

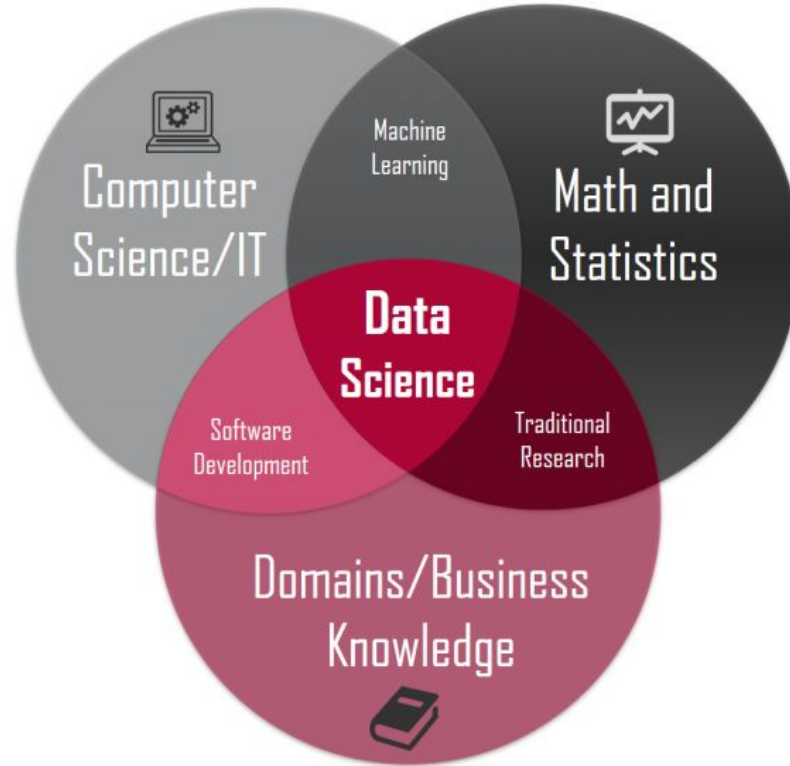
# Class 12

## Wrap-up

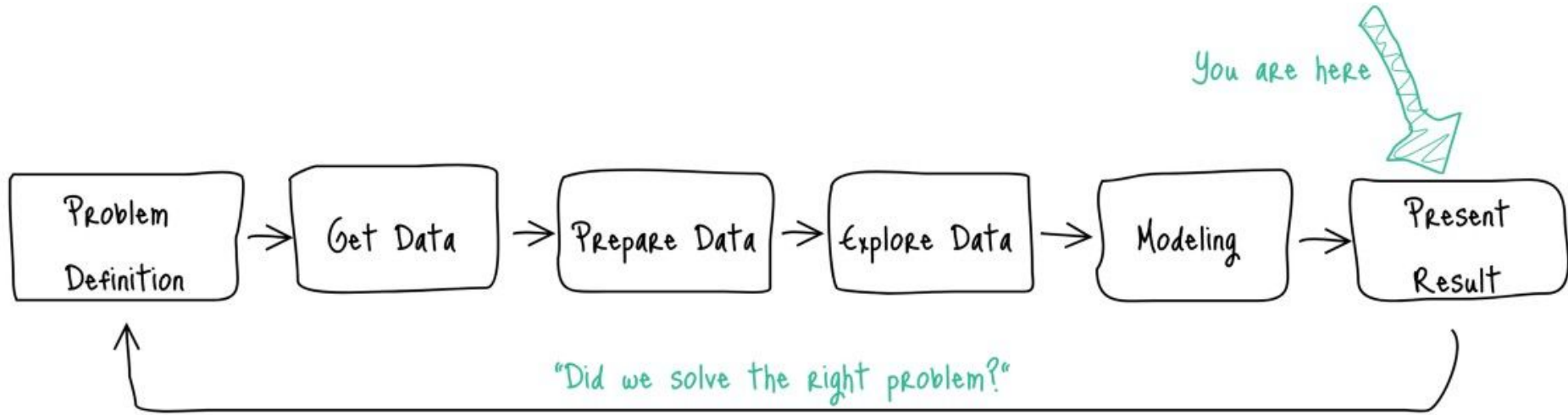
### H Academy

March 24th, 2021 - By Nathan Landman

# Class 1 - What is Data Science?



# Class 1 - What is Data Science?



# Class 1 - What is Data Science?

## THE DATA SCIENCE HIERARCHY OF NEEDS

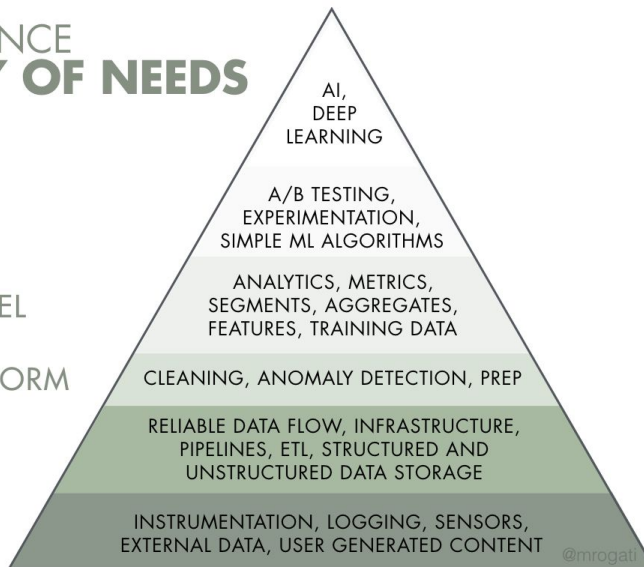
LEARN/OPTIMIZE

AGGREGATE/LABEL

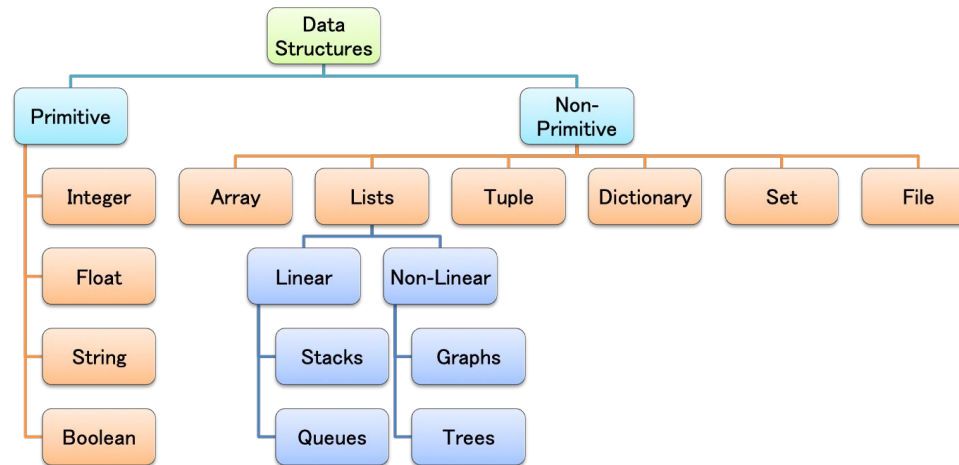
