Demystifying Machine Learning

Presented by

F. Basheer Ahamed & A.Mohamed Noordeen

Trainer’s Profile

F. Basheer Ahamed

- Coming from ECE background, I have around 12 years of software design and development experience in domains like Equity, Airlines, Digital Security and AdTech etc.
- While progressing my career with some of the tech companies like Cognizant, Symantec, HID Global & Ooyala, I gained fair knowledge in Product development and design, AGILE, Threat Modelling & Advance Security and Machine Learning etc.
- Renowned professional trainer (both technical and soft skills) and have conducted various organisational level trainings.
- Currently solving some of the challenges in AdTech (online advertising)

Hobbies:

Reading books - currently catching up with "Zero To One" By Peter Thiel.

What Machine Learning is all about?

- **Humans can learn from past experience.**
- **Machines need to be told what to do (via program, instructions).**
- **Can machines learn from the past experience/data ?**
- **Lets discuss couple of use cases**

What Machine learning did in past and doing today and in future

? ? ?

Father of Machine learning

1956 – Trains computer to play checkers

1962 – Computer defeated the state champion

Machine Learning ..

- **As stated by Arthur Samuel : Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed**
- **Machine learning and artificial intelligence have phenomenal potential to simplify, accelerate, and improve many aspects of our lives.**
- **Computers can ingest and process massive quantities of data and extract patterns and useful information at a rate exponentially faster than humans, and that potential is being explored and developed around the world.**

software

algorithms in decision making

fuzzy boundaries

rule-based decision making

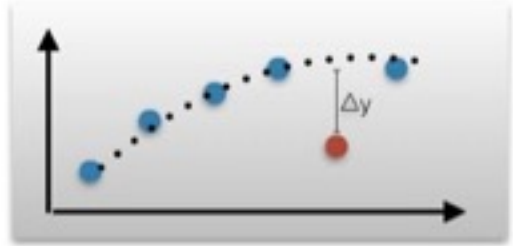
```
if condition fulfilled then
    activity 1
else
    activity 2
```

boolean data
(yes or no)

Examples:

- phone notification
- time- or threshold-based alarms
- simple pattern matching

statistical reasoning



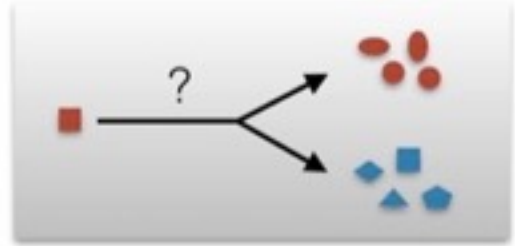
simple regression

numerical data
allowing for curve fitting

Examples:

- extra- and interpolation
- outlier detection
- predictive maintenance

machine learning



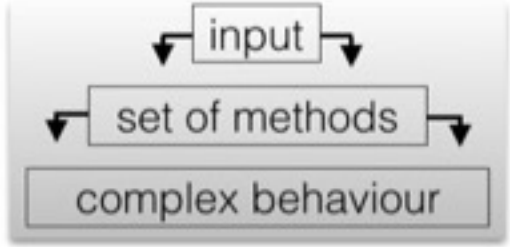
classification tasks

arbitrary data
that needs to be abstracted into numbers

Examples:

- identification of relevant features from large input datasets

artificial intelligence



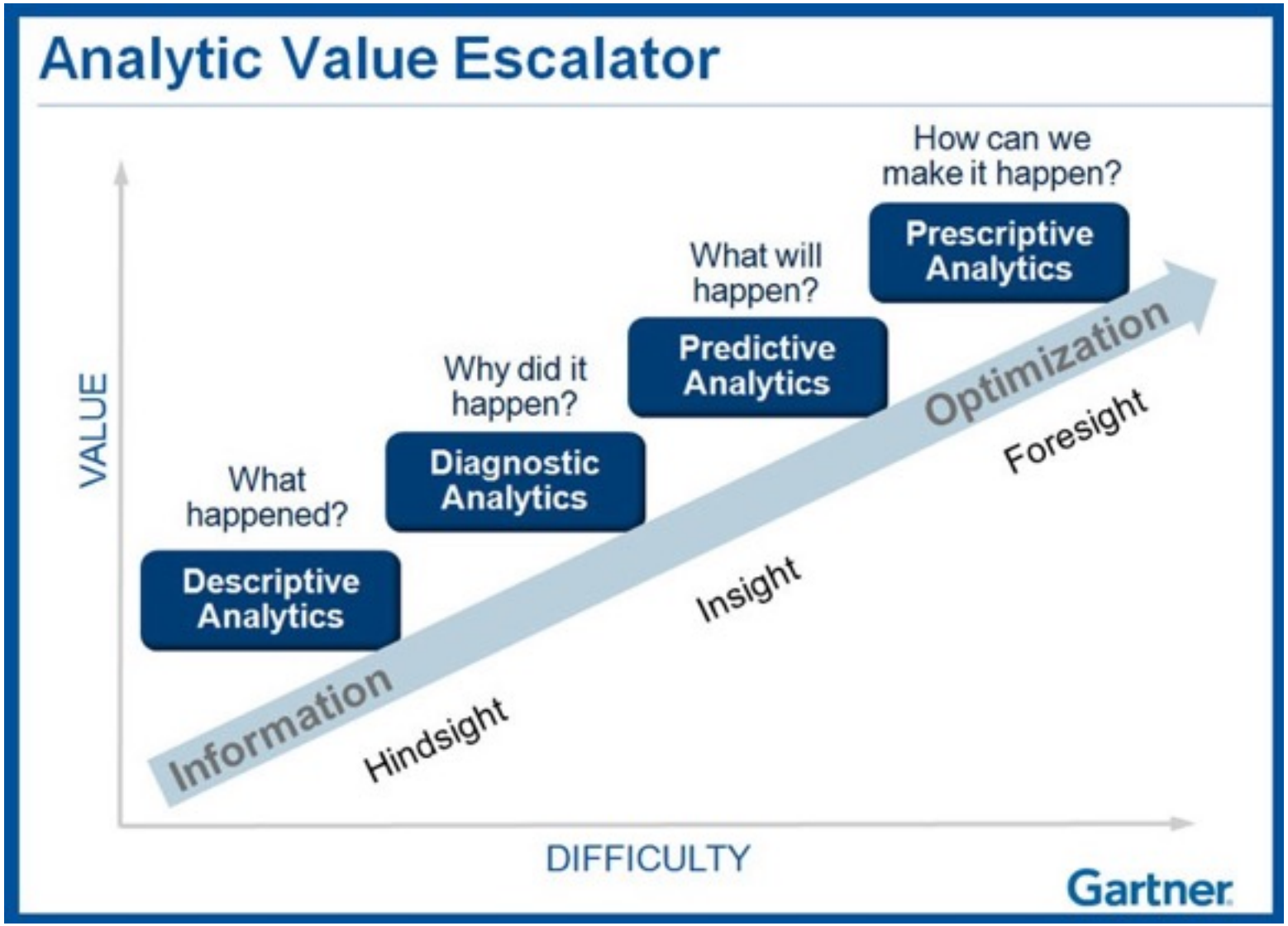
dynamic adaptation to novelty

autonomous selection of best methodology when presented with **arbitrary data**

Examples:

- autonomous vehicles
- human-like

Trend of Analytics:



Real life usage of Machine Learning

- **Self driving cars - Tesla, capable of learning from other connected cars**
- **SIRI - Automatic speech recognition**
- **CORTANA - Virtual assistance**
- **Facebook Face Recognition Software**
- **Google Translator - Powered by (Guess)?:** However, without Google's substantial mountains of data (in every possible language) there's no way someone can compete with Google Translate. What makes Google so powerful? It's DATA. Google has a lot of data it can leverage, and all that data can be adapted and taught to perform specific tasks.

