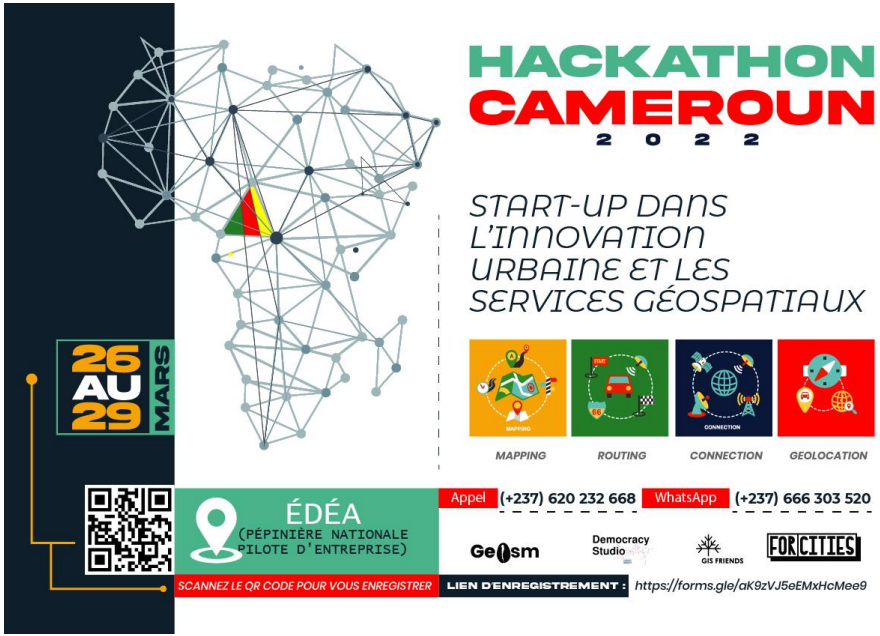


DEEP LEARNING INDABA

UNIVERSITÉ DE DSCHANG
UNIVERSITY OF DSCHANG

Machine Learning

Volviane MFOGO

Introduction to Machine Learning

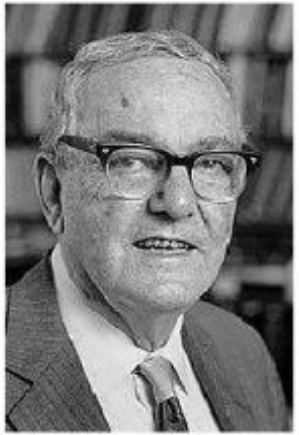
About Me

Volviane Saphir MFOGO
PhD. Candidate, Machine learning and Cyber-security
Game theory and Machine learning for Cyber Deception, Resilience and Agility (GMC-DRA) project
University of Dschang, Cameroon.
IndabaX community lead

Email: smfogo@aimsammi.org / volviane@deeplearningindaba.com

Machine Learning

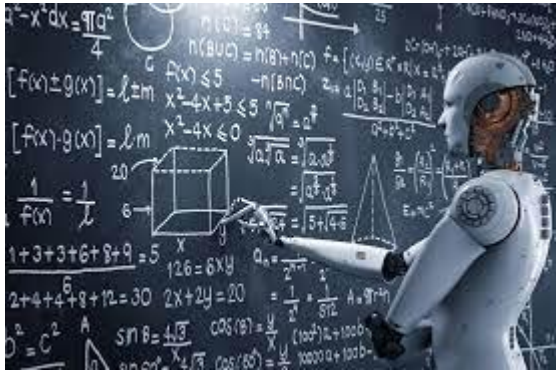
- **Herbert Alexander Simon**: "Learning is any process by which a system improves performance from experience."
- "Machine Learning is concerned with computer programs that automatically improve their performance through experience. "



Herbert Simon
Turing Award 1975
Nobel Prize in Economics 1978

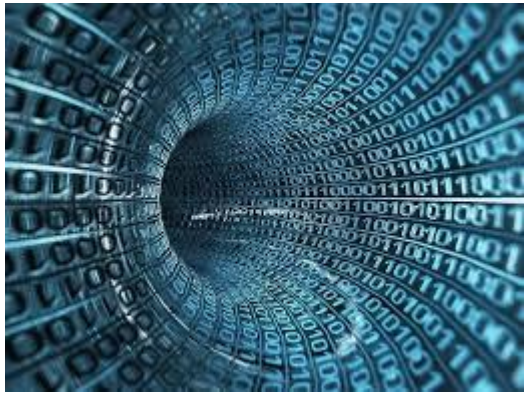
Lior Rokach's slide

Why Machine Learning?



- Ability to mimic human and replace certain monotonous tasks which require some intelligence;
  Like face recognition system

- Automate numerous tasks without human intervention
  Recommended system



- Discover new knowledge from large dataset
- Develop system that are too difficult/expensive to construct manually because the require specific detailed skill or knowledge

Machine learning is important because **it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products**. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations.

Why now?

- Flood of available data (especially with the advent of the Internet)
- Increasing computational power
- Growing progress in available algorithms and theory developed by researchers
- Increasing support from industries

Lior Rokach's slide

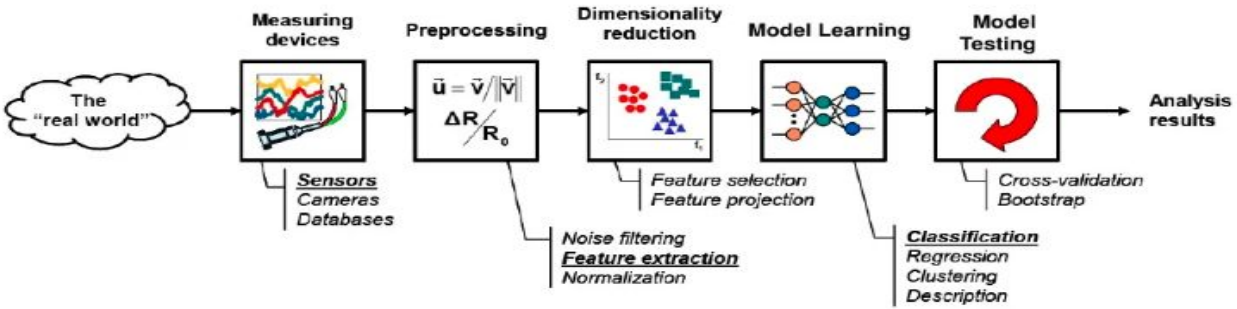
Machine Learning Application

Lior Rokach's slide

The concept of learning in a ML system

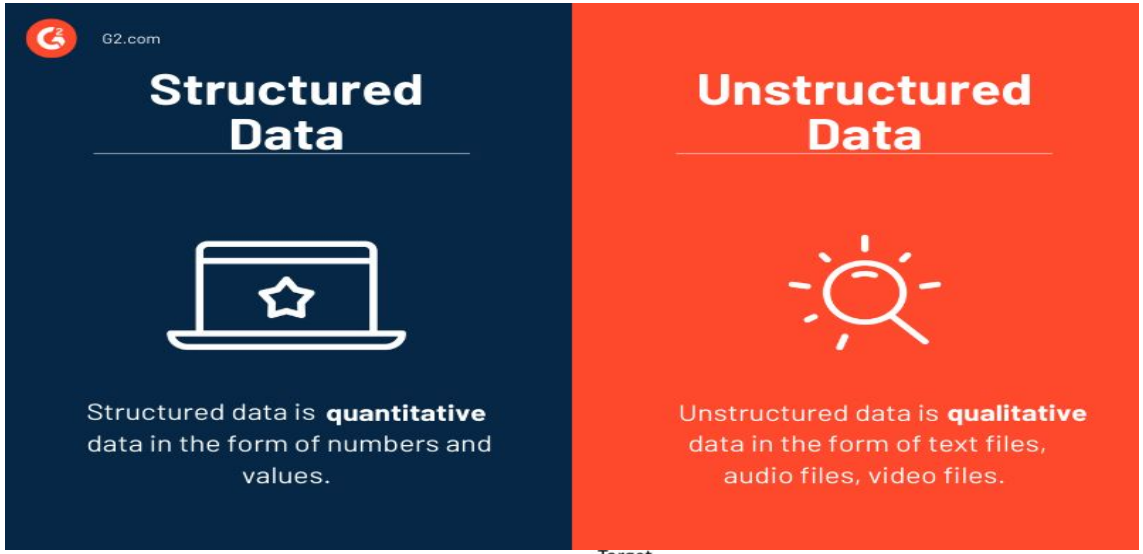
- Learning = Improving with experience at some task
  - Improve over task *T*,
  - With respect to performance measure, *P*
  - Based on experience, *E*.