EXPLORE/TRANSFORM

MOVE/STORE

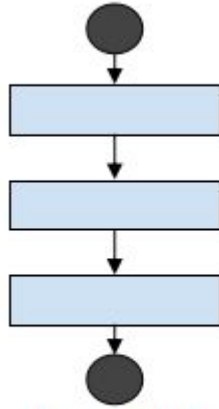
COLLECT



# Class 2 - Python Basics - Data Structures

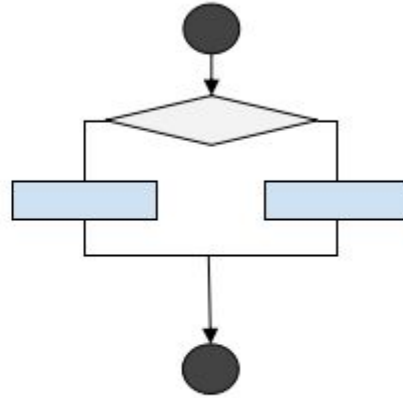


# Class 2 - Python Basics - Control Flow



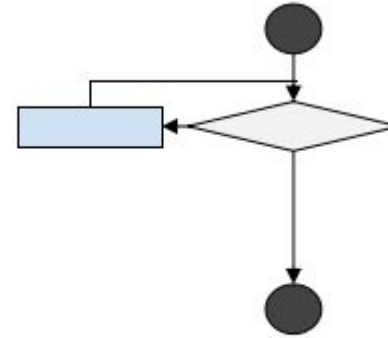
**Sequential**

*Lines of Code*



**Conditional**

*If/Then*

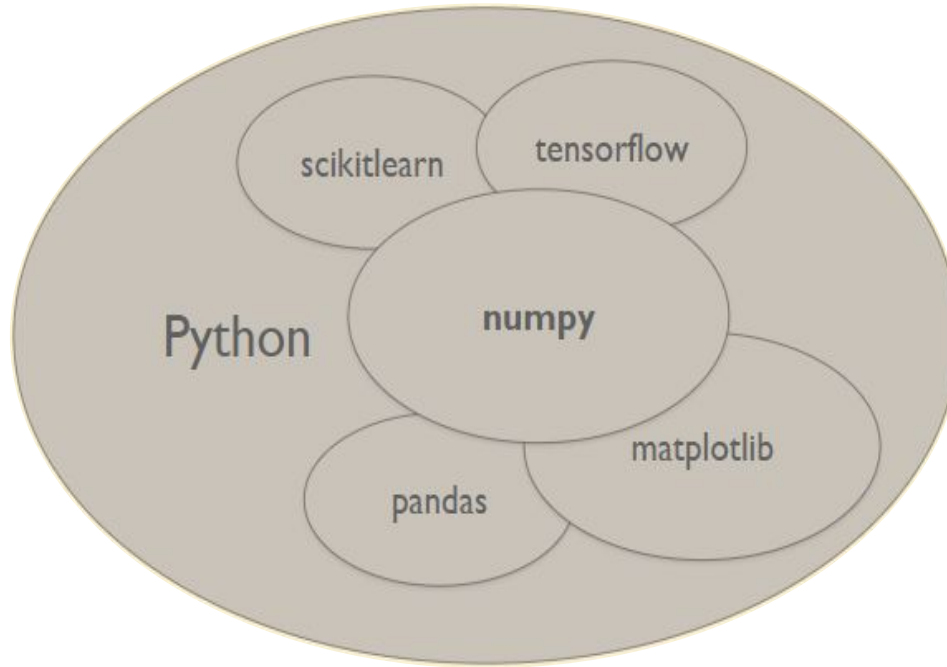


**Looping**

*For Loop*

*While Loop*

