Artificial Intelligence (in Pharmacology)

Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

Dr Varun Ojha

Lecturer, Computer Science

v.k.ojha@reading.ac.uk

University of Reading

Workshop Questions?

Artificial Intelligence

1. What is it?
2. How is it done?
3. What has revolutionised it?
4. How to know it is working well?
5. Where is it in Pharmacology?

PART 1

Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

What is it?

12:25 PM

Artificial Intelligence (AI)

to create intelligent machine that think (react) and act (work) like human beings

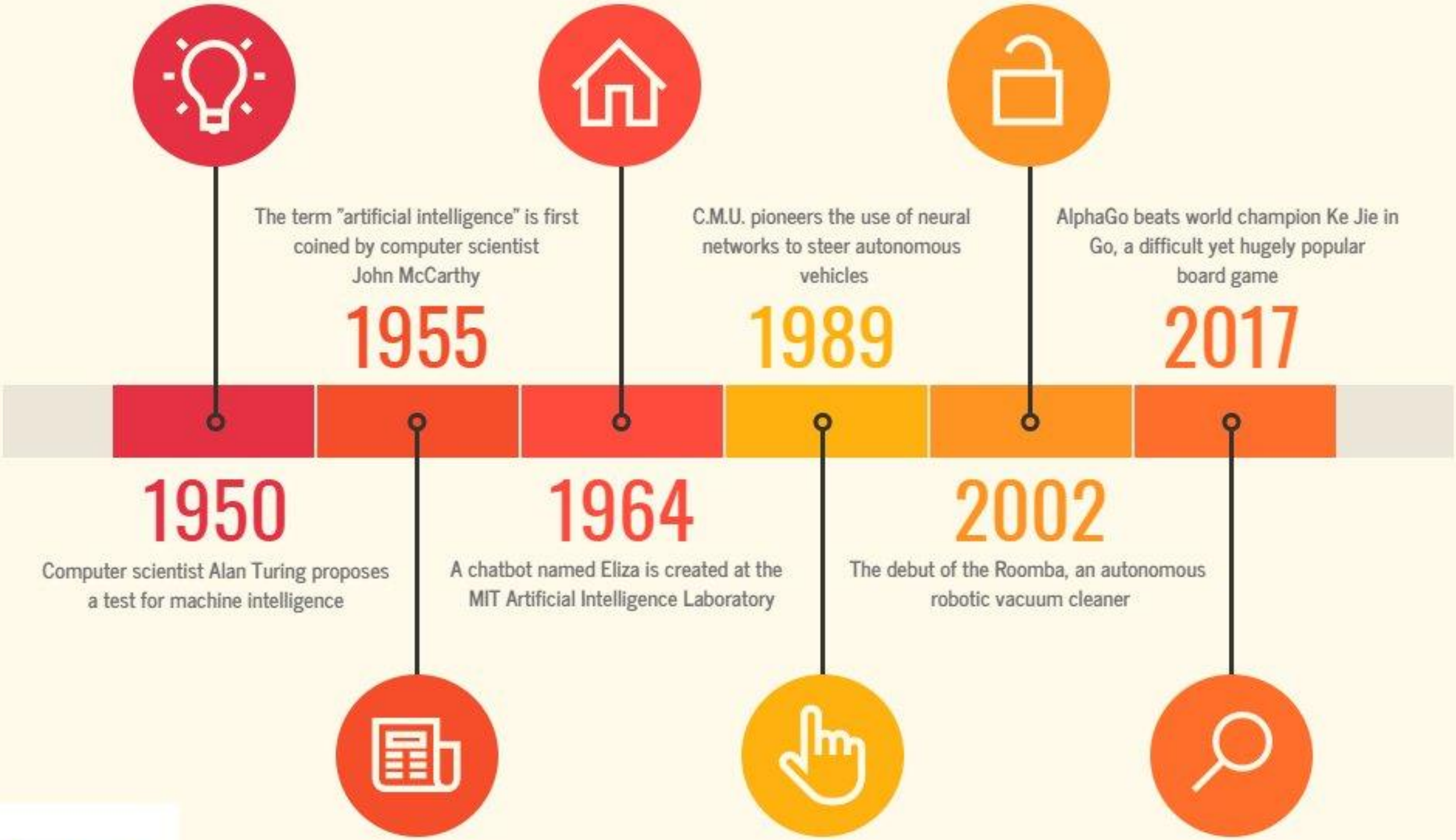
Alan Turing

23 June 1912 – 7 June 1954

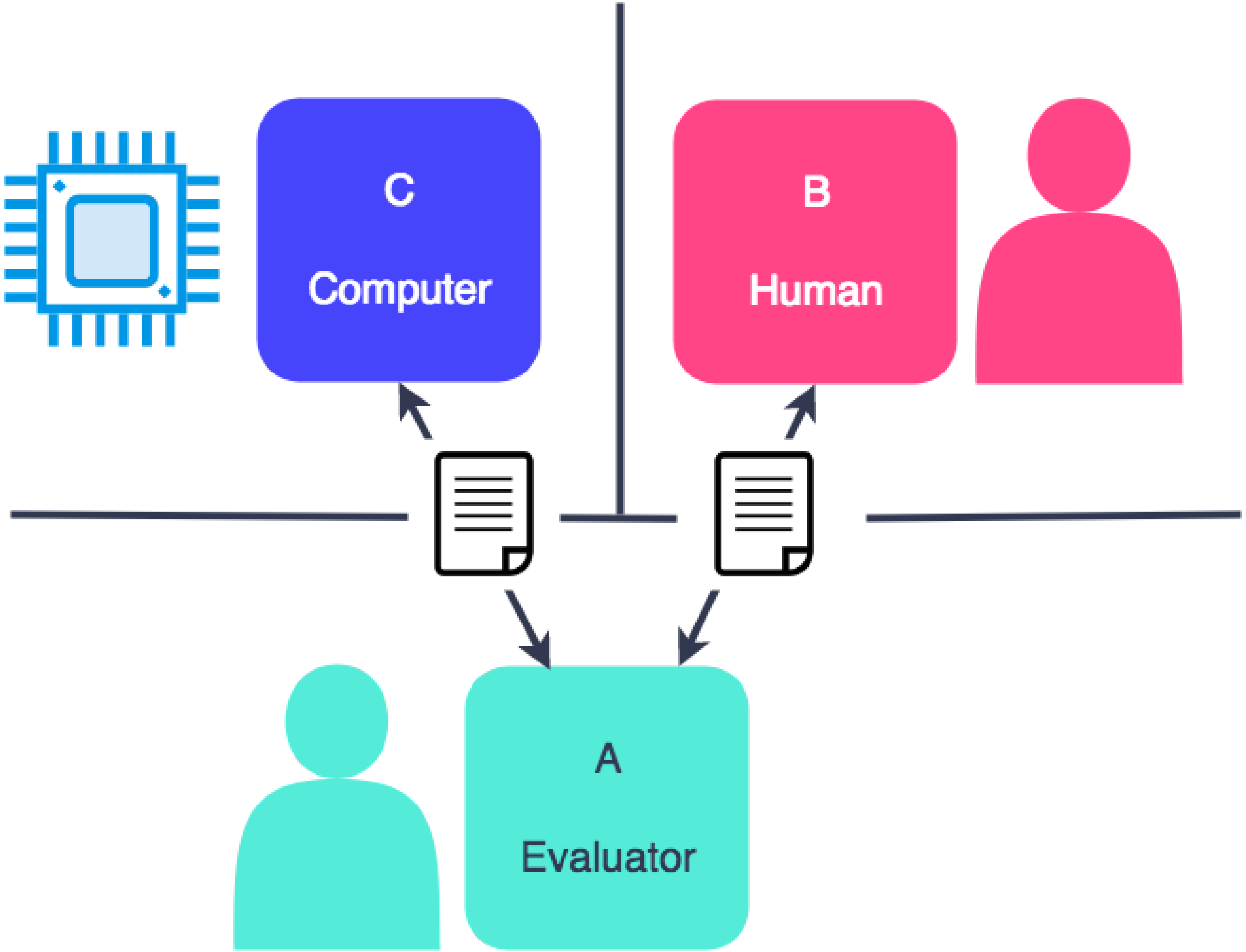
Consider as the Father of Computer Science and Artificial Intelligence

7 Decades of Artificial Intelligence History

Artificial Intelligence, or AI, is revolutionizing industries. Business executives, managers, and analysts worldwide see it as strategic priority in an ever shifting Information Age. Here is a timeline of six of AI's big moments.



VTB Capital



The **Turing test**

AI need to have:

natural language processing

knowledge representation

automated reasoning

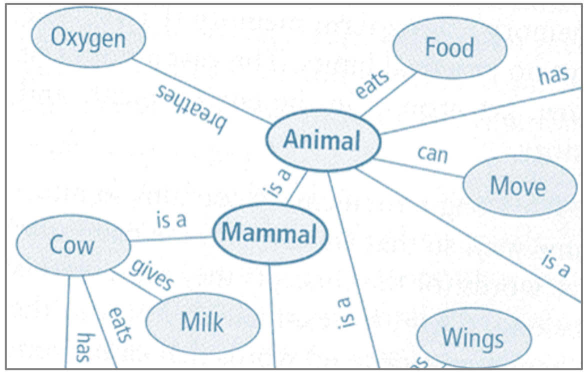
machine learning

computer vision

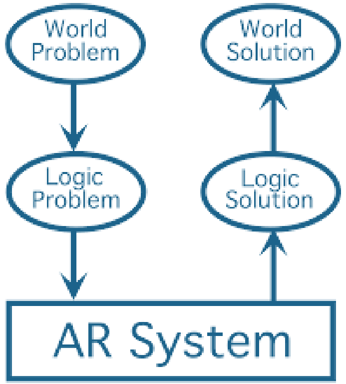
robotics

logic

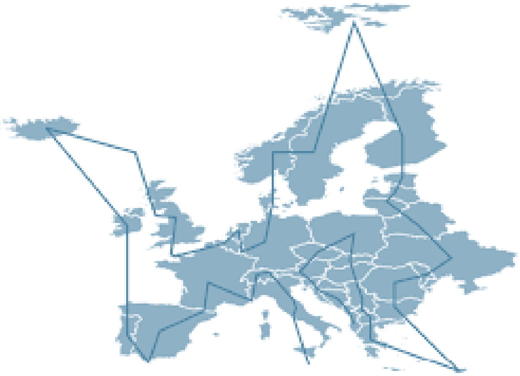
What Does AI Really Do?



