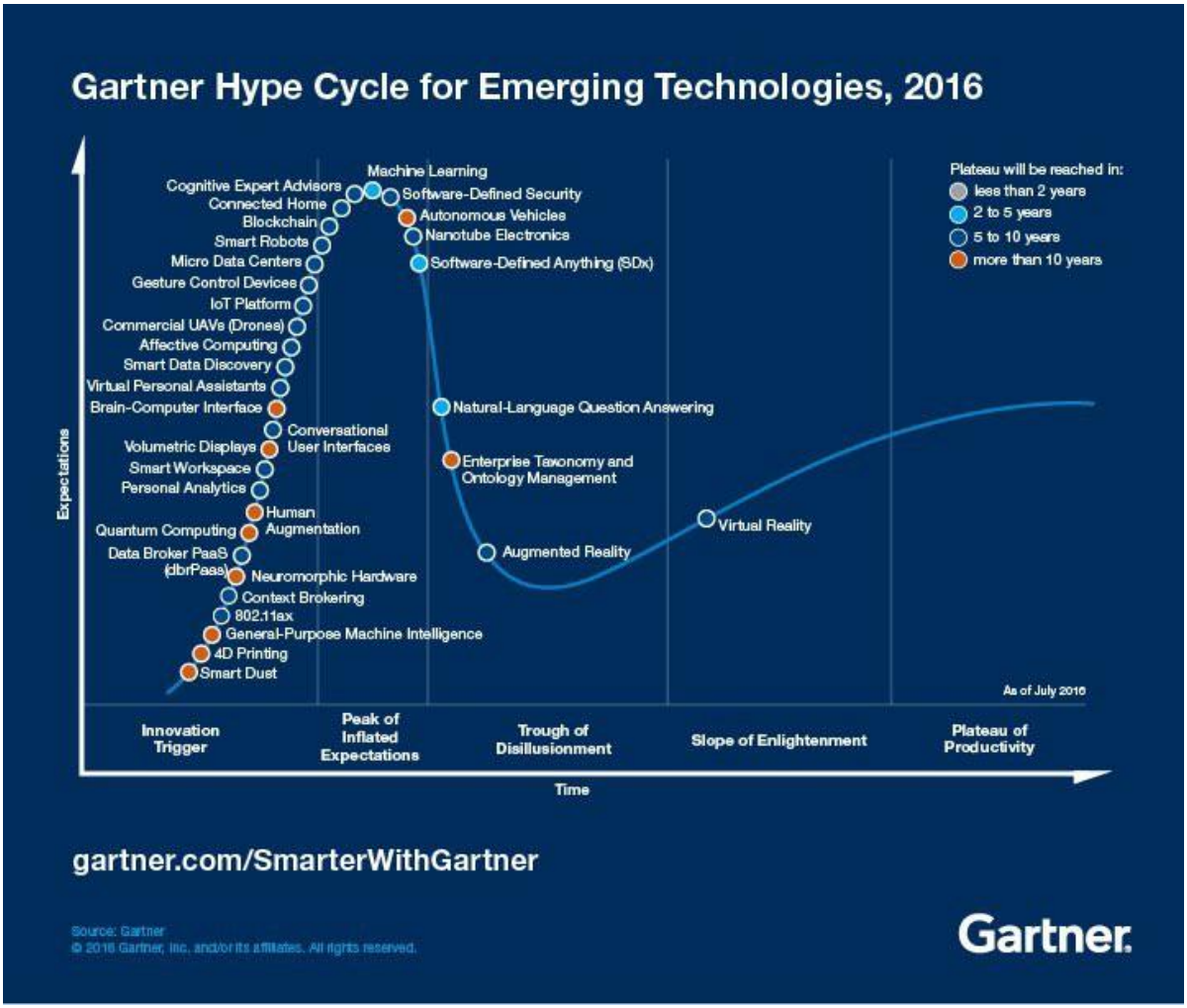
When Machine Learning meets Big Data

[Krishna Kumar Tiwari]
Jul 30, 2017

First of all let us see why everyone around is interested in machine learning now a days, I believe ***Hype Cycle Plot of Emerging Technologies by Gartner*** answers the same.



