if not accurate volume from the user entered sides of a cuboid.

And thus, this is just one the of the ways in which we can use Machine Learning for future applications.