# Class 3 - The Python Data Science Ecosystem



# Class 4 - Pandas and Numpy

**Series**

	apples
0	3
1	2
2	0
3	1

+

**Series**

	oranges
0	0
1	3
2	7
3	2

=

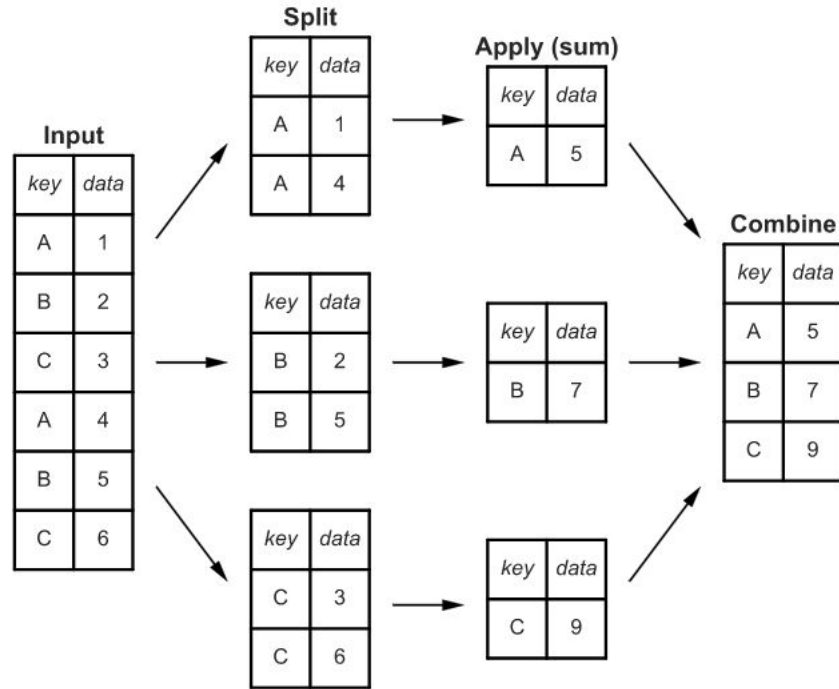
**DataFrame**

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

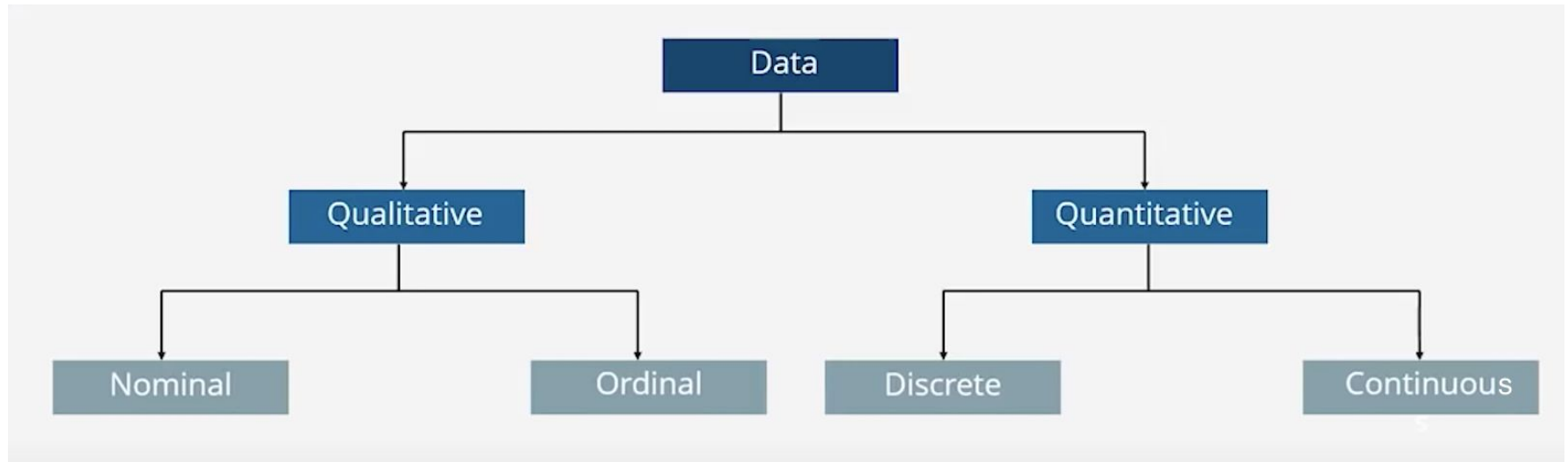




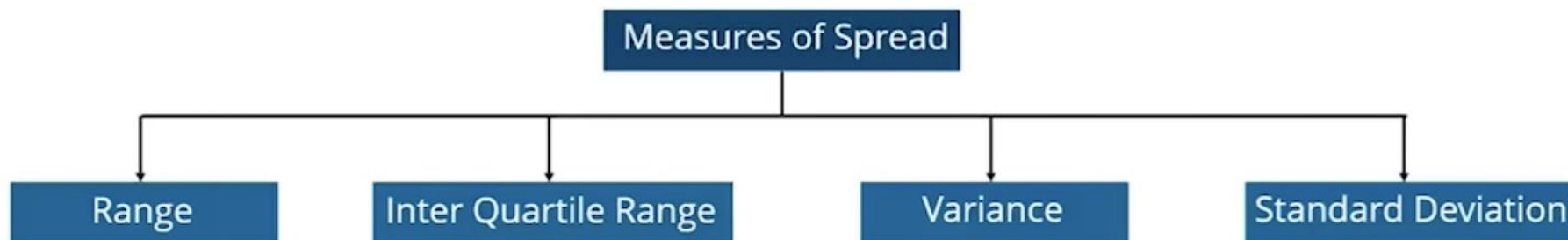
# Class 4 - Pandas Split-Apply-Combine (groupby)