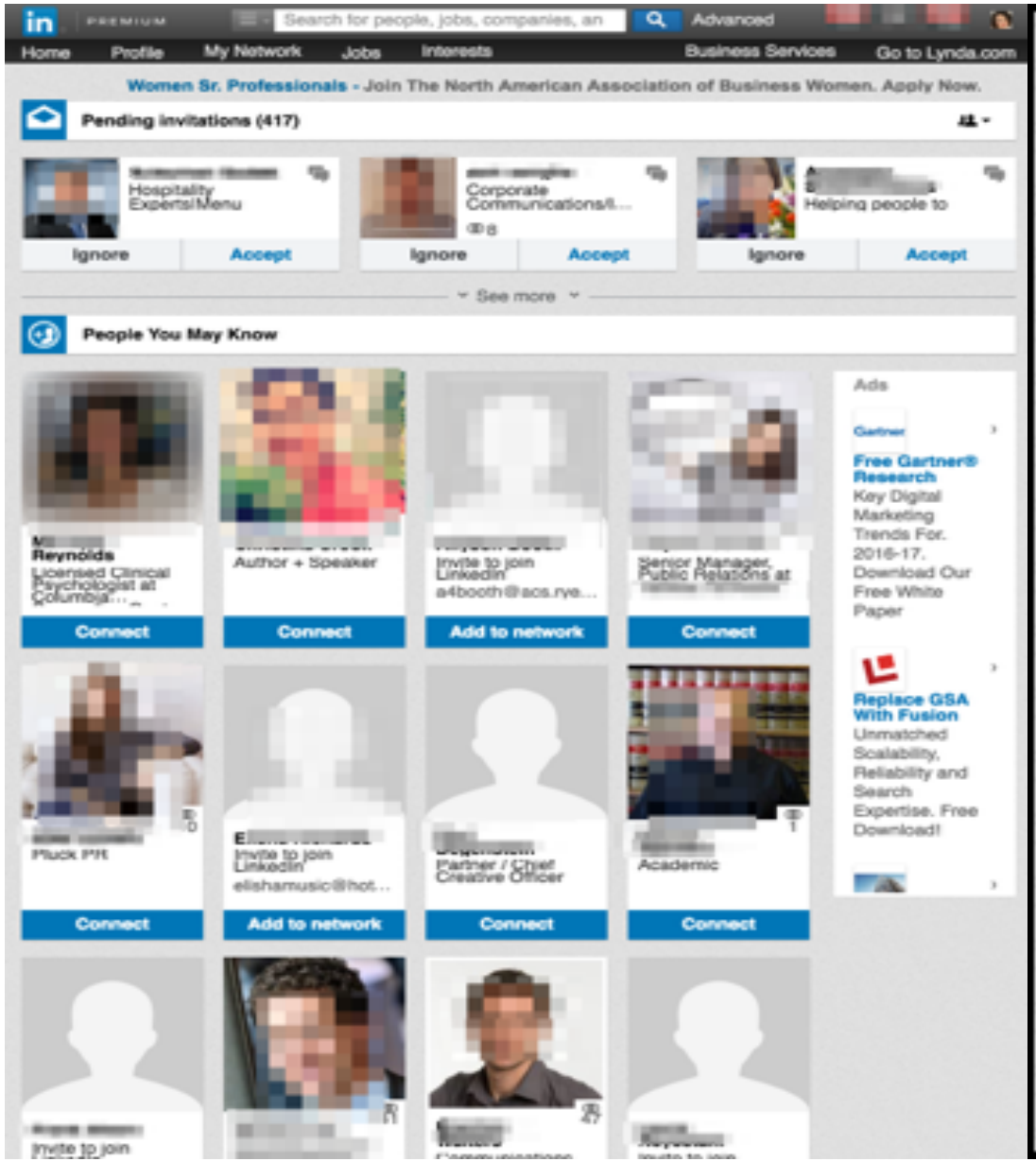
Real life usage of Machine Learning

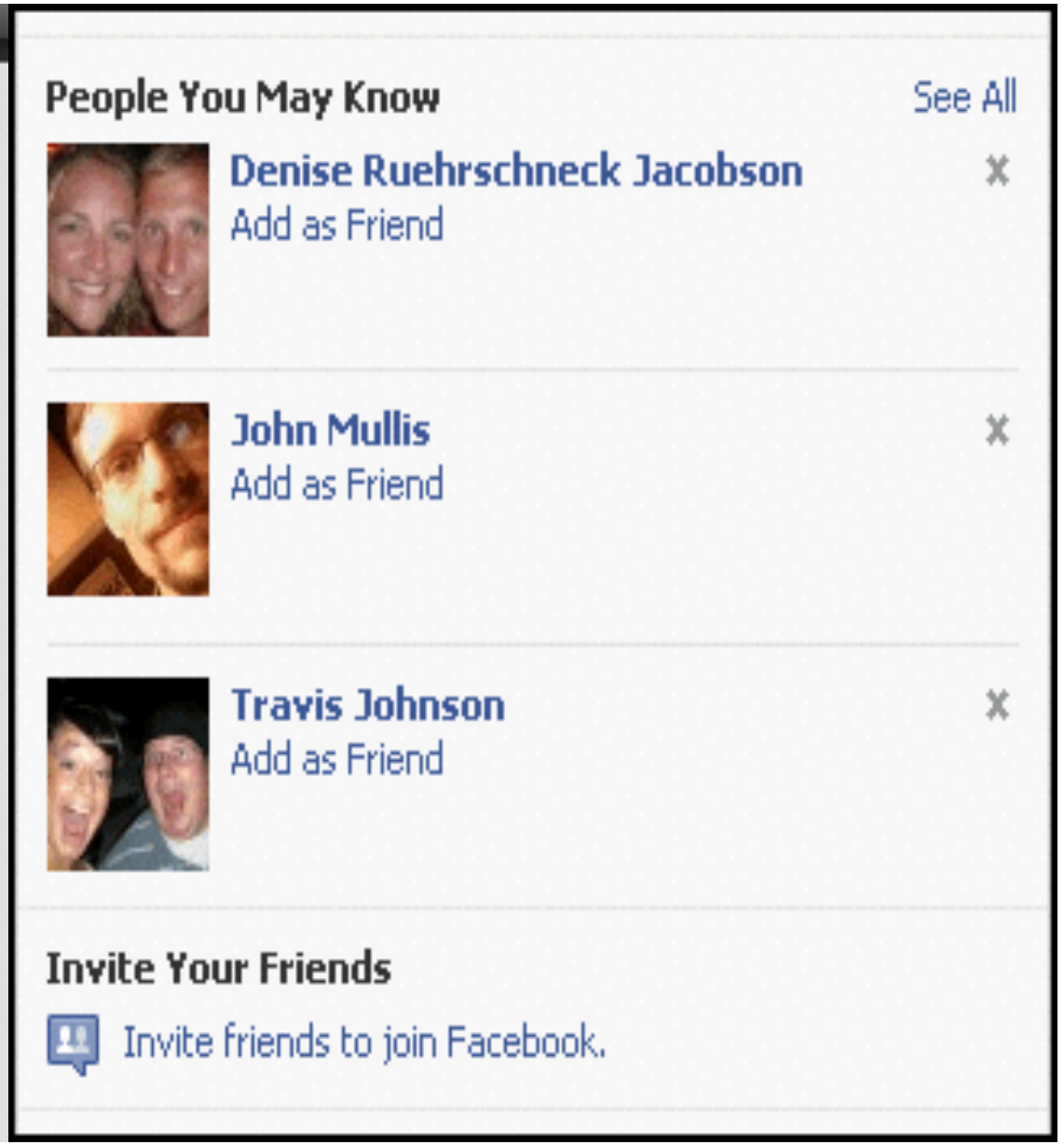
- **Accurate Diagnosis of early stages of cancers**: A system designed by IBM correctly picked the cancerous lesions(damage) in the images with 95% accuracy where a doctor's accuracy is usually between 75% - 84% using manual methods. So, the computing approach will help the doctors make more informed decisions by increasing the efficiency to recognise melanoma and spot the cases where it is difficult for the doctors to identify.

- **Digit Recognition**: A model of this problem would allow a computer program to read and understand handwritten zip codes and sort envelops by geographic region.

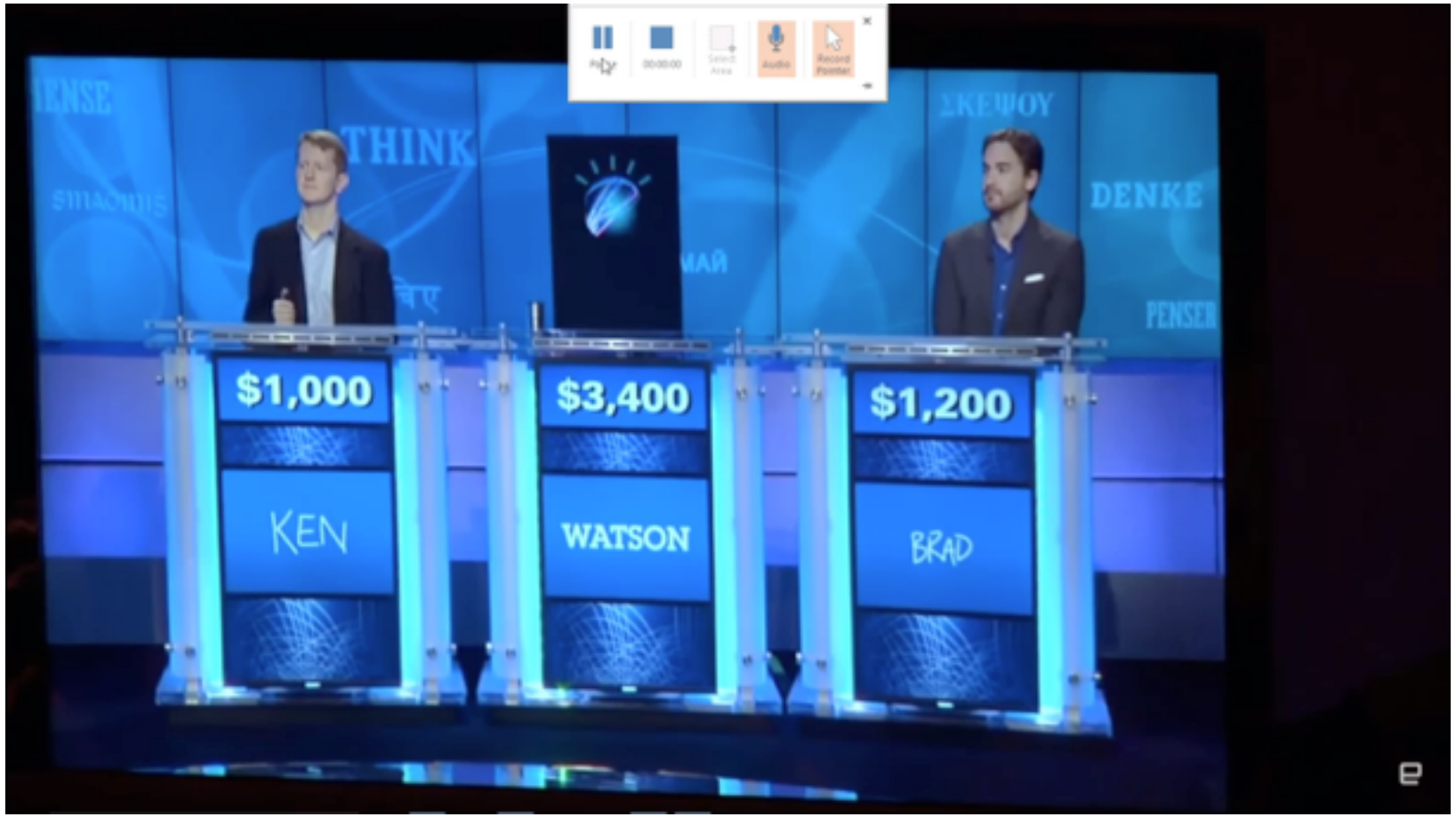
- **Product Recommendation**: A model of this decision process would used the historical data and allow a program to make recommendations to a customer and motivate product purchases. Amazon has this capability. Also think of Facebook, GooglePlus and Facebook that recommend users to connect with you after you sign-up.

- **Credit Card Fraud Detection**: Given credit card transactions for a customer in a month, identify those transactions that were made by the customer and those that were not. A program with a model of this decision could refund those transactions that were fraudulent.


Search for people, jobs, companies, an
Advanced
Home
Profile
My Network
Jobs
Interests
Business Services
Go to Lynda.com
Women Sr. Professionals - Join The North American Association of Business Women. Apply Now.
Pending invitations (417)
Hospitality Experts/Menu
Corporate Communications/I...
Helping people to
Ignore
Accept
See more
People You May Know
Reynolds
Licensed Clinical Psychologist at Columbia...
Author + Speaker
Invite to join LinkedIn
a4booth@aol.rye...
Senior Manager, Public Relations at
Connect
Add to network
Pluck Pit
Invite to join LinkedIn
elisharnusic@hot...
Partner / Chief Creative Officer
Academic
Ads
Gartner
Free Gartner® Research
Key Digital Marketing Trends For 2016-17. Download Our Free White Paper
Replace GSA With Fusion
Unmatched Scalability, Reliability and Search Expertise. Free Download!


People You May Know
See All
Denise Ruehrschneck Jacobson
Add as Friend
John Mullis
Add as Friend
Travis Johnson
Add as Friend
Invite Your Friends
Invite friends to join Facebook.


PIENSE
THINK
ΣΚΕΨΟΥ
DENKE
PENSER
$24,000
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$1,000
$77,147
Who is Bram Stoker?
$ 17,973
$21,600
WHO IS BRAM STOKER?
$5600


THINK
DENKE
PENSER
$1,000
KEN
$3,400
WATSON
$1,200
BRAD


Google
self-driving car

Why Machine Learning is so easy now?