The Learning Process



Lior Rokach's slide

Type of dataset



Input Attributes

Target Attribute

Number of new Recipients	Email Length (K)	Country (IP)	Customer Type	Email Type
0	2	Germany	Gold	Ham
1	4	Germany	Silver	Ham
5	2	Nigeria	Bronze	Spam
2	4	Russia	Bronze	Spam
3	4	Germany	Bronze	Ham
0	1	USA	Silver	Ham
4	2	USA	Silver	Spam

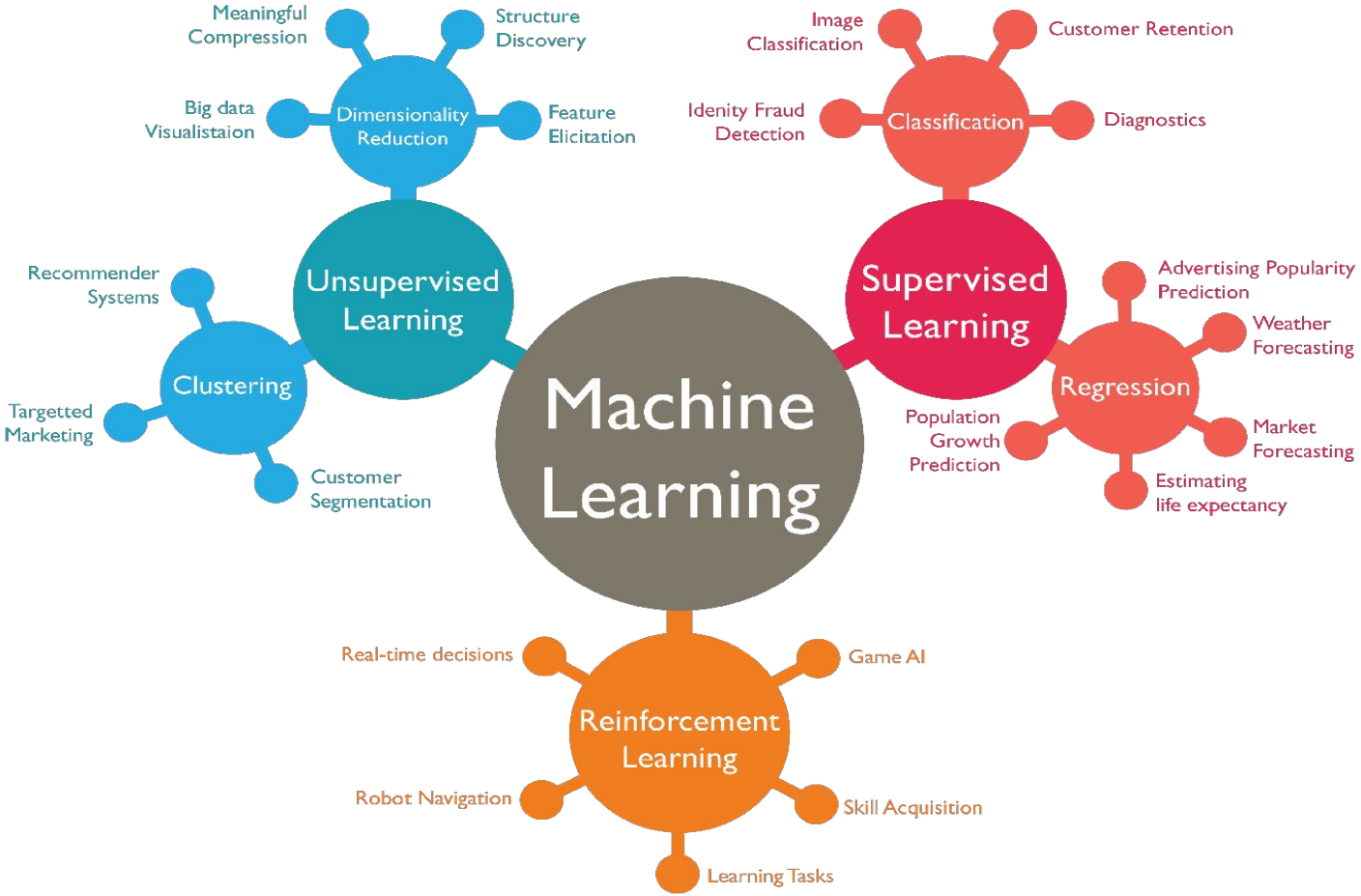
Instances

Numeric

Nominal

Ordinal

Type of Learning

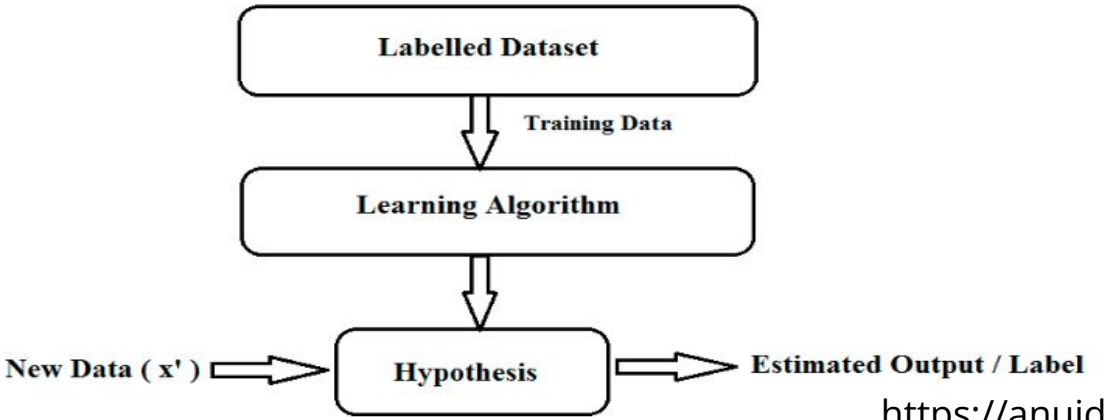


Supervised Learning

- The majority of practical machine learning uses supervised learning.
- Supervised learning is where you have input variables X and an output variable Y and you use an algorithm to learn the mapping function from the input to the output.