# Class 5 and 6 - Exploring Datasets



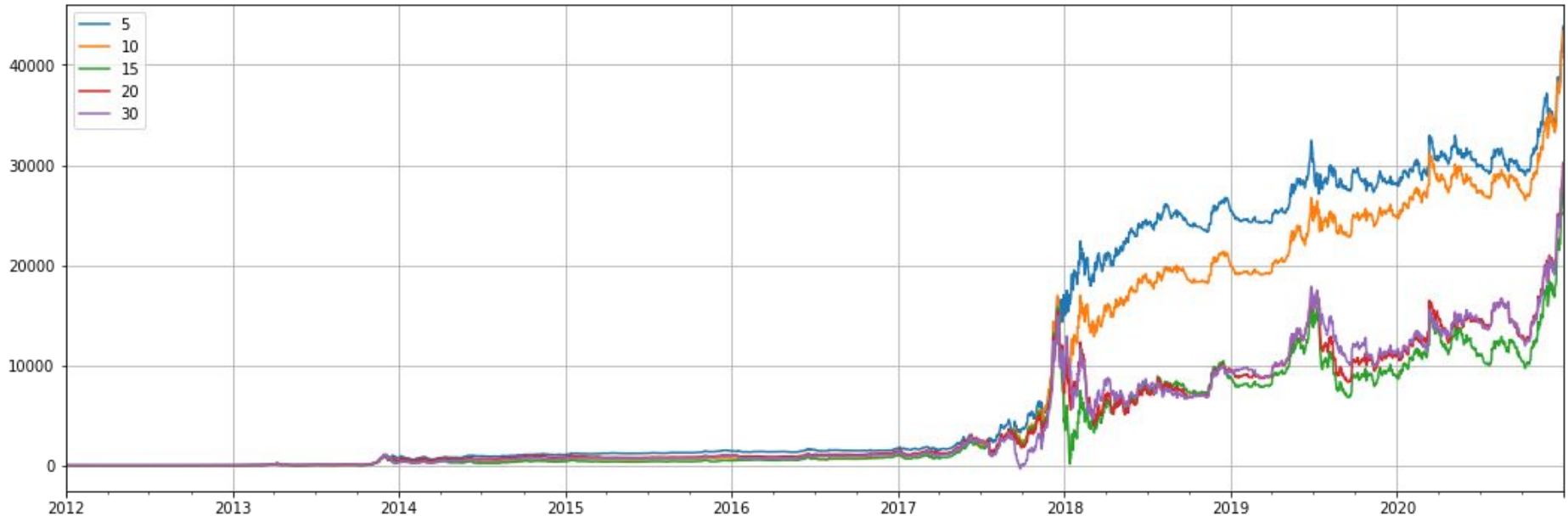
# Class 5 and 6 - Exploring Datasets



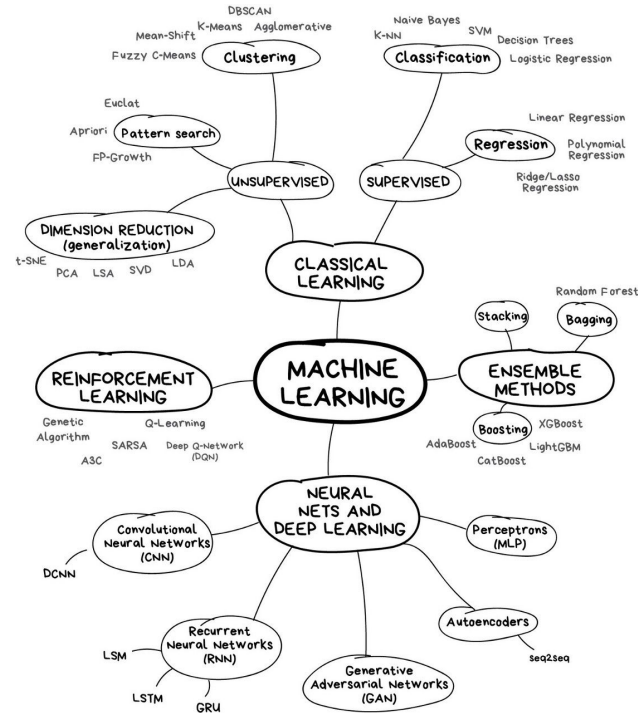
$$S^2 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})^2}{n}$$