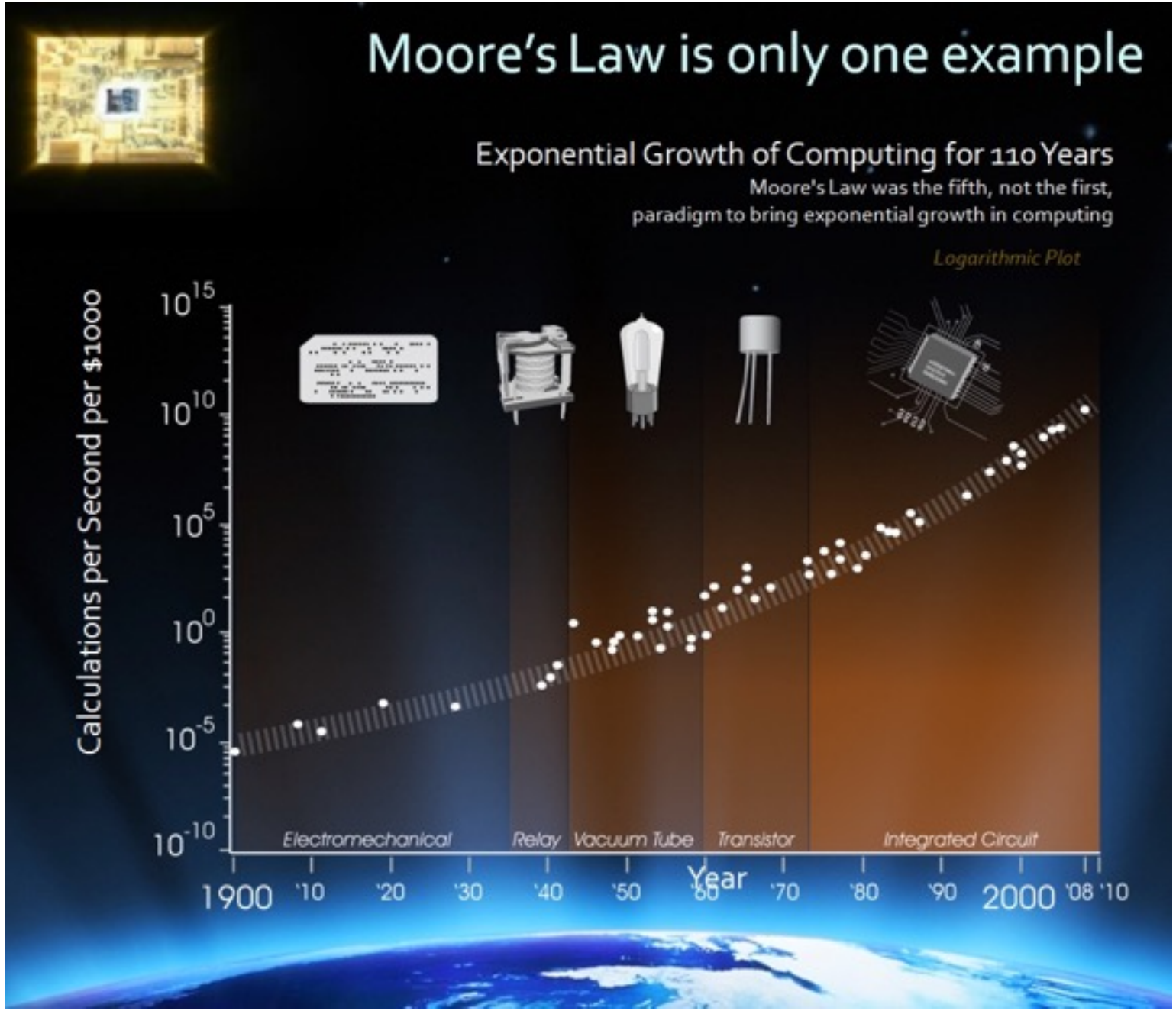
ML has been in the buzz from the 1970s, but why it suddenly started getting more weightage now?

Challenges so far:

Computation : The real challenge, instead, has been the computing horsepower. It takes a long time for machines to learn, to go through all these steps. But as our computers have gotten faster and bigger, machine learning that seemed impossible years ago is now becoming almost commonplace.

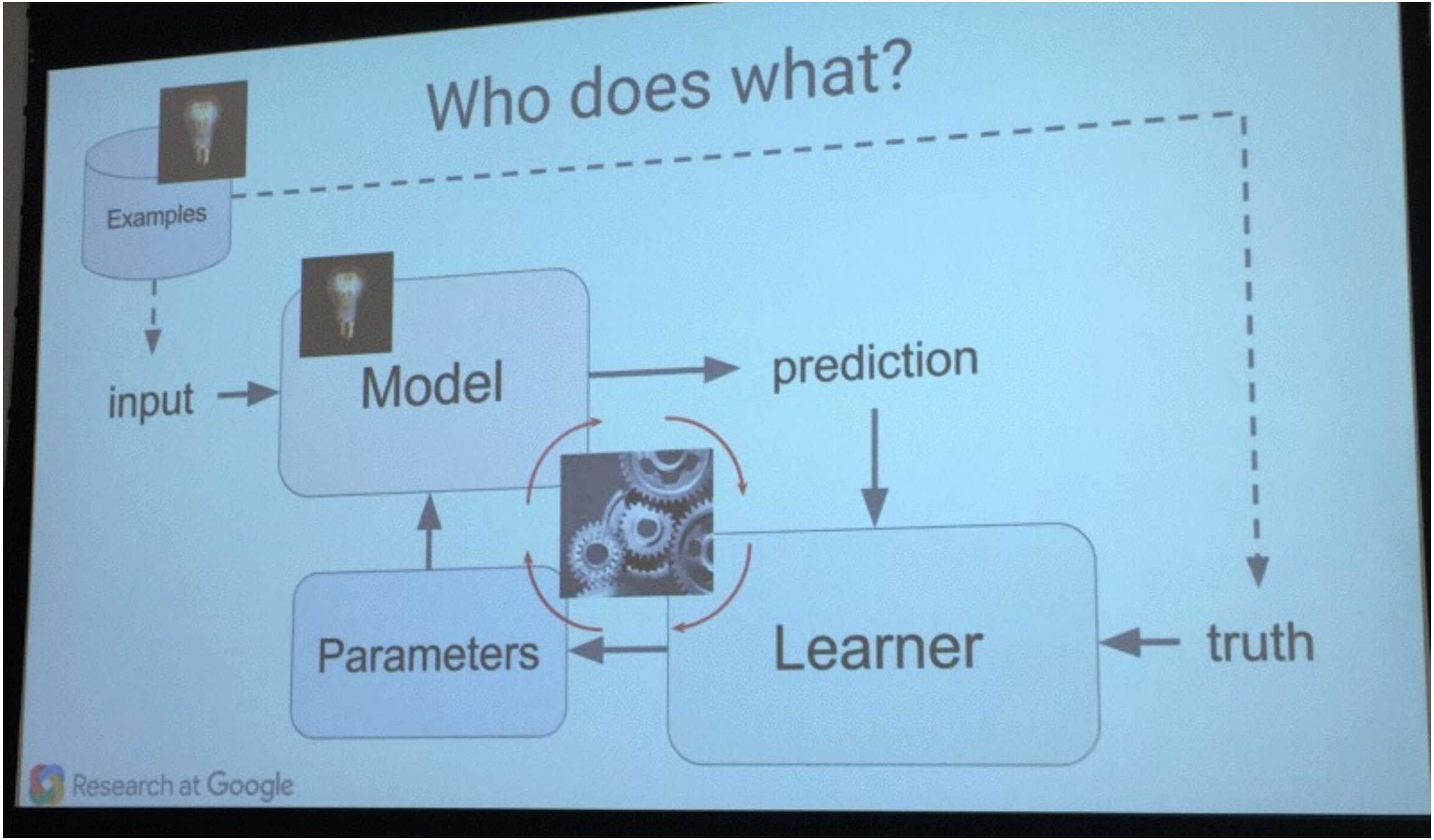
Data storage: We need tons of data to train the machines, with the evolution of High storage devices it has been possible now

Data processing systems: With the help of Big Data and Distributed Processing systems, it has been possible now.


Moore's Law is only one example
Exponential Growth of Computing for 110 Years
Moore's Law was the fifth, not the first,
paradigm to bring exponential growth in computing
Logarithmic Plot
Calculations per Second per $1000
10^15
10^10
10^5
10^0
10^-5
10^-10
Electromechanical
Relay
Vacuum Tube
Transistor
Integrated Circuit
Year
1900
'10
'20
'30
'40
'50
'60
'70
'80
'90
2000
'08
'10

How do Machines Learn ?

Who does what?



Machine learning systems are made up of three major parts:

- **Model**: the system that makes predictions or identifications.

- **Parameters**: the signals or factors used by the model to form its decisions.

- **Learner**: the system that adjusts the parameters — and in turn the model — by looking at differences in predictions versus actual outcome.

Secret of Learning

Gradient Descent or Gradient Learning:

It means that the system makes those little adjustments over and over, until it gets things right. If you likened it to climbing down a steep mountain. You don't want to jump or run, because that's dangerous. You'll more likely make a mistake and fall. Instead, you inch your way down, carefully, a little at a time.

Example:

https://martechtoday.com/how-machine-learning-works-150366

- The learner will again and again adjust the parameters, to reshape the model. Another set of test data will be inputted. A comparison will happen again, and the learner will again adjust the model.

- The cycle will keep repeating until there's a high degree of confidence in the ultimate model, that it really is predicting the outcome of scores based on hours of study. Nearly all successful system uses it, works for 2 parameters or 2 billion

Problem set

- A Specific metallic component is wearing out with use in an acidic environment

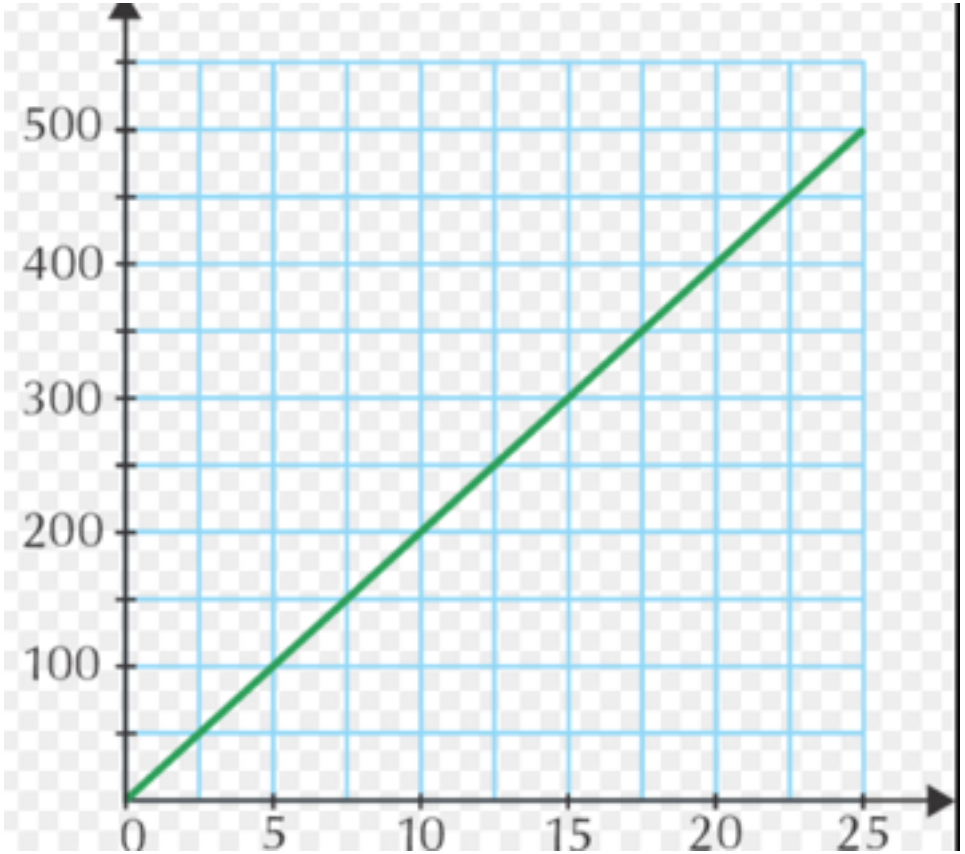
- How do we understand and use this component?