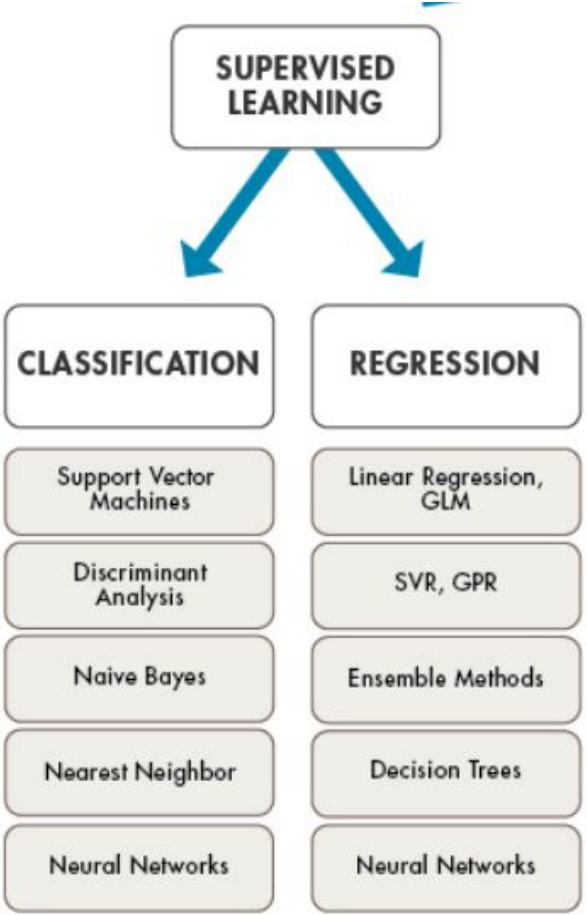
$$Y = f(X)$$

- The goal is to approximate the mapping function f so well that when you have new input data X that you can predict the output variables Y for that data.



https://anujdutt9.github.io/ML Basics.html

Supervised Learning Problem



Regression

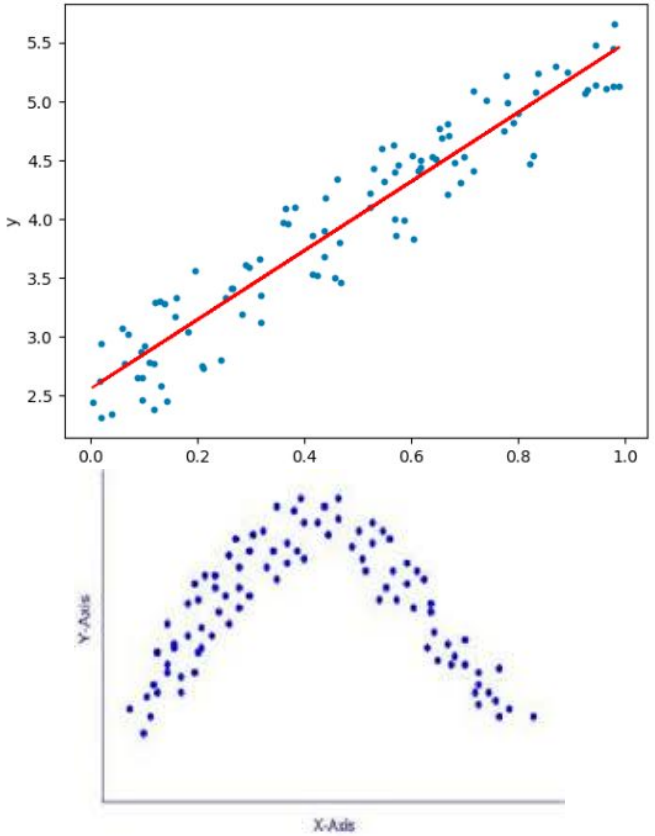
- The goal of regression tasks is to predict a continuous number, or a floating-point number in programming terms (or real number in mathematical terms).
- Predicting a price of fruit from their weight and their color is an example of a regression task

Classification

- In classification, the goal is to predict a class label, which is a choice from a predefined list of possibilities. (discret value in mathematical terms)
- Predicting a spam or non-spam email

Linear Regression

- Linear Regression is a machine learning algorithm based on supervised learning.
- It is mostly used for finding out the relationship between variables and forecasting. It performs the task to predict a dependent variable value Y based on a given independent variable X.
- This regression technique finds out a linear relationship between X(fea-
- tures) and Y(target). Hence, the name is Linear Regression.
- The regression line is the best fit line for the model.



https://towardsdatascience.com/linearregression-using-python-b136c91bf0a2

Linear Regression

- In regression, the relationship between Y and X is modelled in the following form:

  Y= **a***X + **b** + **E**

  Where:
  - Y is the dependent variable (a.k.a label or target)
  - X is the independent variable (a.k.a features)
  - **a** is the coefficient (a.k.a parameter)
  - b is an intercept
  - E is an error term for each observation (since there is additional variation not explained by the target)

- **Assumption**

The errors E are normally distributed.
This can be tested by plotting an histogram of the residuals of the regression and checking that they all have a bell shape. Alternatively, you could use the Shapiro-Wilk test for normality.

Linear Regression : Formally

Given a set of data point $D = \{(x_i, y_i)\}_{i=1}^{n}$ where x_i are the features and $y_i \in \mathbb{R}^d$ are the target corresponding to the features.

- The hypothesis function for linear regression also called model is :

$$h_\theta(X) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_n x_n = \sum_{i=0}^{n} \theta_i x_i \qquad (x_0 = 1)$$

- Criteria:

For the class of hypothesis above, the loss function is given by the ordinary means square error.

$$l(\theta) = \frac{1}{2} \sum_{i=1}^{n} (h_\theta(x_i) - y_i)^2 \qquad (2)$$

There is many approach to implement linear regression:

Numerical approach : This approach is based on **gradient descent**., Analytical approach(With closed form solution)

The criteria above can be derived using Maximum Likelihood Estimation (MLE)

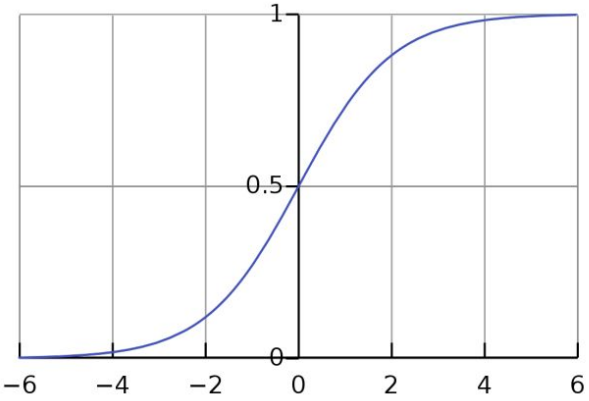
Logistic Regression

- Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.

$$D = \{(x_i, y_i)\} \qquad y_i \in \{0, 1\}$$

- We can call a Logistic Regression a Linear Regression model but the Logistic Regression uses a more complex cost function, this cost function can be defined as the 'Sigmoid function'.