$$\sigma = \sqrt{S^2}$$

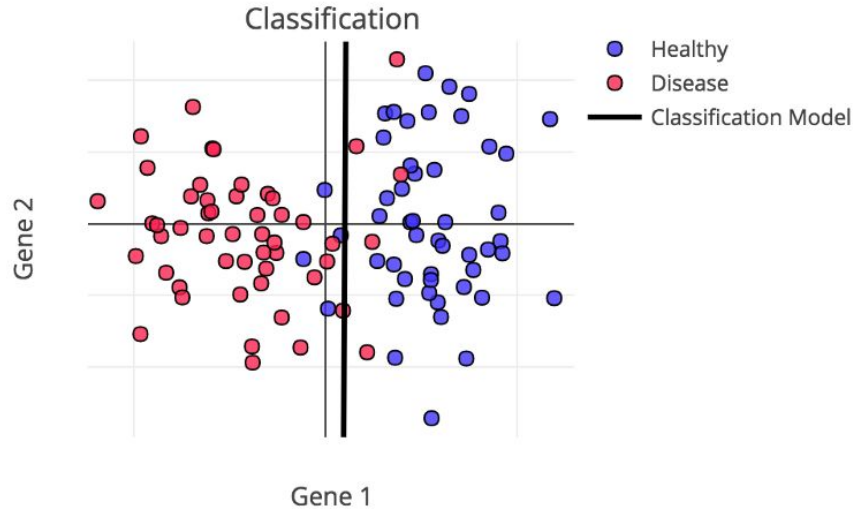
# Class 7 - Time Series Analysis - Trading BTC



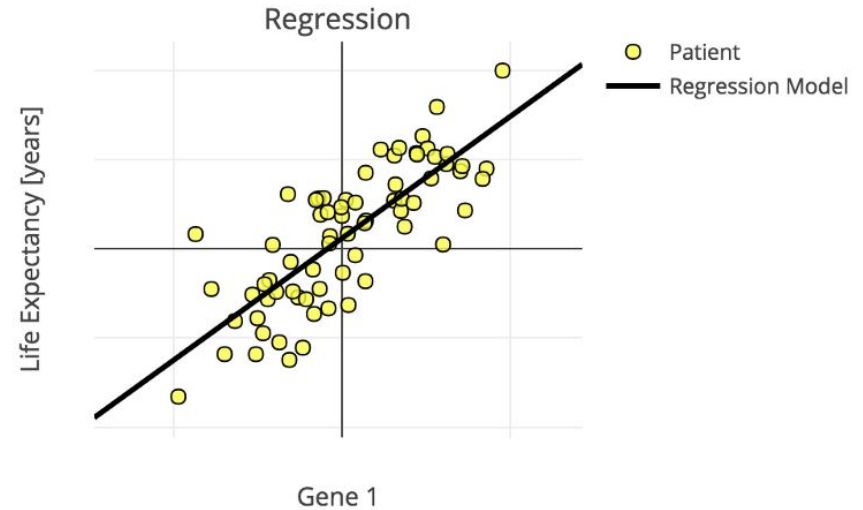
# Class 7 - A Machine Learning Engineer's Toolbox