Approach 1…… Science

- A chemist studies what happens to the material in acid and understands that the metal is reacting with the hydrogen in the acid to form a vapor.

- Comes up with a equation and science behind

Approach 2…….Engineering



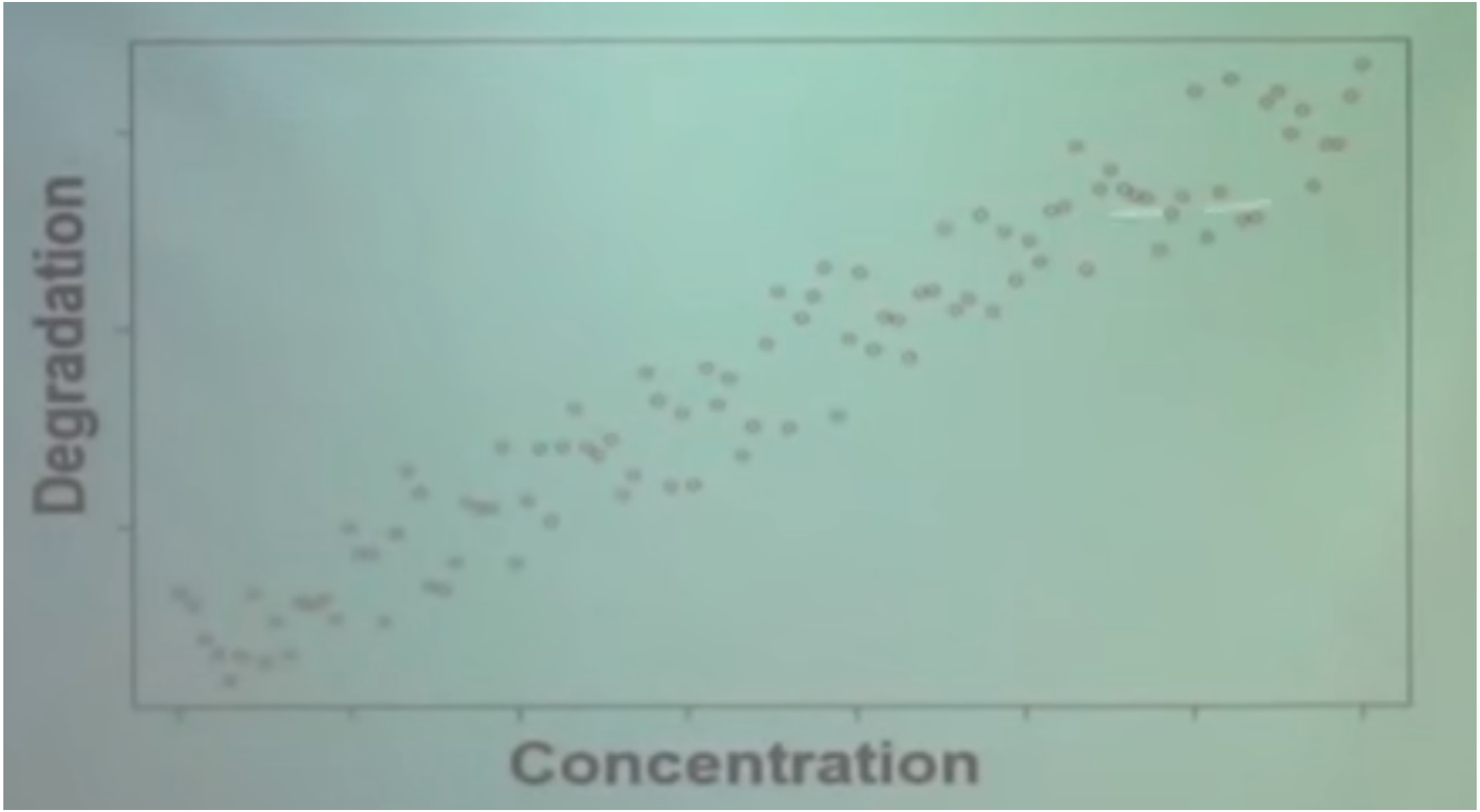
- An engineer tests the degradation in a few concentrations and temperature, plots the degradation, validate the theory and comes up with the thumbs up rule

Deductive learning

- A few hypothesis that cannot be proved
- Rest of science is developed by extending these using mathematics and experiments

Inductive Learning
Approach 2

- Measure the degradations in hundreds of conditions (temperatures and concentrations for various times)


Degradation
Concentration

Result is same but

- Do not know the reasons(causations are not known and only correlations are identified)
- Walmart example(does not tell how to act)
- Marketing customer(does not tell the reason)

Where it does not work

- Fails when randomness prevails
- Need a lot of data to come to the correct conclusions

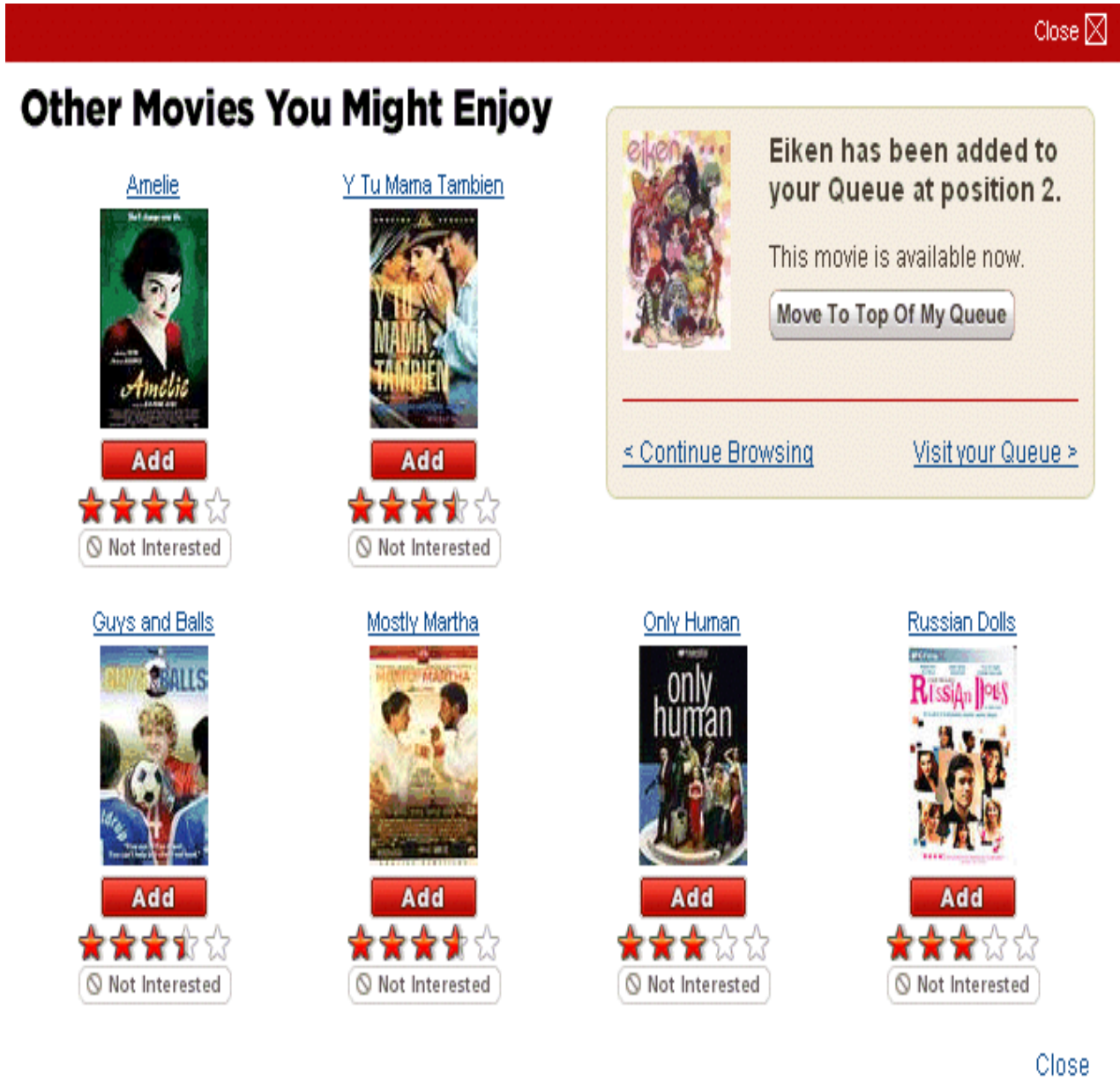
“In God we trust.
All others must bring data.”

- Dr. W. Edwards Deming

Machine learning wonders at past


Google
Google Search
I'm Feeling Lucky


amazon.com
Shop All Departments
Search All Departments
Your Amazon.com › Your Amazon Facebook Page
Facebook Profile Info
Edit your Facebook profile
Birthday:
June 10
Current City:
Chicago, Illinois
You don't have any information about favorite books, music, or movies on Facebook. Edit your Facebook profile and add your favorites to get personalized recommendations on this page.
Birthday and Gift Suggestions for Your Friends on Facebook
November 1
(in 3 weeks)
See gift suggestions
November 23
See gift suggestions
December 8
See gift suggestions
December 27
See gift suggestions
› See all friends on Facebook and their birthdays
Popular Among Your Friends on Facebook
The Godfather DVD Collection ... DVD ~ Marlon Brando
$41.49
Back in Black ~ AC/DC
$3.99
Goldfinger DVD ~ Sean Connery
$12.49
The Beatles Stereo Box Set ~ The Beatles
$188.88
As a Man Thinketh by James Allen
$3.50
› See more


Close
Other Movies You Might Enjoy
Amelie
Add
Not Interested
Y Tu Mama Tambien
Add
Not Interested
Eiken has been added to your Queue at position 2.
This movie is available now.
Move To Top Of My Queue
< Continue Browsing
Visit your Queue >
Guys and Balls
Add
Not Interested
Mostly Martha
Add
Not Interested
Only Human
Add
Not Interested
Russian Dolls
Add
Not Interested
Close

Computers can Learn

Automatic drug discover

Geoffrey Hinton

From Wikipedia, the free encyclopedia

Geoffrey Everest Hinton FRS[12] (born 6 December 1947) is a British-born Canadian[13] cognitive psychologist and computer scientist, most noted for his work on artificial neural networks. As of 2015 he divides his time working for Google and University of Toronto.[14] He was one of the first researchers who demonstrated the use of generalized backpropagation algorithm for training multi-layer neural nets and is an important figure in the deep learning community.[15][16][17]

Contents [hide]

1 Education
2 Career
3 Research
4 Honours and awards
5 Personal life
6 References

Education [edit]

Hinton was educated at King's College, Cambridge graduating in 1970, with a Bachelor of Arts in experimental psychology.[1] He continued his study at the University of Edinburgh where he was awarded a PhD in artificial intelligence in 1977 for research supervised by Christopher Longuet-Higgins.[3][18]

Career [edit]

After his PhD he worked at the University of Sussex, the University of California, San Diego, Carnegie Mellon University.[1] He was the founding director of the Gatsby Charitable Foundation Computational Neuroscience Unit at University College London,[1] and is currently[19] a professor in the computer science department at the University of Toronto. He holds a Canada Research Chair in Machine Learning. He is the director of the program on "Neural Computation and Adaptive Perception" which is funded by the Canadian Institute for Advanced Research. Hinton taught a free online course on Neural Networks on the education platform Coursera in 2012.[20] Hinton joined Google in March 2013 when his company, DNNresearch Inc, was acquired. He is planning to "divide his time between his university research and his work at Google".[21]