$$h_\theta(X) = \frac{1}{1 + e^{-\theta^T X}} \qquad 0 \leq h_\theta \leq 1$$

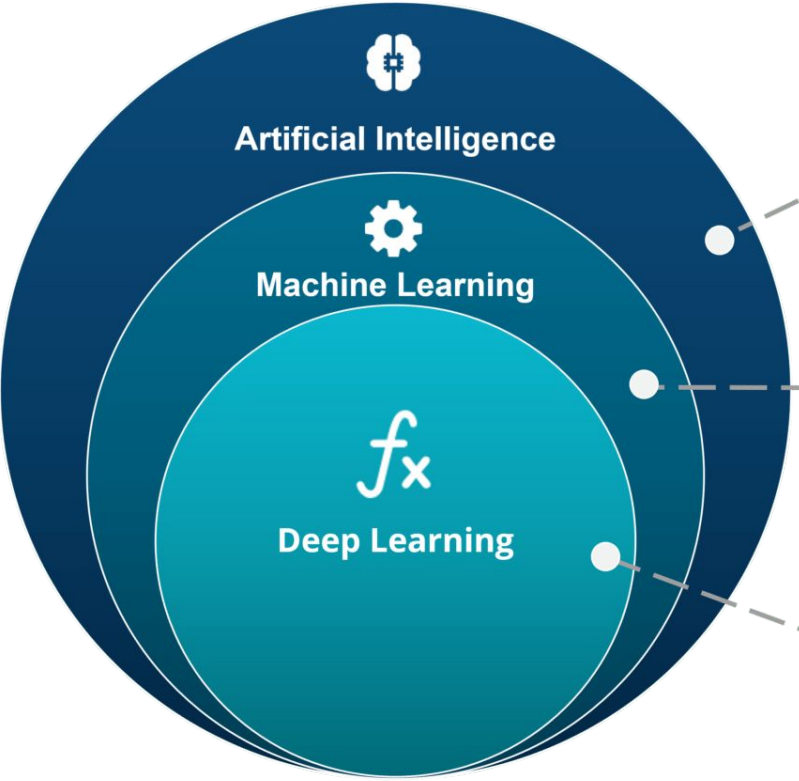


Criteria

Since the target is a discrete variable (binary), we can use a Bernoulli distribution/

$$\begin{cases} p(y = 1|X;\theta) &= h_\theta(X) \\ p(y = 0|X;\theta) &= 1 - h_\theta(X) \end{cases} \implies p(y|X;\theta) = h_\theta^y(X)\,(1 - h_\theta(X))^{1-y}$$

Anything more powerful than Machine learning?