It clearly represents machine learning is now on peak of inflated expectation, and we all can smell this. Irrespective of field we are in, we see individual, industries & academia talking about automation, making machines more intelligent and powerful using AI, machine learning & data

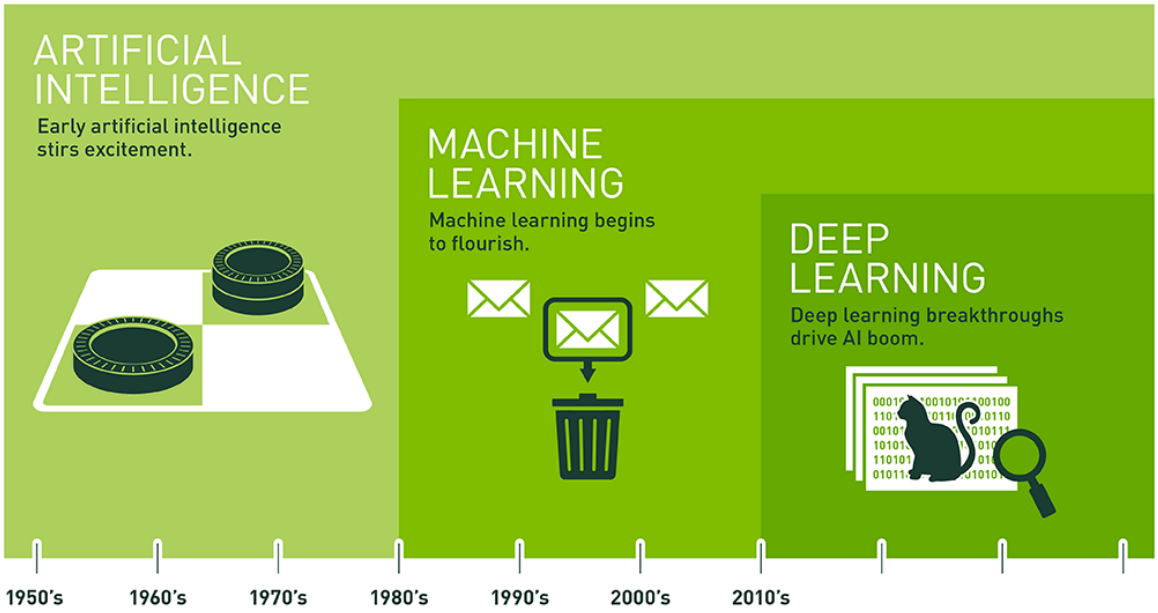
mining. By looking at current applications of machine learning we can confidently say it has more substance than gas *(I remember the days when people says it is GAS which smell good)*

But the question is -Did machine learning suddenly become a hype?

To understand this we need to look back to the basics of machine learning, below is the definition of machine learning provided by ExpertSystems.

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning** focuses on the development of computer programs that can access data and use it learn for themselves.

It seems very exciting to imagine programs which can learn on their own but they need data, huge amount of data which can be used in learning.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Tough Machine learning is more than five decades old and have been in use for about four decades ago but the world has come to admire it only in the last two decades only. Most of the big giants like Google, IBM, Microsoft & Baidu have already started shifting their main focus to machine learning and deep learning application.

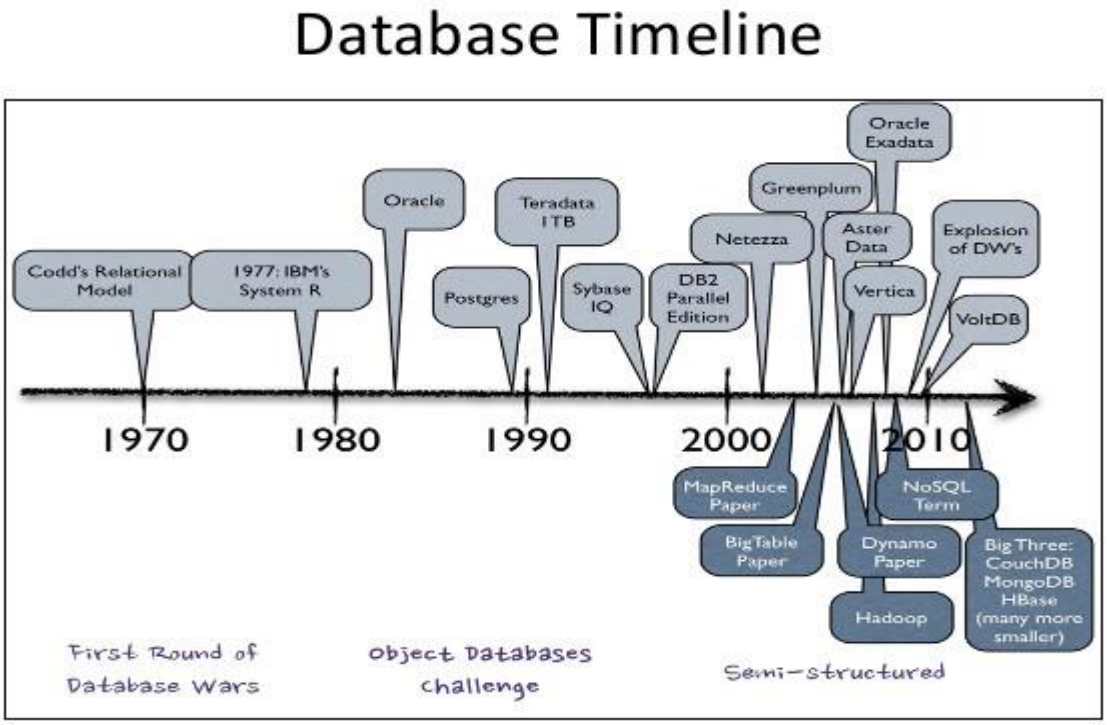
So what happened in the last two decades that caused this change?

