Nitty-gritty of Data and Exploratory Analysis with Python 3

Dev Skill
Class 16

Jayanta Sarker Shuva

Machine Learning:

Q1 **1, 1, 2, 3, 5, ___?**

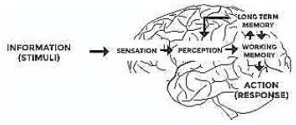
What is the number in the blank?

Answer1: 8, Fibonacci series

Q2

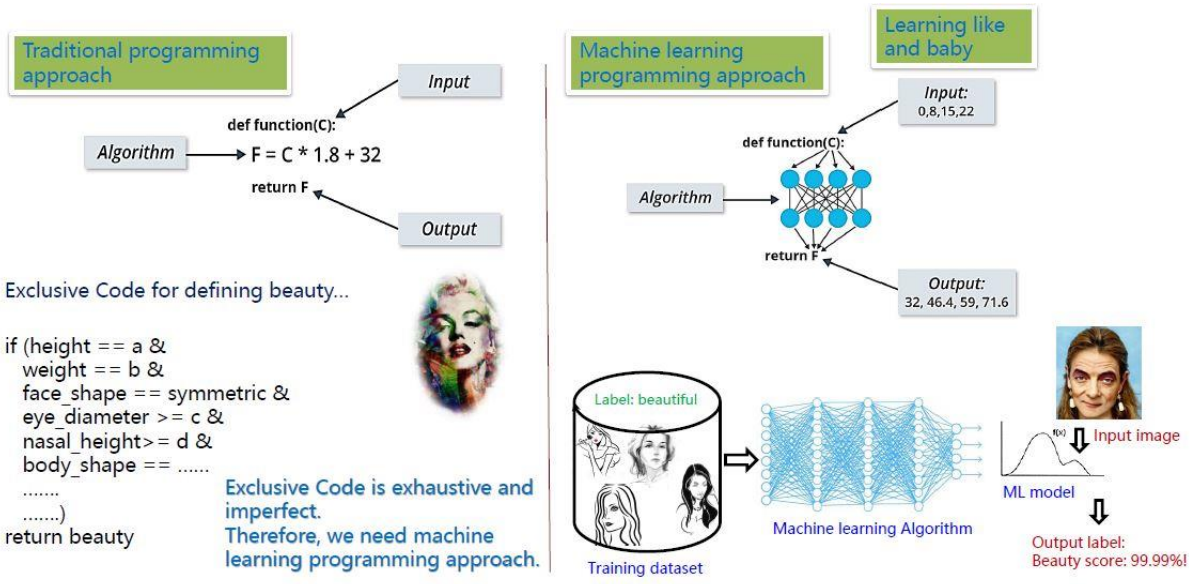
Input:	0,	8,	15,	22,	38
Output:	32,	46.4,	59,	71.6,	?

Question: What will be the Output value for an Input value of 38?

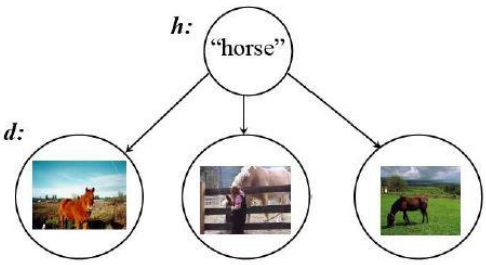


Answer2: 100.4,
Celsius to Fahrenheit; F = C * 1.8 + 32

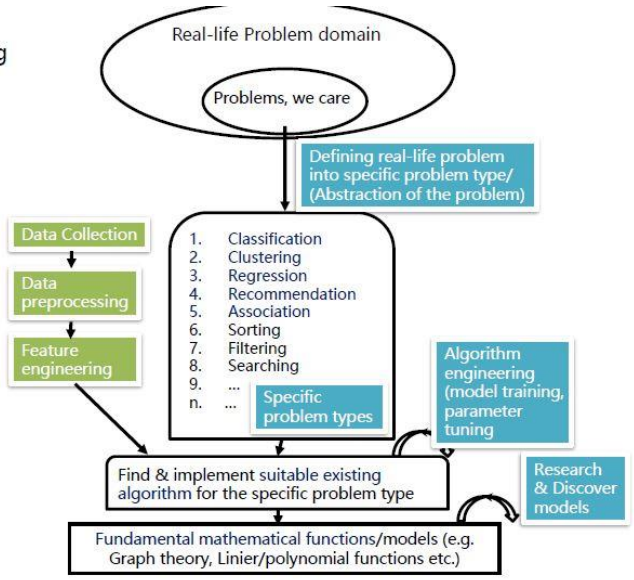
Think about how do we know the right distance, force, angle etc. to score a basket?