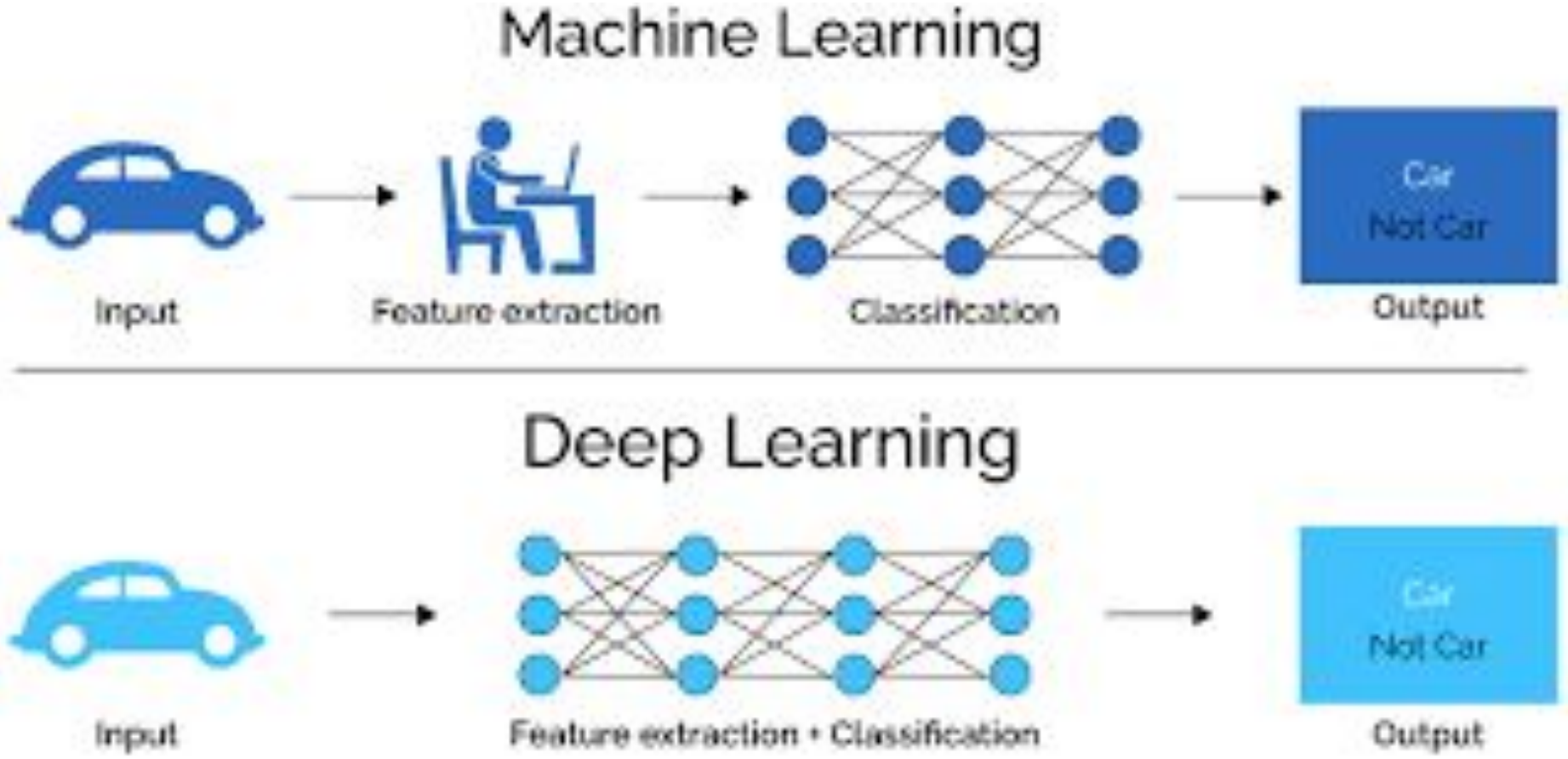


ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

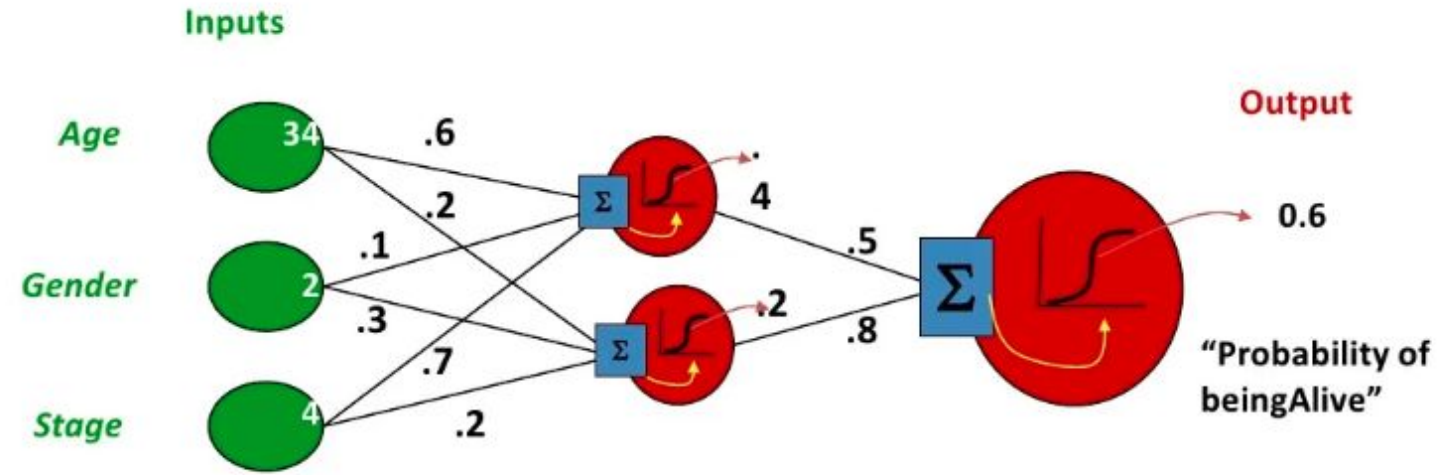
Deep learning vs. machine learning



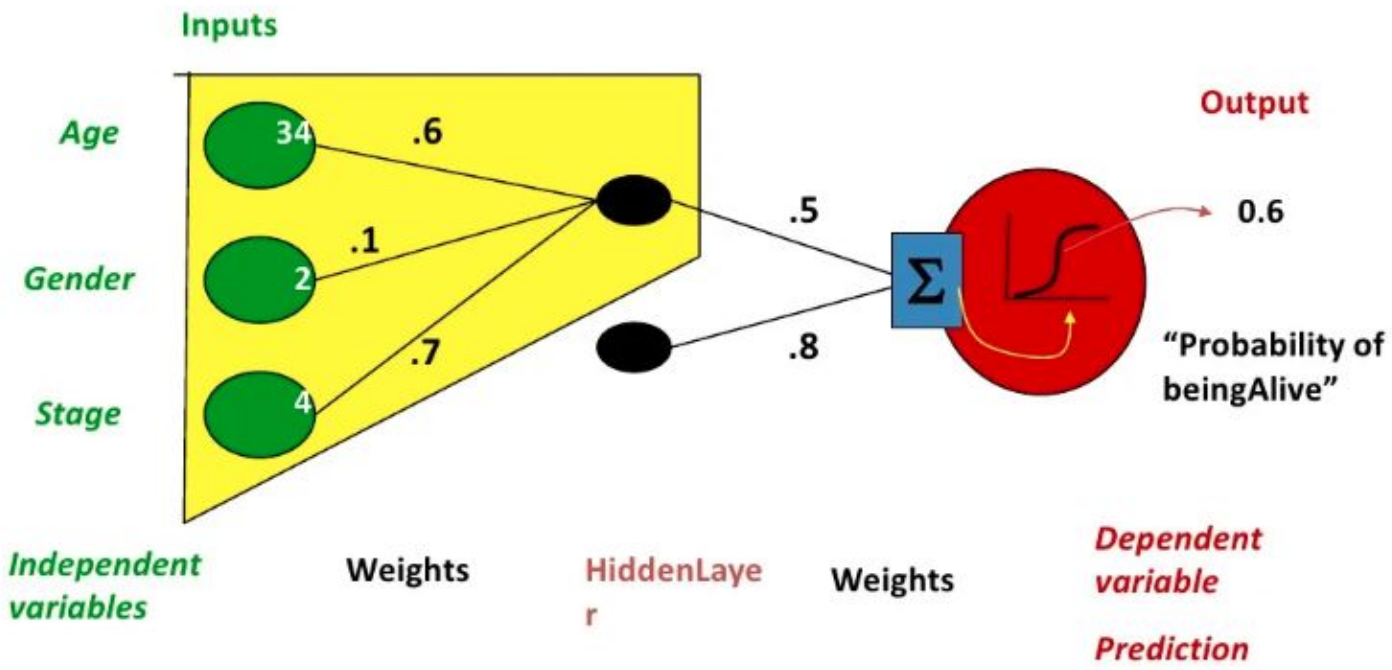
Deep learning vs. machine learning

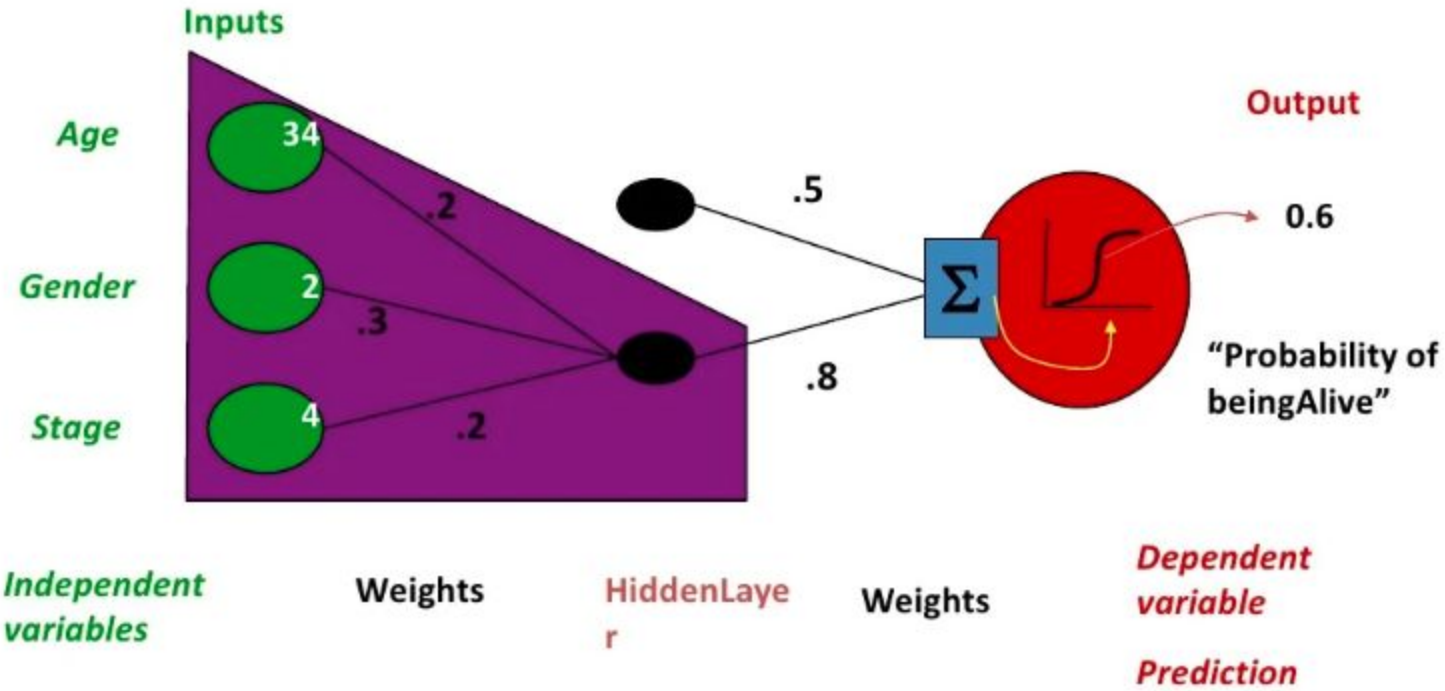
- Deep learning distinguishes itself from classical machine learning by the type of data that it works with and the methods in which it learns.
- Machine learning algorithms leverage structured, labeled data to make predictions. This doesn’t necessarily mean that it doesn’t use unstructured data; it just means that if it does, it generally goes through some pre-processing to organize it into a structured format.
- Deep learning eliminates some of data pre-processing that is typically involved with machine learning. These algorithms can ingest and process unstructured data, like text and images, and it automates feature extraction, removing some of the dependency on human experts.
- For example, let’s say that we had a set of photos of different pets, and we wanted to categorize by “cat”, “dog”, “hamster”, et cetera. Deep learning algorithms can determine which features (e.g. ears) are most important to distinguish each animal from another. In machine learning, this hierarchy of features is established manually by a human expert.