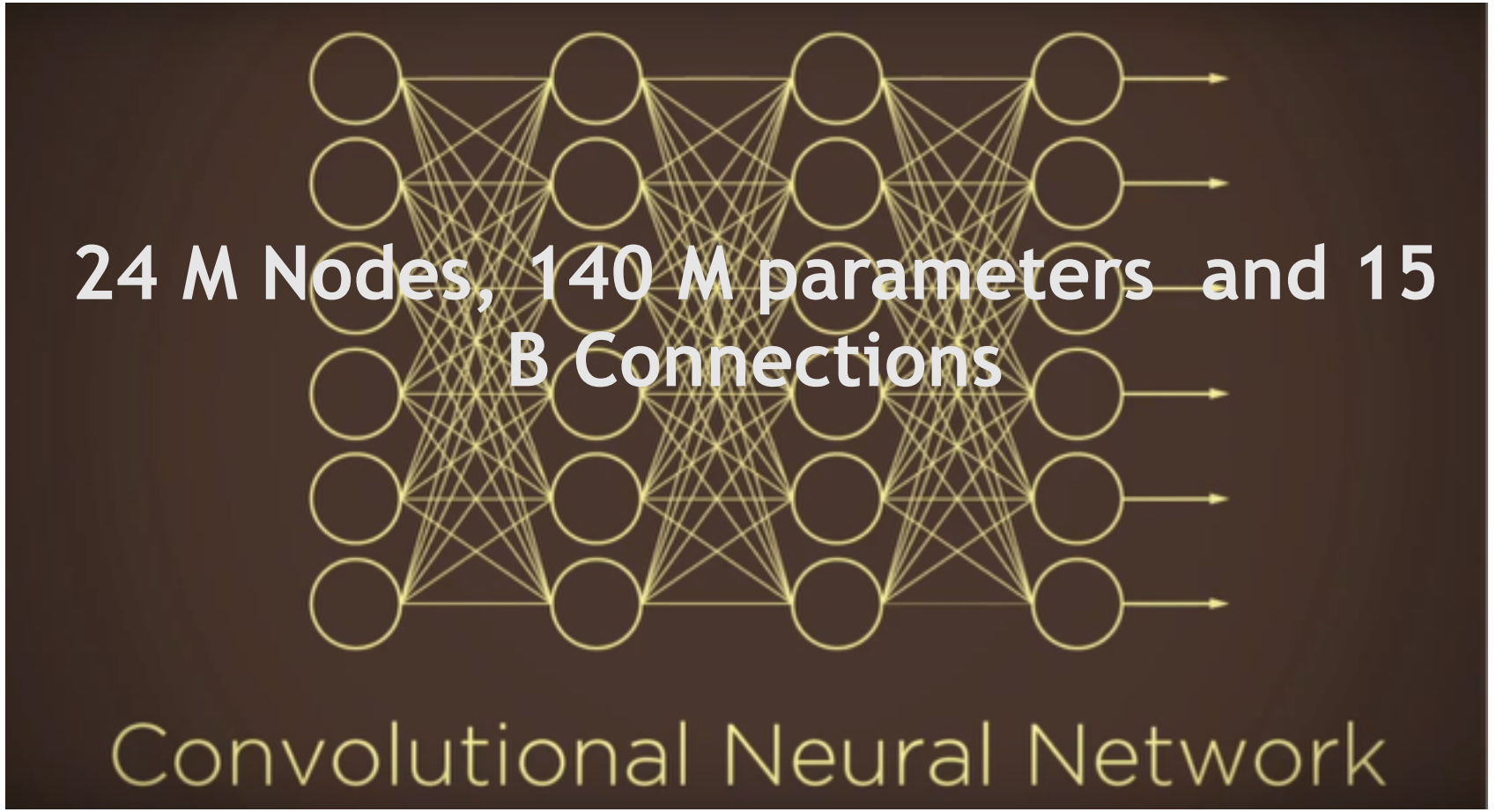
Geoffrey Hinton

Born	Geoffrey Everest Hinton 6 December 1947 (age 69)[1] Wimbledon, London
Residence	Canada
Fields	Machine learning Neural networks Artificial intelligence Cognitive science Object recognition[2]
Institutions	University of Toronto Google

Google DeepMind

Computers can listen and understand

Computers can see


24 M Nodes, 140 M parameters and 15 B Connections
Convolutional Neural Network

Some Interesting things around AI

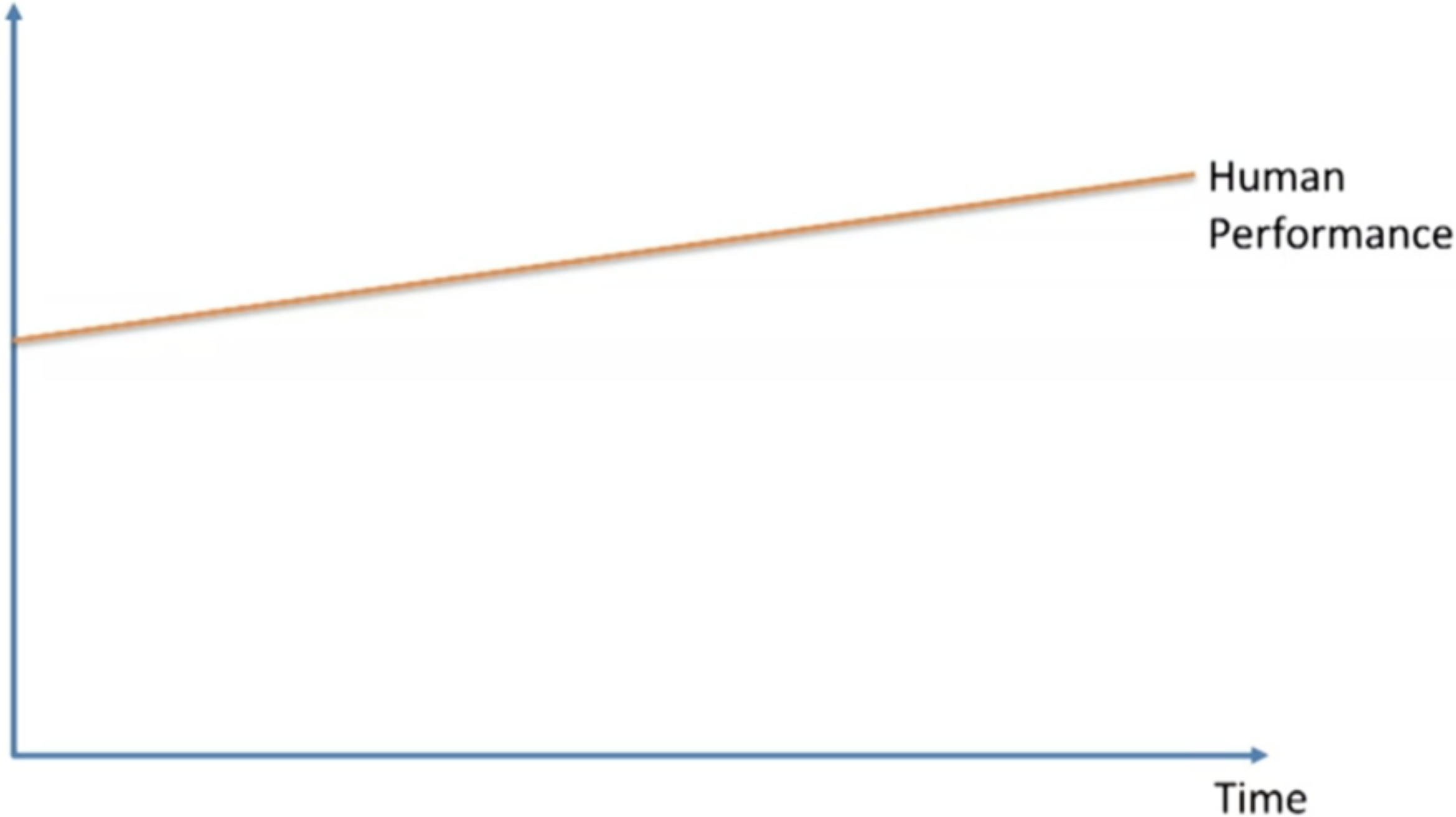
- Artificial Intelligence created its own Movie

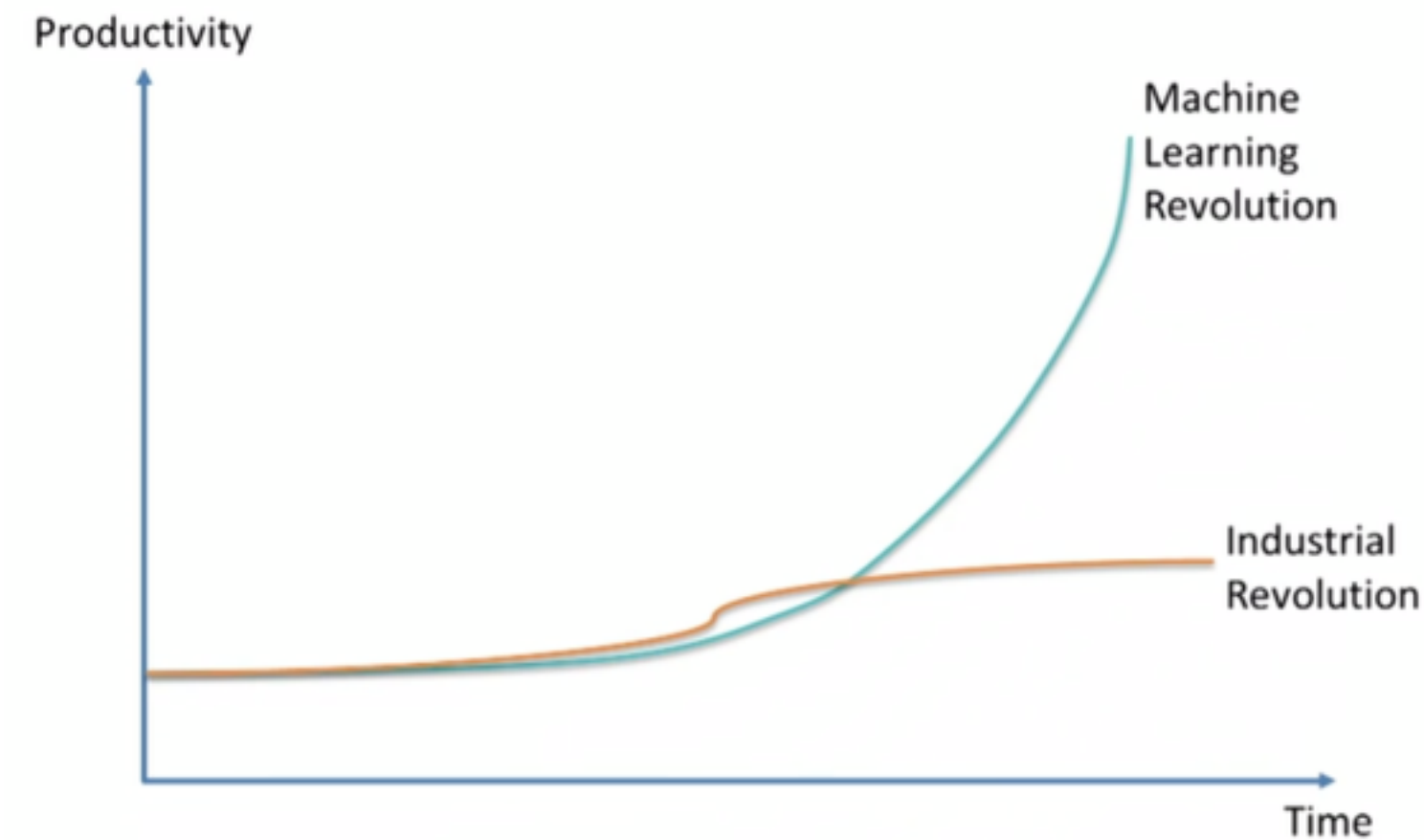
https://arstechnica.com/gaming/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/