Knowledge Representation

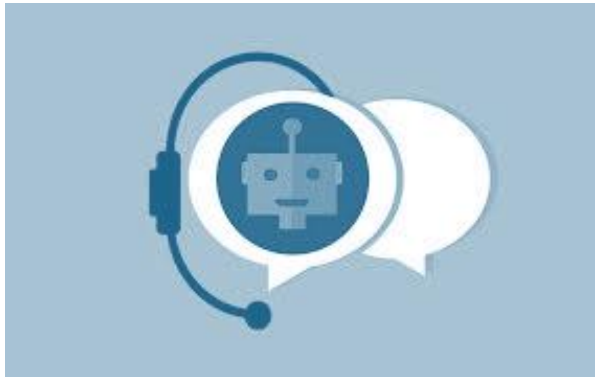


Automated reasoning

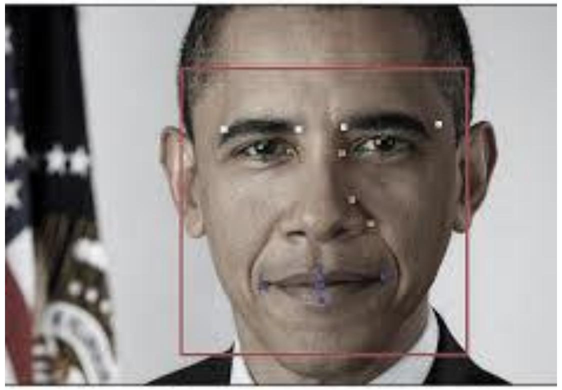


Planning

Machine Learning



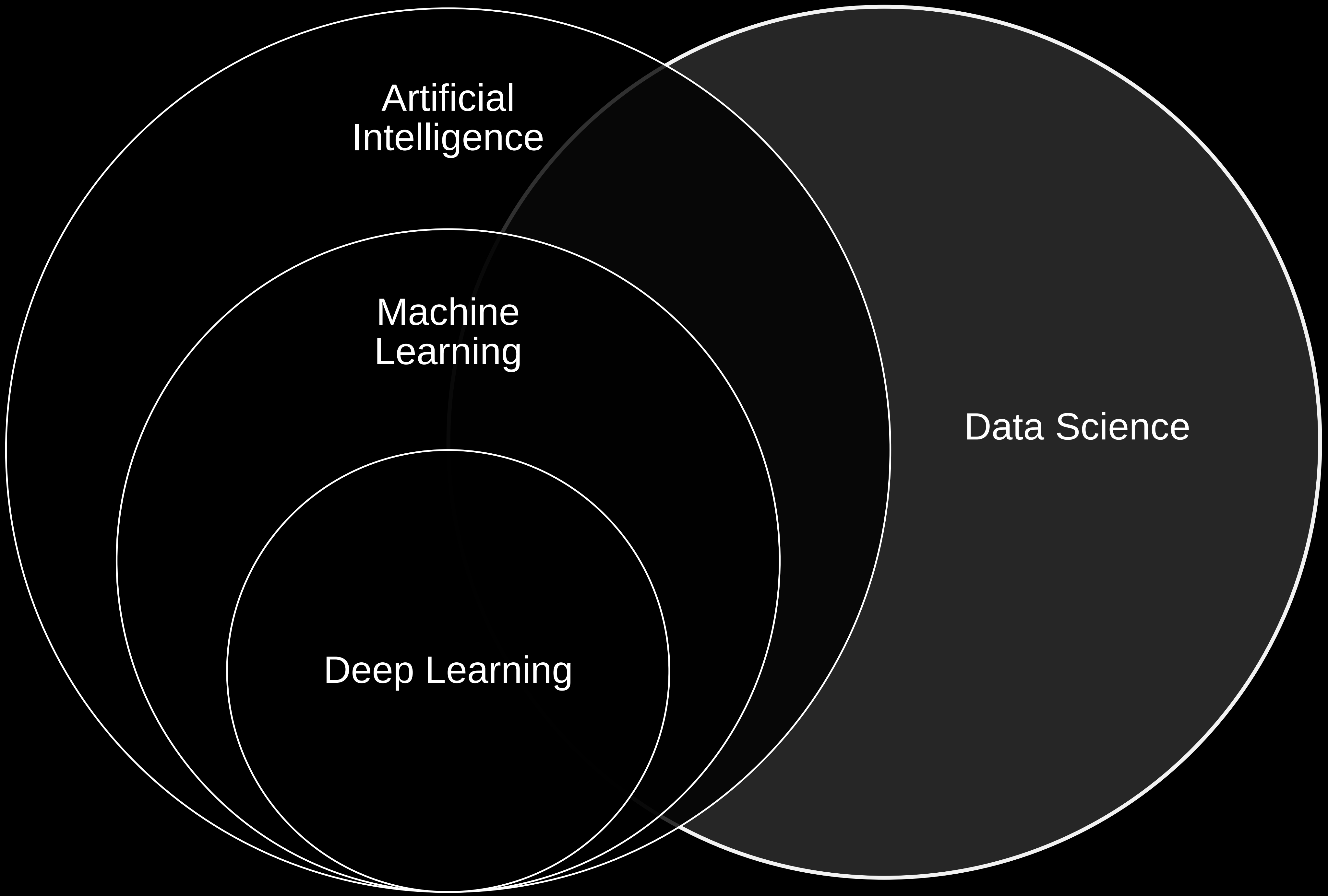
Natural language understanding



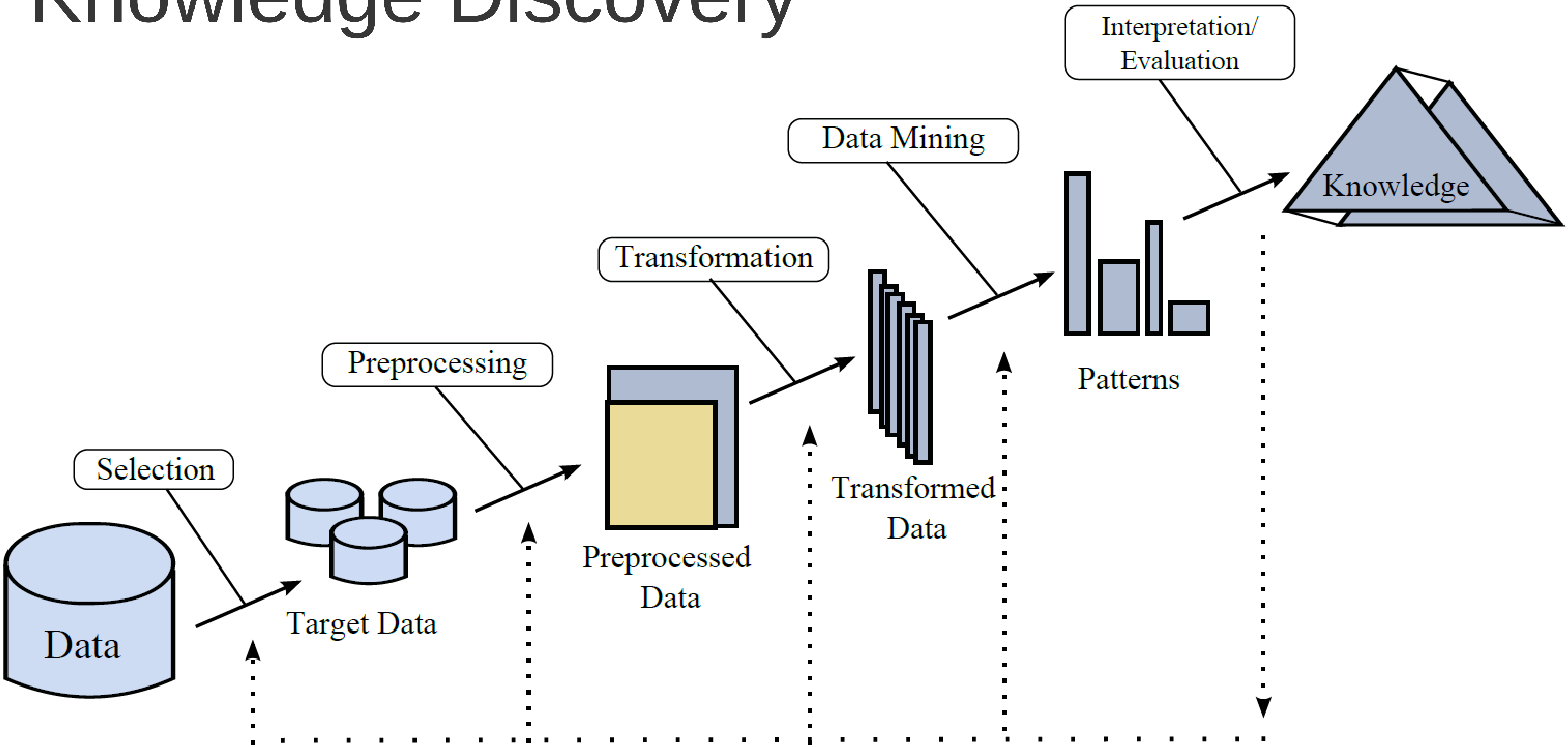
Machine vision

Robotics

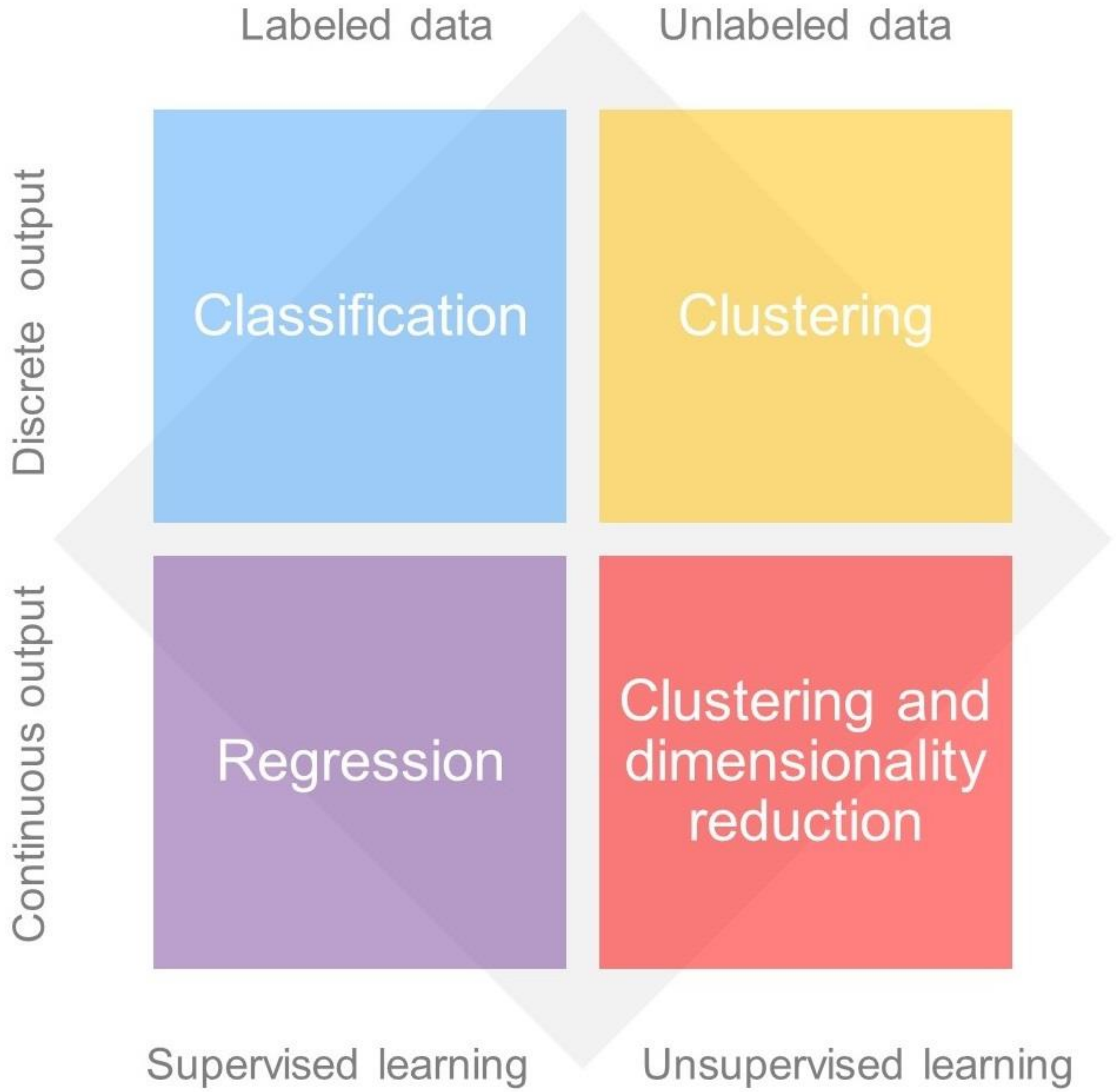
Web Search


Artificial
Intelligence
Machine
Learning
Deep Learning
Data Science

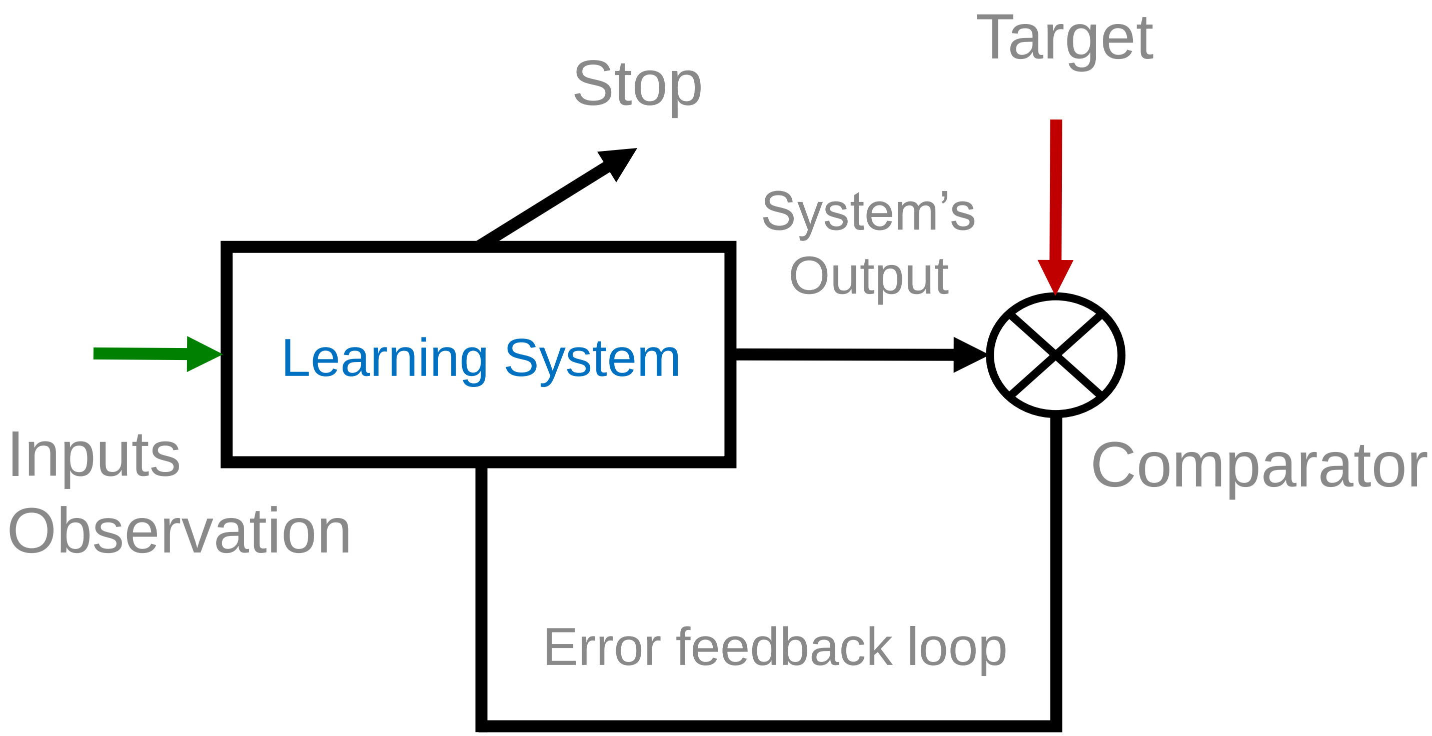
Knowledge Discovery



Paradigms of extracting Knowledge from data



Learning



Learning Cycle:

Input (Continuous input)

Target (Known output)

Output (System's output)

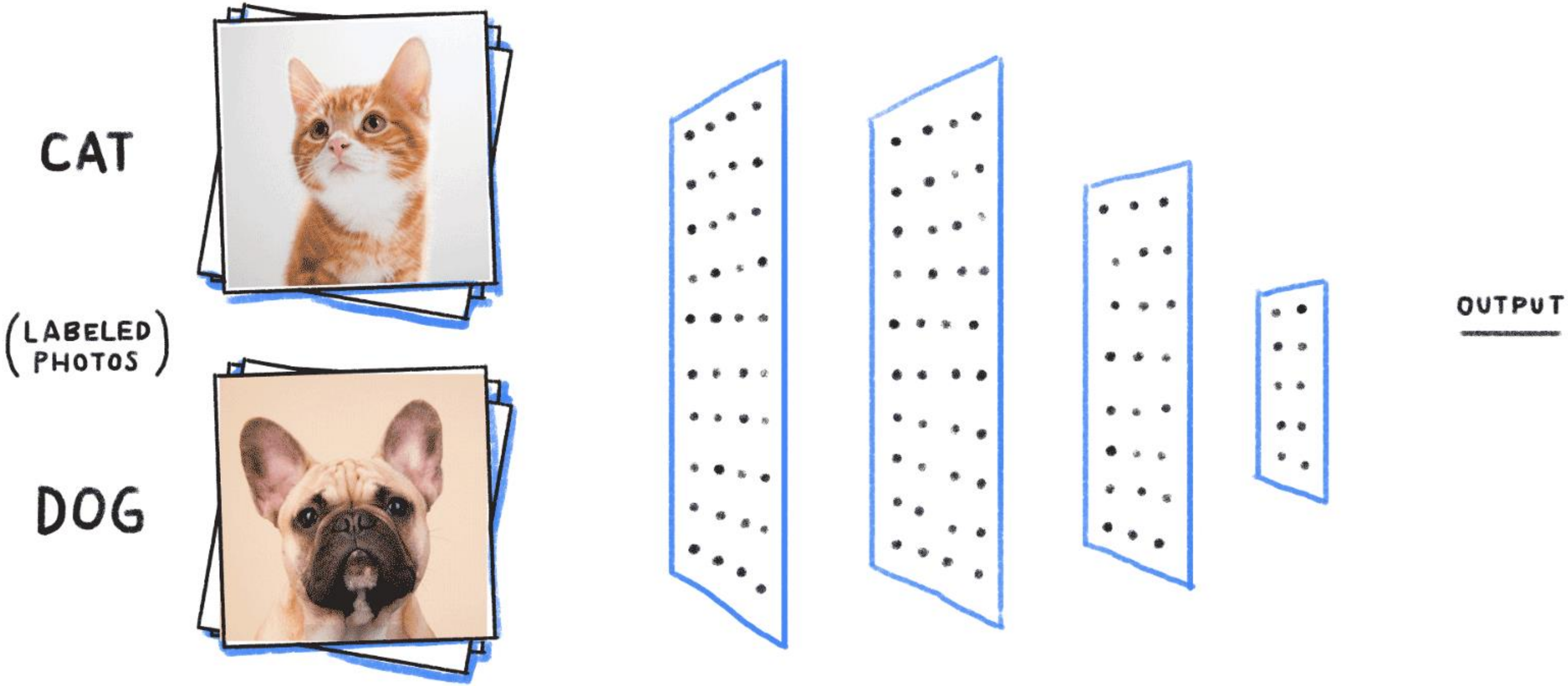
Feedback Loop (Training iteration)

Learning System (Model)

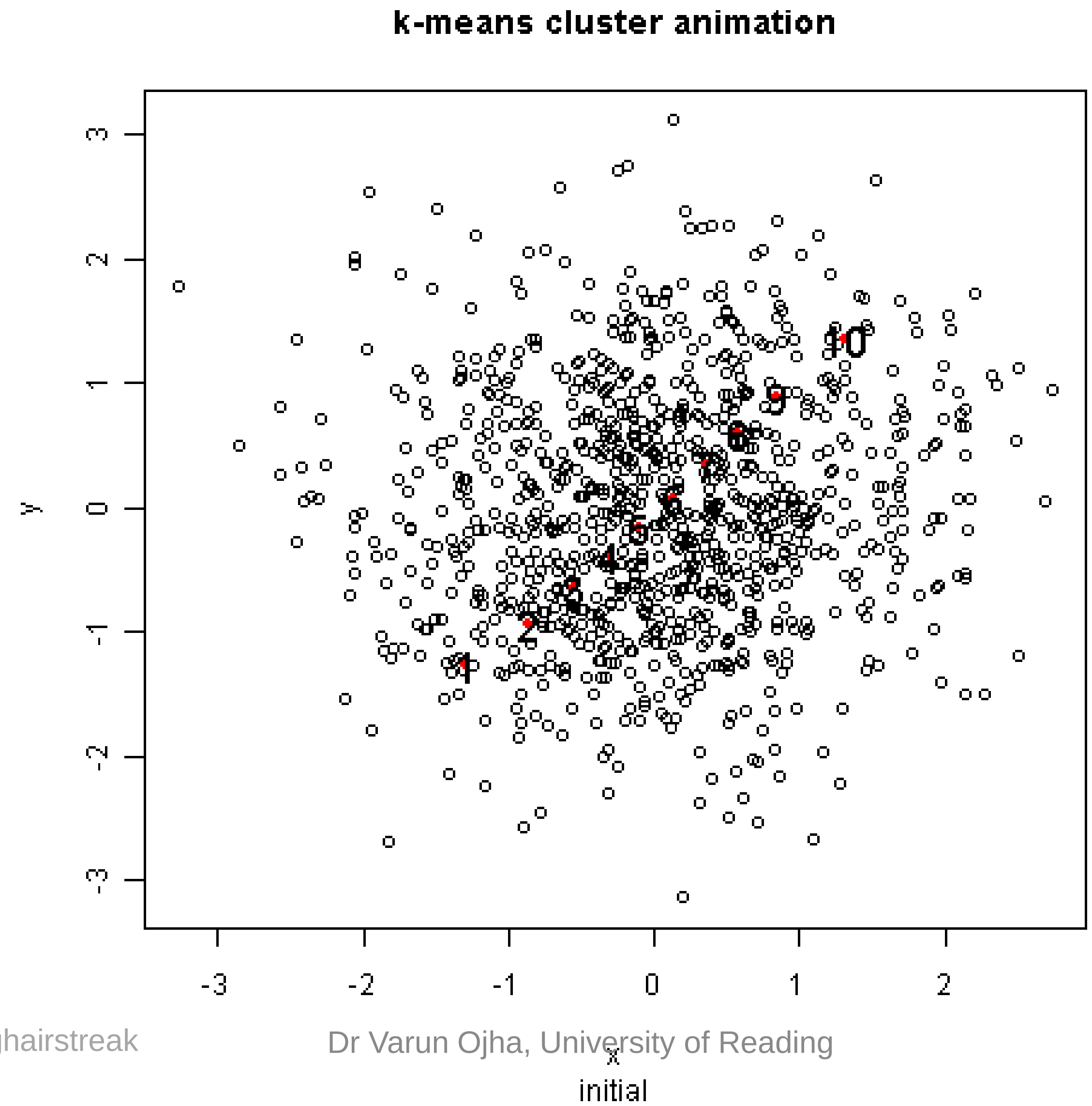
Stop (When to stop learning)

Supervised Learning

Source: https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

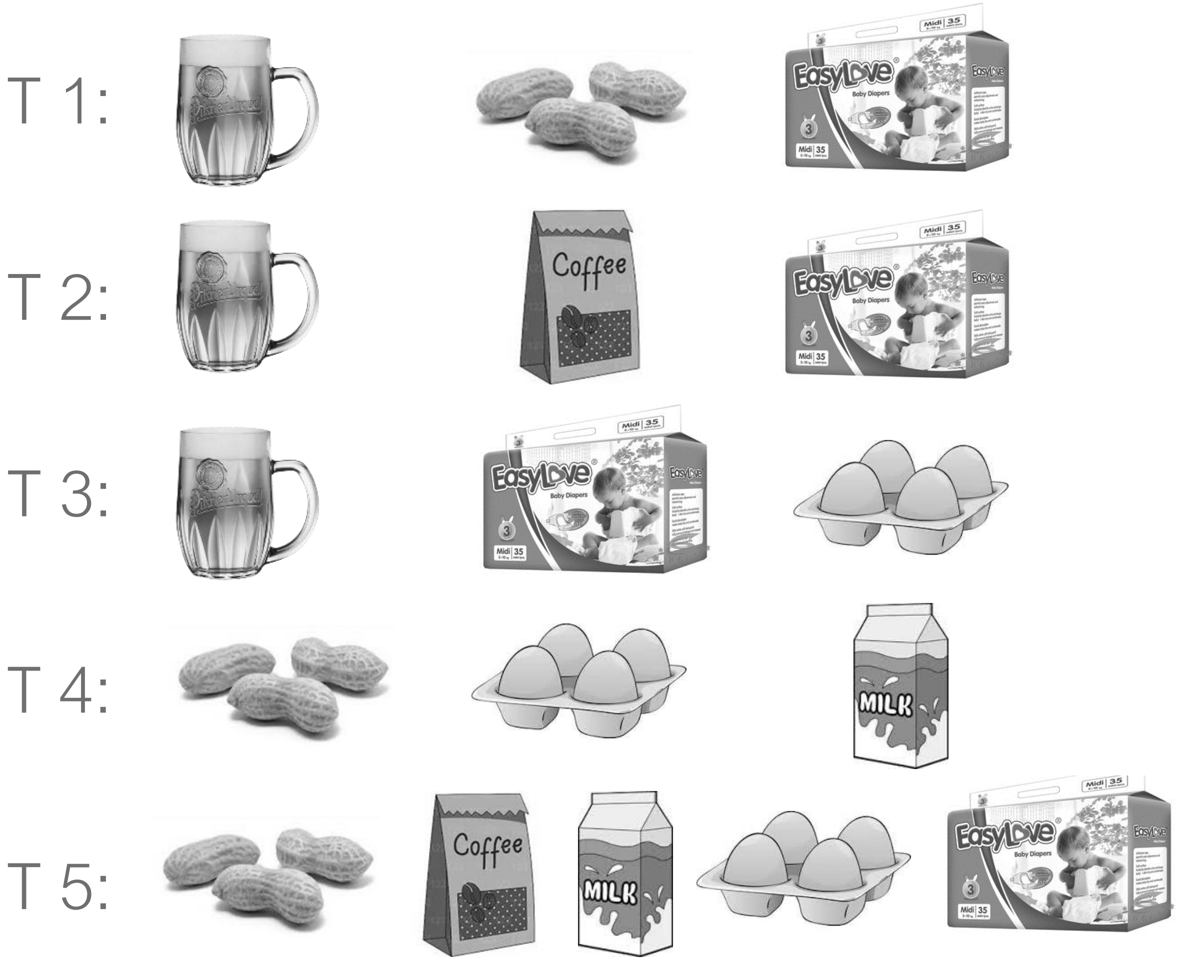


Unsupervised Learning



Source: https://gfycat.com/basicenchantinghairstreak

Product recommendation



Associated Rule Mining: to find frequent pattern (rule) in dataset. (The Impossible Correlation - 1992)

- **IF** someone buy an item X **THEN** what is the possibility that the person will by the item Y
  - So compute **X → Y** (Support, Confidence)
  - Support: The probability that transaction contain both X and Y
  - Confidence: The conditional probability that the transaction containing X also contain Y.
  - **Beer → Diapers (60%, 100%)**
  - **Diapers → Beer (60%, 75%)**