# Class 7 - A Machine Learning Engineer's Toolbox



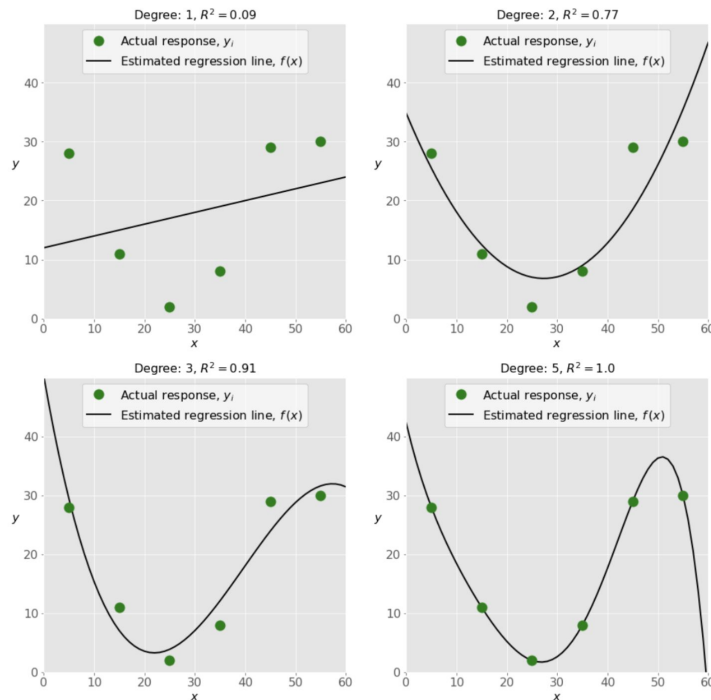
Logistic Regression



Linear Regression

# Class 7 - A Machine Learning Engineer's Toolbox

Bias vs.  
Variance

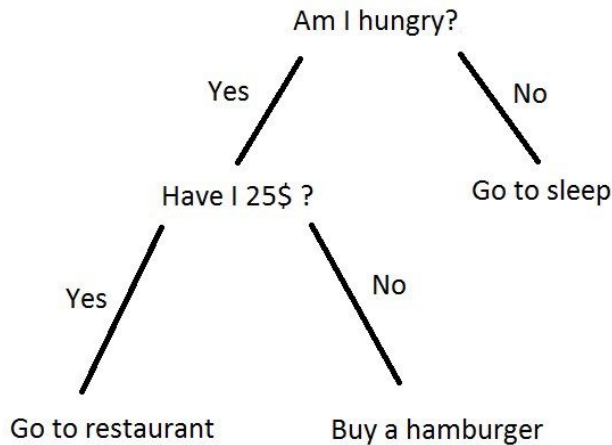


Example of underfitted, well-fitted and overfitted models

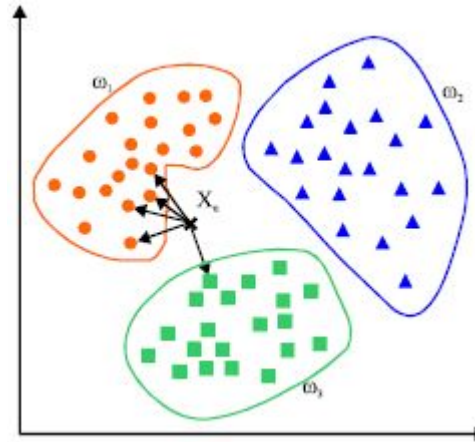
Overfit vs.  
Underfit

# Class 8 - A Machine Learning Engineer's Toolbox

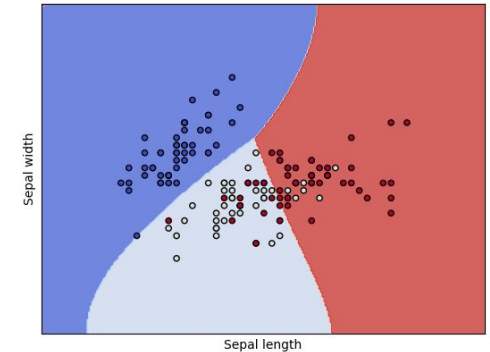
## Decision Trees



## kNN



## SVMs

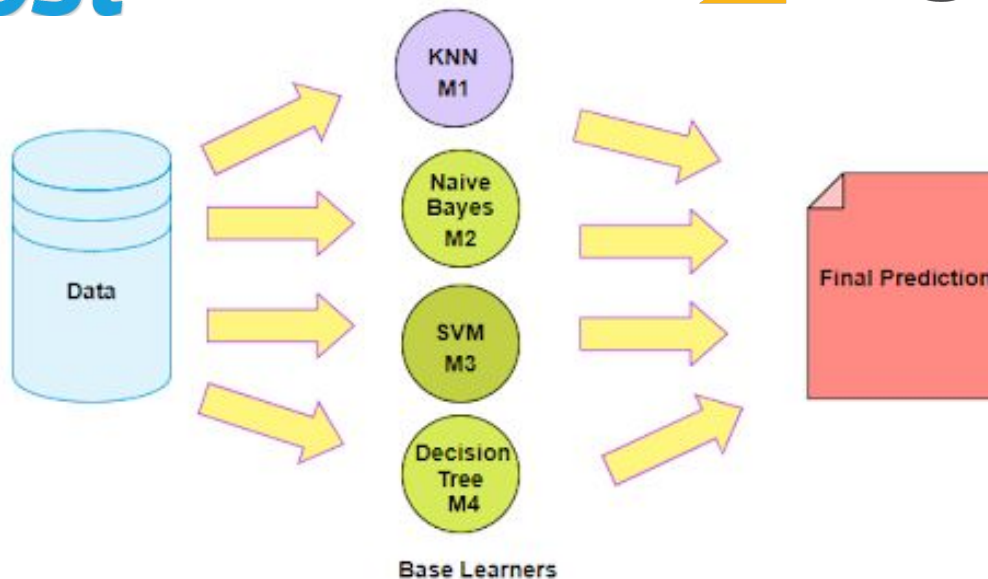




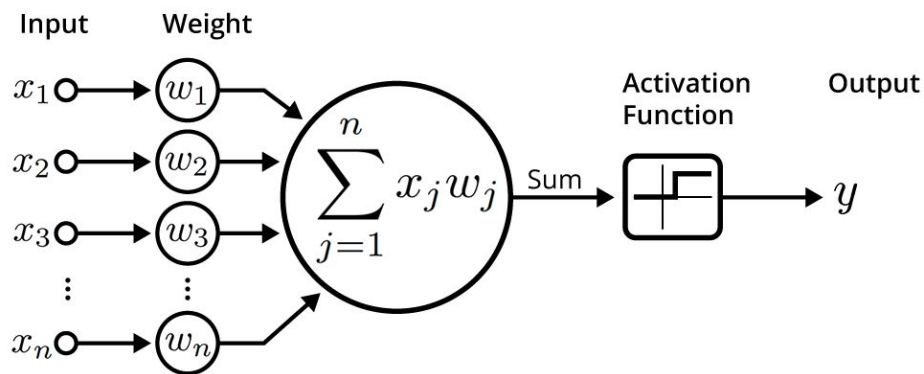
# Class 9 - A Machine Learning Engineer's Toolbox

*dmlc*  
**XGBoost**

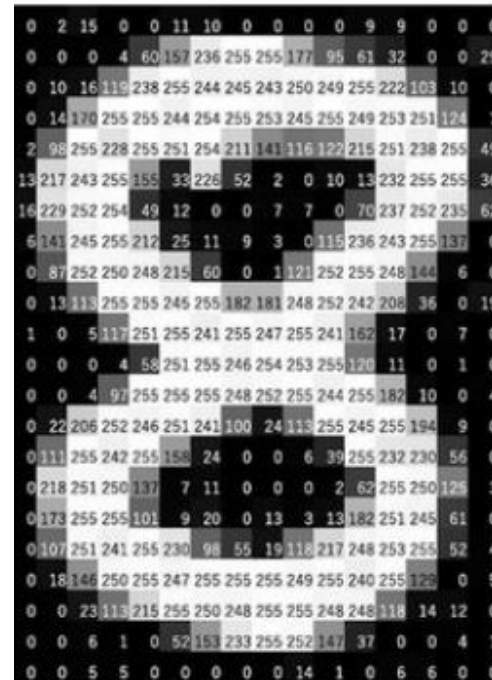
 **LightGBM**



# Class 10 - Deep Learning in Machine Vision



An illustration of an artificial neuron. Source: Becoming Human.