- Artificial Intelligence created its own Music
- https://www.youtube.com/watch?v=3OEmzI52stk
- Artificial Intelligence created its own Art
- https://www.youtube.com/watch?v=Sbd4NX95Ysc

For Future


Reading & Writing
Speaking & Listening
Looking at things
Integrating knowledge
• Driving cars
• Preparing food
• Diagnosing disease
• Finding legal precedents
• …


Human
Performance
Time


Productivity
Machine
Learning
Revolution
Industrial
Revolution
Time

Lets Discuss Pros and Cons of ML

Pros & Cons

Advantages of ML?
- Speed, Accuracy, upgradations etc

Disadvantages or limitations of ML
- Can only learn from huge amount of data, for new things, it needs to be educated.

Will it take our job ?

Perception makes the difference :

How long will you have to work to buy BMW?

Doctor: I think I can buy one in 6-8 months of my practice.

MBA: I need about 11-12 months of hard work.

Engineer: At least 2-3 years of very hard work.

Mr Ratan Tata : I think... about 5 years.

Interviewer: Why so long, Mr Tata?

Mr Ratan Tata: Well, its a big _____ to buy

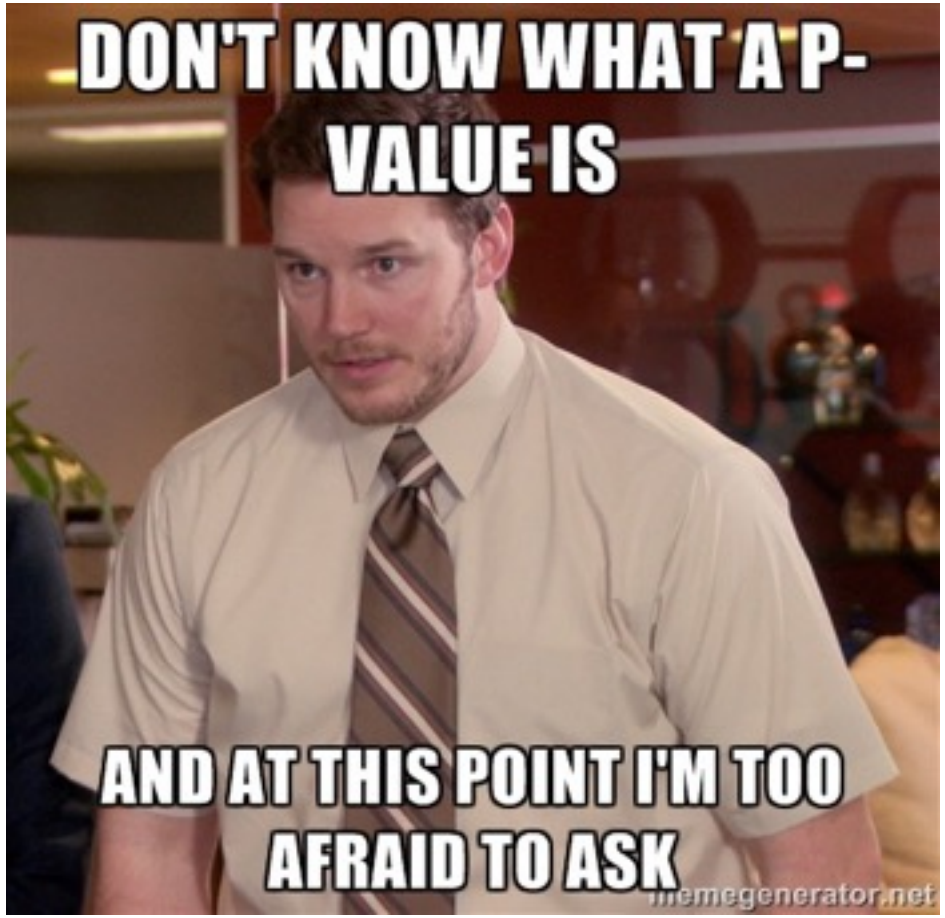
Future Jobs - Michio Kaku - Physicst

- Those of Artificial intelligence can’t do
- Pattern recognition and common sense are big problem of AI
- Repetitive jobs will be wiped off

- Jobs involved in intellectual capitalism will shine like
- Leadership
- Creativity
- Analysis
- Imagination
- Telling joke
- Writing a story, script

How to tame it?

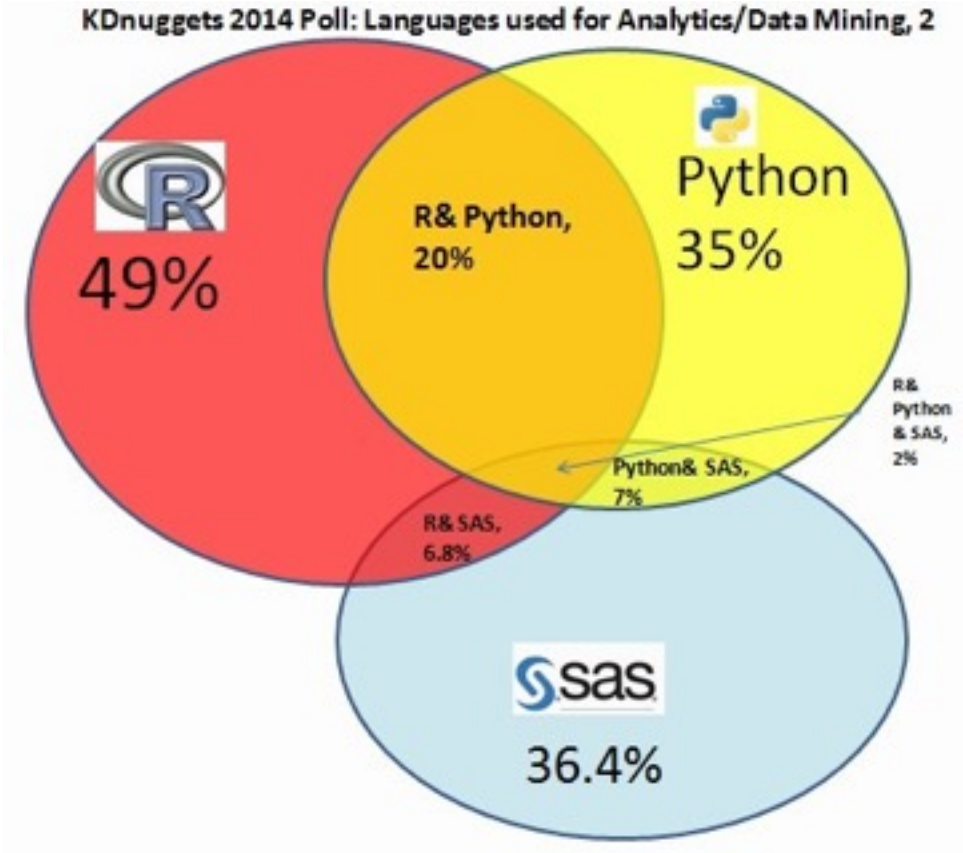
Revise Statistics and Probability



Learn algorithm how it works rather concentrating on output

HOW IT WORKS

Learn and practice a programming language



Big data and Data science are happily married

Practice all algorithm with each use case

Make use of Kaggle and Analytics vidhya

Contribute your work to Opensource

Write a blog

References

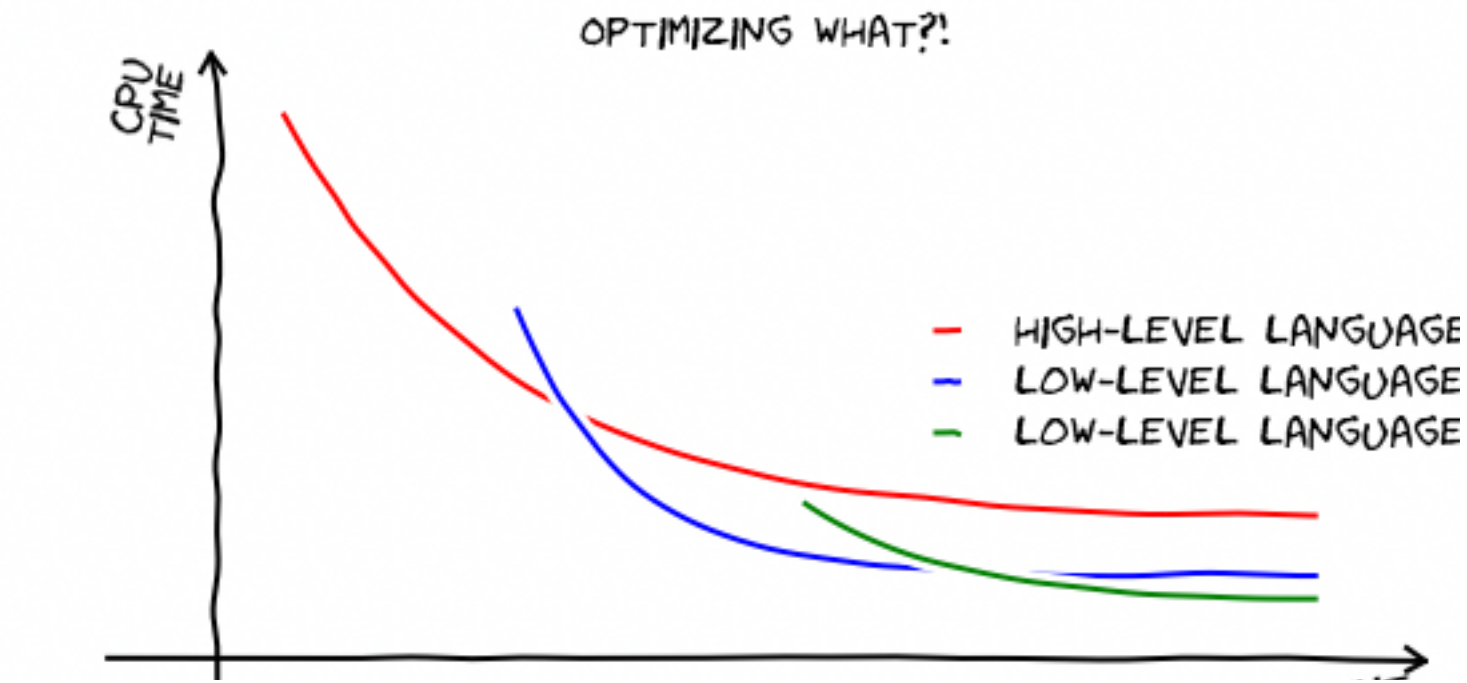
- https://www.ted.com/playlists/310/talks_on_artificial_intelligen
- http://www.kdnuggets.com/
- https://www.youtube.com
- https://www.cubs100.org/GetImage.aspx?IDMF=ea11fbcd-1eb1-4829-9276-41f91b5bd75a&w=900&h=600&src=mc
- https://securityledger.com/wp-content/uploads/2016/05/IBM-Watson.jpg

Python Introduction

Python ?