Fraud Detection

Detect unusual transaction

Image source: https://goo.gl/images/kUzUho

Trend/ Forecast

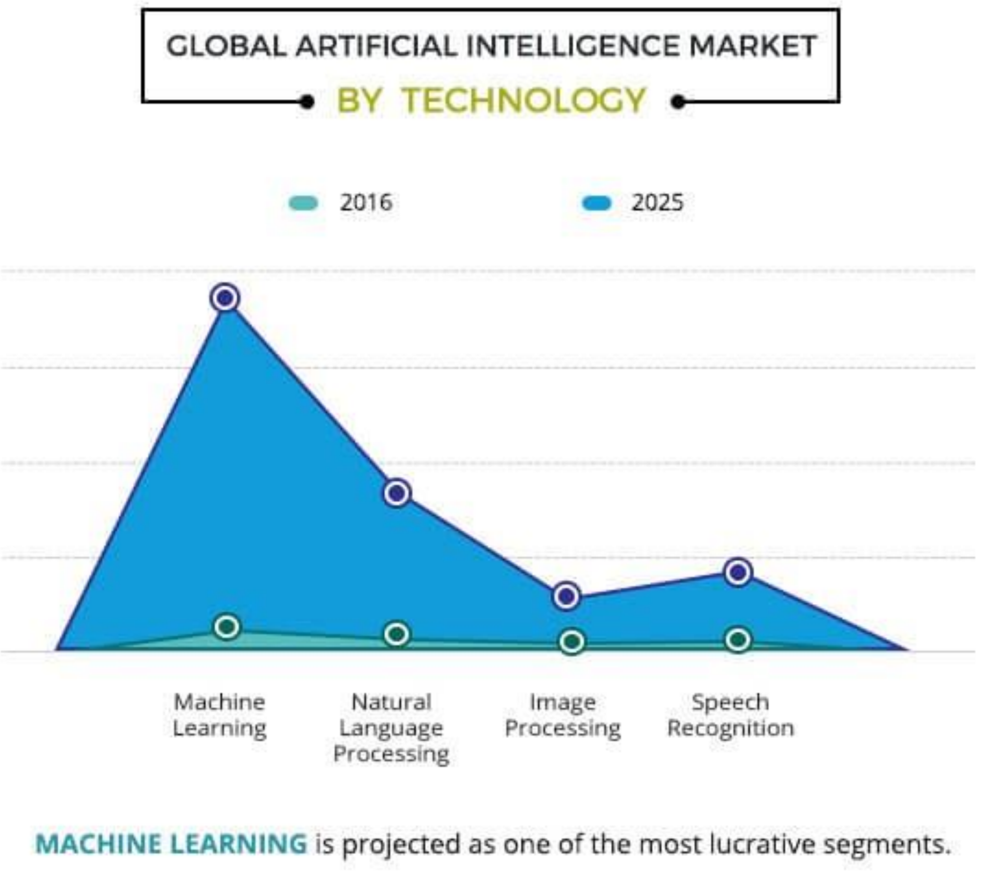
1 Pound sterling equals

1.30 United States Dollar

10 Mar, 13:02 UTC · Disclaimer

1	Pound sterling
1.30	United States Dolla

Which AI will be used most



Source: https://www.alliedmarketresearch.com/artificial-intelligence-market

PART 2
Artificial Intelligence (in Pharmacology)

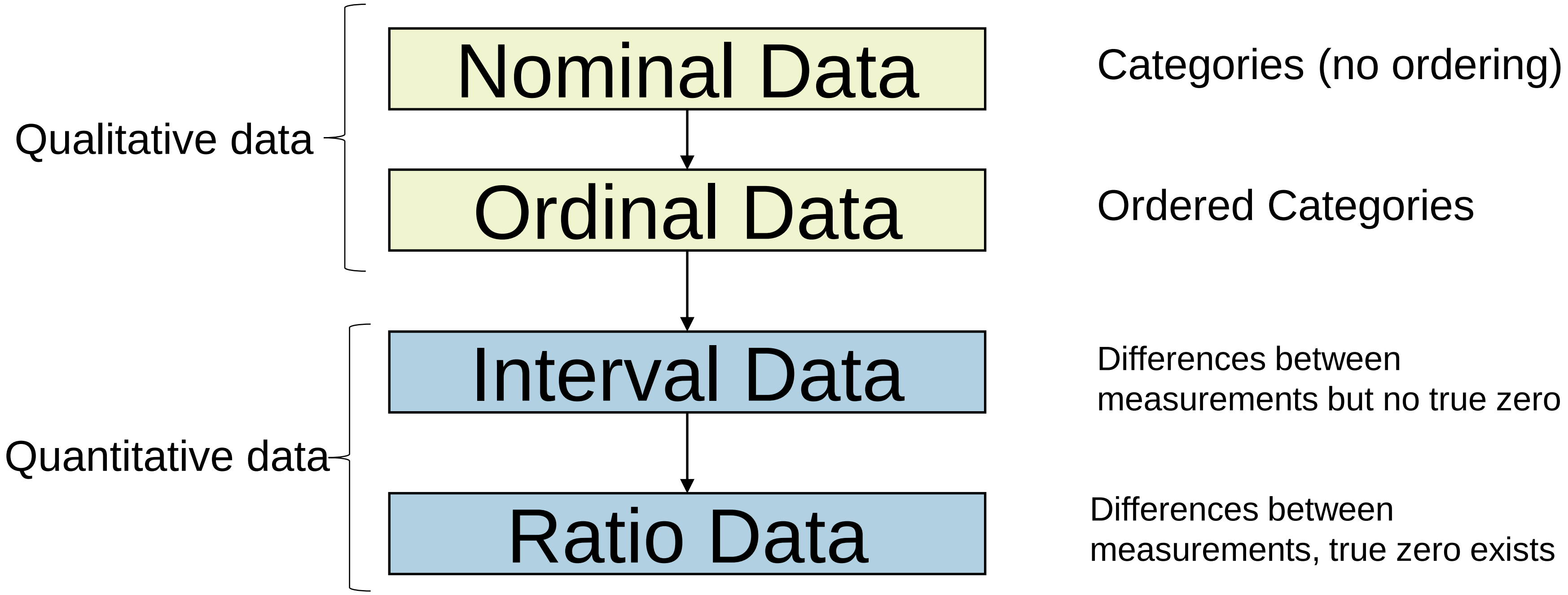
Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

How is it done?

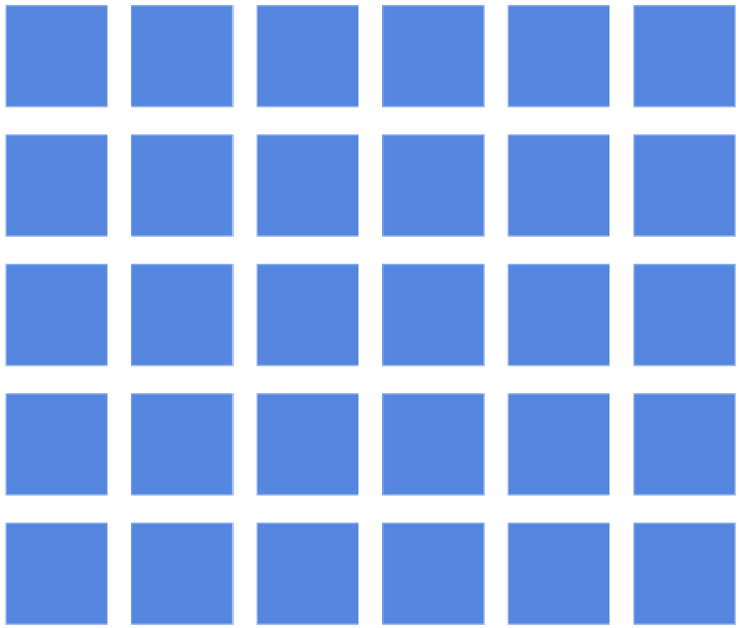
12:25 PM

Data-Driven
Artificial Intelligence

Types of Data

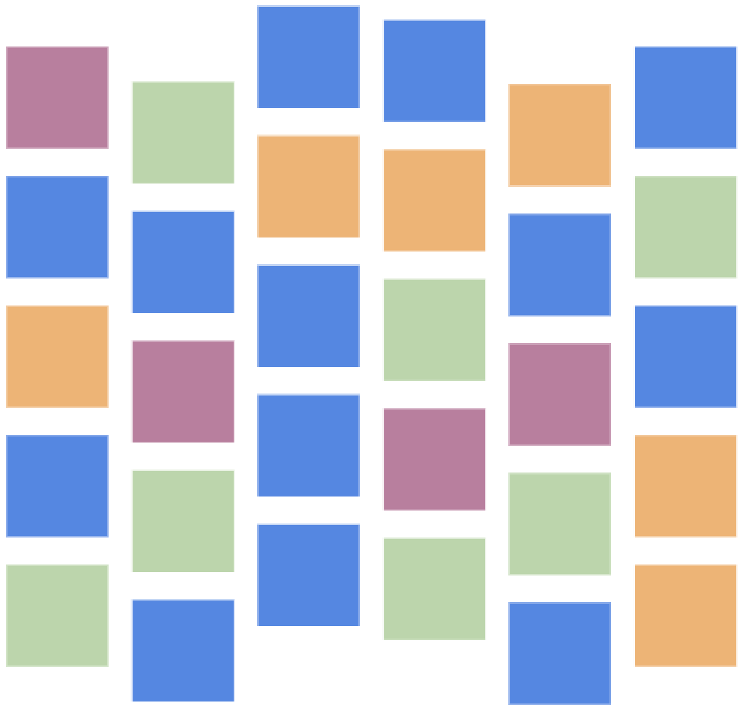


Structured data



Data stored in databases and tables

Unstructured data



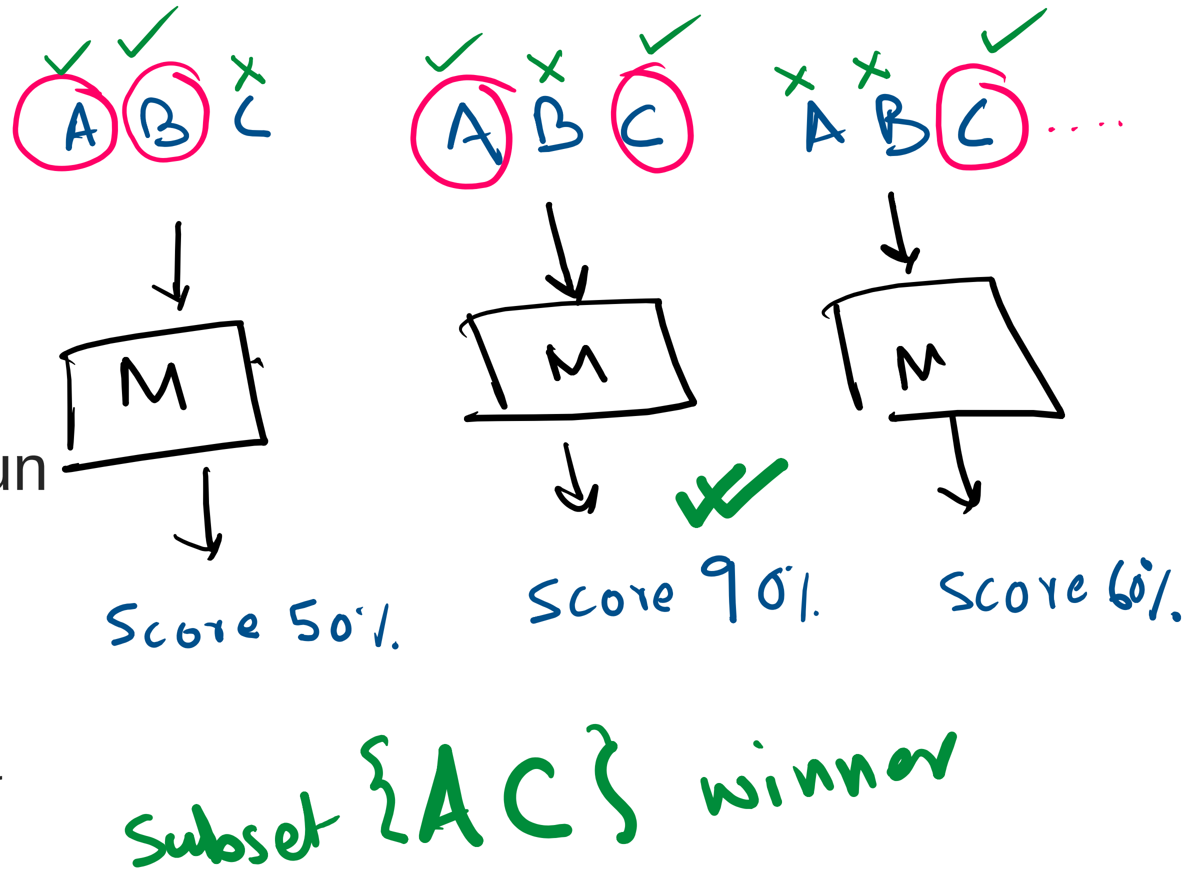
Images, text, audio, video, documents

Data Quality

- Noise
- Outliers
- Missing values

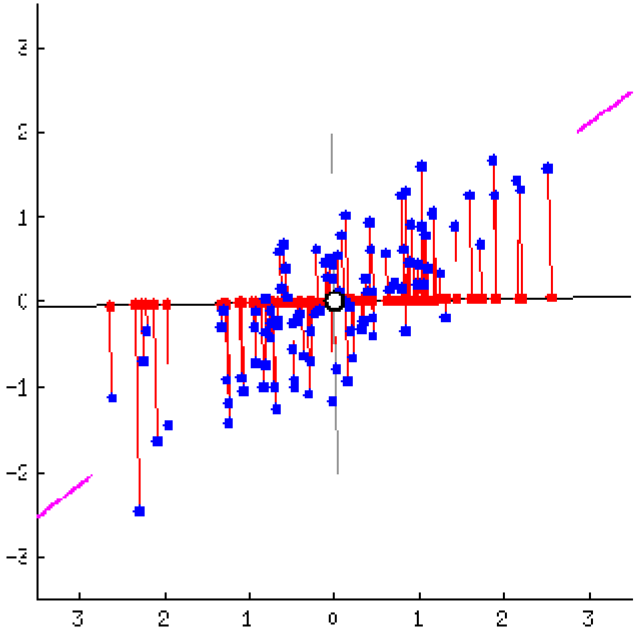
Feature Subset Selection (Techniques)

- **Brute-force approach:**
  - Try all possible feature subsets
- **Filter approaches:**
  - Features are selected before the run
- **Wrapper approaches:**
  - Use the data mining algorithm as a black box to find the best subset



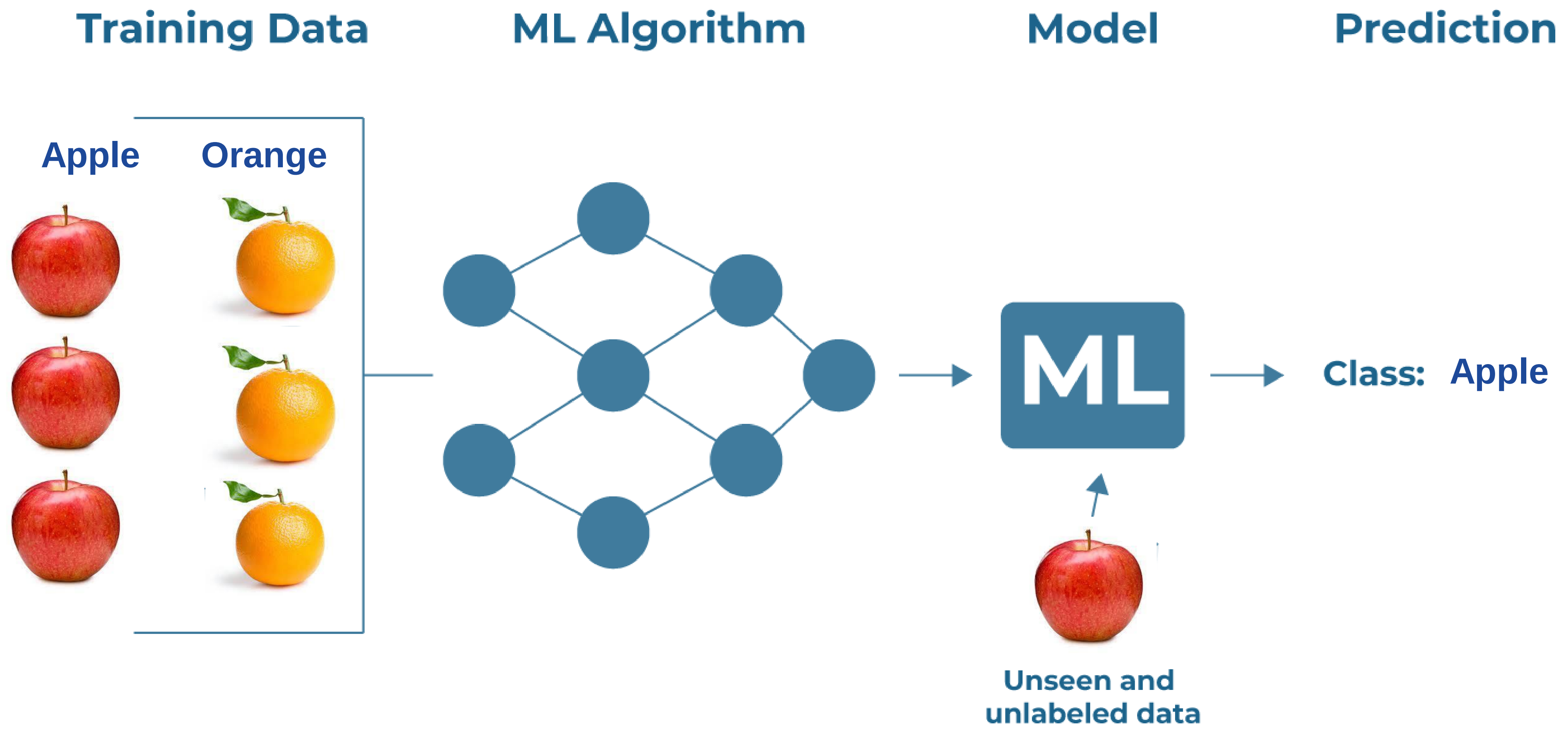
Dimensionality Reduction with PCA

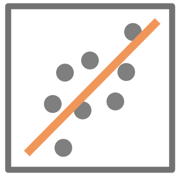
- Principle Component Analysis (PCA):
  - Goal is to find a projection that captures the largest amount of variation in data



Source: https://www.r-bloggers.com/2017/01/principal-component-analysis-in-r/

Supervised Learning (Classification)



Regression and Classification

Class/Traget attribute

	#	Inputs Attributes (Independent)		Target/Class/Output Attributes (Dependent)
		A1	**A2**	**A3**
Records	**Ex. 0**	$A1_0$	$A2_0$	$A3_0$
	Ex. 1	$A1_1$	$A2_1$	$A3_1$
	Ex. 2	$A1_2$	$A2_2$	$A3_2$
	Ex. 3	$A1_3$	$A2_3$	$A3_3$
	Ex. 4	$A1_4$	$A2_4$	$A3_4$
	Ex. 5	$A1_5$	$A2_5$	$A3_5$
	Ex. 6	$A1_6$	$A2_6$	$A3_6$
	Ex. 7	$A1_7$	$A2_7$	$A3_7$
	Ex. 8	$A1_8$	$A2_8$	$A3_8$
	Ex. 9	$A1_9$	$A2_9$	$A3_9$

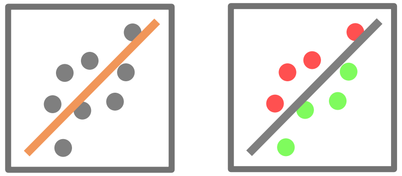
Target (Class) Attributes (A3)