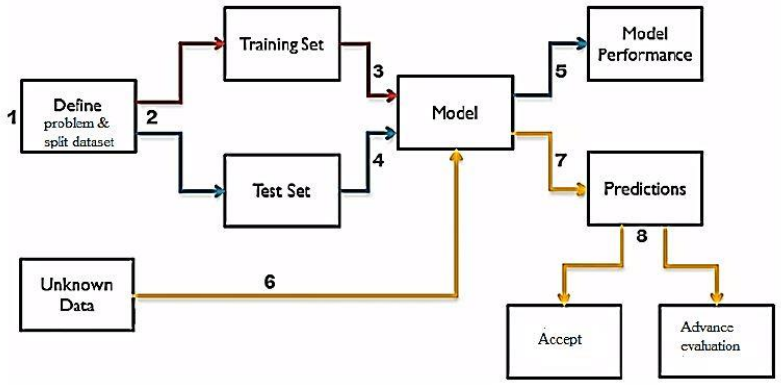
Machine learning studies the process of constructing abstractions (features, concepts, functions, relations and ways of acting) automatically from data.



How can a piece of software determine the above inputs are the images of horse

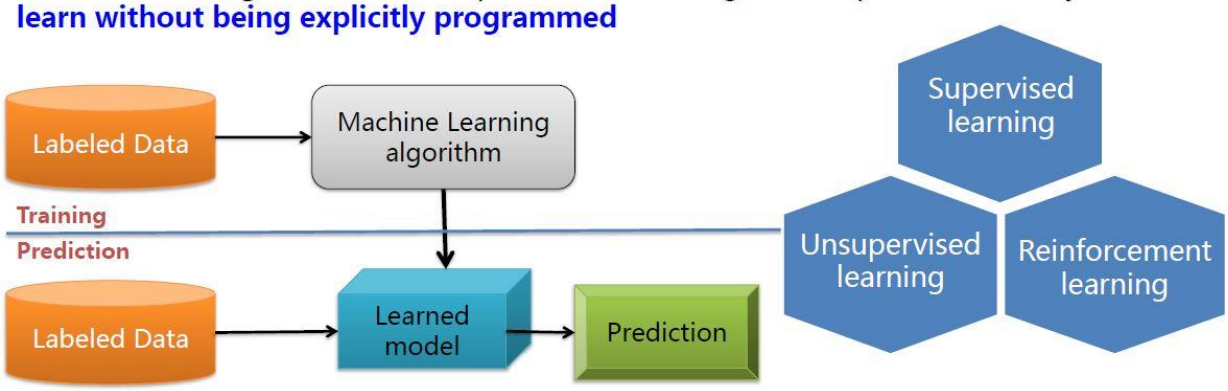


1.1 Machine Learning Concept (Procedure)



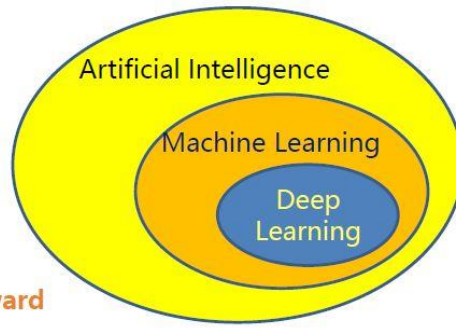
1) Problem abstraction
2) Data collection
3) Data Preprocessing
3.1 Extracting source data
3.2 Parsing data
3.3 Cleaning data
3.4 Extracting data
3.5 Labeling data
3.6 Generating training & test dataset
4) Feature Engineering
4.1 Finding single point outlier
4.2 Finding Quartile of the dataset
4.3 Finding steep drop
4.4 .....
4.5 New feature creation
5) Training model
4.1 Finding suitable algorithm
4.2 Train and tuning parameter
4.3 Output model
6) Model evaluation & Deployment

Machine learning is a field of computer science that gives computers the ability to **learn without being explicitly programmed**