# Class 10 - Deep Learning Concepts

## Training in Batches

### Epoch :

An Epoch represent one iteration over the entire dataset.



### Batch :

We cannot pass the entire dataset into the Neural Network at once. So, we divide the dataset into number of batches.

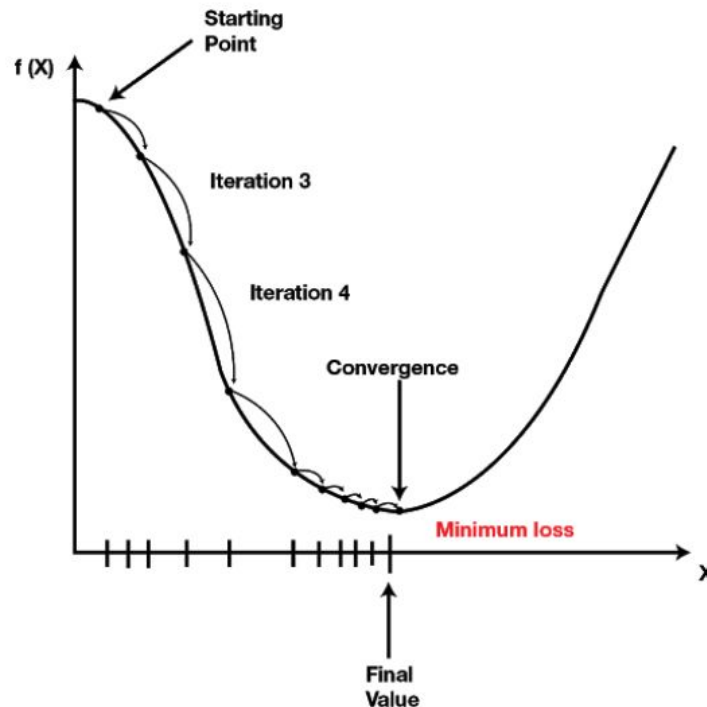


### Iteration :

If we have 1000 images as Data and a batch size of 20, then an Epoch should run  $1000/20 = 50$  iteration.

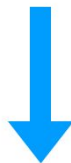


## Learning Rate



# Class 11 - Advances in NLP - Training

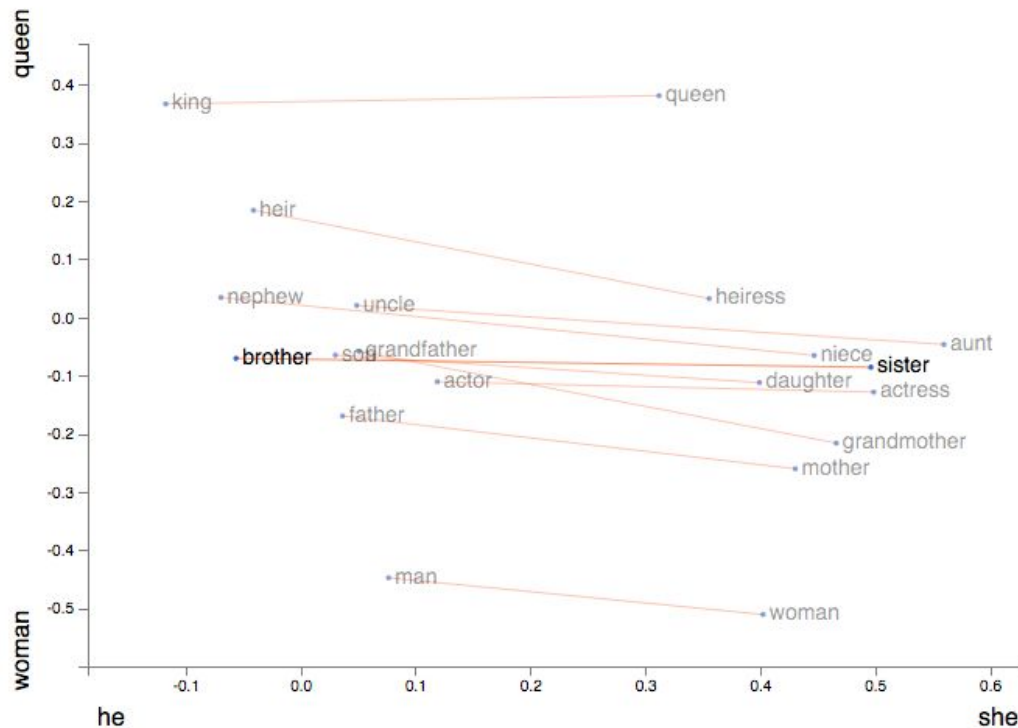
**Text:** Second Law of Robotics: A robot must obey the orders given it by human beings



**Generated training examples**

Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot
3	Second law of robotics : a robot	must
...		