Regression

Continuous **(Numerical)** labeled data

Classification

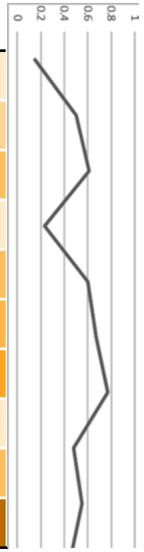
Discrete **(Categorical)** labeled data



Tasks: Regression and Classification

Continuous labeled data

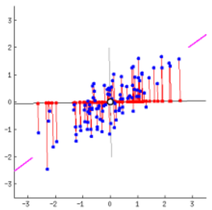
#	Inputs (X)		Target (Y)
	Area (m^2)	Distance(mile)	Price (£Bn)
Ex. 0	76.85	17.27	0.15
Ex. 1	76.97	19.54	0.5
Ex. 2	77.10	18.51	0.76
Ex. 3	85.28	46.09	0.23
Ex. 4	85.42	35.83	0.6
Ex. 5	88.02	2.59	0.67
Ex. 6	77.25	6.34	0.89
Ex. 7	77.49	6.98	0.2
Ex. 8	85.81	12.18	0.55
Ex. 9	98.81	2.18	9.45



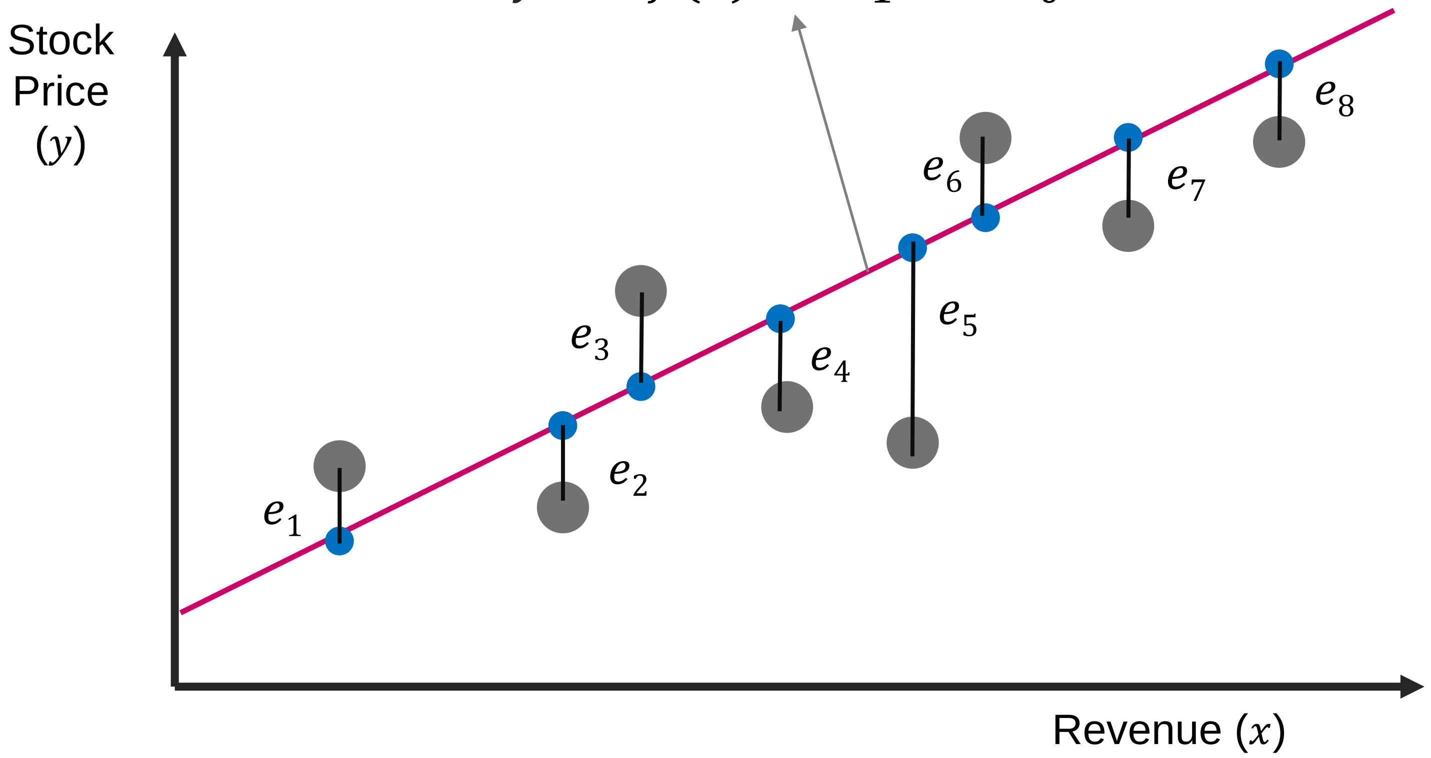
Discrete labeled data

#	Inputs (X)		Class (Y)
	Length (cm)	Weight (kg)	Sales
Ex. 0	23.2	3.2	Good
Ex. 1	70.9	19.5	Bad
Ex. 2	60.5	18.51	Bad
Ex. 3	24.5	4.6	Good
Ex. 4	110.0	35.83	Bad
Ex. 5	23.8	3.7	Good
Ex. 6	25.8	4.5	Good
Ex. 7	24.7	4.9	Good
Ex. 8	85.8	25.6	Bad
Ex. 9	78.8	20.33	Bad

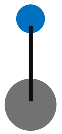
Regression: Linear function



$$\hat{y} = f(x) = w_1 x + w_0$$



- ✓ Best Fit
- ✓ Find the line (parameters of a line equation) that minimize the norm of the *y* errors
- ✓ (sum of the squares)



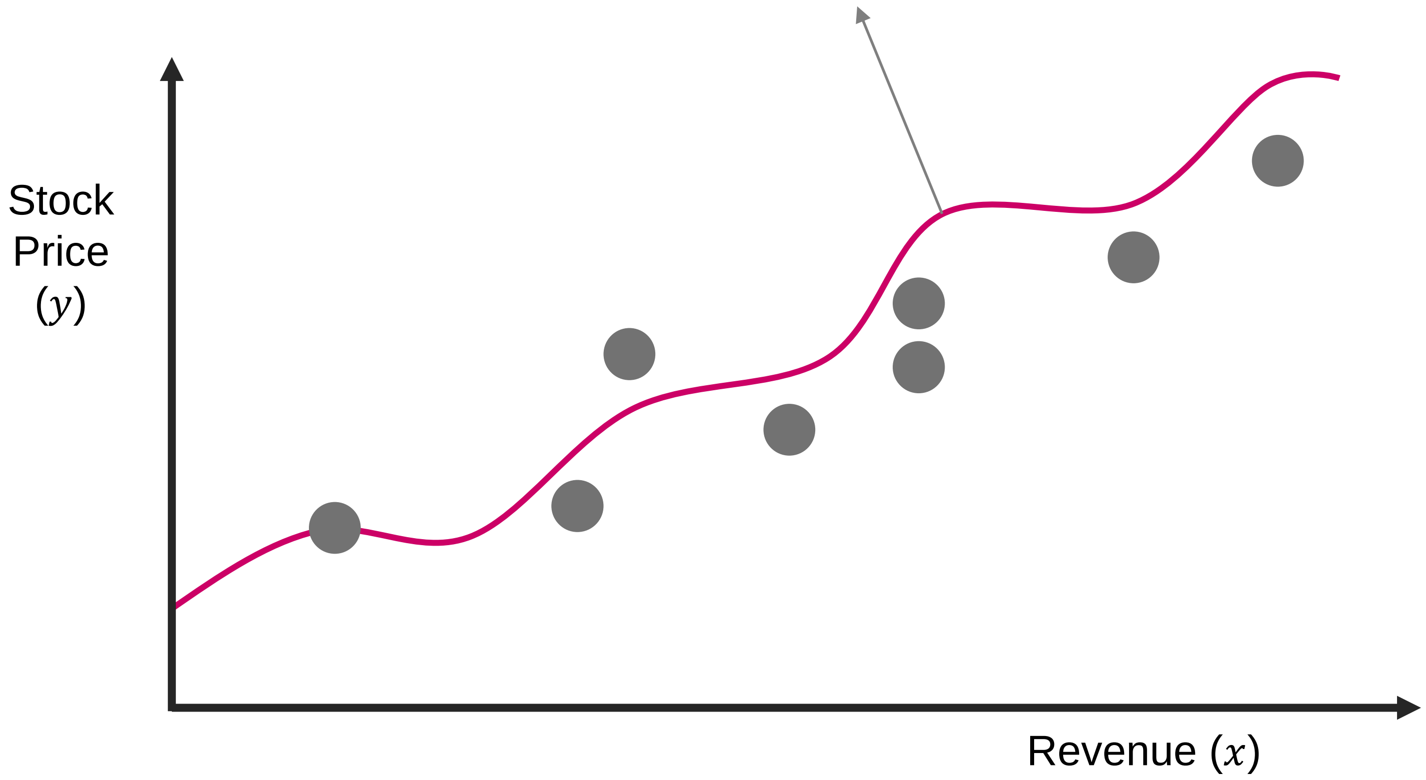
Error

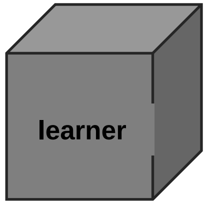
$$e_i = \hat{y}_i - y_i$$

$$\boldsymbol{e} = \sum_{i=1}^{8} (\hat{y}_i - y_i)^2$$

Regression: Non-Linear function

$$\hat{y} = f(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m$$

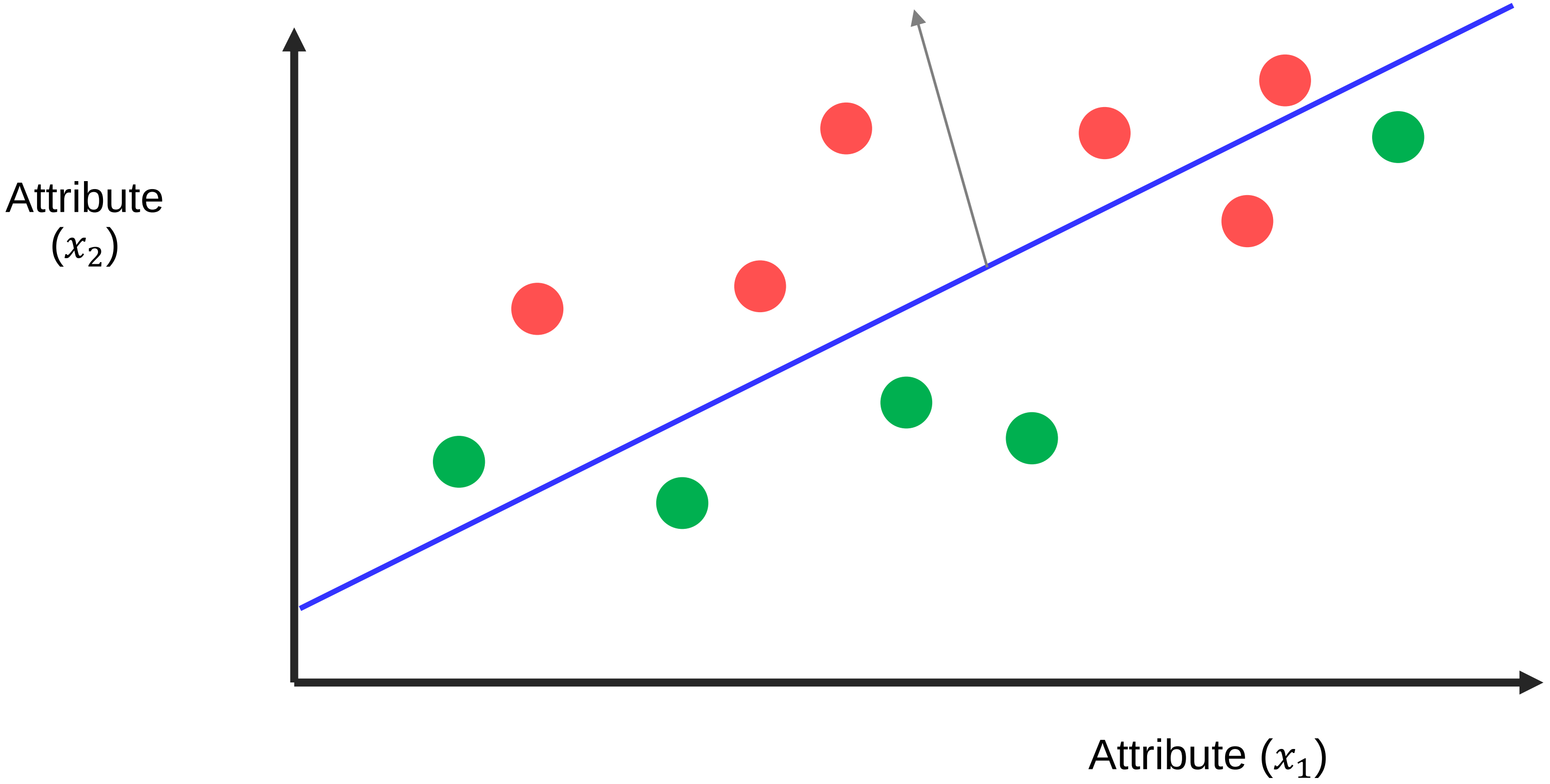




Find values of the weight: $w_1, w_2, \ldots w_m$

Classification: Linear function

$$\hat{y} = f(x) = w_1 x + w_0$$

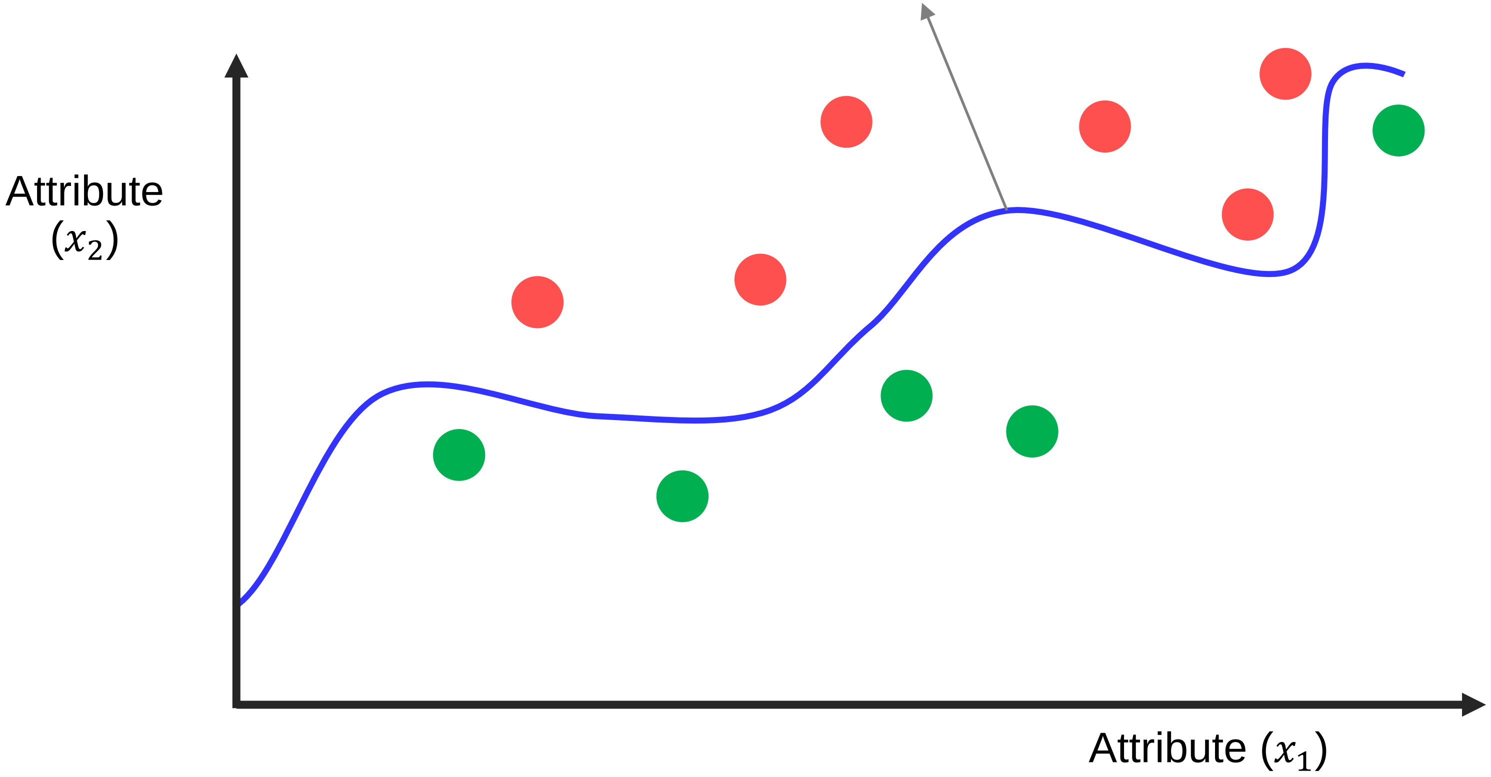


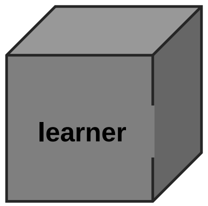
- ✓ Best Fit
- ✓ Find the line (parameters of a line equation) that minimize the error (misclassification) rate

$$\mathbf{e} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \neq y_i$$

Classification: Non-Linear function

$$\hat{y} = f(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m$$





Find values of the weight: $w_1, w_2, \ldots w_m$

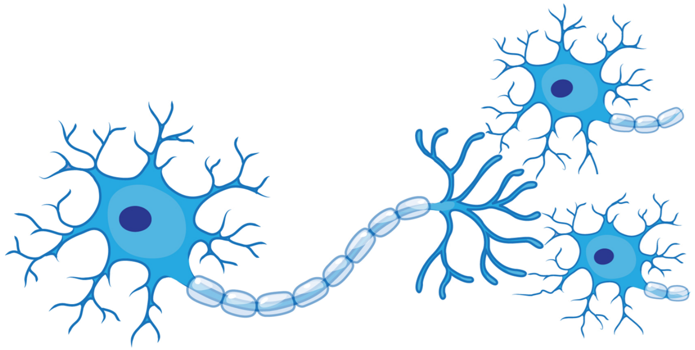
PART 3
Artificial Intelligence (in Pharmacology)

Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

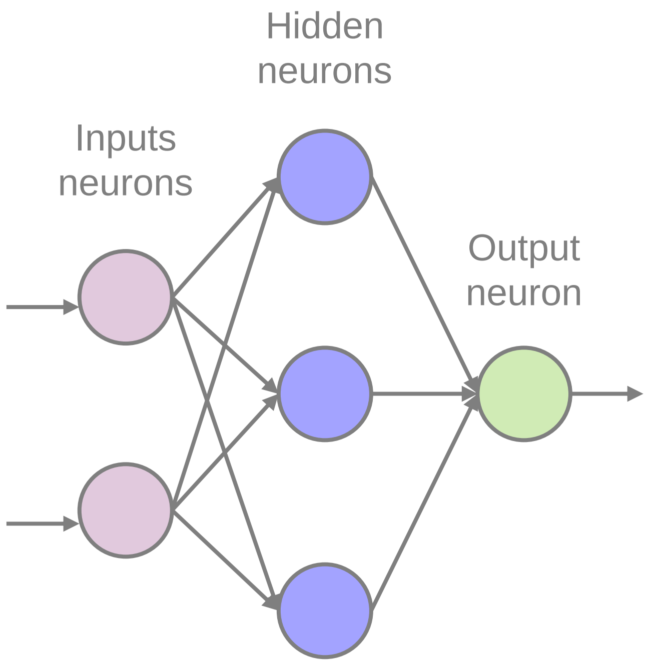
What has revolutionised it?