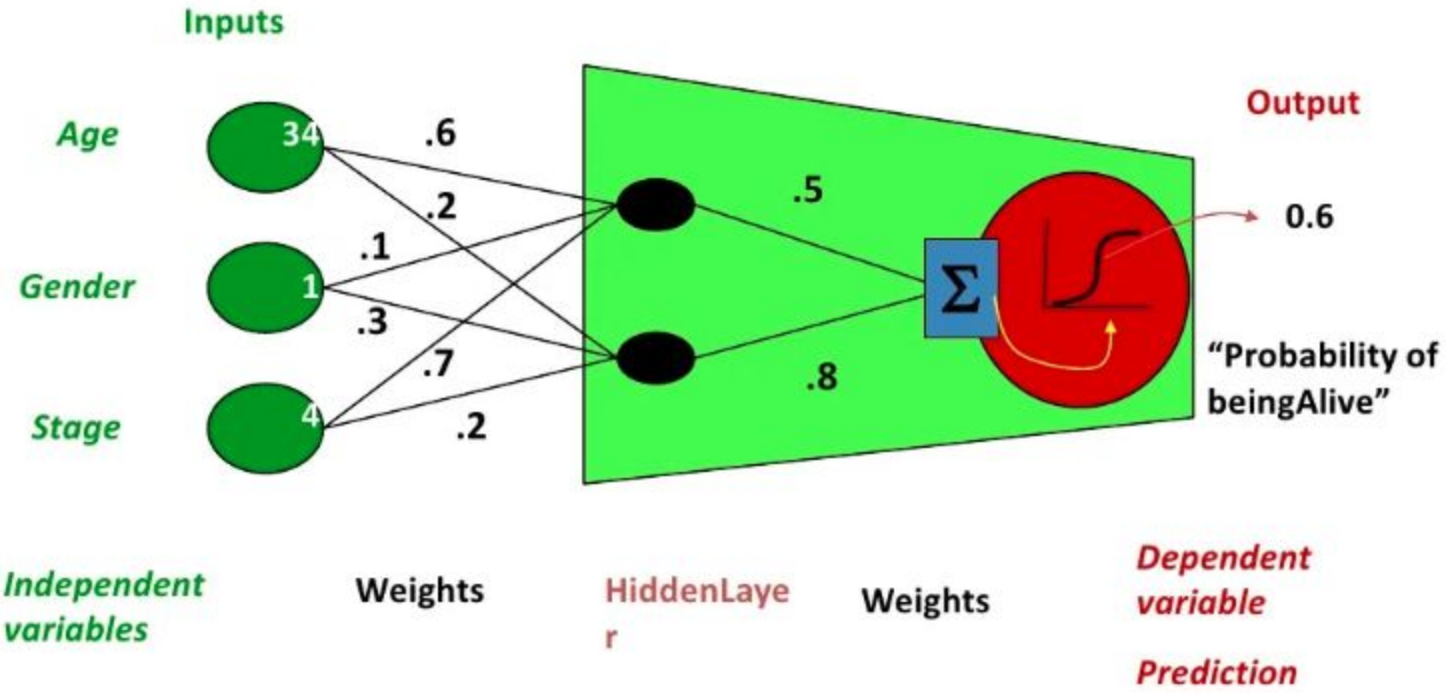
Neural Network Model

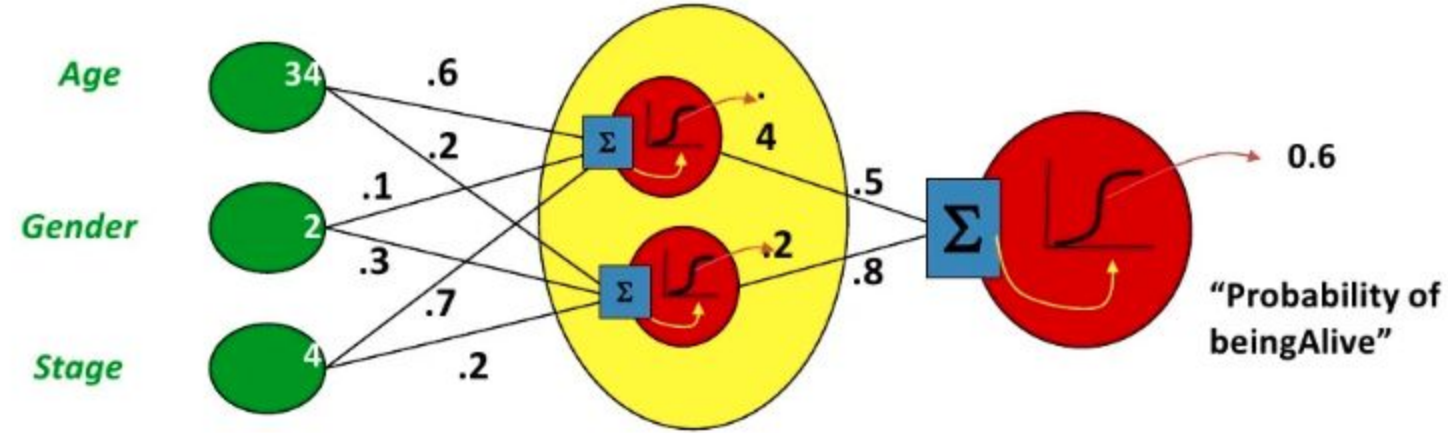


“Combined logistic models”




Inputs
Age
Gender
Stage
34
2
4
.2
.3
.2
.5
.8
Σ
Output
0.6
"Probability of beingAlive"
Independent variables
Weights
HiddenLayer
Weights
Dependent variable
Prediction


Inputs
Age
34
Gender
1
Stage
4
.6
.2
.1
.3
.7
.2
.5
.8
Σ
Output
0.6
"Probability of beingAlive"
Independent variables
Weights
HiddenLayer
Weights
Dependent variable
Prediction