# Class 11 - Advances in NLP - Embeddings



# Class 11 - Advances in NLP - Modern Models

Input Prompt:

Recite the first law of robotics



Output:

# Class 11 - Advances in NLP - Modern Models

Input Prompt:

Recite the first law of robotics



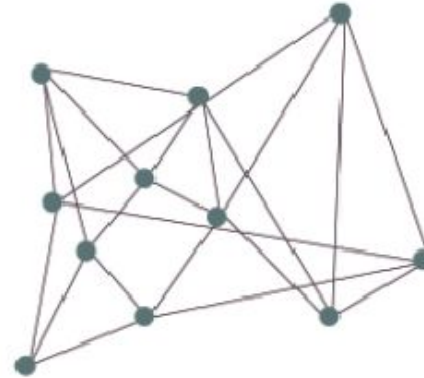
Output:

# Web3.0

## *Centralized vs Decentralized*



Unique Point of Failure



No unique Point of Failure > more secure

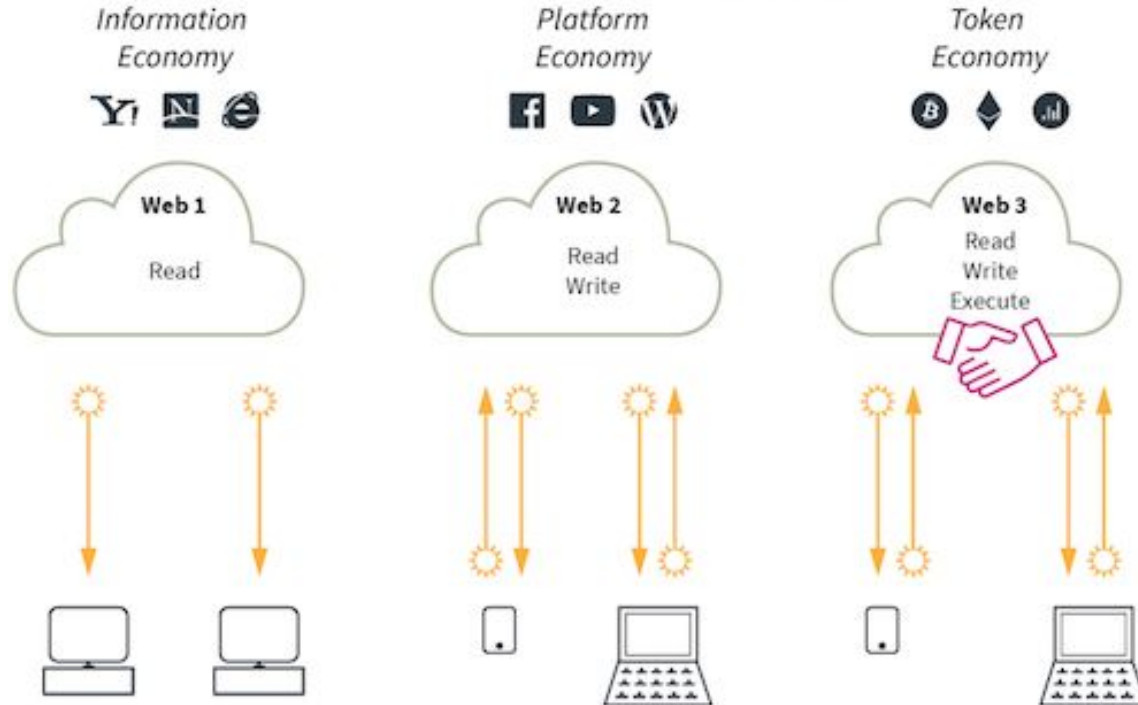
From the Book “**Token Economy**” by Shermin Voshmgir, 2019  
Excerpts available on <https://blockchainhub.net>



# Web3.0

## History of the Web

From the Book "**Token Economy**" by Shermin Voshmgir, 2019  
Excerpts available on <https://blockchainhub.net>



# Where to go Next

Refresher - <https://flatironschool.com/free-courses/data-science-bootcamp-prep>

Computer Vision - <https://cs231n.github.io/>

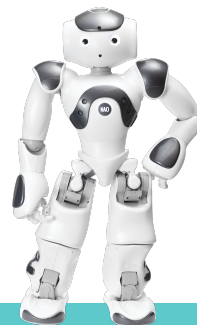
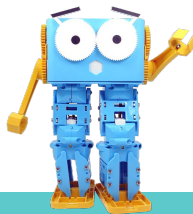
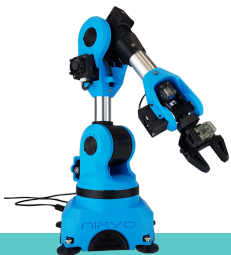
NLP - <http://web.stanford.edu/class/cs224n/>

Statistics - <https://www.coursera.org/learn/statistics-for-data-science-python>

Algorithms - <https://www.coursera.org/browse/computer-science/algorithms>

Reinforcement Learning - <https://www.coursera.org/courses?query=reinforcement%20learning>

General Computer Science - <https://www.edx.org/course/subject/computer-science/>



# Thank you, and I hope you enjoyed the course!