12:25 PM

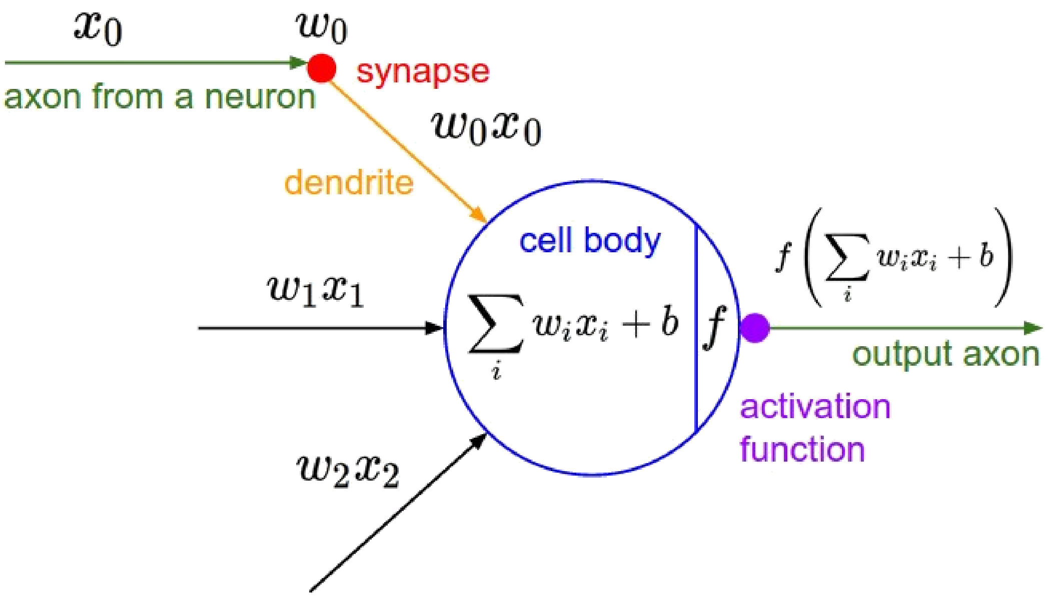
Learning Systems: Neural Networks



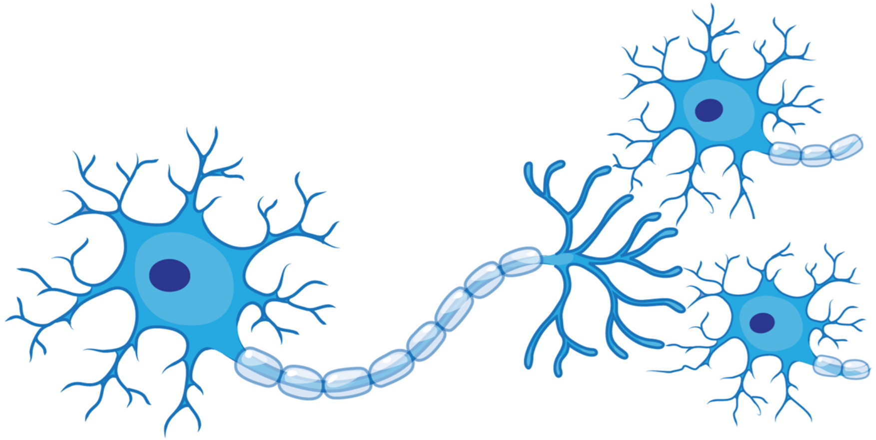
Biological networks of neurons in human brains

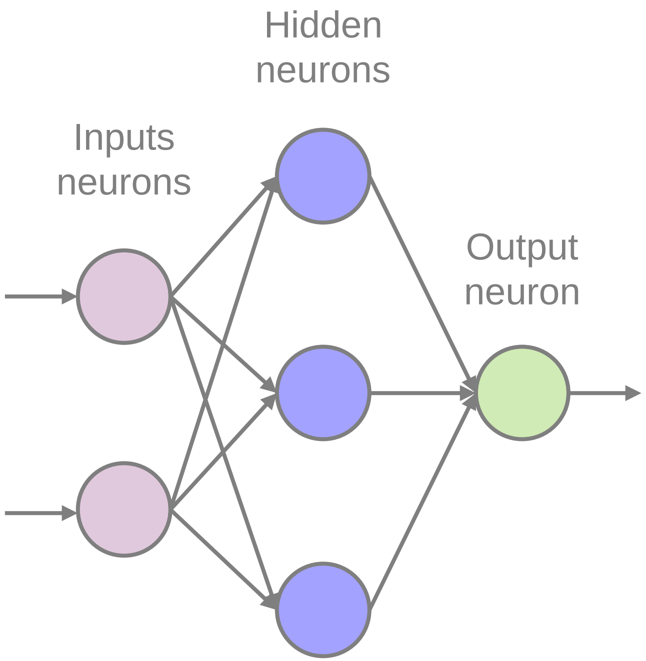


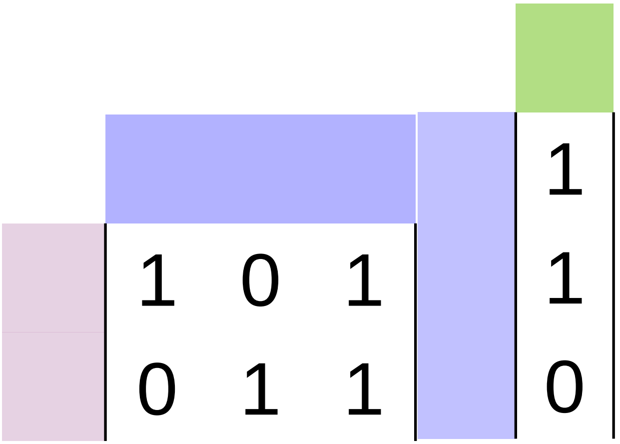
AI representation of biological neural networks



Learning Systems: Neural Networks







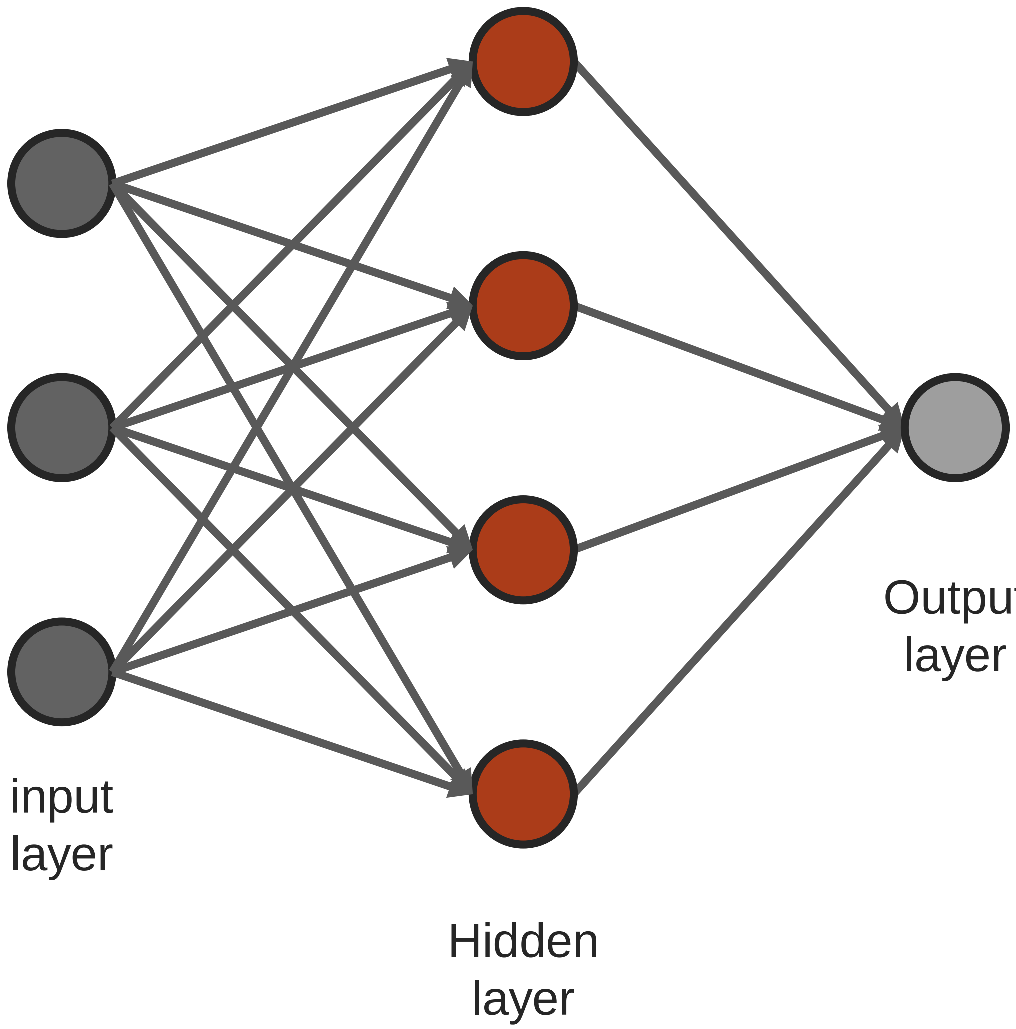
1 Biological networks of neurons in human brains

2 AI representation of biological neural networks

3 Mathematical representation of the neural networks

NEURAL NETWORK

Architecture



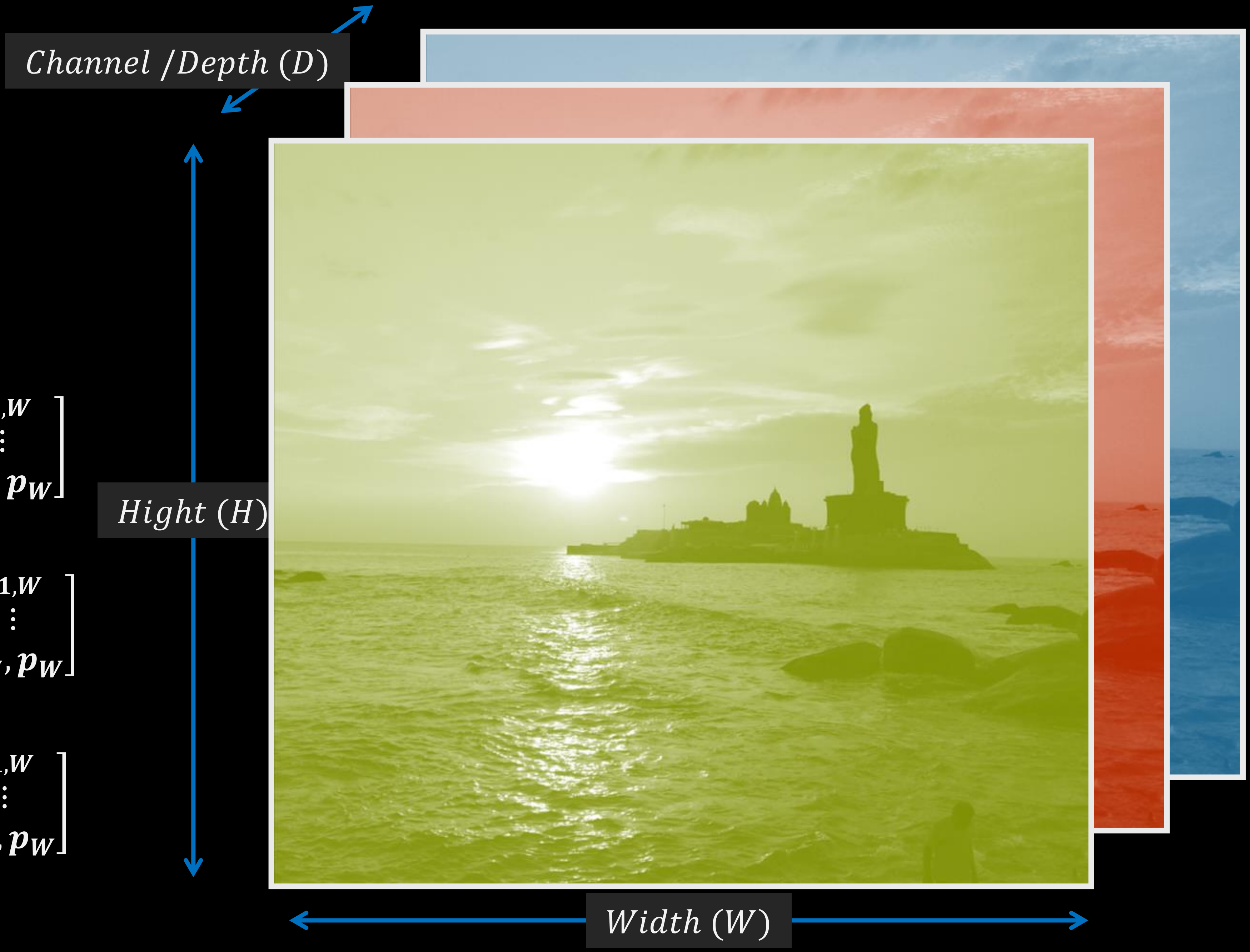
Data

Image: Colour

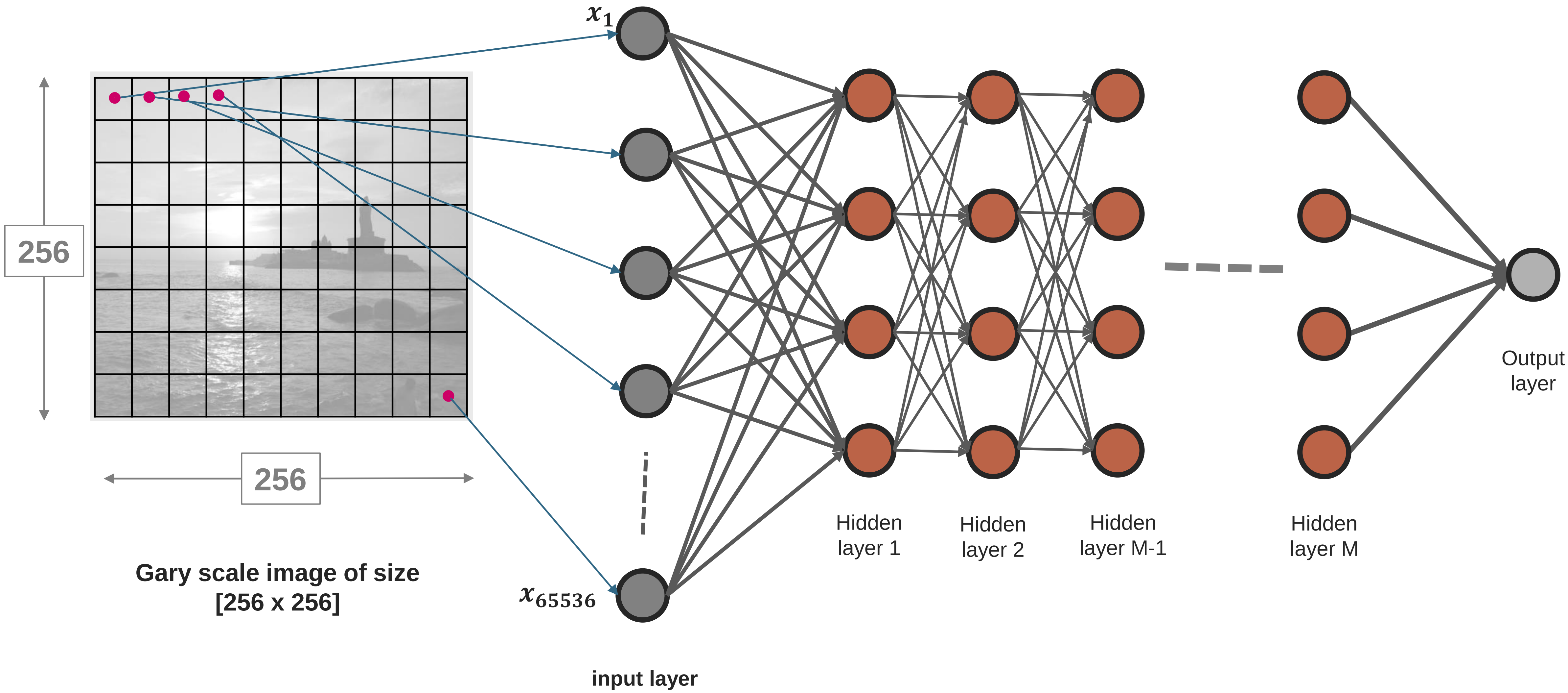
$$I_{RED} = \begin{bmatrix} p_{11} & \cdots & p_{1,W} \\ \vdots & \ddots & \vdots \\ p_{H,1} & \cdots & p_{H},p_{W} \end{bmatrix}$$

$$I_{Green} = \begin{bmatrix} p_{11} & \cdots & p_{1,W} \\ \vdots & \ddots & \vdots \\ p_{H,1} & \cdots & p_{H},p_{W} \end{bmatrix}$$

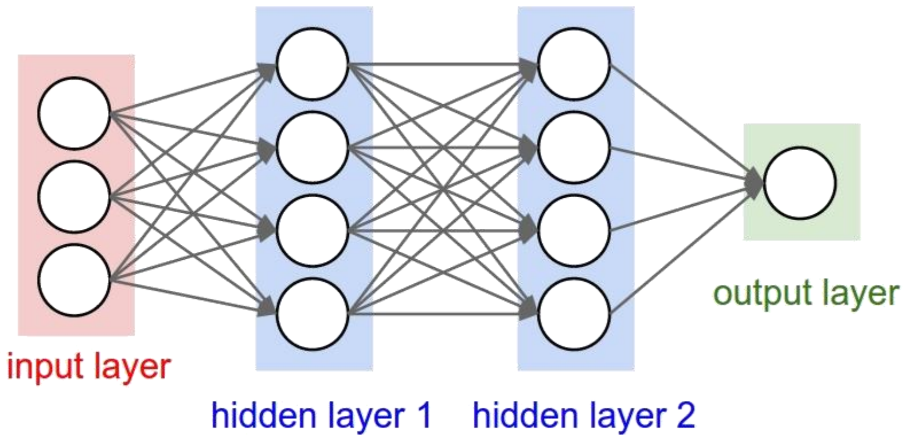
$$I_{Blue} = \begin{bmatrix} p_{11} & \cdots & p_{1,W} \\ \vdots & \ddots & \vdots \\ p_{H,1} & \cdots & p_{H},p_{W} \end{bmatrix}$$