Independent variables

Weights

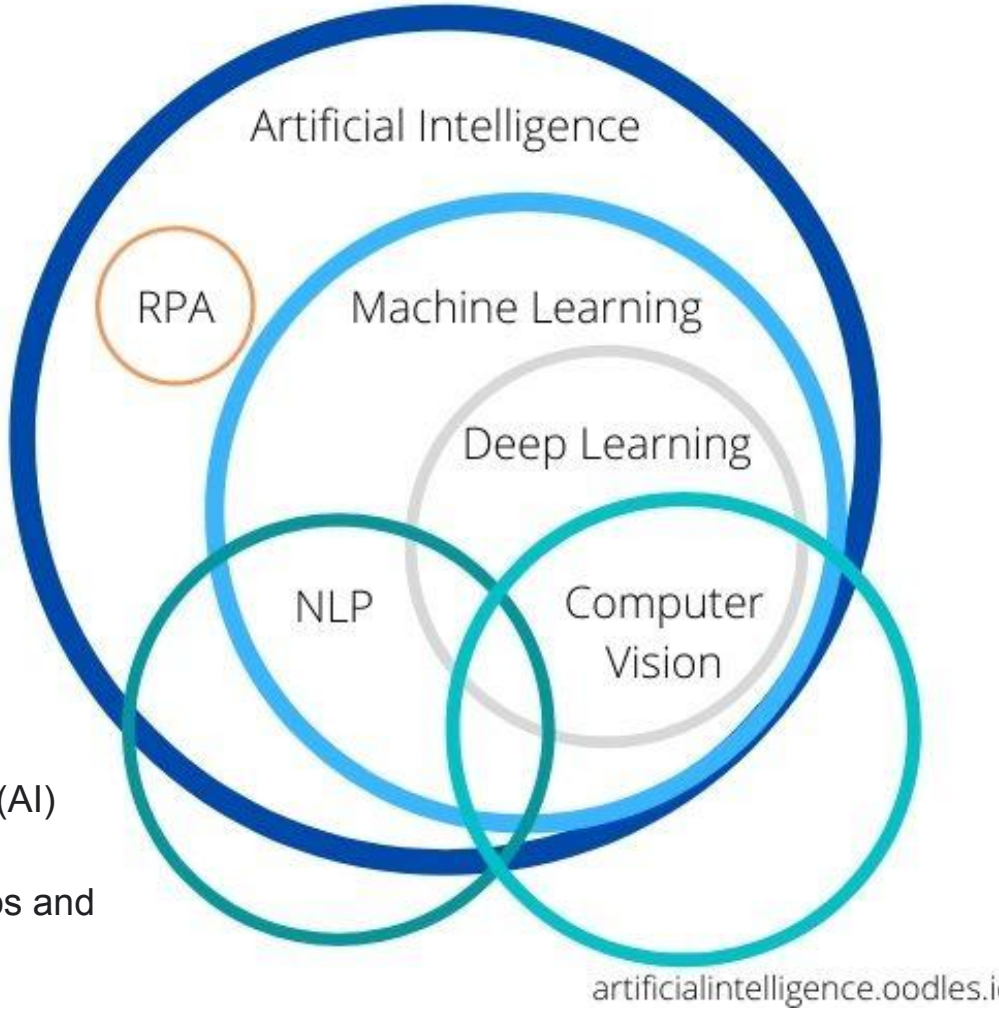
HiddenLayer

Weights

Dependent variable

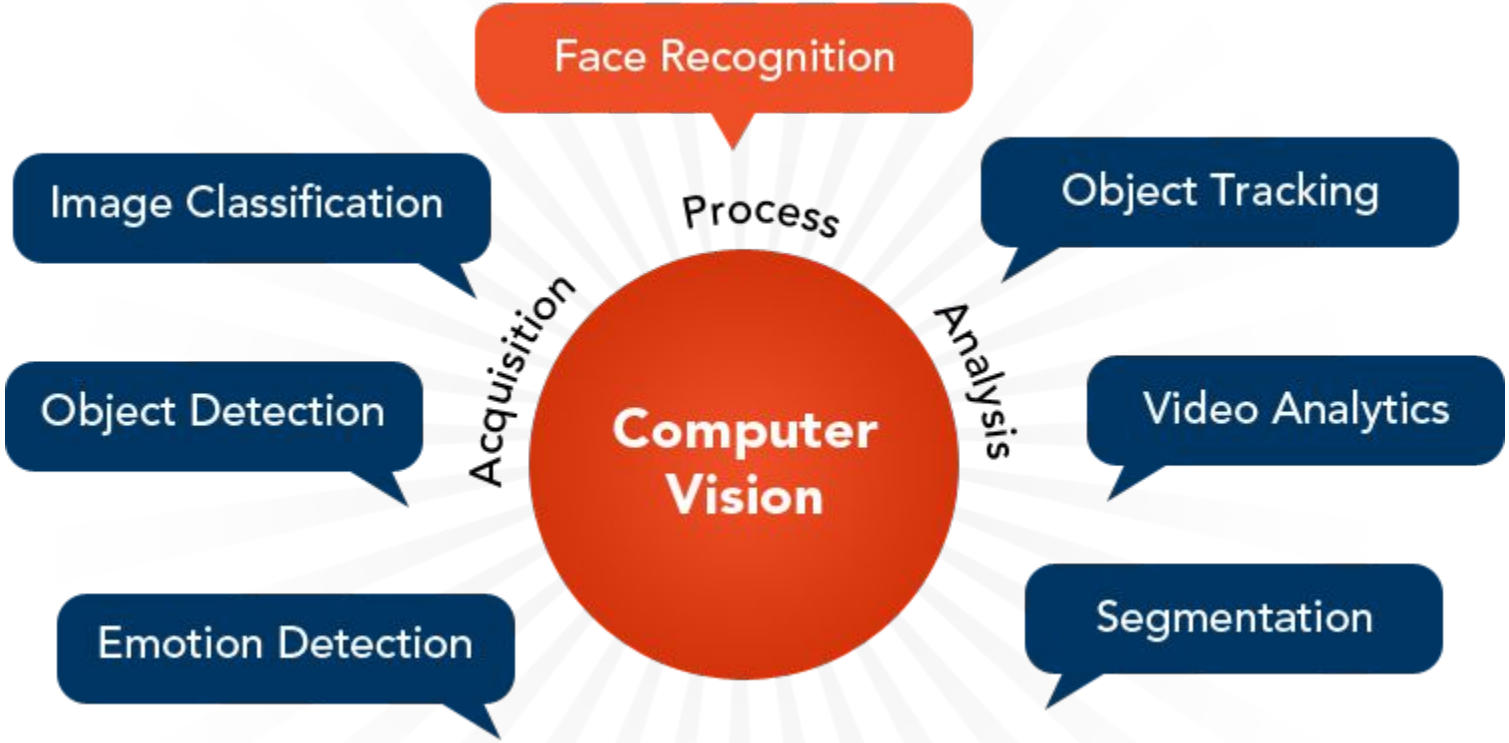
Prediction

Computer Vision

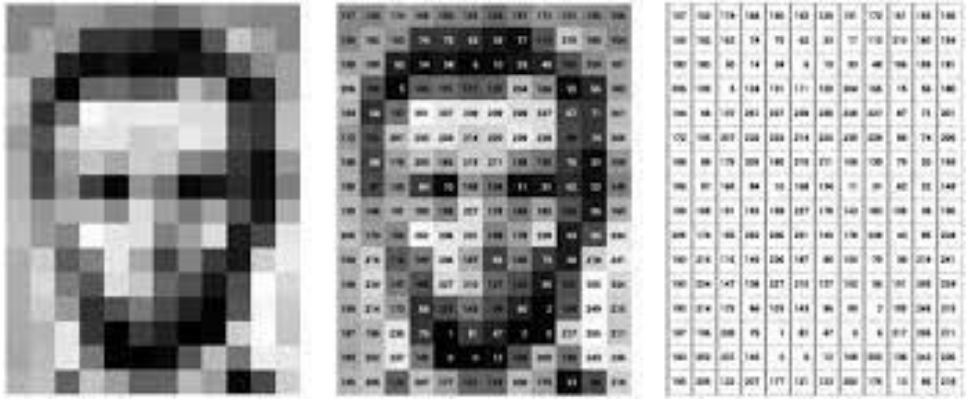


- Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs .

Computer Vision

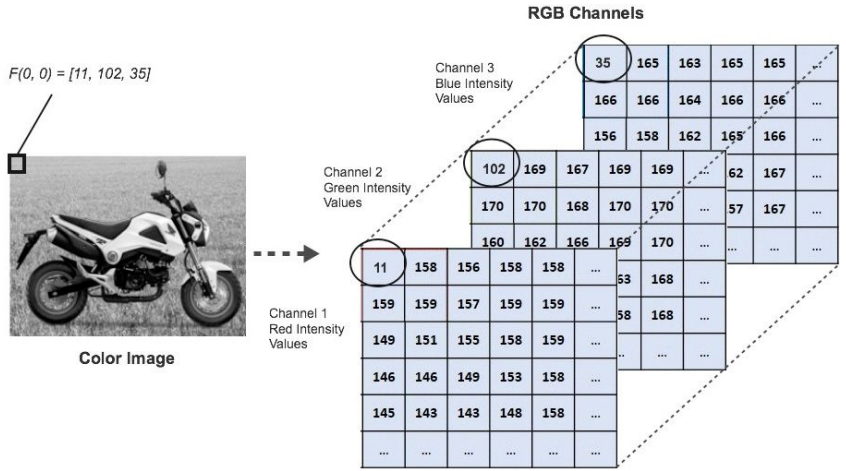


Image



- An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows.

- A normal greyscale image has 8 bit colour depth = 256 greyscales. A “true colour” image has 24 bit colour depth = 8 x 8 x 8 bits = 256 x 256 x 256 colours = ~16 million colours.