- Interpretation language
- Intuitive and minimalistic code
- Expressive language
- Dynamically typed (Duck type)
- Automatic memory management

Python Optimizes



Programming syntax

- Variables
- Operators
- Collections [List, set, Dictionary]
- Conditional Statements
- Loops
- Functions

Python differences

- Major difference from other High level language (Java)
  - Duck type programming:
    - Python is strongly but dynamically typed. This means names in code are bound to strongly typed objects at runtime. The only condition on the type of object a name refers to is that it supports the operations required for the particular object instances in the program.
    - For example, I might have two types Person and Car that both support operation "run", but Car also supports "refuel". So long as my program only calls "run" on objects, it doesn't matter if they are Person or Car. This is called "duck typing" after the expression "if it walks like a duck and talks like a duck, it's a duck”.
    - In the real world I don't generally care if I have a rock or a hammer: both have "hit()" interfaces that result in similar consequences when called

Differences

- Indentation instead of braces for blocks:
    - One of the most distinctive features of Python is its use of indentation to mark blocks of code. Consider the if-statement from our simple password-checking program
    - if pwd == 'apple':
    - print('Logging on ...')
    - else:
    - print('Incorrect password.')
    - print('All done!')

NumPy

- **NumPy :**
  - which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed.
    - Using NumPy, a developer can perform the following operations −
    - Mathematical and logical operations on arrays.
    - Fourier transforms and routines for shape manipulation.
    - Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

ndArray

- The most important object defined in NumPy is an N-dimensional array type called ndarray. It describes the collection of items of the same type. Items in the collection can be accessed using a zero-based index.

- Every item in an ndarray takes the same size of block in the memory. Each element in ndarray is an object of data-type object (called dtype).

Mathematical functions

- NumPy has quite a few useful statistical functions for finding minimum, maximum, percentile standard deviation and variance, etc. from the given elements in the array. The functions are explained as follows –

- numpy.amin() and numpy.amax()

PANDAS

- Pandas provides high level data manipulation tools built on top of NumPy.
- Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language
- NumPy by itself is a fairly low-level tool, and will be very much similar to using MATLAB.
- Pandas on the other hand provides rich time series functionality, data alignment, NA-friendly statistics, group by, merge and join methods, and lots of other conveniences. It has become very popular in recent years in financial applications.

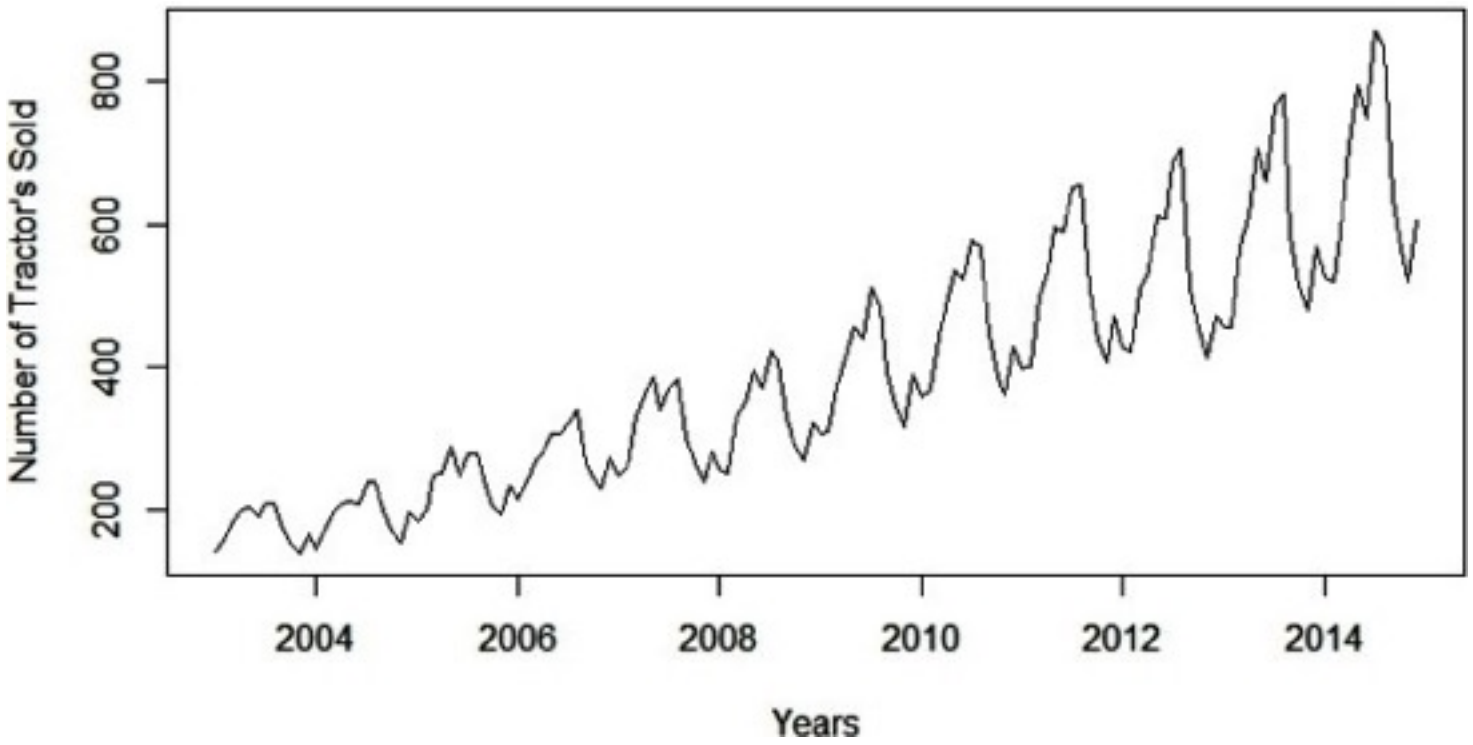
Pandas structure

- Pandas deals with the following three data structures −
  - Series. [1D labeled homogeneous array, size immutable.]
  - DataFrame [General 2D labeled, size-mutable tabular structure with potentially heterogeneously typed columns.]
  - Panel [General 3D labeled, size-mutable array.]

- These data structures are built on top of Numpy array, which means they are fast

- The best way to think of these data structures is that the higher dimensional data structure is a container of its lower dimensional data structure. For example, DataFrame is a container of Series, Panel is a container of DataFrame.

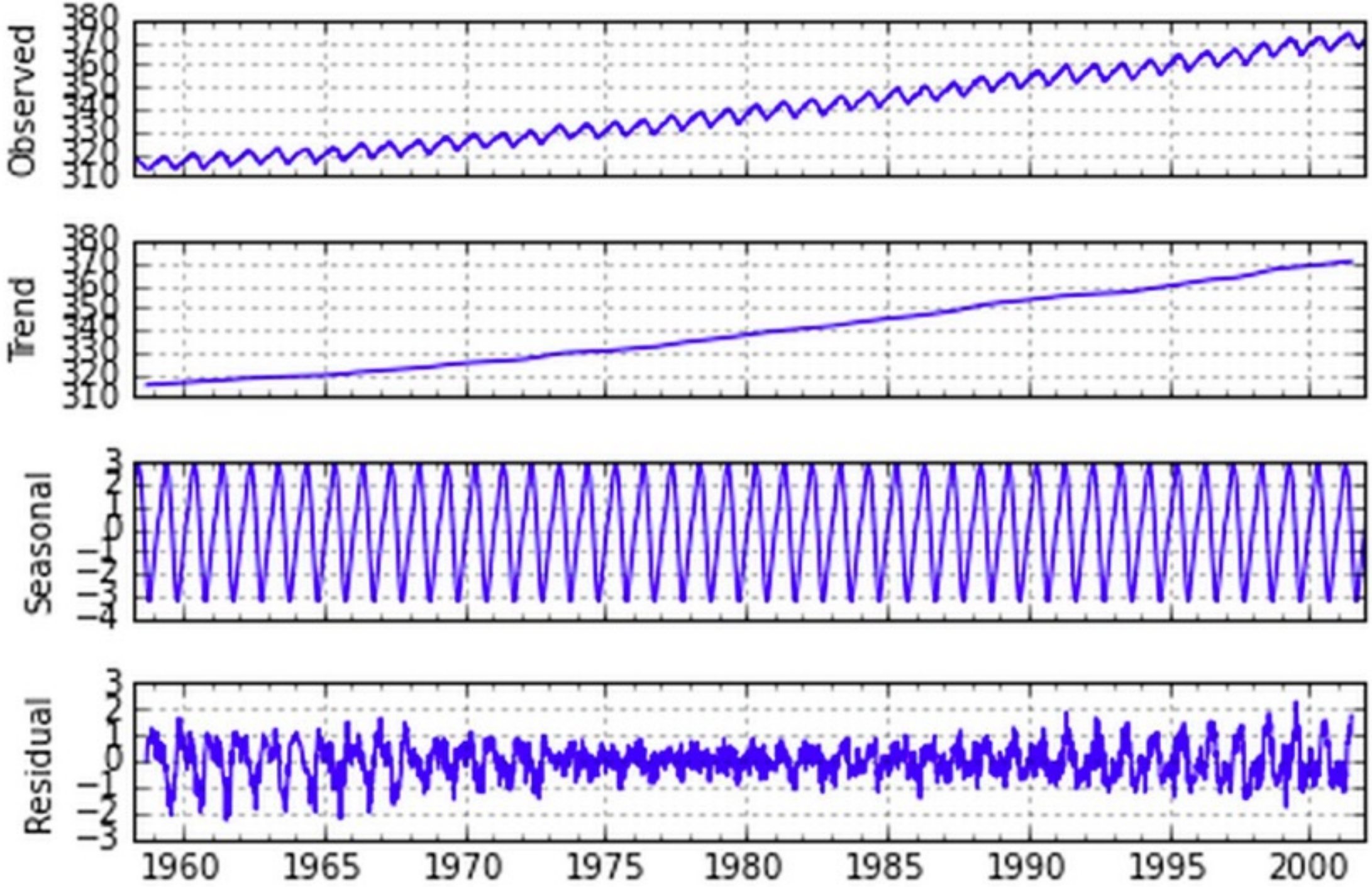
Time Series Plot of Tractors sales



Time Series Forecasting

- **WHAT IS A TIME SERIES? :**
  - **A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year would comprise a time series. This is because sales revenue is well defined, and consistently measured at equally spaced intervals. Data collected irregularly or only once are not time series.**

- The fundamental idea for time series analysis is to decompose the original time series (sales, stock market trends, etc.) into several independent components. Typically, business time series are divided into the following four components:
  - **Trend – overall direction of the series i.e. upwards, downwards etc**
  - **Seasonality – monthly or quarterly patterns**
  - **Cycle – long-term business cycles**
  - **Irregular remainder / Residuals – random noise left after extraction of all the components**


Observed
Trend
Seasonal
Residual
1960
1965
1970
1975
1980
1985
1990
1995
2000

Stationary Time series

- The mean of the series should not be a function of time rather should be a constant.
- The variance of the series should not a be a function of time.
- The covariance of the i th term and the (i + m) th term should not be a function of time.