Many awesome things came into the world in last two decades ,but what fueled the growth (read hype) around machine learning and other related fields is nothing but DATA. Due to innumerable new technologies, devices (desktop, laptop, mobile etc), social networking sites and the penetration of technology in the world, the amount of data produced by mankind is in gigabytes and that creates opportunity of data bases, data mining, information retrieval, machine learning, deep learning, AI and many more.

The amount of data produced by us from the beginning of time till 2003 was 5 billion gigabytes. If you pile up the data in the form of disks it may fill an entire football field. The same amount was created in every two days in 2011, and in every ten minutes in 2013. This rate is still growing enormously. Though all this information produced is meaningful and can be useful when processed, it is being neglected.

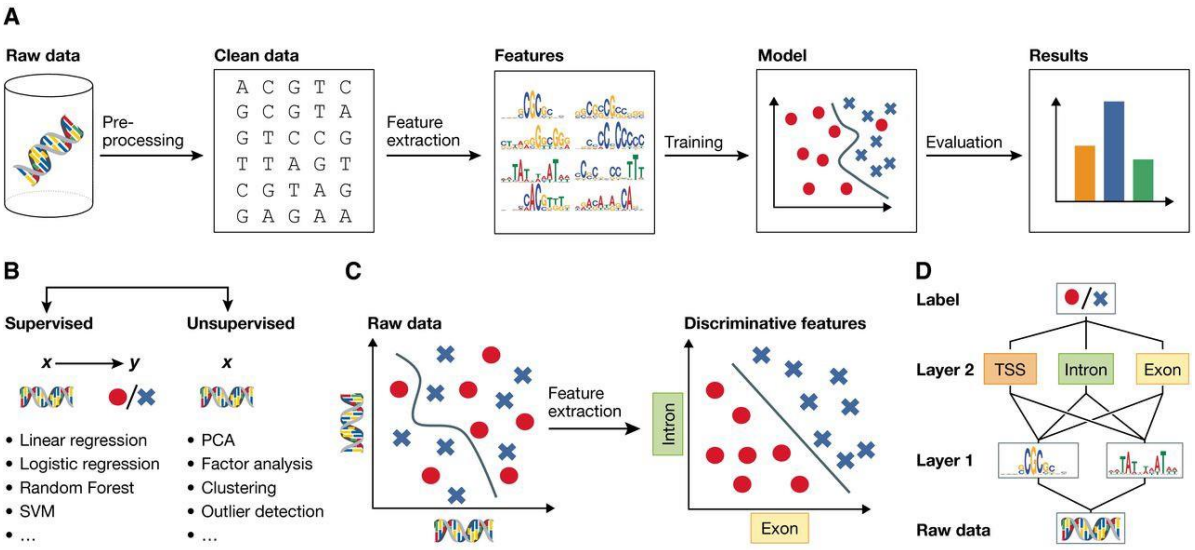
90***% of the world's data was generated in the last few years only. Data is the new oil and fueling the information revolution.***

This data growth from inception gave opportunity to new data storage systems. Till Yr 2000 we cared mostly about the relational data, mostly structured and we have many ways to store it. Checkout the below timeline of data bases, after Yr 2000 data was unstructured, semi-structured, huge and people understand that keeping them can help in future hence the BigData came into picture and many technologies on top on it.