Cassava Disease Classification

Classify pictures of cassava leaves into 1 of 4 disease categories or healthy

[iCassava 2019 Fine-Grained Visual Categorization Challenge]

- Dataset
- Approach
- Results and Experiment

Dataset

Data acquisition Together with the National Crops Resources Research Institute (NaCRRI) in Uganda developed and deployed a crowdsourcing system where small-holder farmers in disparate places in Uganda were given smartphones with an application used to collect images of the crops in the farmers' fields. (Approximately 200 farmers sent images of plants from their gardens over the course of 1 year.)

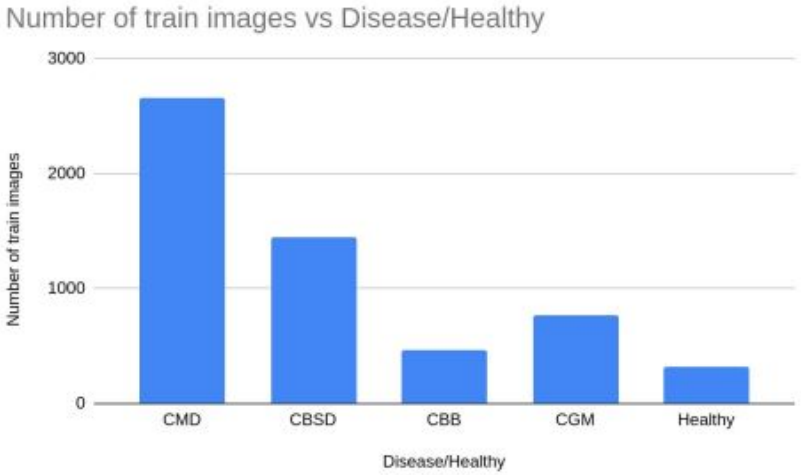
We build a machine learning model able to classify between four of the most common cassava diseases: Cassava Brown Streak Disease (CBSD), Cassava Mosaic Disease (CMD), Cassava Bacterial Blight (CBB) and Cassava Green Mite (CGM).

(a) Healthy (b) CBB (c) CGM (d) CMD (e) CBSD

Figure 1: Prototypical images associated with the five classes in our dataset covering healthy cassava leaves as well as 4 common diseases.

Dataset

Dataset consists of **9,436** labeled and **12,595** unlabeled images of cassava plant leaves.



Statistics of the Cassava Dataset. Paper

The graph shows training examples for CMD is twice more than every of the other classes. This creates a class imbalance.

Dataset

Data Preprocessing

- We split train data into 80% train and 20% validation using 5-fold stratified k-fold cross validation. The stratification was to cater for the imbalance in the dataset.
- Images were resized to 224px for small models and 500px for larger models.
- Different transforms such as centre cropping, rotation/affine transformation, horizontal and vertical flips.

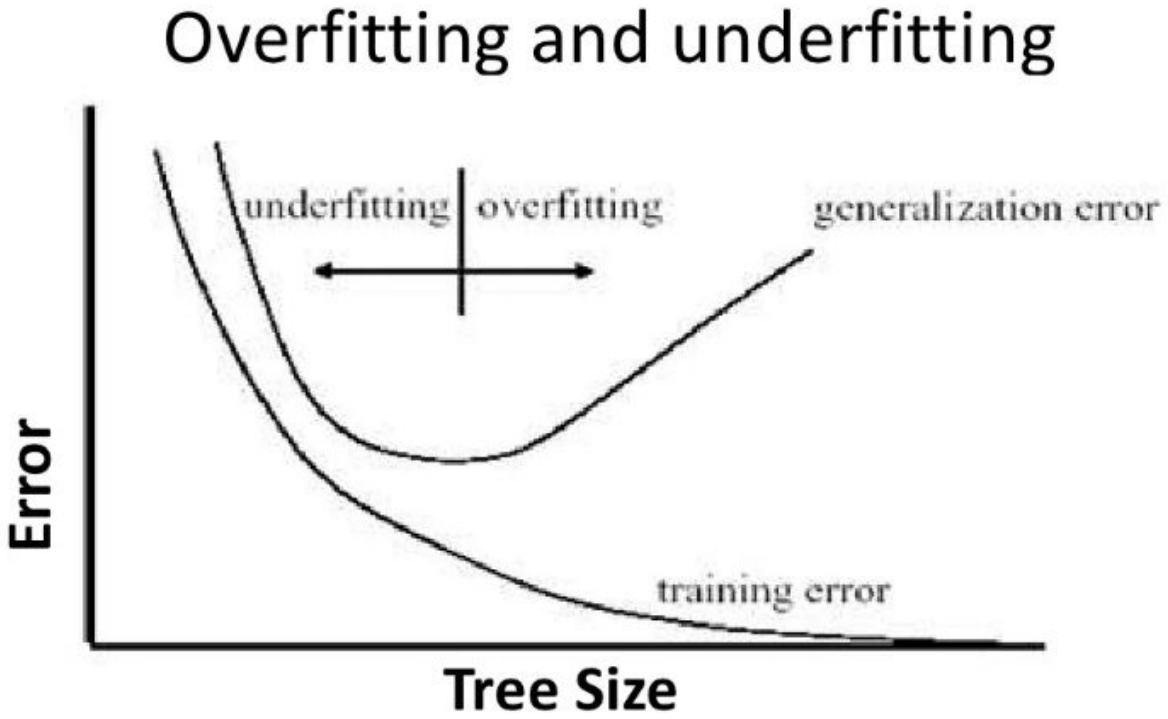
Experiments and results

No	Model	Train acc	Valid acc	Public LB	Parameters						
					optimizer	Batch-size	Learning rate	Epochs	Drop-out	Folds	Extra-images
1	Resnet18	0.87	0.864	0.856	SGD	16	0.0002	22	No	5	No
3	Resnet34	0.8723	0.8736	0.8735	Adam	16	0.0002	13	No	2	No
4	Resnet50	0.8992	0.886	0.8741	SGD	16	0.001	9	No	5	No
5	Resnet50	0.9072	0.8993	0.90794	SGD	16	0.001	20	0.5	5	No
6	Resnext101	0.9266	0.906	0.912	SGD	16	0.0002	3	0.5	2	No
7	Resnext50	0.9651	0.909	**0.91655**	Adam	8	0.001	10	No	5	No
8	Ensemble of 5, 6, 7			**0.92185**							
9	Resnet50	0.9172	0.9108	0.910594	SGD	16	0.001	20	0.5	5	Yes
10	Ensemble of 5, 6, 9			**0.92384**							yes

Take away

- Cross validation is important to understand how the model generalizes, especially for small datasets like this one.
- Dropout can help to prevent overfitting, especially in models with large capacity such as ResNet50.

Be careful!



Overtraining: means that it learns the training set too well – it overfits to the training set such that it performs poorly on the test set.

Underfitting: when model is too simple, both training and test errors are large

Want to learn more?

- El Naqa, Issam, and Martin J. Murphy. "What is machine learning?." *machine learning in radiation oncology*. Springer, Cham, 2015. 3-11.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
- Mwebaze, Ernest, et al. "iCassava 2019 fine-grained visual categorization challenge." *arXiv preprint arXiv:1908.02900* (2019).


Thank
You
Maake
Dhanyavadagalu
Biyan
Mauruuru
Dank Je
Juspaxar
Grazas
Arigato
Gracias
Danke
Merci
Kia Ora
Matondo
Grazie
Muchchakkeram
Gratias Tibi
Ohrigado
ありがとう
Nirringrazzjak
Suksama
Dziekuje
Asante
Kiitos
Dankon
Raibh Maith Agat
Djiere Dieuf
Welalin
Di Ou Mèsi
谢
谢