- However, we often do look for stationarity. Why?

- **Consider the forecasting problem. How do you forecast? If everything's different tomorrow then it's impossible to forecast, because everything's going to be different. So the key to forecasting is to find something that will be the same tomorrow, and extend that to tomorrow. That something can be anything. I'll give you a couple of examples.**

- For forecasting we absolutely need to find the constant (time invariant) component in the series, otherwise it's impossible to forecast by definition. Stationarity is just a particular case of the invariance.

How to make time series stationary

- **How to make a Time Series Stationary:** Though stationarity assumption is taken in many TS models, almost none of practical time series are stationary. So statisticians have figured out ways to make series stationary, which we'll discuss now. Actually, its almost impossible to make a series perfectly stationary, but we try to take it as close as possible.

- Lets understand what is making a TS non-stationary. There are 2 major reasons behind non-stationary of a TS:
  - Trend – varying mean over time. For eg, in this case we saw that on average, the number of passengers was growing over time.
  - Seasonality – variations at specific time-frames. eg people might have a tendency to buy cars in a particular month because of pay increment or festivals.

- **The underlying principle is to model or estimate the trend and seasonality in the series and remove those from the series to get a stationary series. Then statistical forecasting techniques can be implemented on this series. The final step would be to convert the forecasted values into the original scale by applying trend and seasonality constraints back.**

ARIMA

- The ARIMA model is all about differencing the series to make it stationary.

- ARIMA is a broader class of time series models. It helps to model by eliminating the non-stationary elements and forecast and then apply the trend and seasonality to it.

- ARIMA is a combination of 3 parts i.e. AR (AutoRegressive), I (Integrated), and MA (Moving Average). A convenient notation for ARIMA model is ARIMA(p,d,q). Here p,d, and q are the levels for each of the AR, I, and MA parts.

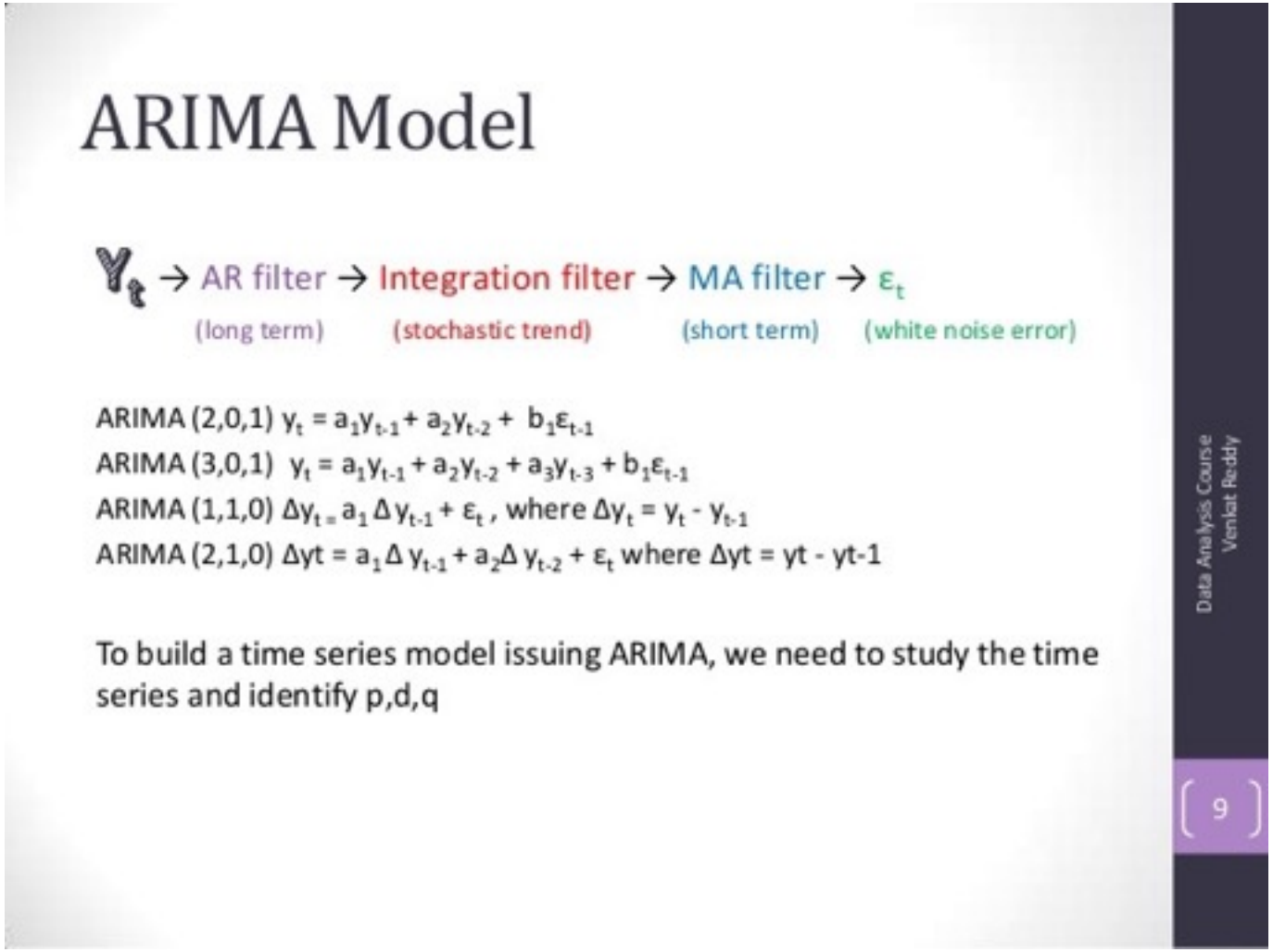
- Each of these three parts is an effort to make the final residuals display a white noise pattern (or no pattern at all). In each step of ARIMA modeling, time series data is passed through these 3 parts

- **Integrated (I) – subtract time series with its lagged series to extract trends from the data**

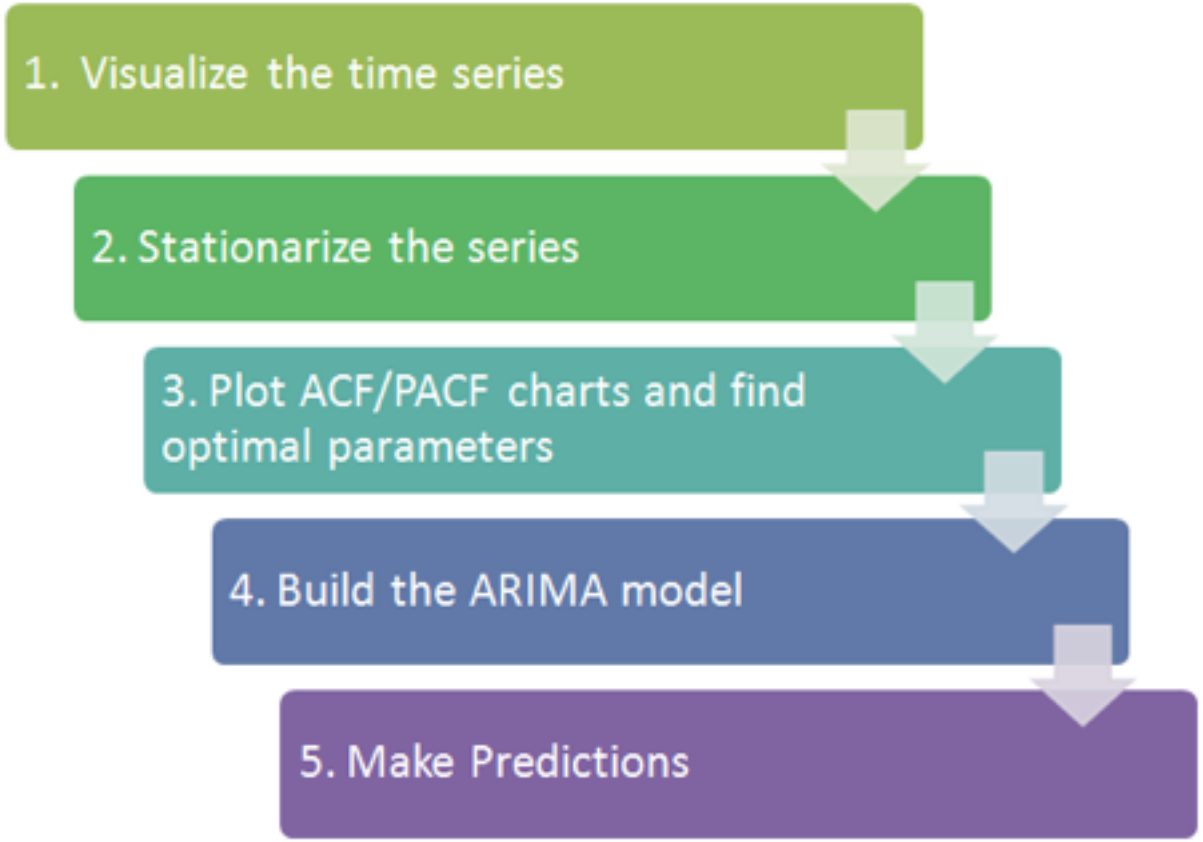
- **AutoRegressive (AR) – extract the influence of the previous periods' values on the current period**

- **Moving Average (MA) – extract the influence of the previous period's error terms on the current period's error**

AR, I & MA


ARIMA Model
Y_t → AR filter → Integration filter → MA filter → ε_t
(long term)
(stochastic trend)
(short term)
(white noise error)
ARIMA (2,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1}$
ARIMA (3,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 \varepsilon_{t-1}$
ARIMA (1,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$
ARIMA (2,1,0) $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \varepsilon_t$ where $\Delta yt = yt - yt\text{-}1$
To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q
Data Analysis Course
Venkat Reddy
9

Flow chart to forecasting



- ACF: Since autocorrelation is the linear correlation of a signal with itself at two different points in time, ACF (autocorrelation function) is just such correlation as a function of the lag hh between two points of time, like acf(h)=corr(xt,xx+h)

- PACF: (partial autocorrelation function) is essentially the autocorrelation of a signal with itself at different points in time, with linear dependency with that signal at shorter lags removed, as a function of lag between points of time. Informally, the partial correlation between xtxt and xt+hxt+h is the autocorrelation between xtxt and xt+hxt+h without the contribution of xt+1,xt+2,....,xt+h−1xt+1,xt+2,....,xt+h−1.

PACF

- Suppose you have 3 points in a time series x3 - x2 - x1. Using ACF you would generally find the correlation between x1 and x2 , for example. The value of correlation thus obtained is technically not true value of correlation, because the value of x2 is likely to be inspired by the value of x3.
- So PACF is that portion of the correlation between x1 and x2, which is not explained by the correlation between x3 in x2.
- Graphically, lets say, the correlation between x1 and x2 is given by the PINK box(including the GREEN box), and the correlation between x2 and x3 is given by the GREEN box. So the partial correlation of x1 and x2 is the original correlation minus the correlation of x2 and x3(GREEN box), which is just the PINK Pentagon.
- 

Q & A