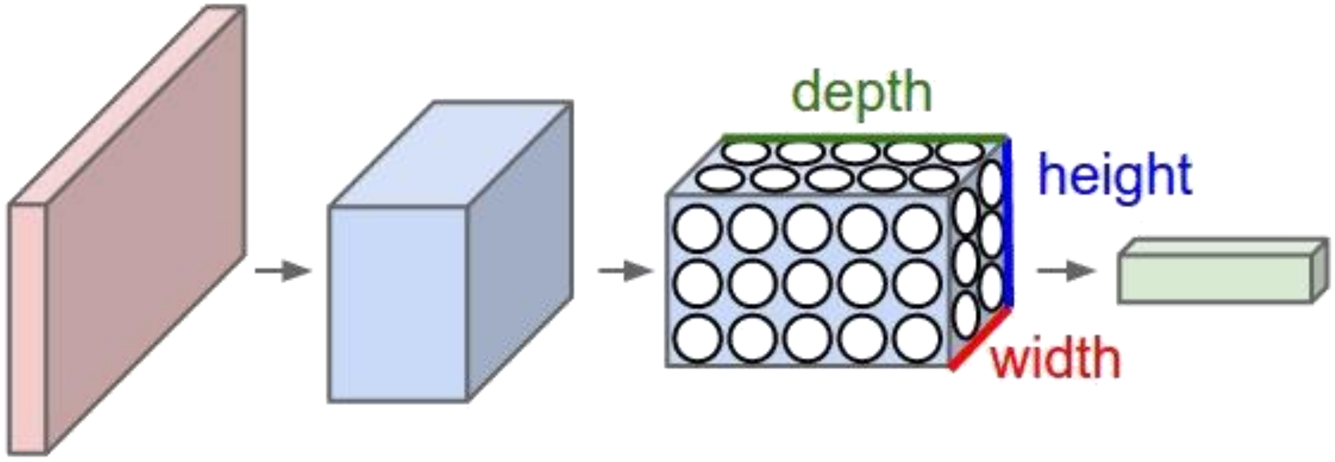


Deep Learning



Convolutional Neural Network (CNN)



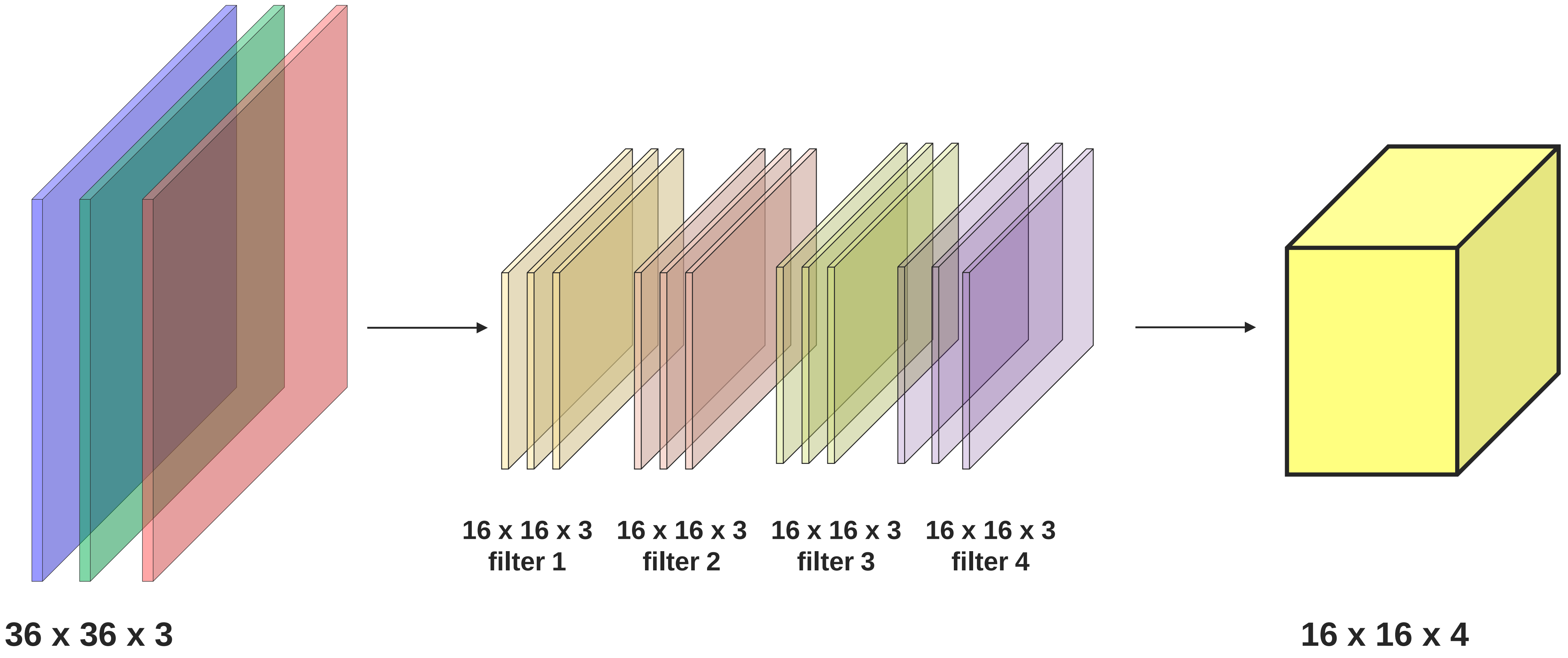


A regular Deep Neural Network

Convolutional Neural Network

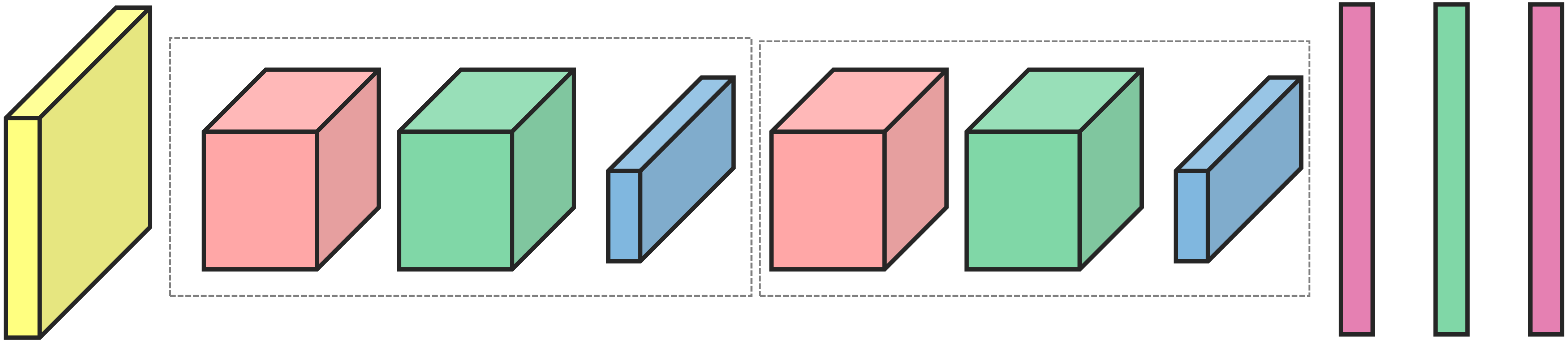
A very good source: http://cs231n.github.io/convolutional-networks/

Convolution



Convolutional Net Architecture

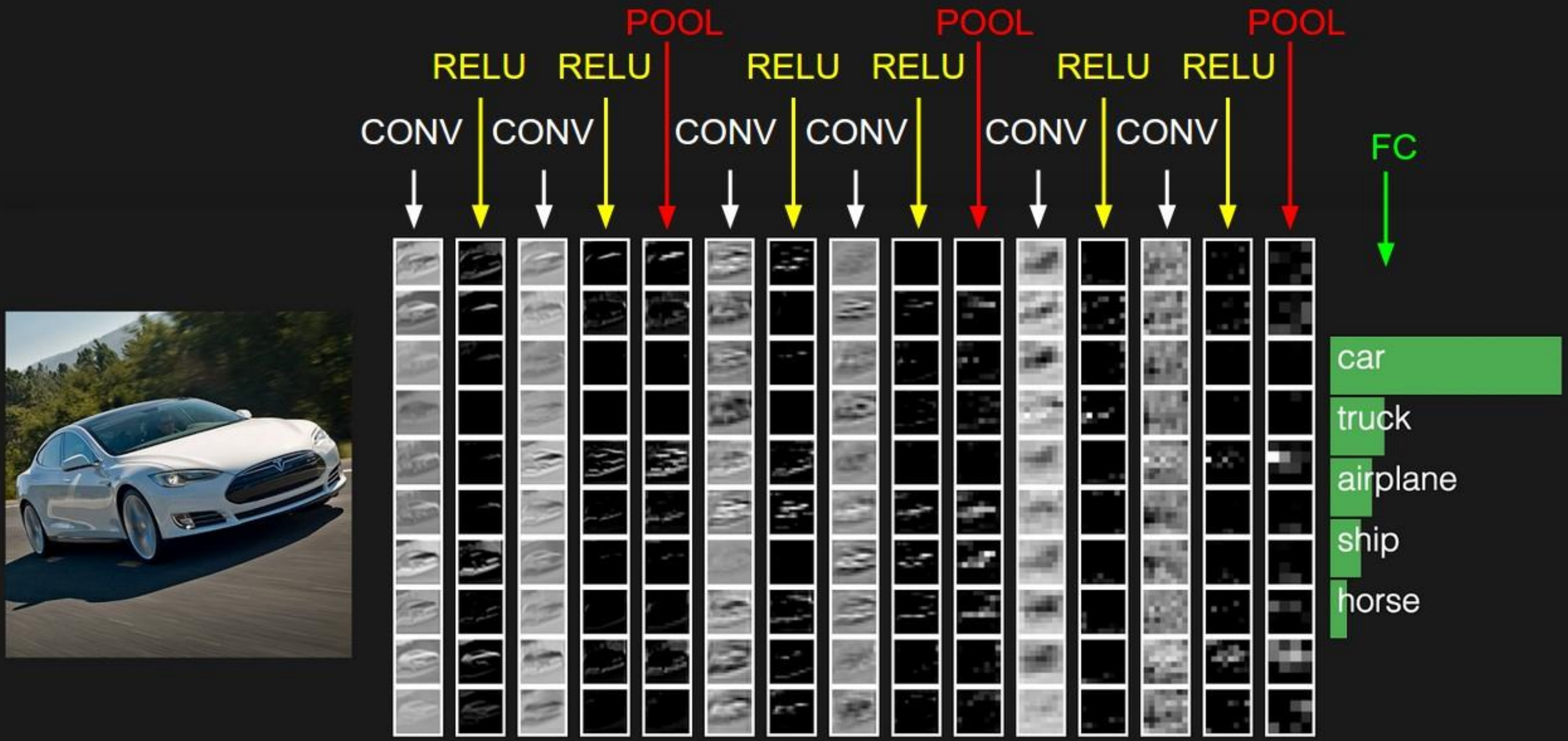
$$\mathbf{INPUT} \longrightarrow [\mathbf{CONV} \longrightarrow \mathbf{RELU} \; -> \; \mathbf{POOL}] * \mathbf{2} \longrightarrow \mathbf{FC} \longrightarrow \mathbf{RELU} \longrightarrow \mathbf{FC}$$



A very good source: http://cs231n.github.io/convolutional-networks/

ConvNet/ CNN: A Simple Example

Live demo http://cs231n.stanford.edu/



PART 4
Artificial Intelligence (in Pharmacology)

Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

How to know it is working well?

12:25 PM

Loss function: Mean Squared Error, $\boldsymbol{E}$

$$\boldsymbol{E} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

$\hat{y}_i$ - predicted output
y_i - target output
n - number of examples in training/test set

Loss function: Mean Absolute Error, $\boldsymbol{E}$

$$\boldsymbol{E} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|$$

$\hat{y}_i$ - predicted output
y_i - target output
n - number of examples in training/test set

Loss function: Misclassification rate, $\boldsymbol{E}$

$$\boldsymbol{E} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i \neq y_i)$$

$\hat{y}_i$ - predicted output
y_i - target output
n - number of examples in training/test set

Loss function: Log loss

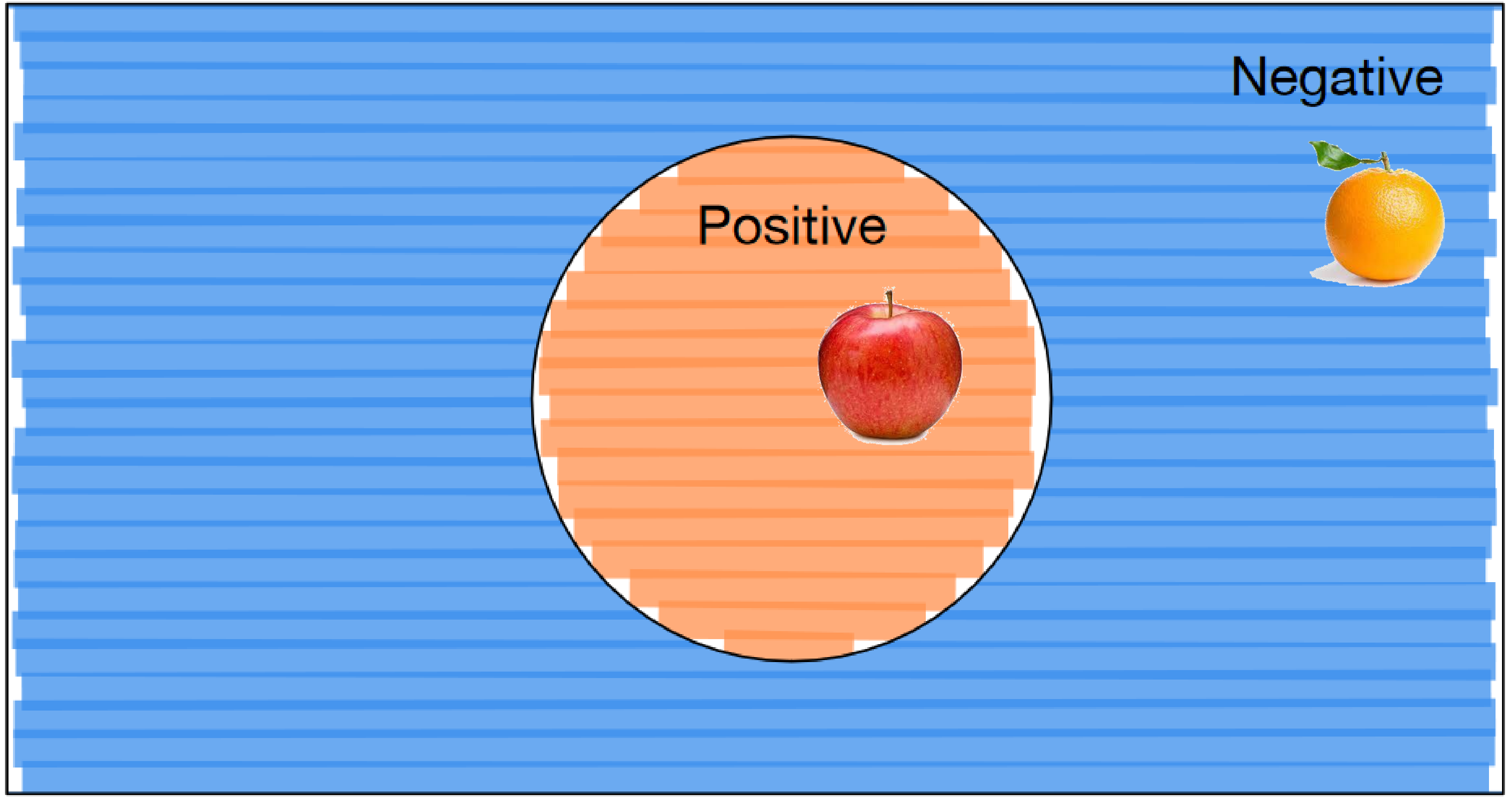
This part will be zero if $y_i = 0$

This part will be zero if $y_i = 1$

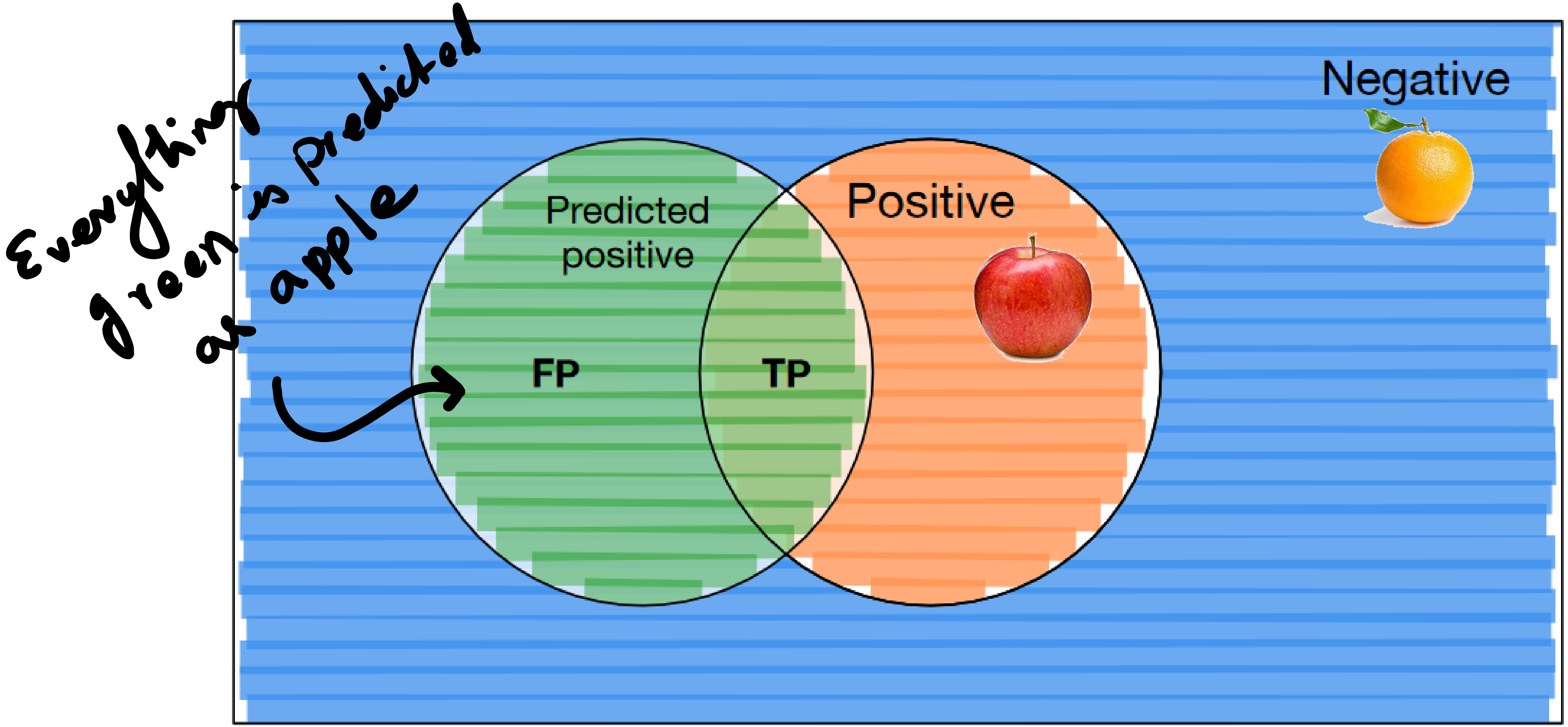
$$-\log \mathbf{P}(y_i \mid \hat{y}_i) = -\left((\boldsymbol{y_i})\mathbf{log}(\hat{y}_i) + (\mathbf{1} - \boldsymbol{y_i})\mathbf{log}(\mathbf{1} - \hat{y}_i) \right)$$

$\hat{y}_i$ - predicted output
y_i - target output
n - number of examples in training/test set

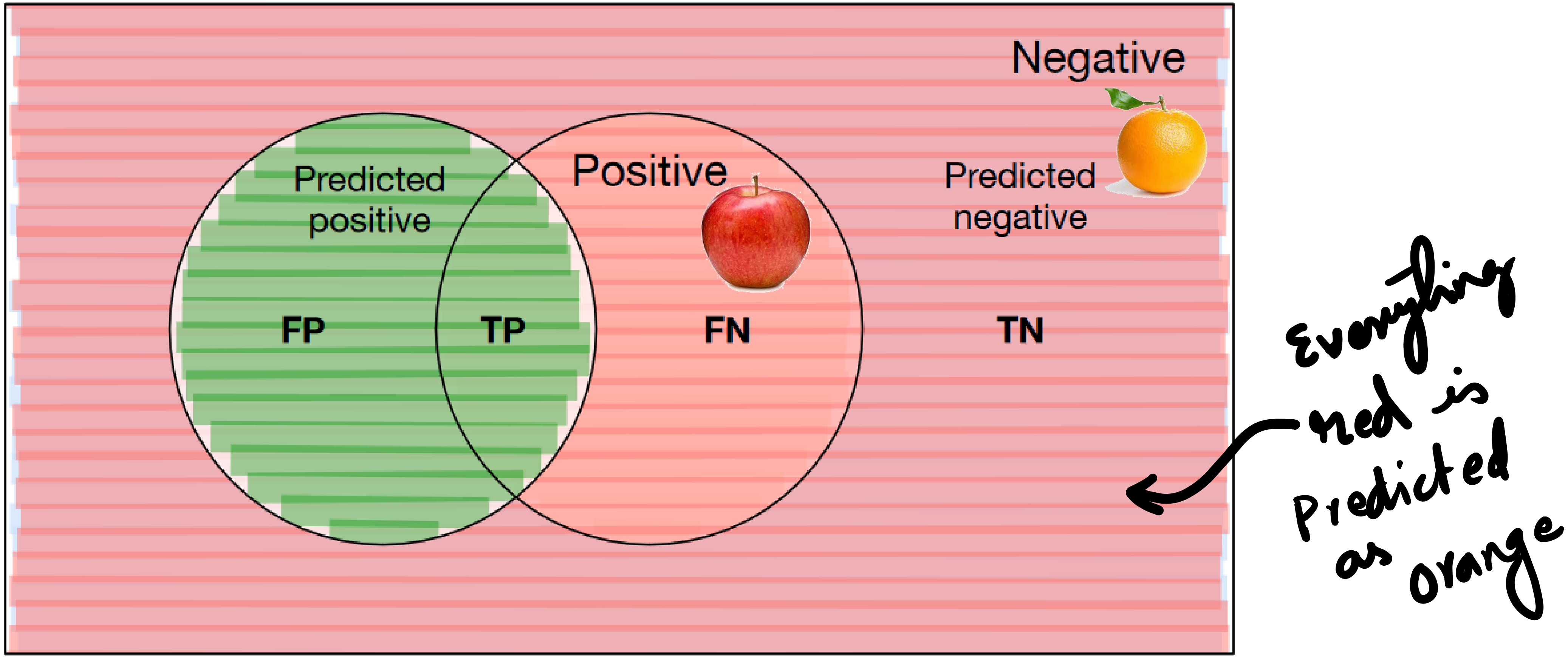
A Binary Classification Problem



False Positives and True Positives



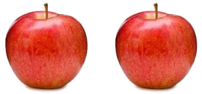
False Negatives and True Negatives



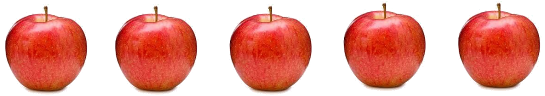
Evaluation Metrics: Precision

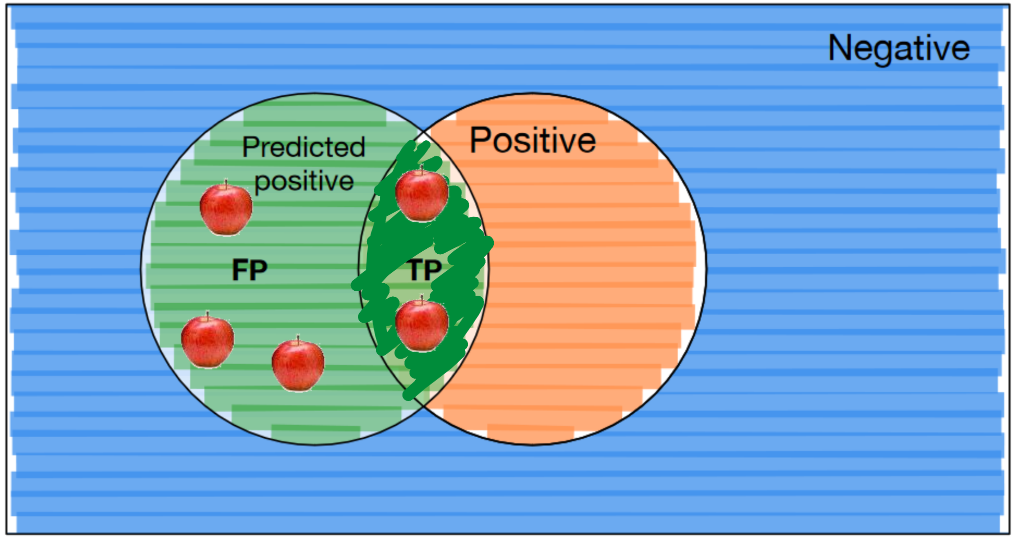
- Precision is the fraction of positive predictions for the respective class that is correct: *How well you guess*

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$



$$= \frac{\text{Total } \textit{apples} \text{ correctly predicted as } \textit{apples}}{\text{Total number of postive prediction in the data}}$$

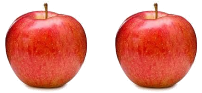




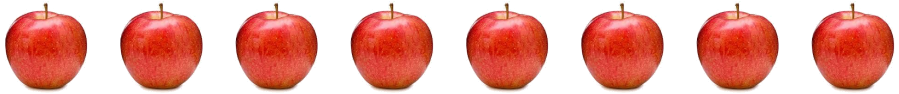
Evaluation Metrics: Recall

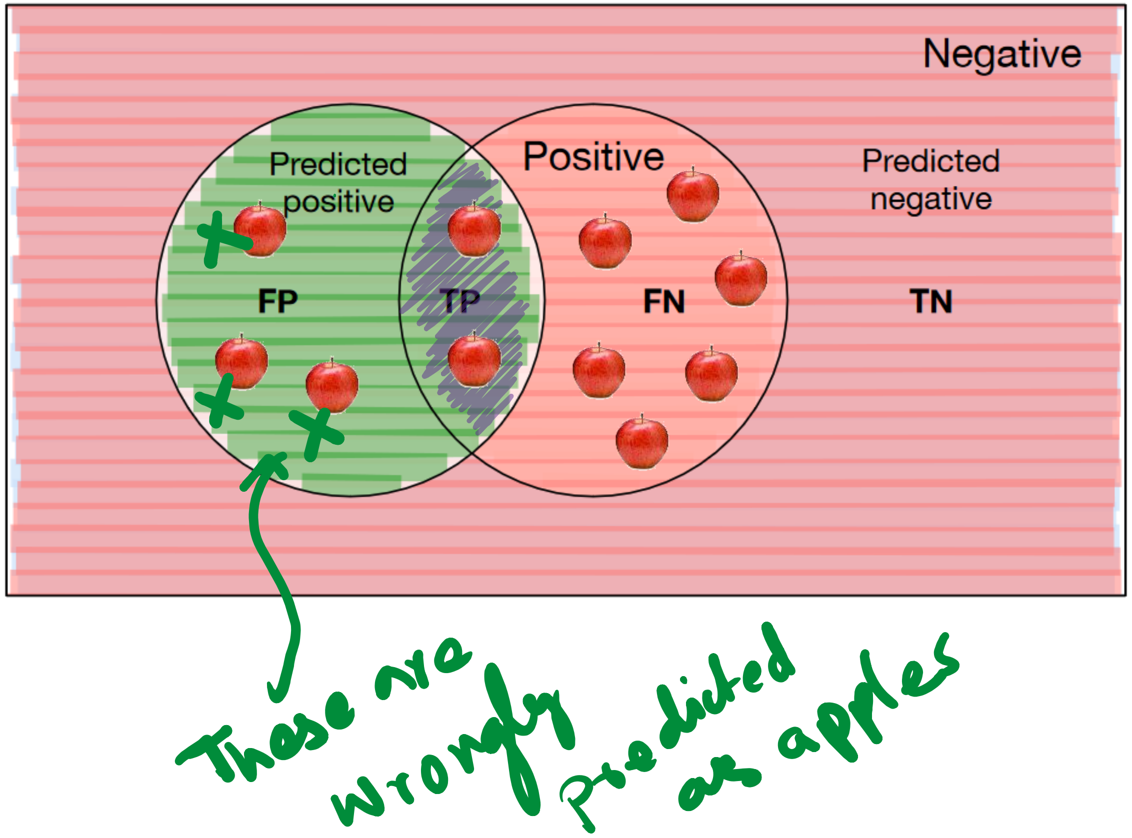
Recall is the fraction of positive values in the data that we correctly predict:
How complete is the prediction?

$$\text{Recall}=\frac{True\ Positive}{True\ Positive+False\ Negative}$$



$$=\frac{\text{Total } \textit{apples} \text{ correctly predicted as } \textit{apples}}{\text{Total } \textit{apples} \text{ in the data}}$$





PART 4
Artificial Intelligence (in Pharmacology)

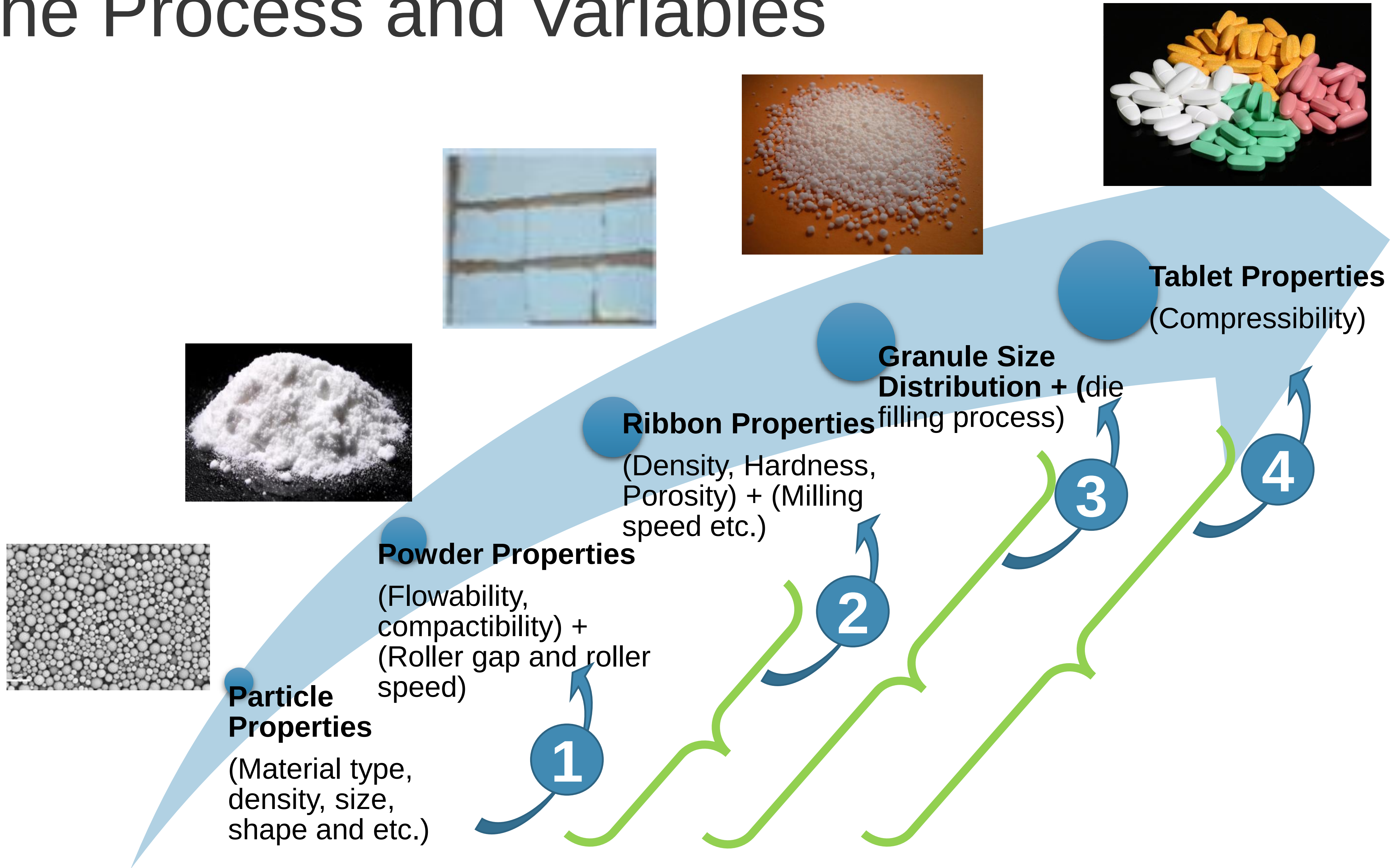
Workshop 04 March 2022
Mathematical Modelling for Pharmacology
University of Reading

Where is it in Pharmacology?

12:25 PM

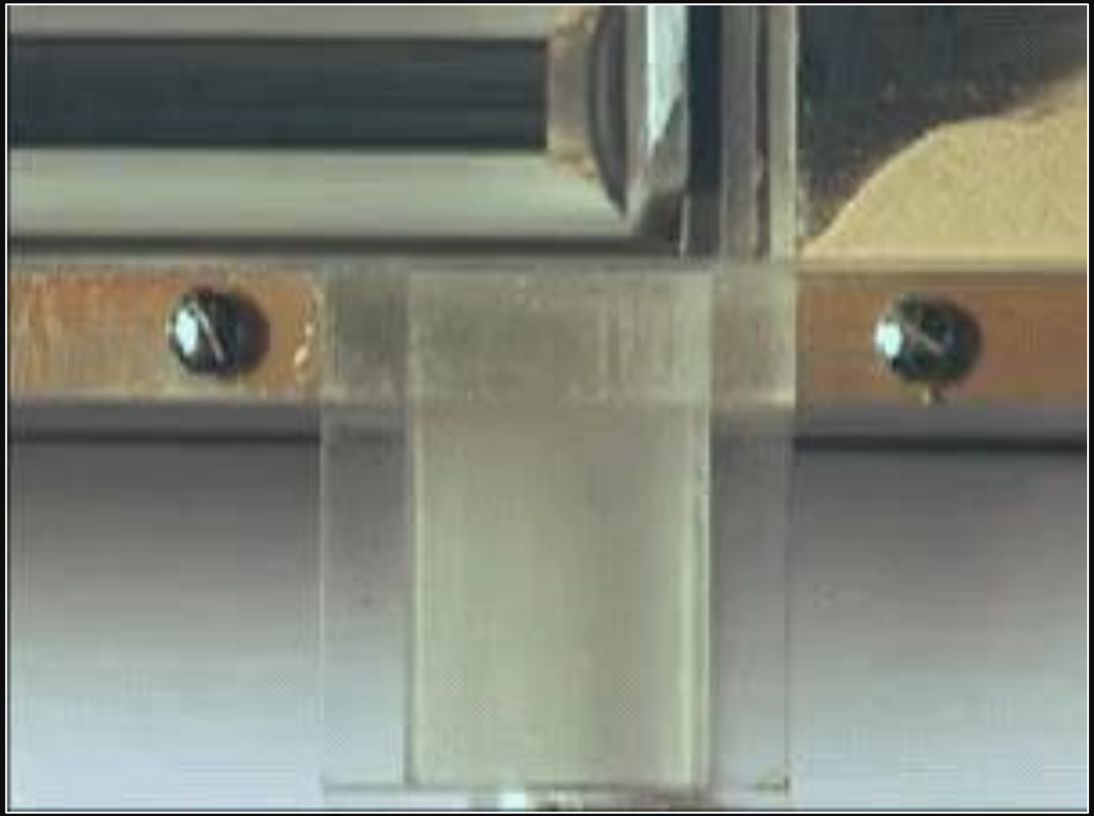
Pharmaceutical (Drugs Production)

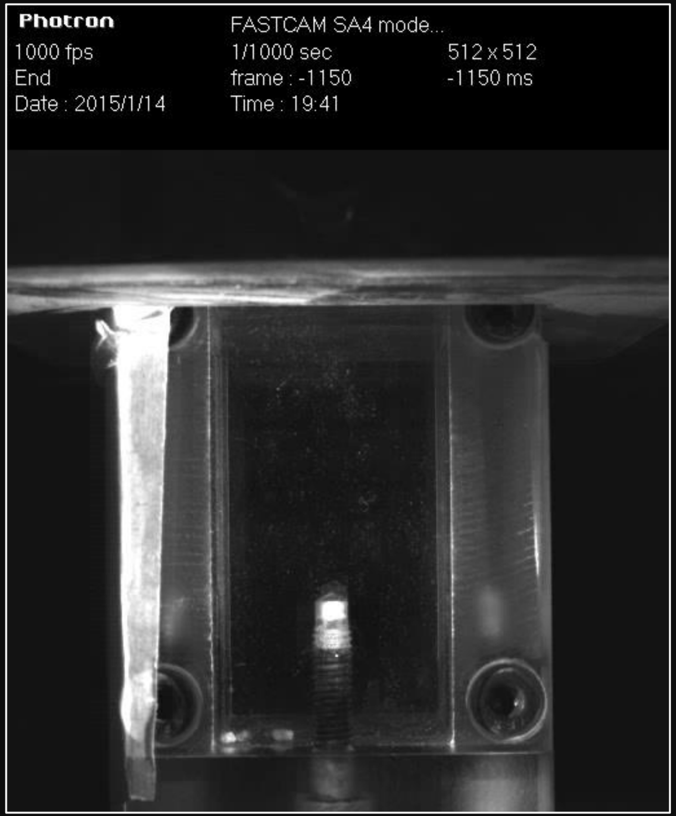
The Process and Variables



Computational intelligence modelling of pharmaceutical industrial processes

Die Filling





Ojha VK et al. (2017) Neural Computing and Applications

Data Collection for Modelling Die Filling

#	Samples Name	Input: True density Feature #1	d50 (µm) Feature #2	Granules size Feature #3	Shoe speed Feature #4	Output Mass (g)
1	MCC PH 101	1581	59.83	90	10	12.81
2	MCC PH 101	1581	59.83	90	10	12.78
:	:	:	:	:	:	:
5	MCC PH 101	1581	59.83	90	20	12.3
6	MCC PH 101	1581	59.83	90	30	9.55
:	:	:	:	:	:	:
135	MCC PH 102	1570.3	94.7	250	50	13.45
136	MCC PH 102	1570.3	94.7	250	60	13.5
:	:	:	:	:	:	:
388	MCC DG	1785.6	52.33	2360	400	9.51
389	MCC DG	1785.6	52.33	2360	400	9.3

Ojha VK et al. (2017) Neural Computing and Applications

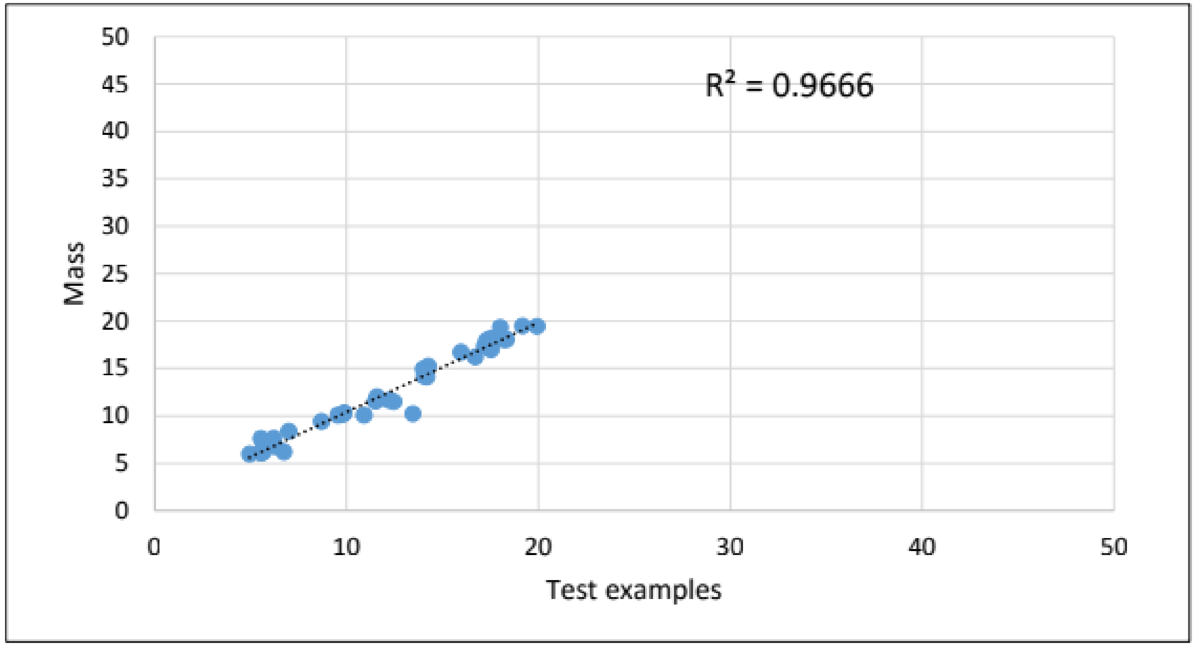
Performance of Algorithms on Prediction and Feature Selection

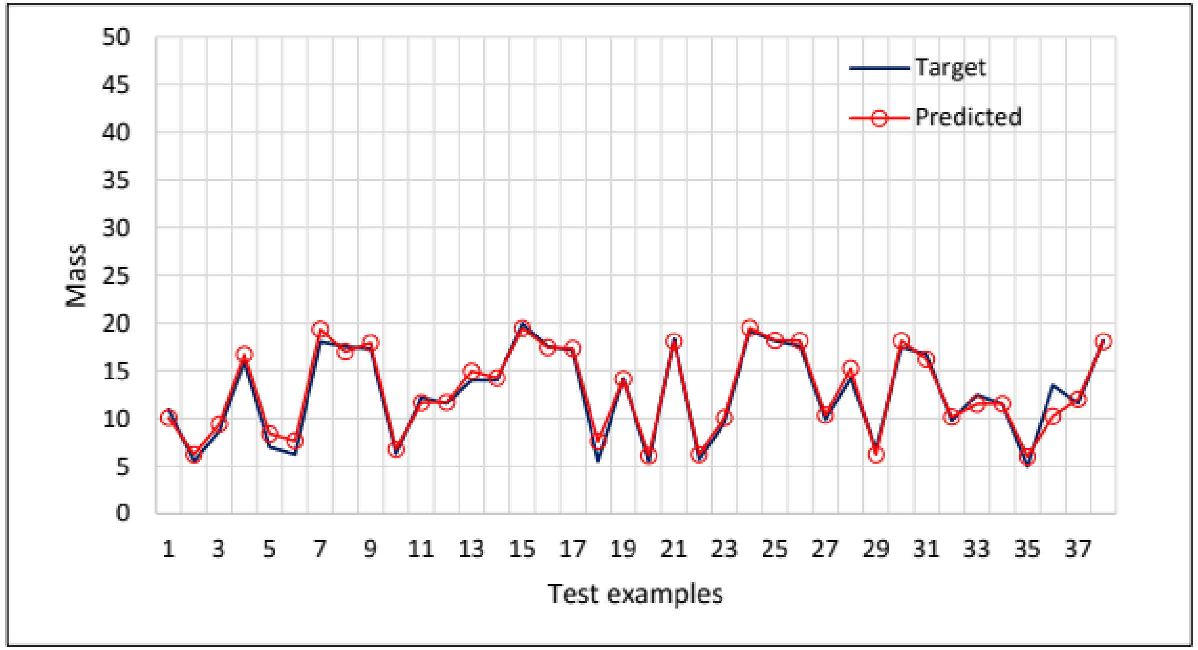
Model No.	Model Type	Mean of RMSEs Train	Mean of RMSEs Test	Mean of r Train	Mean of r Test	Std over r Train	Std over r Test	Model Complexity[1]	Selected Features[2]
1	FNT	2.0206	2.0571	0.93	0.95	0.0087	0.0383	43	1, 2, 3, 4
2		2.3891	2.3934	0.91	0.91	0.0083	0.0617	34	2, 3, 4
3		2.5491	2.2618	0.88	0.91	0.0078	0.0563	32	3, 4
4	REP-Tree	2.5751	3.1637	0.88	0.82	-	-	99	1, 2, 3, 4
5	GPR	2.9632	3.4023	0.86	0.79	-	-	-	1, 2, 3, 4
6	MLP	3.3687	3.4427	0.81	0.79	-	-	-	1, 2, 3, 4

Note: [1]Complexity is the sum of total nodes in the created tree-model. [2]Features Nos are assigned in Table 1

Ojha VK et al. (2017) Neural Computing and Applications

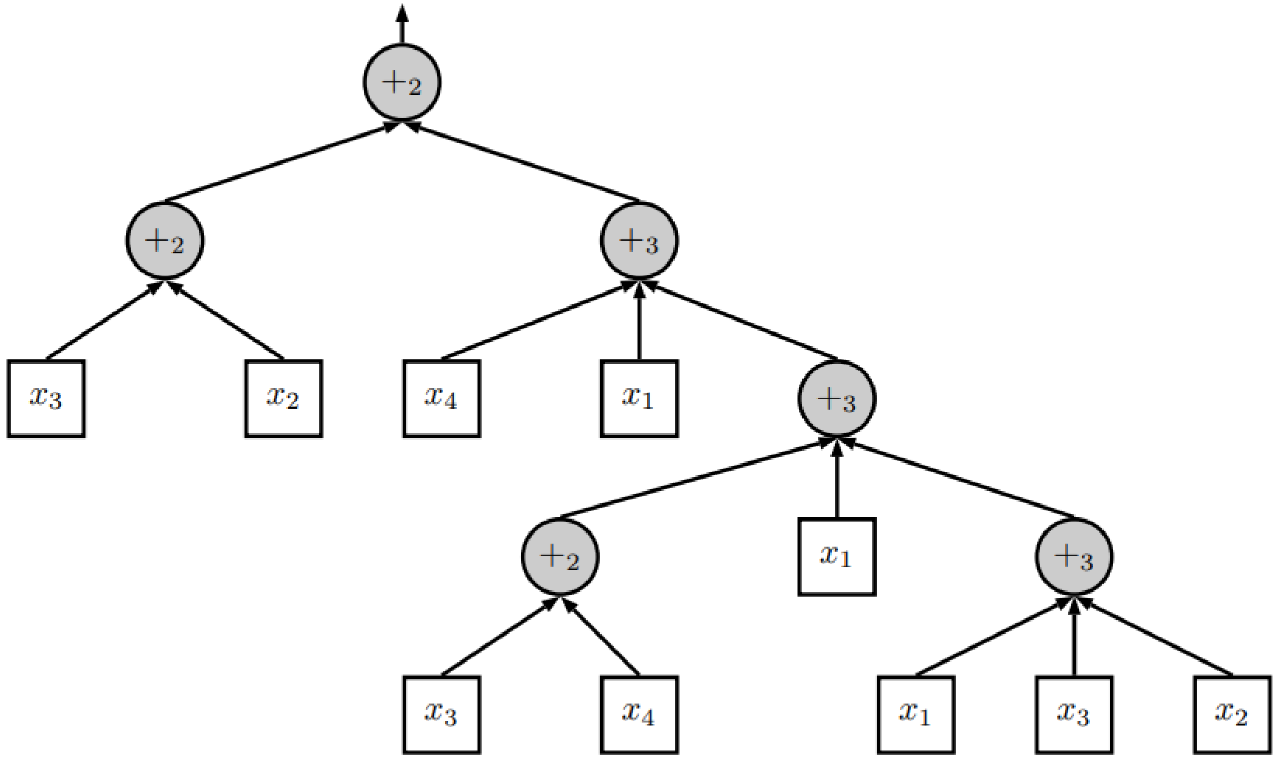
Assessing Performance Visually





Ojha VK et al. (2017) Neural Computing and Applications

Neural Tree Model for Future Use



Ojha VK et al. (2017) Neural Computing and Applications

Quality of each Feature

#	Input Features set	Selection Rate (R)	Predictability Score (P)
1	A_1= True density	0.55173	0.541356
2	A_2= d50	0.62069	0.586262
3	A_3= Granule size	1	1
4	A_4= Shoe speed	0.86207	0.92563

Ojha VK et al. (2017) Neural Computing and Applications

Quality of the Subset

#	Input Feature set	Selection Rate (R)	Predictability Score (P)
1	A_1= True density, d50, Granule size, Shoe speed	0.31035	0.969497
2	A_2= d50, Granule size, Shoe Speed	0.17242	0.941601
3	A_3= True density, Granule size, Shoe speed	0.13793	1
4	A_4= Granule size, Shoe speed	0.24138	0.979663
5	A_5= True density, d50, Granule size	0.10345	0.493741
6	A_6= d50, Granule size	0.03448	0.470451

Ojha VK et al. (2017) Neural Computing and Applications

Drug Dissolution

Ojha VK et al. (2015) International Journal of Nanomedicine

The PLGA dataset description

PLGA: poly(lactic-co-glycolic acid)

Sl No	Group name	No of features	Importance
1	Protein descriptors	85	Describes the type of molecules and proteins used
2	Formulation characteristics	17	Describe the molecular properties such as molecular weight, particle size, etc
3	Plasticizer	98	Describe the properties such as fluidity of the material used
4	Emulsifier	99	Describe the properties of stabilizing/increase the pharmaceutical product life
5	Time in days	1	Time taken to dissolve
6	% of molecules dissolved	1	Output

Abbreviations: PLGA, poly(lactic-co-glycolic acid); Sl, serial; No, number.

Ojha VK et al. (2015) International Journal of Nanomedicine

Prediction of Dissolution Rate

Selection method	Selected features	GPReg	LReg	MLP	REP	SMOReg
No selection	300	16.81	17.07	18.57	13.05	17.95
BFE	1	27.47	26.61	28.33	24.37	26.97
BFE	5	17.11	23.45	23.11	14.23	23.38
CFS	5	20.80	25.08	22.41	18.31	25.42
Class-MLP-greedy	7	17.96	25.03	22.26	14.96	25.35
BFE	10	15.93	19.98	21.00	13.19	19.53
Class-MLP-BFS	15	15.88	22.90	16.83	13.91	24.23
Wrapper-GPReg-greedy	15	14.88	20.22	15.20	13.34	20.86
Class-GPReg-BFS	16	18.46	23.07	19.71	14.19	23.69
Class-GPReg-greedy	19	15.06	19.05	15.61	14.03	19.68
Wrapper-MLP-greedy	19	16.44	24.01	20.42	14.26	24.85
Wrapper-LReg-greedy	24	15.91	17.46	17.03	13.54	18.02
BFE	Optimal*	15.71	17.85	17.82	13.90	17.88
Class-LReg-BFS	31	15.95	16.92	15.63	14.00	17.58
Class-LReg-greedy	37	16.31	17.14	16.27	14.02	17.69

Notes: Values are the average of ten RMSE. *Optimal set of attributes for the GPReg, LReg, MLP, REP and SMOReg regression models are 18, 32, 31, 31, and 30, respectively.

Abbreviations: 10-CV, ten-fold cross-validation; GPReg, Gaussian process regression; LReg, linear regression; MLP, multilayer perception; REP, reduced error pruning; SMOReg, sequential minimal optimization; No, number; BFE, backward feature elimination; CFS, correlation-based feature selection; BFS, best fit search; wrapper, wrapper feature selection; greedy, greedy search; class, classifier-based feature selection.

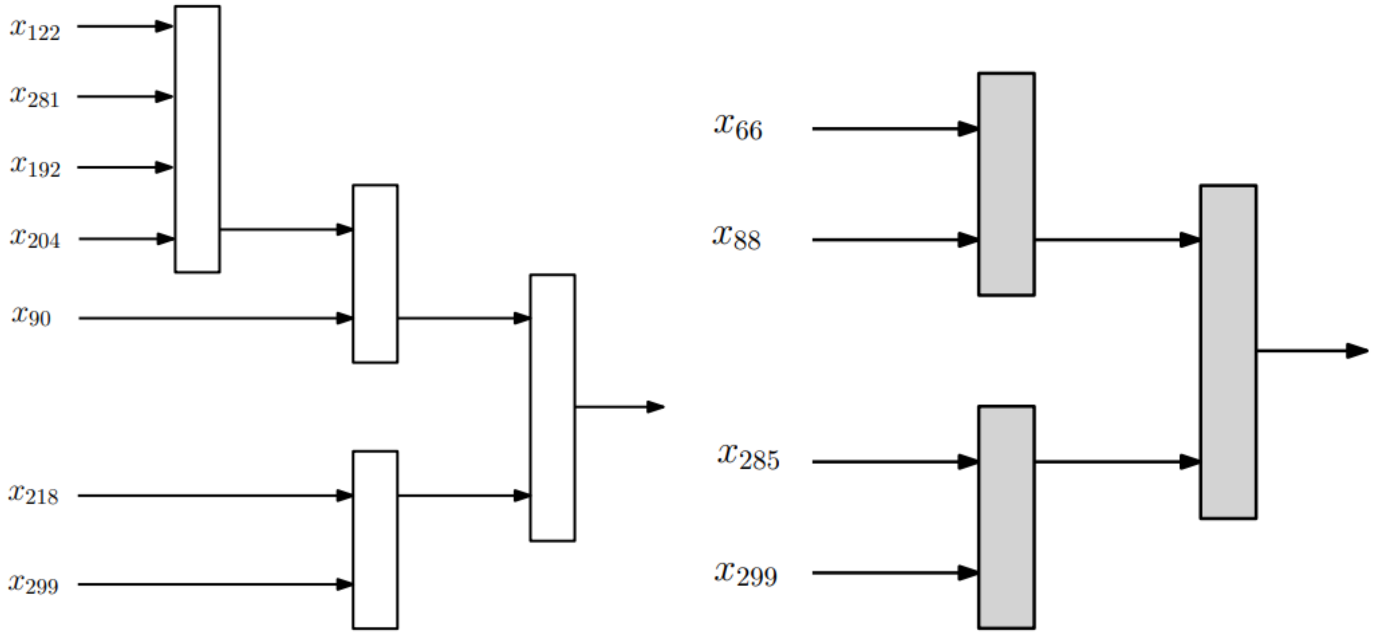
Performance of Algorithms on Prediction and Feature Selection

Algorithm	RMSE E_t	No. of features
MLP	14.3	17
HFIT	13.2	15
REP Tree	13.3	15
GPR	14.9	15
MLP	15.2	15
MLP	15.4	11
T1HFIT$^{\mathrm{M}}$	18.6	7
T2HFIT$^{\mathrm{M}}$	15.2	4

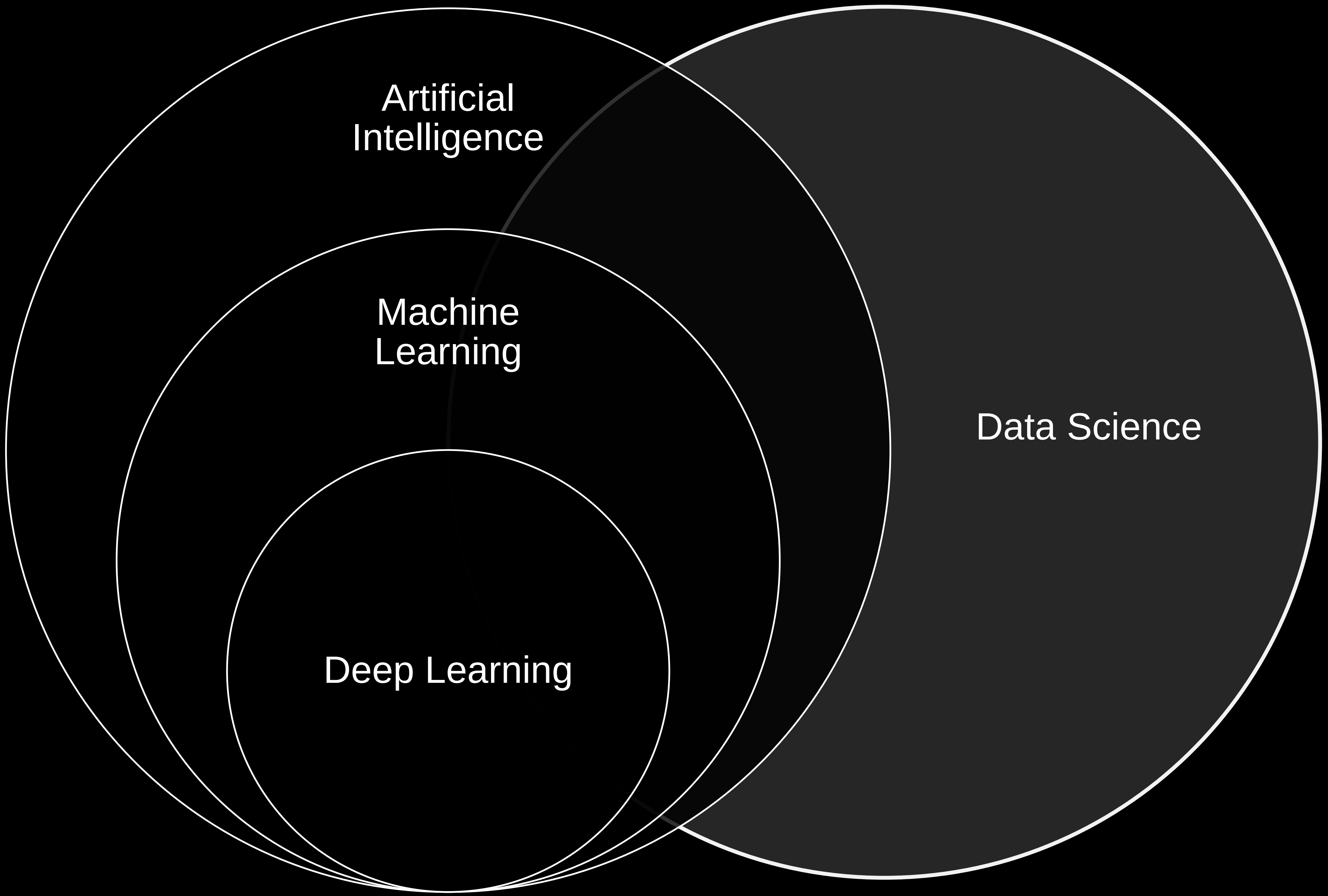
Ojha VK et al. (2018) IEEE Transactions on Fuzzy Systems

Fuzzy Tree Model for Future Use

(Type 1 T1HFIT^{M} and Type 2 T2HFIT^{M})



(a) T1HFIT^{M}: $E_t = 18.66$ (b) T2HFIT^{M}: $E_t = 15.25$


Artificial Intelligence
Machine Learning
Deep Learning
Data Science