Methods that can learn from and make predictions on data

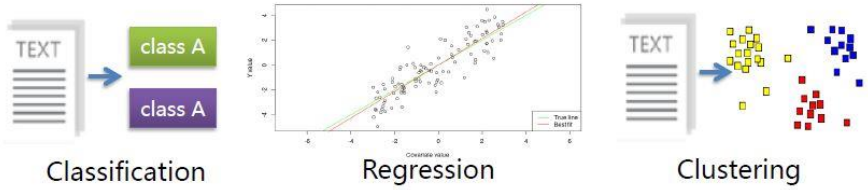
1.1 Machine learning concepts



Types of Learning:
Supervised: Learning with a **labeled training** set
Example: email *classification* with already labeled emails

Unsupervised: Discover **patterns** in **unlabeled** data
Example: *cluster* similar documents based on text

Reinforcement learning: learn to **act** based on **feedback/reward**
Example: learn to play Go, reward: *win or lose*



Classification Regression Clustering

Anomaly Detection, Sequence labeling,...

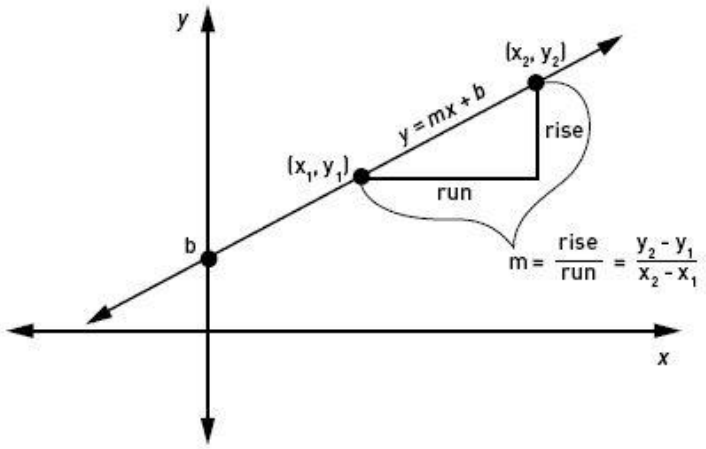
Linear Regression:

Golden equation for straight line is –

$$y = mx + b$$

This is the base equation for linear Regression. Let's do some codes to better understanding.

Finding Slope (m) and y-intercept (b):



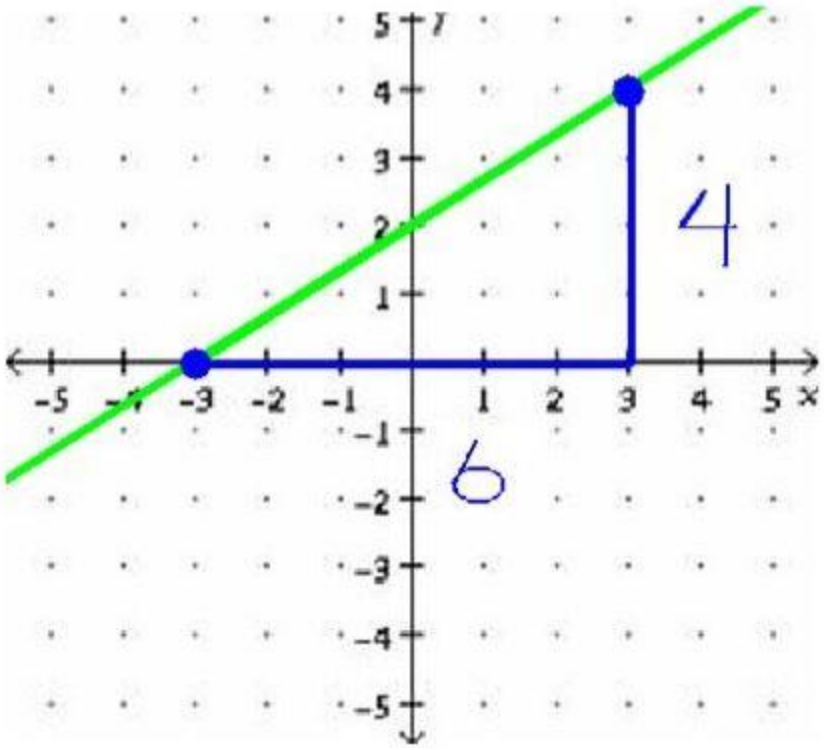
Let's consider equation - y = 2/3x + 2

slope= rise/run

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{y2-y1}{x2-x1}$$

Ex. (-3,0) (3,4)

$$\text{slope} = \frac{4 - 0}{3-(-3)} = \frac{2}{3}$$



Cost Function:

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

$$c = \bar{y} - m\bar{x}$$

Here ***m*** is slope and ***c*** is y-intercept or ***b.***