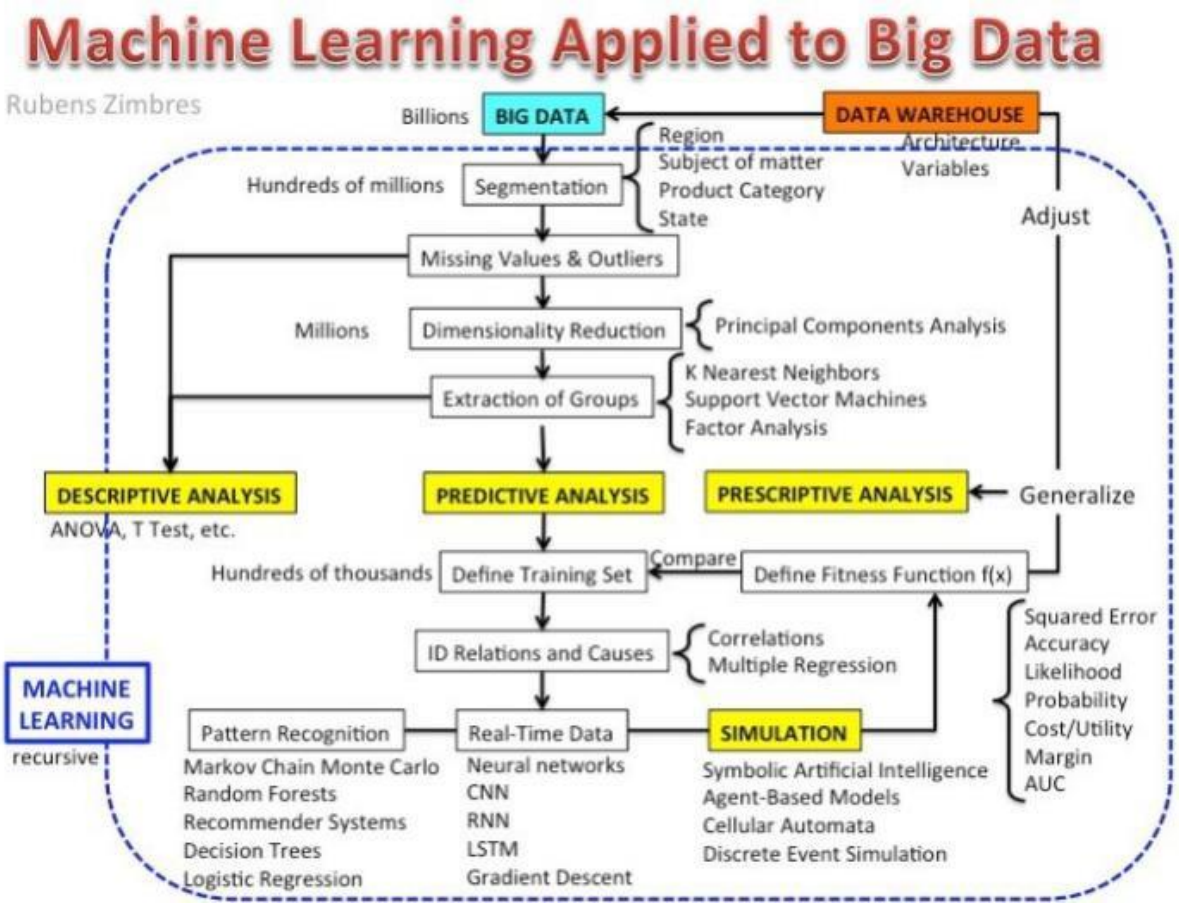
Cool, so many Big Data technologies like Hadoop makes the job simple and creates opportunity for the AI & machine learning models.

Lets see in short how the machine learning is applied on general data. Capturing the raw data & cleaning is most tedious job in the steps mentioned below. Cleaning huge amount of data is not possible without distributed technologies like map-reduce.



Now we have fair idea on how the machine learning needs big data but still it is becoming challenging to come up the design and architecture on how to marry these two.

Rubens Zimbres and well known data scientist define the general architecture of How to apply machine learning on big data. Below plot gives a crystal clear view of how to marry these two. Most challenging part here is to maintain Data Warehouses with a proper design so that same data can be consumed in seamless fashion by machine learning models.



We will explore each and every step involved here in my next post. Till now we got a view on how to marry big data & machine learning.

Thanks for reading this, please share your thoughts, feedback &ideas in comments. You can also reach out me on @simplykk87 on twitter and linkedin.

References

Data Science & Machine Learning Newsletter # 55
*I share the articles took my attention and publish 'Data Science & Machine Learning Newsletter' every Friday. If you…*www.necatidemir.com.tr

https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm