Machine Learning

Artificial Intelligence

School of Computing
Universiti Teknologi Malaysia

Machine Learning

How do we teach a machine to learn?

Content

- Machine learning process
- From data to features
- Learning types
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning
- Machine learning techniques
  - Clustering, classification, regression
- Evaluation metrics
- Computational intelligence

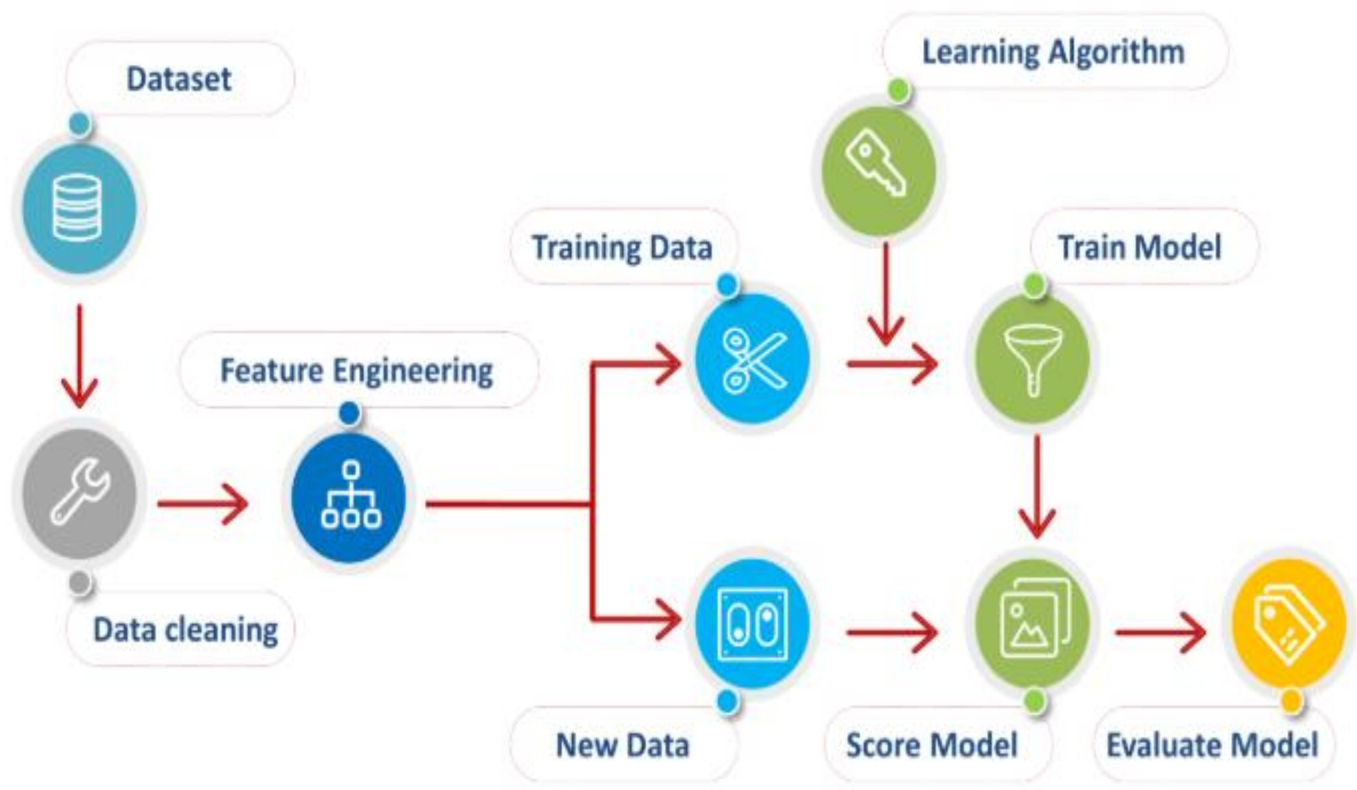
Machine Learning

Learning Process

Machine Learning Process

What is Machine Learning?

Computational learning using algorithms to learn from and make predictions on data.



How machine learns?

- Machine learning utilizes techniques of understanding **patterns** from input **data**
- Data is used for **training** by learning algorithm to **learn** and develop a specific computational **model**
- **Pre-processing** involves extracting important features of data
- **Model** refers to predefined prototype problem solver, produced based on patterns of input data

Machine Learning

From Data to Features

Some real data

- *In the 1920's, botanists collected measurements on the*
  1) sepal length
  2) sepal width
  3) petal length
  4) petal width

of 150 iris, 50 from each of three species (setosa, versicolor, virginica)
The measurements became known as ***Fisher's iris data***

Iris data

https://www.kaggle.com/arshid/iris-flower-dataset

Data Sources

IRIS.csv 5 columns

About this file

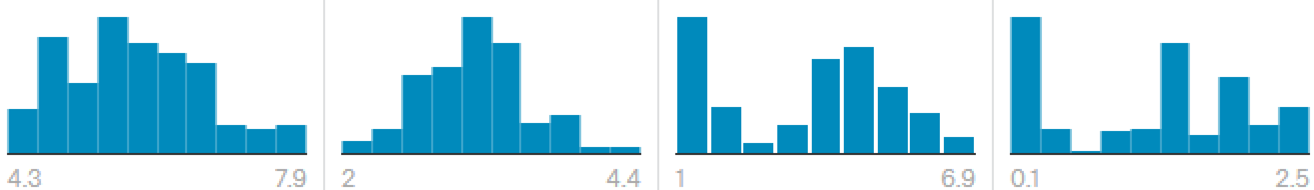
The dataset is a CSV file which contains a set of 150 records under 5 attributes - Petal Length, Petal Width, Sepal Length, Sepal width and Class(Species)

Columns

- # sepal_length
- # sepal_width
- # petal_length
- # petal_width
- A species

IRIS.csv (4.51 KB) 5 of 5 columns Views

	# sepal_length	# sepal_width	# petal_length	# petal_width	A species



					Iris-setosa 33%
					Iris-versicolor 33%
					Other (1) 33%
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.8	4	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa

Data to Features

- Collection of data observations and their attributes/features

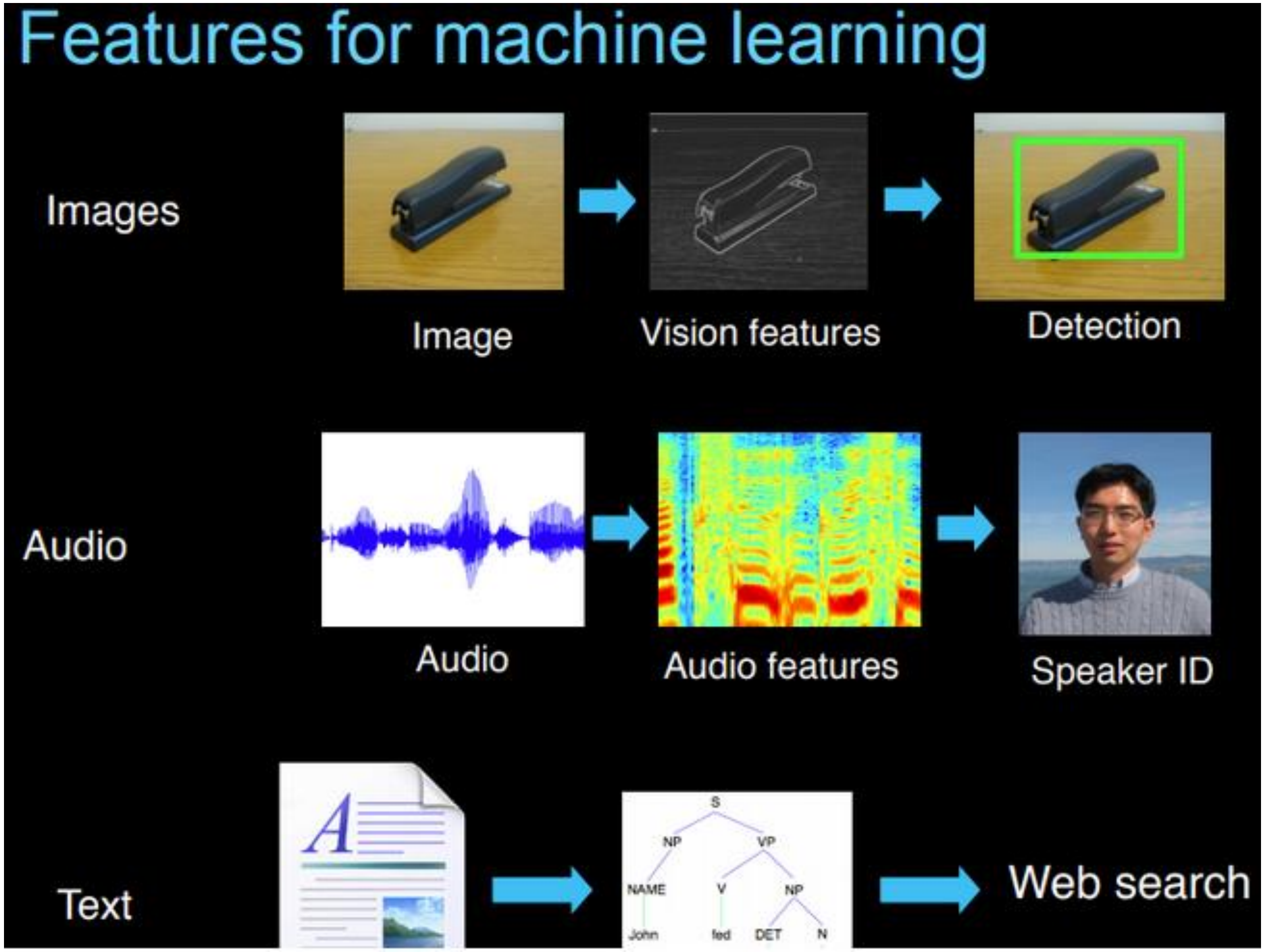
▪ **Attribute is the properties or characteristics of the data object**
▪ **Attribute is aka feature / field / characteristic / variable**

# sepal_length	# sepal_width	# petal_length	# petal_width	species
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
				Iris-

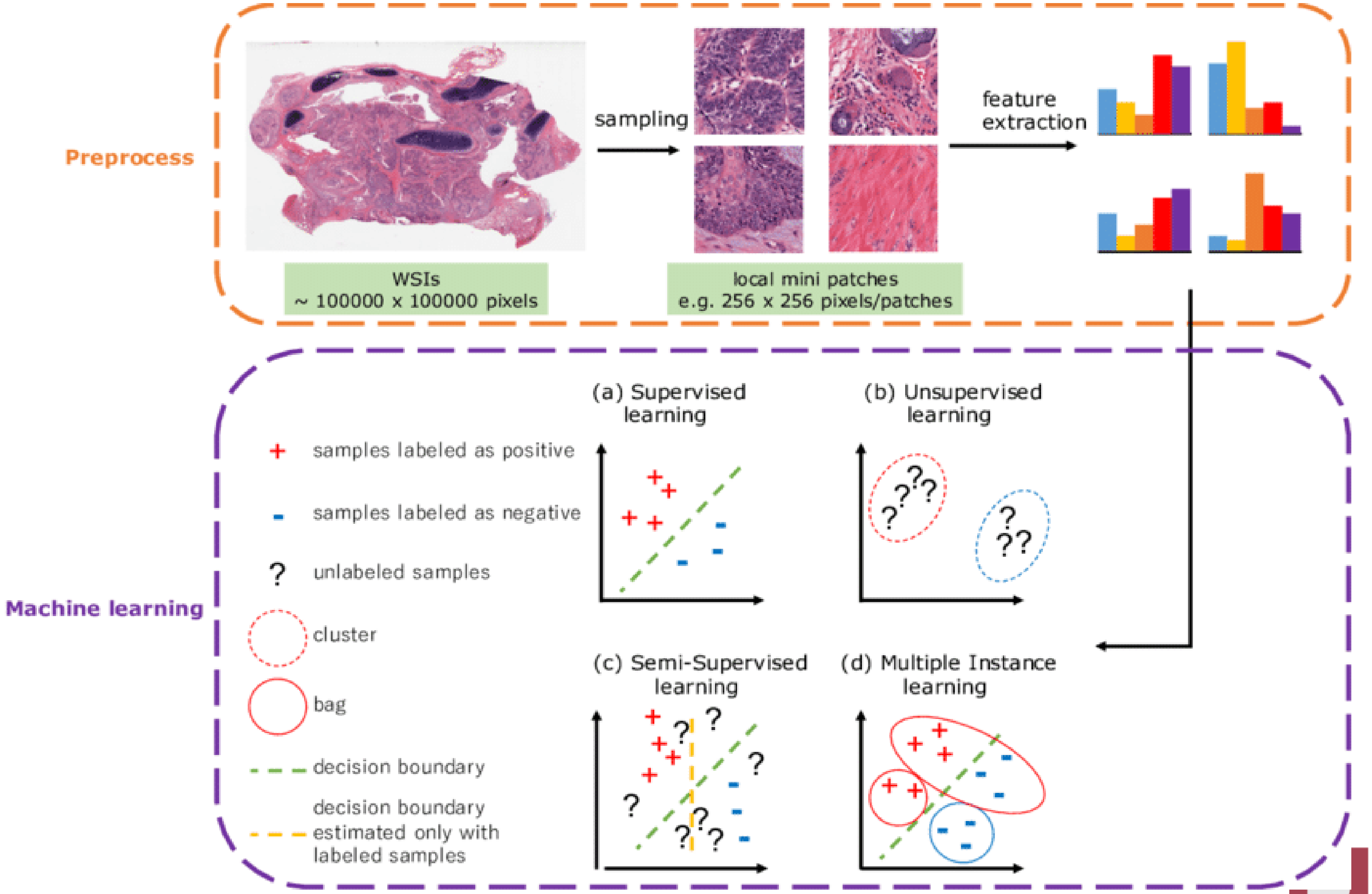
Observation/instances/examples/samples/objects/records

Class

Features can be from many forms of data; images, audio or text



Real world example: Steps of ML in digital pathological image analysis



Machine Learning

Learning Types

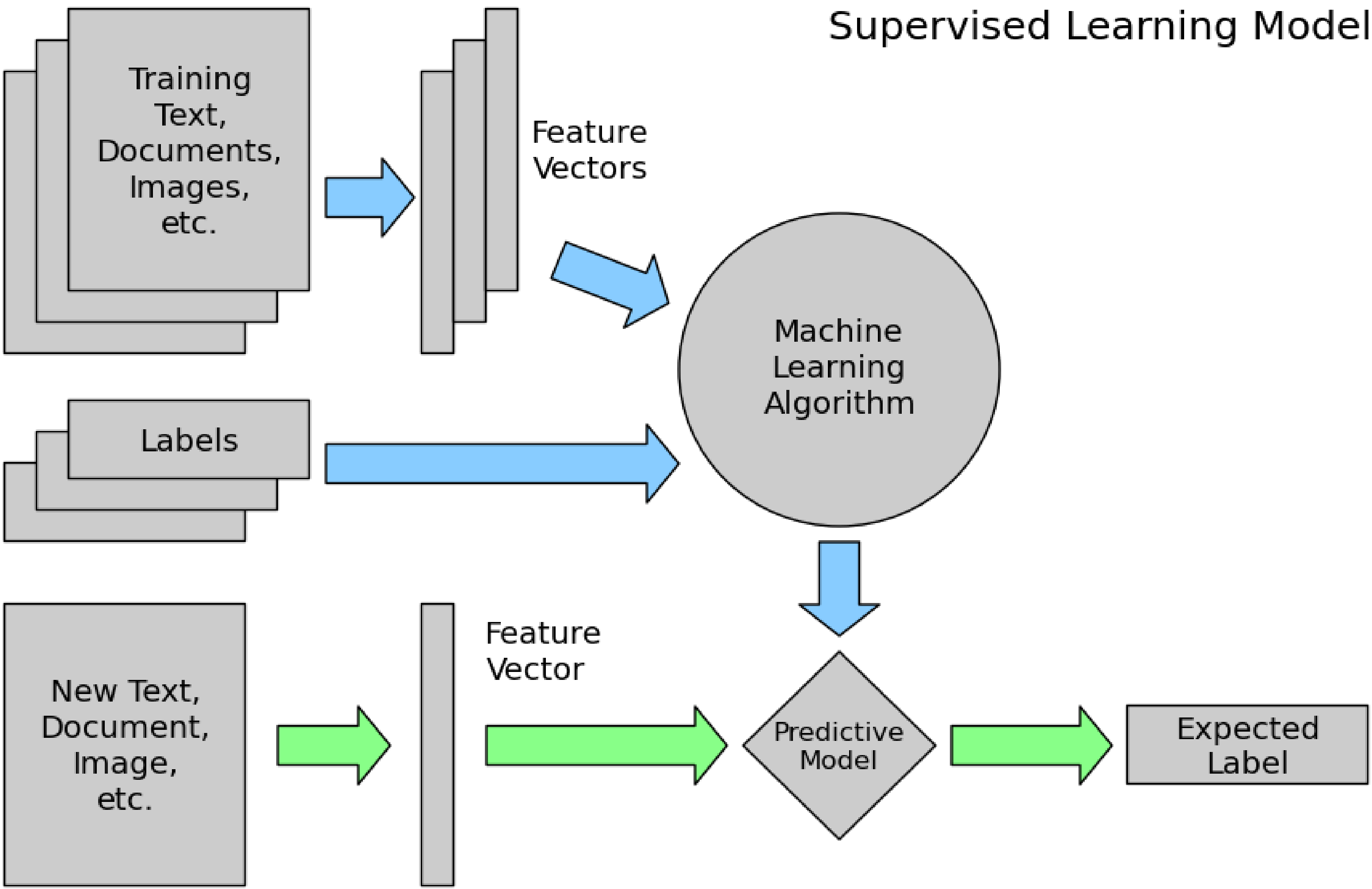
Learning types

- Supervised learning
  – The correct classes of the training data are known
- Unsupervised learning
  – The correct classes of the training data are NOT known
- Reinforcement learning
  – Machine learn its behavior based on feedback from environment

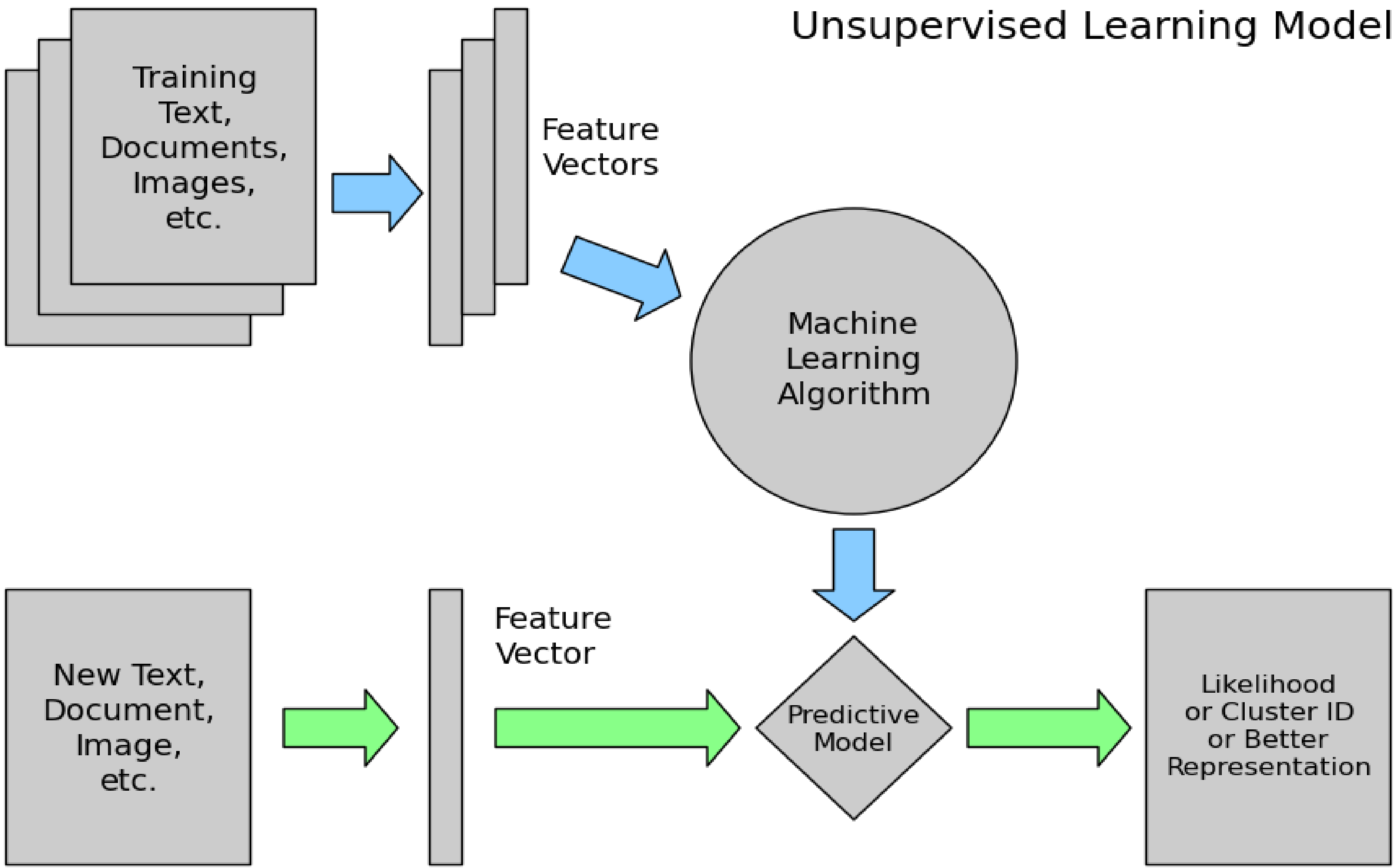
Learning types

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	The machine learns by using labelled data	The machine is trained on unlabelled data without any guidance	An agent interacts with its environment by producing actions & discovers errors or rewards
Type of problems	Regression & Classification	Association & Clustering	Reward based
Type of data	Labelled data	Unlabelled data	No pre-defined data
Training	External supervision	No supervision	No supervision
Approach	Map labelled input to known output	Understand patterns and discover output	Follow trail and error method
Popular algorithms	Linear regression, Logistic regression, Support Vector Machine, KNN, etc	K-means, C-means, etc	Q-Learning, SARSA, etc

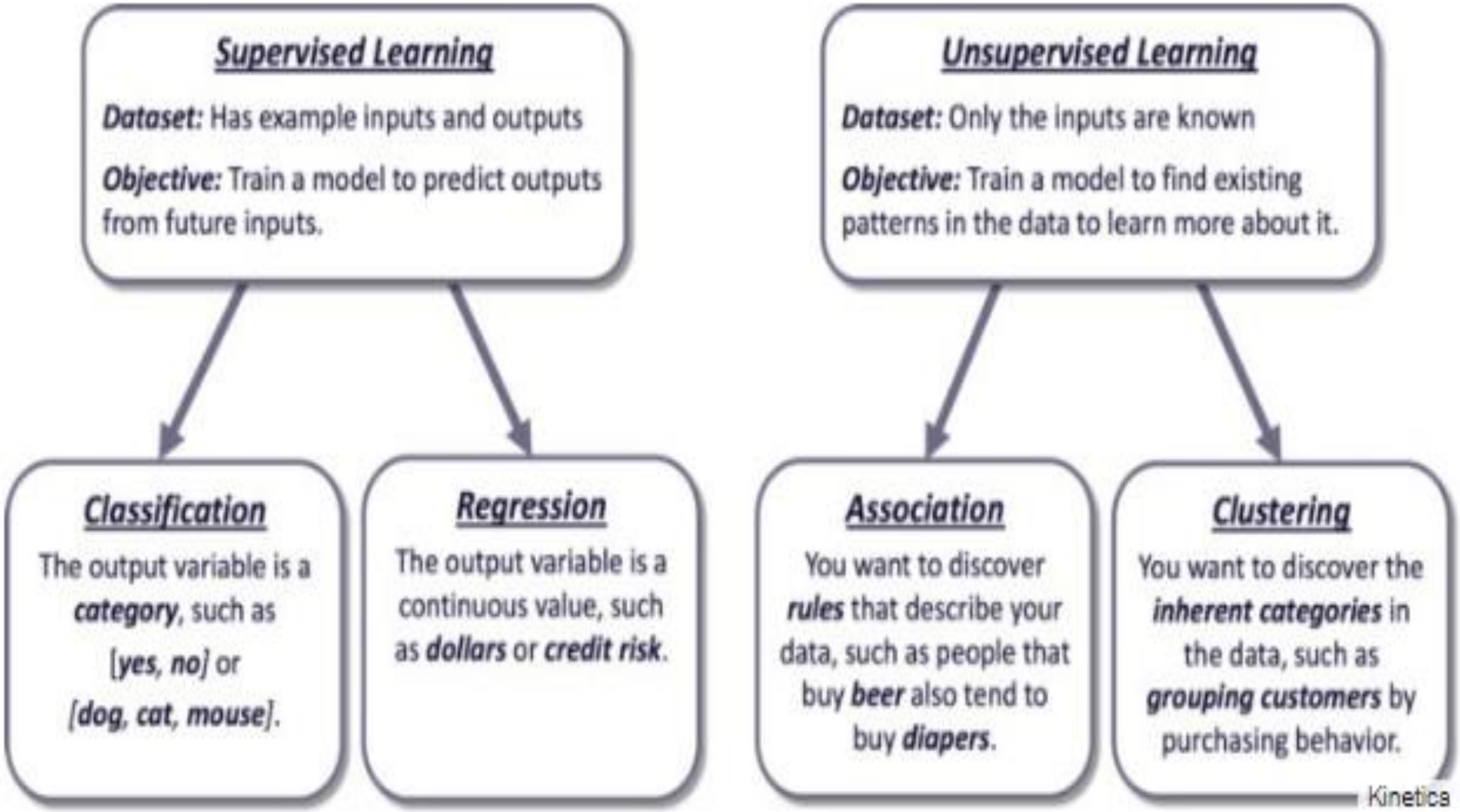
1. Supervised Learning



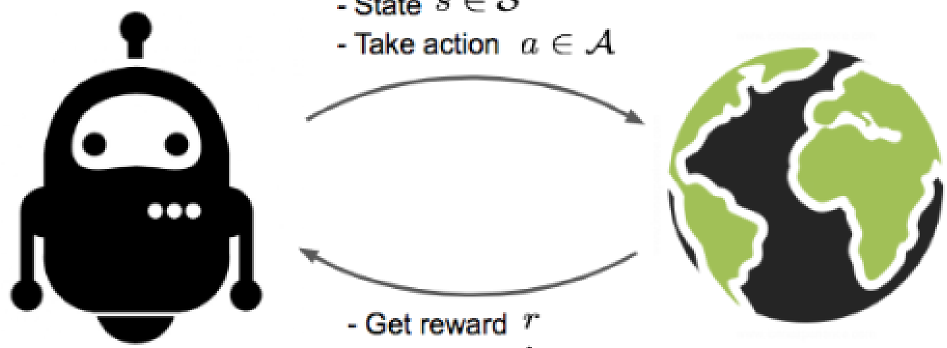
2. Unsupervised Learning



Supervised vs unsupervised



3. Reinforcement learning

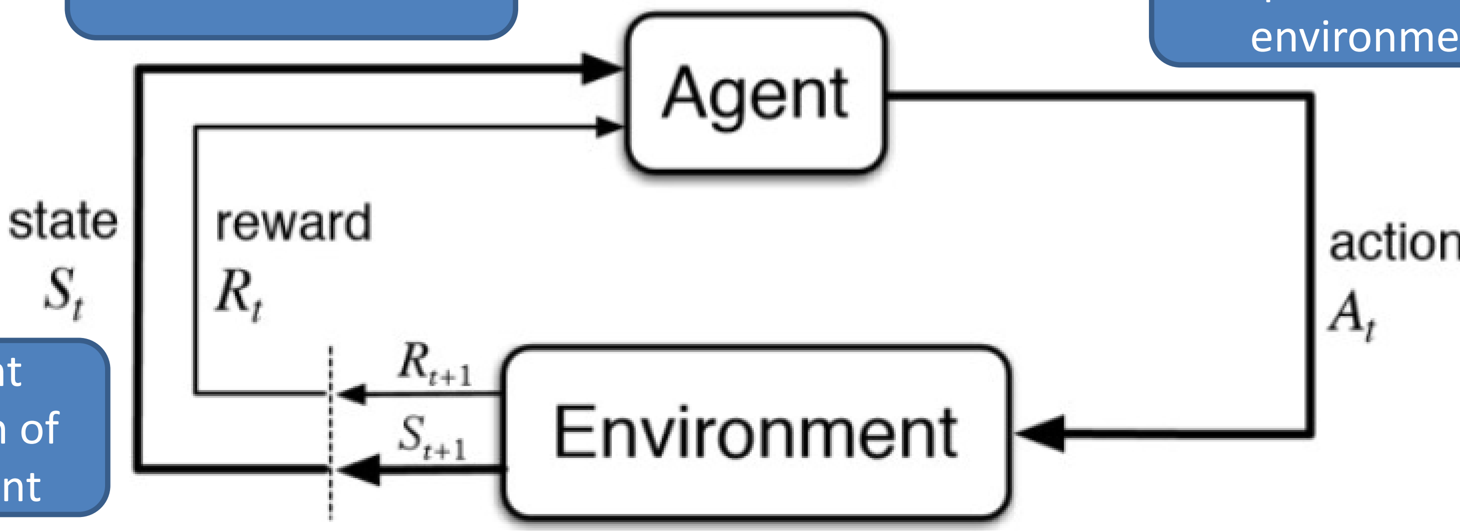


- Reinforcement learning is **a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones**. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error
- An agent interacts with the environment trying to take smart actions to maximize cumulative rewards
- The environment is formulated using Markov decision process to search for solutions

Markov Decision Process (MDP)

Feedback from the environment

Action that agent performs in environment



Current situation of the agent

Physical world in which the agent operates

Markov Decision Process (MDP)

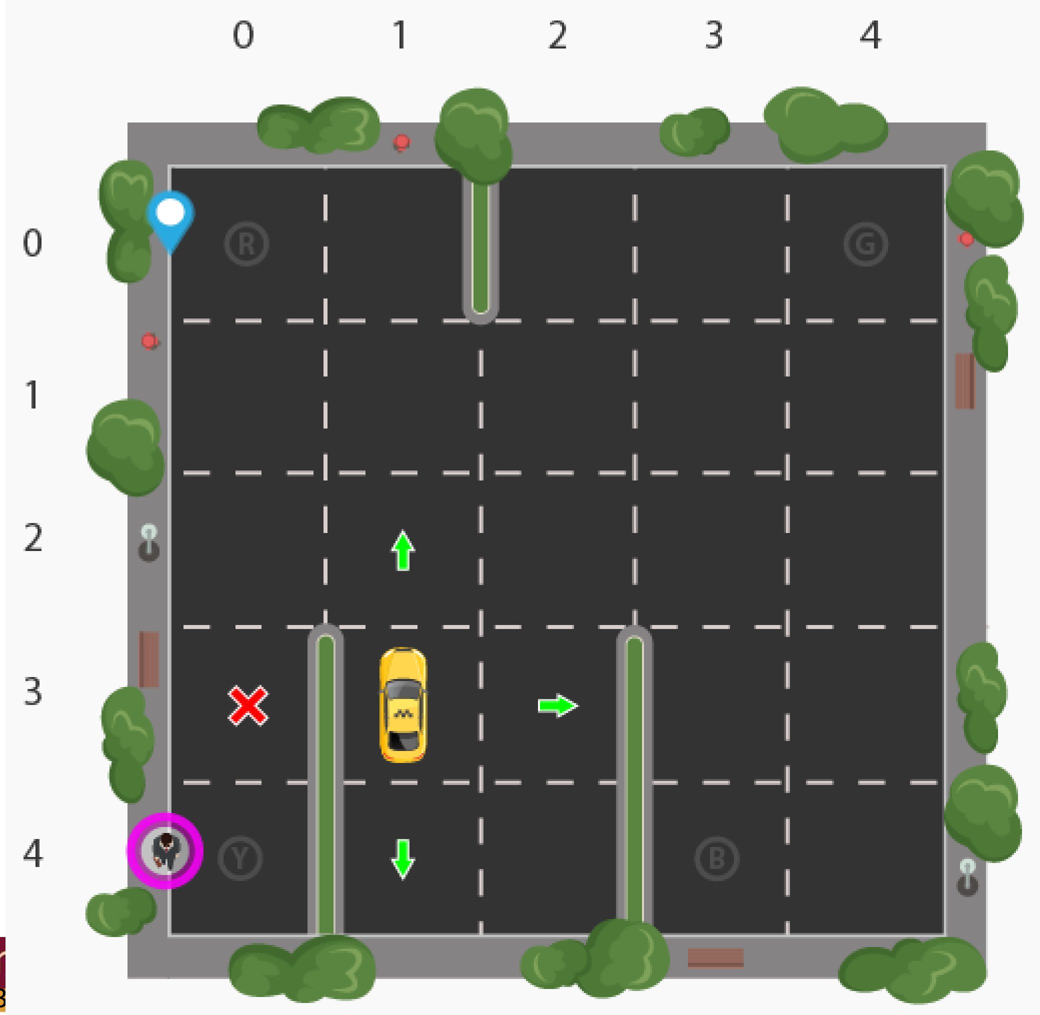
- The agent is acting in an **environment**
- How the environment reacts to certain actions is defined by a **model** which we may or may not know
- The agent can stay in one of the many **states** S of the environment and choose to take one of the many **actions** A so that it can switch from one state to another
- Which state the agent will arrive is determined by transition **probabilities** P between states
- Once an action is taken, a **reward** R is delivered by the environment as feedback

Example of RL for self-driving cab

- Self-driving cab (agent) must drop off passenger at the right location (R,G,B,Y)
- Passenger is at Y and wants to go to R
  - **Rewards**; successful dropoff by agent will get +reward, otherwise agent will be penalized/-reward
  - **State space;** all possible situations the cab can be in

    5 x 5 x 5 x 4 = 500 possible states

    5x5 (grid), 5 passengers location (4 in the cab, 1 waiting), 4 locations (R,G,B,Y)
  - **Action space**; six possible actions to be south, north, east, west, pickup and dropoff



RL in real world

- Games
  - AlphaGo, AlphaGo Zero etc
- Robotic
  - (remember Spot from BostonDynamics?)
- Traffic light control
- Web system configuration
- And many more

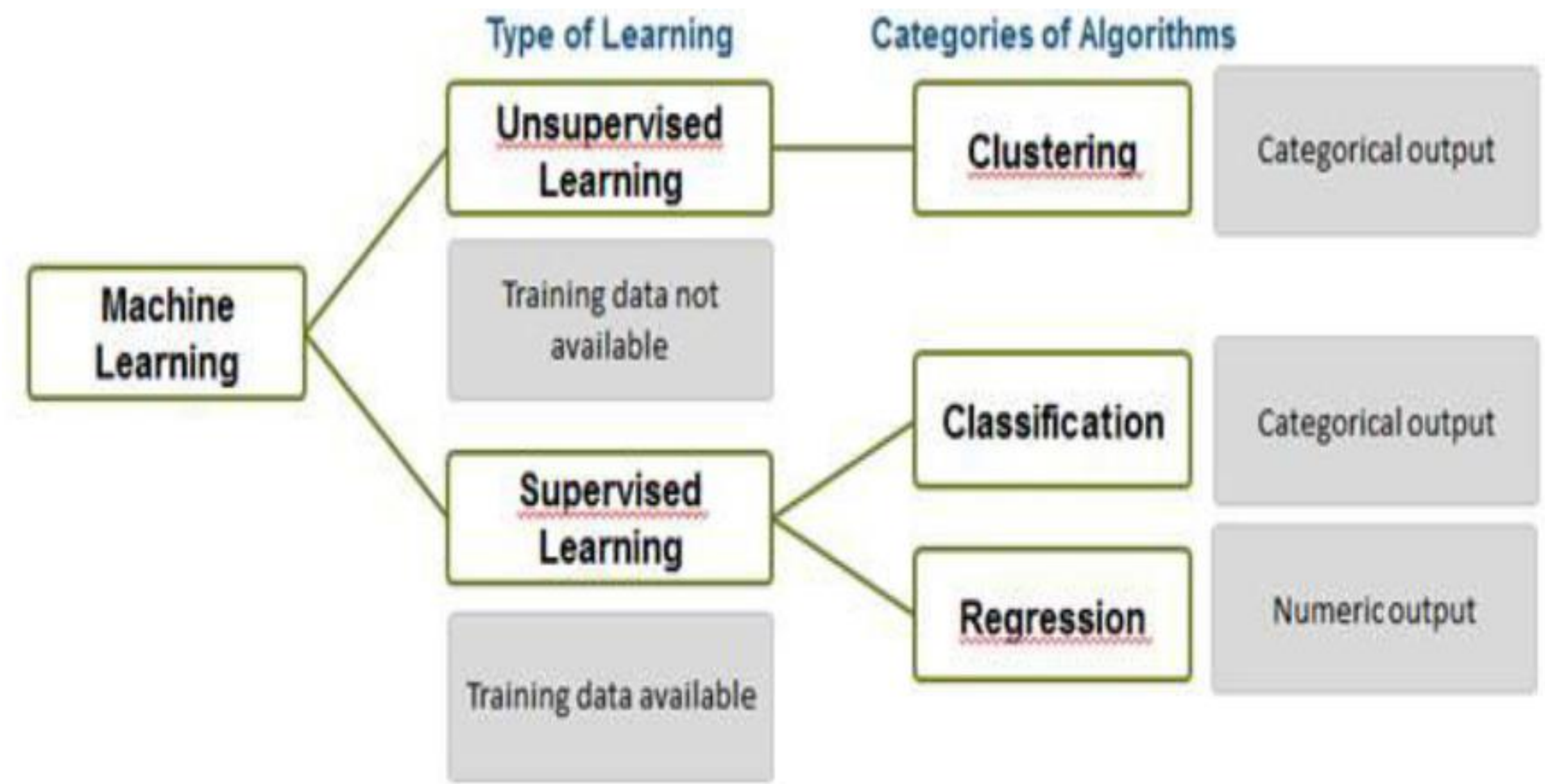
Machine Learning

Learning Techniques

Machine learning techniques

- Classification
  - Predict class from observation
- Clustering
  - Group observations into ‘’meaningful” groups
- Regression
  - Predict value from observations

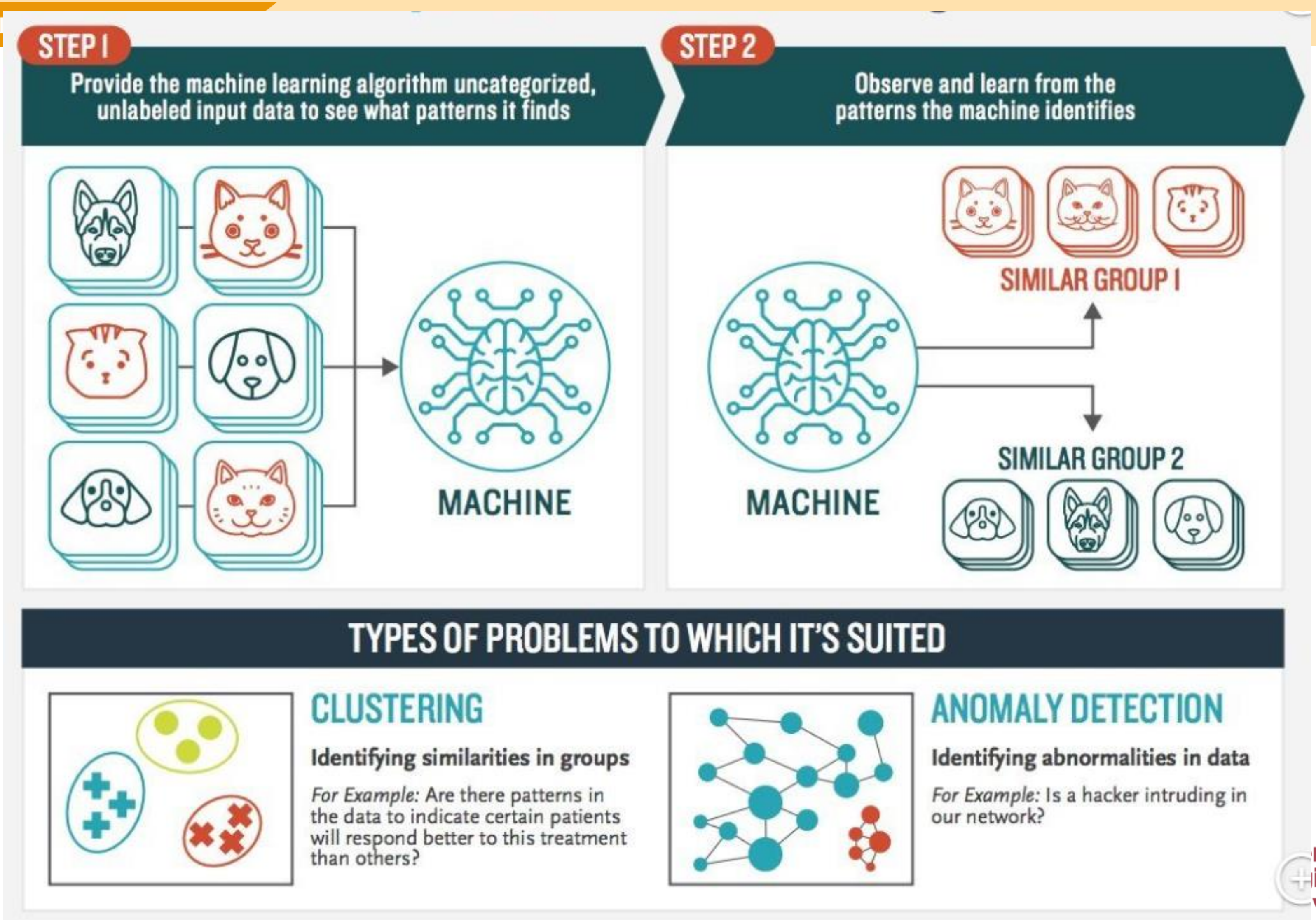
Classification and clustering



Clustering

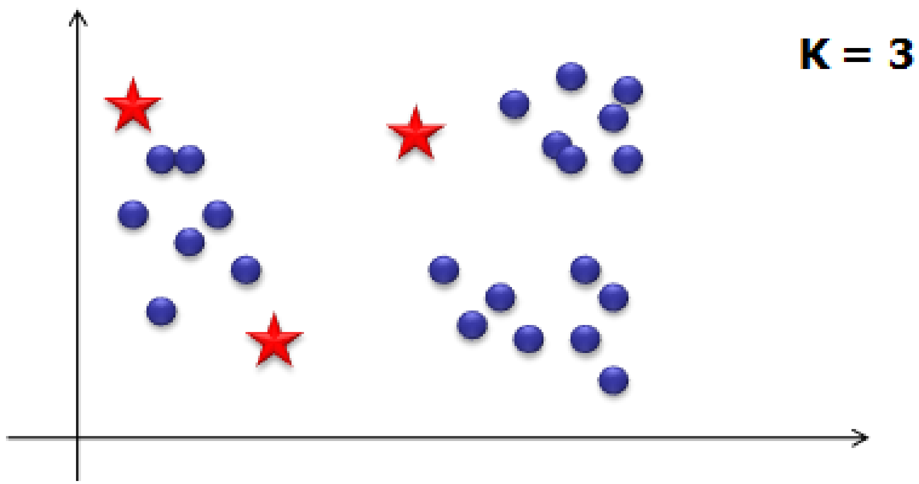
- Clustering is **identifying similar groups** of data in a dataset
- We assume that observations in each group are comparatively **more similar** to observations of that group compared to those of the other groups
- **A cluster refers to a collection of data points aggregated together because of certain similarities**
- E.g. try to understand customers better so that we can market our products to them better. How do we start?
- E.g. try to find your close-knit groups of friends on Facebook? How will you identify them?

How clustering works



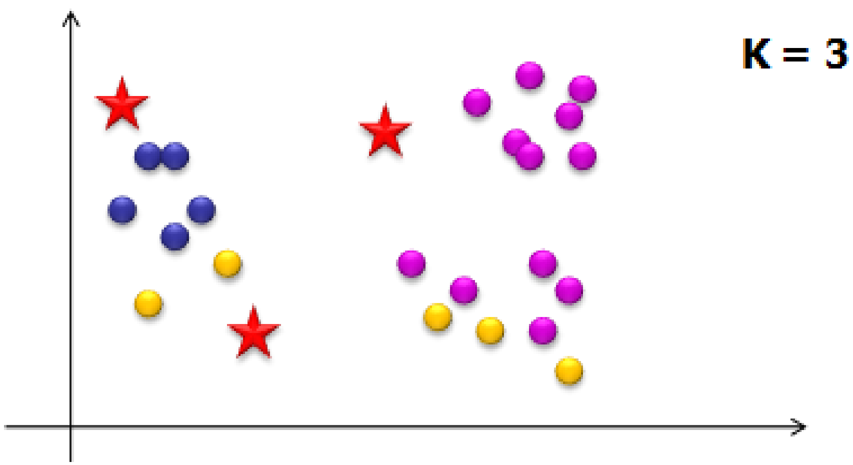
Clustering (*k*-mean algorithm)

- **Algorithm description**
  - Choose the number of clusters, K
  - Randomly choose initial positions of K centroids
  - Assign each of the points to the "nearest centroid" (depends on distance measure)



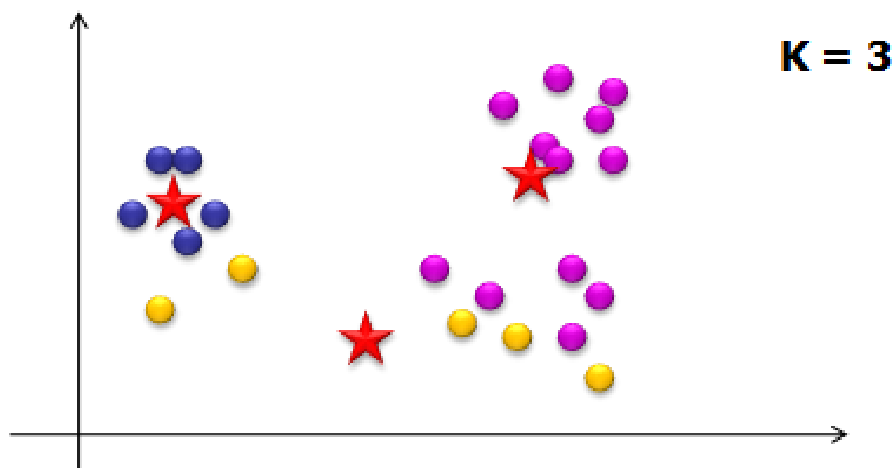
Clustering (*k*-mean algorithm)

- **Algorithm description**
  - Choose the number of clusters - K
  - Randomly choose initial positions of K centroids
  - Assign each of the points to the "nearest centroid" (depends on distance measure)



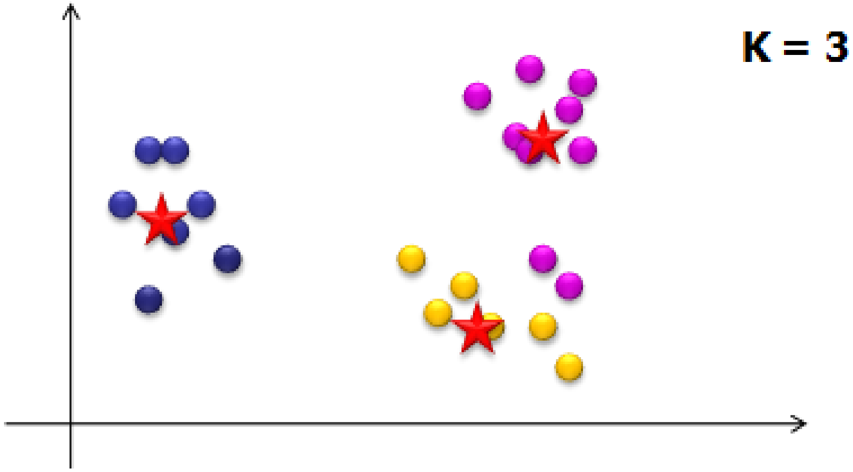
Clustering (*k*-mean algorithm)

- **Algorithm description**
  - Choose the number of clusters - K
  - Randomly choose initial positions of K centroids
  - Assign each of the points to the "nearest centroid" (depends on distance measure)
  - Re-compute centroid positions
  - If solution converges → Stop!



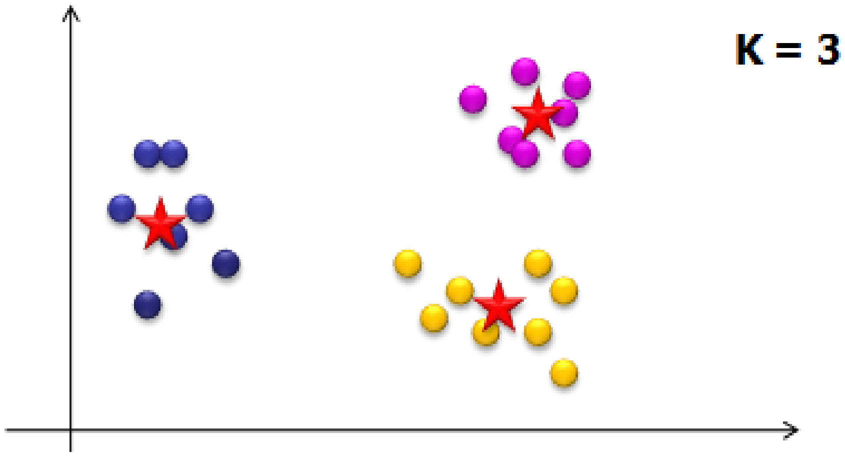
Clustering (*k*-mean algorithm)

- **Algorithm description**
  - Choose the number of clusters - K
  - Randomly choose initial positions of K centroids
  - Assign each of the points to the "nearest centroid" (depends on distance measure)
  - Re-compute centroid positions
  - If solution converges → Stop!



Clustering (*k*-mean algorithm)

- **Algorithm description**
  - Choose the number of clusters - K
  - Randomly choose initial positions of K centroids
  - Assign each of the points to the "nearest centroid" (depends on distance measure)
  - Re-compute centroid positions
  - **If solution converges → Stop!**

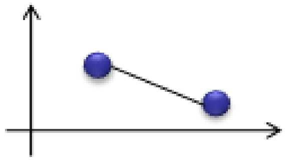


Distance measure for *k*-mean

- How should we choose K?

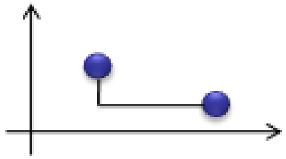
- What type of distance measures can we use, and how to choose between them?
  - Euclidean

    $((x_2 - x_1)^2 + (y_2 - y_1)^2)^{0.5}$

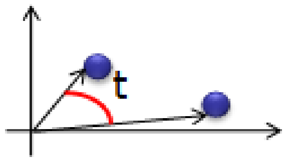


  - Sum of absolute differences

    $|x_2 - x_1| + |y_2 - y_1|$

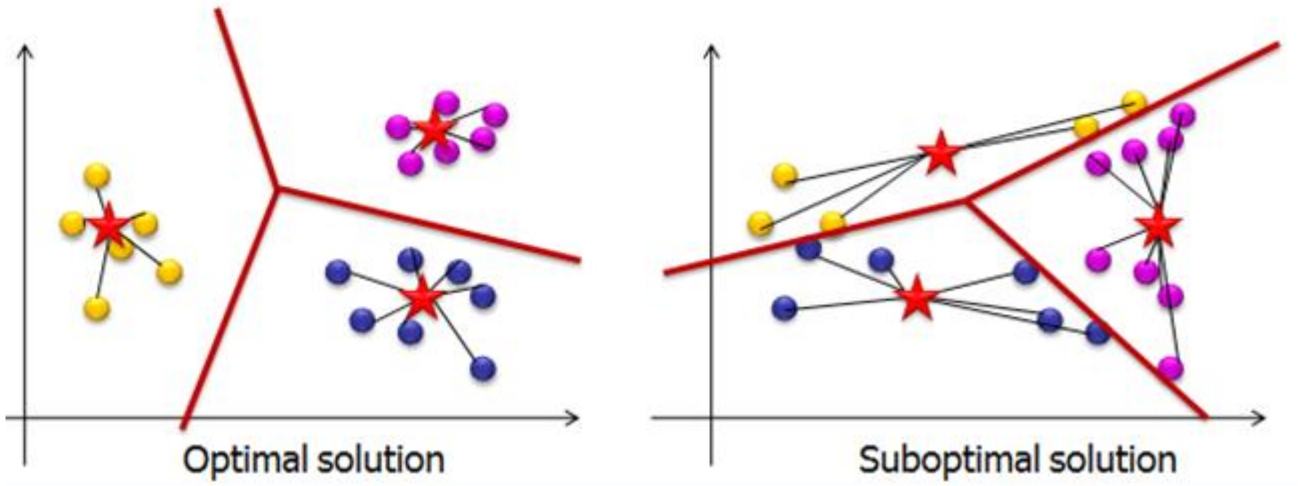


  - 1 – Cos(t)



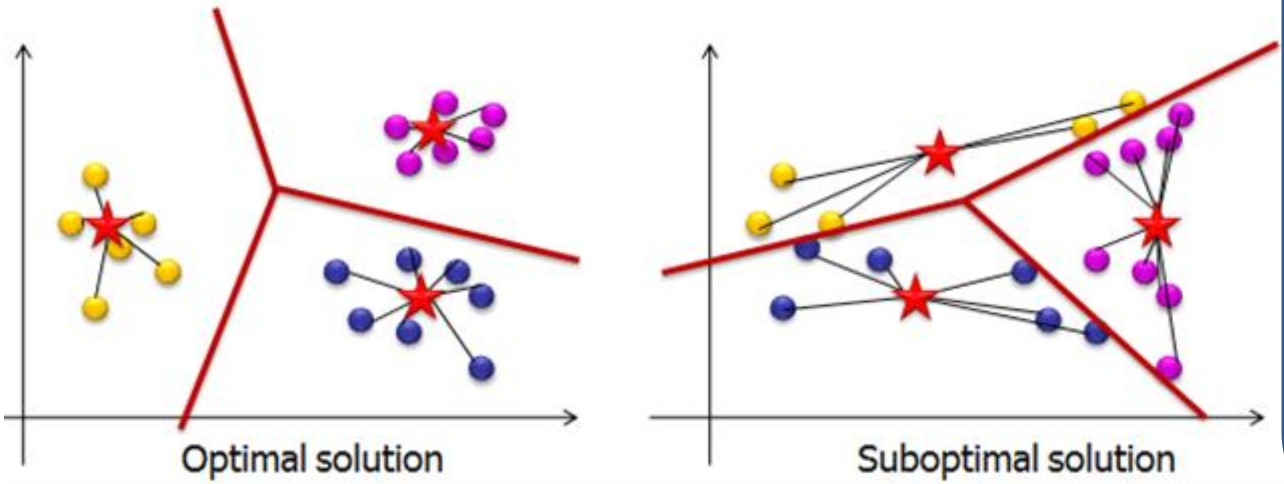
  - And more…

Clustering (*k*-mean algorithm)



- Find similarities among all observations such that similar observations can be put in clusters
- Computing distance from each observation/data point to centroids to obtain well-represented clusters.
- Why are they optimal and suboptimal solution? Can you guess?

Clustering (*k*-mean algorithm)



Centroid == mean i.e. averaging the data

Data points are allocated to the nearest centroid/cluster as to keep the centroids as small as possible

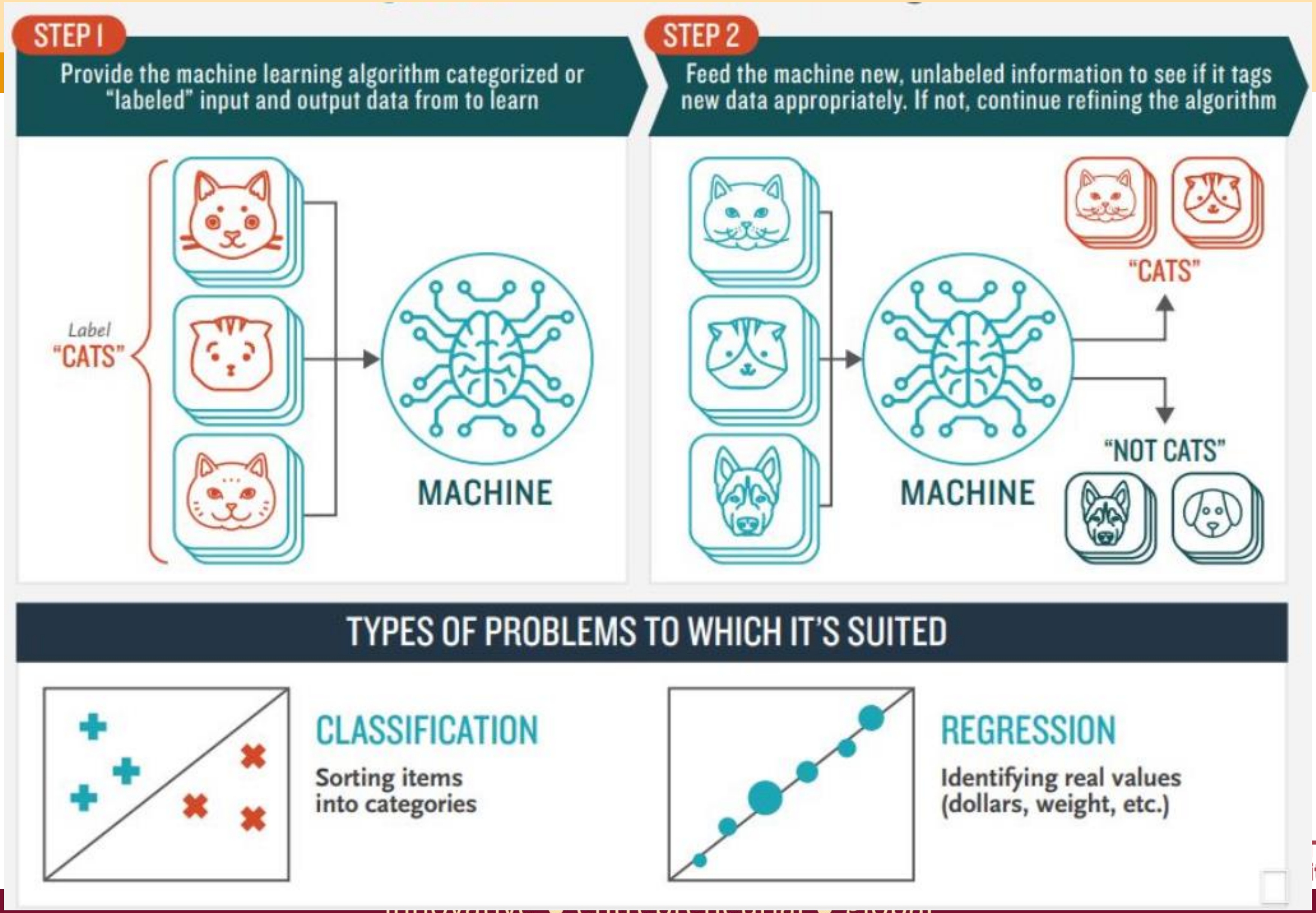
Optimal solution means it is easier to distinguish between each cluster i.e. better boundary separation

- Find similarities among all observations such that similar observations can be put in clusters
- Computing distance from each observation/data point to the centroids to obtain well-represented clusters.
- Why are they optimal and suboptimal solution? Can you guess?

Classification

- Classification refers to predicting the class of given observations/data points
- Class means targets/labels or sometimes categories
- In classification, we are mapping the input data to specific type of class/category

How classification works



Classification

- Classification of waste
- You know beforehand each bin category/class of the waste
- How do you decide which waste goes to which bin?
- We learn from the waste i.e. from the data points, then we can decide to which class it belongs to

Classification

- The data that gets input to the classifier may contains four measurements for waste e.g. 1)what material it is made from, 2)size, 3)weight and its 4)biodegradable level
- The job of the classifier then is to output the correct waste type for every input

Classification algorithms

- Decision tree
- Naïve Bayes
- ANN, now evolving into Deep Learning
- Random Forest
- K-Nearest Neighbour
- Support Vector Machine (SVM)
- And many more!

Machine Learning

Performance Evaluation

How good is the learning model?

- The aim of a model is to be able to produce good results
- Employ another set of data called **testing data** to **evaluate** the learning **model**
  – Can be completely new
  – Blinded dataset from the same resource
  – Smaller part of training data previously separated
  – Rule of thumb for train-test split (holdout method); 70%:30, 80%:20%, 90%:10%

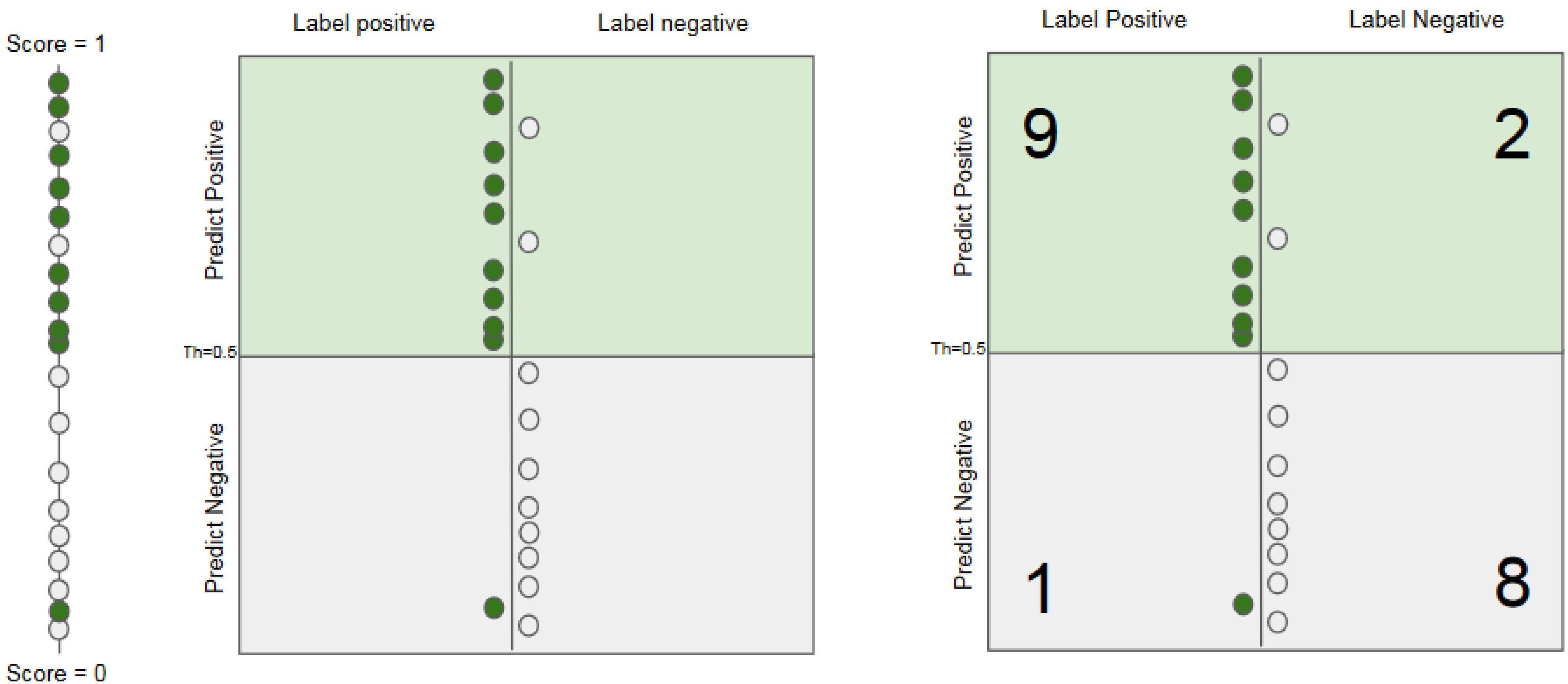
Performance measures / Evaluation metrics

- Confusion matrix to evaluate the performance of ML methods

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

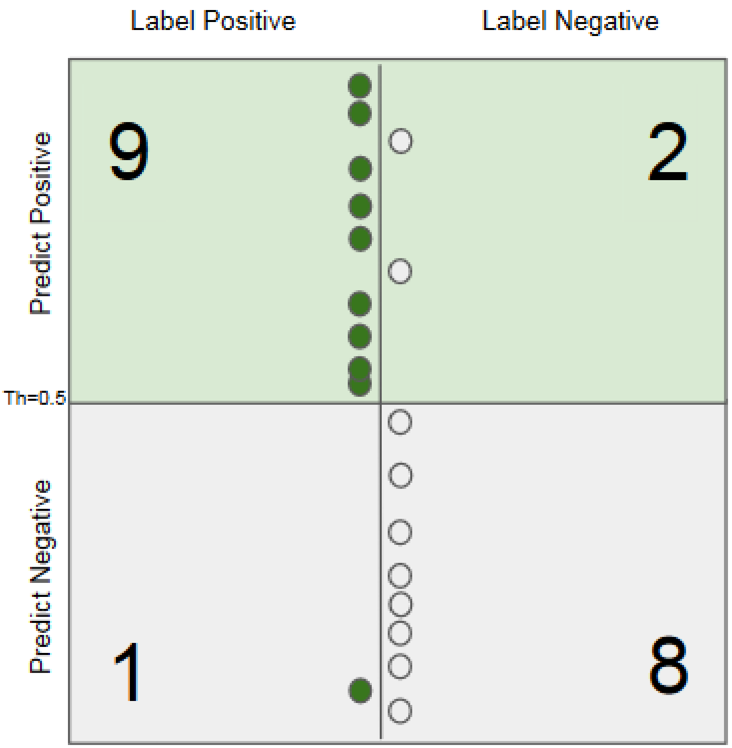
- E.g. of performance measures
  - % game won
  - % correctly classified images
  - % correctly identified text

Confusion matrix



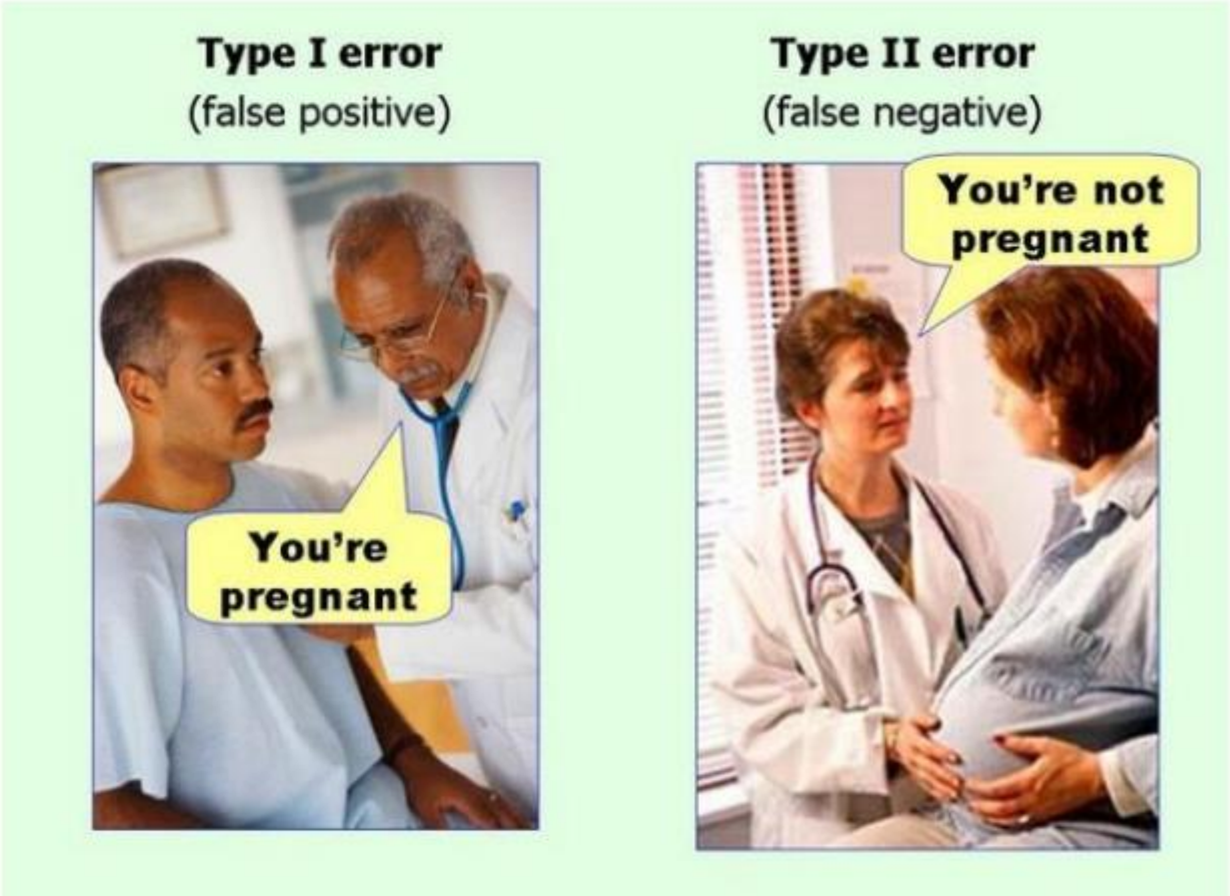
●	Positive labelled example
○	Negative labelled example

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)



- TP = 9 (positive observation correctly identified as positive)
- TN = 8 (negative observation correctly identified as negative)
- FP = 2 (positive that is *false* i.e. negative observation incorrectly identified as positive)
- FN = 1 (negative that is *false* i.e. positive observation incorrectly identified as negative)

Type I error = FP and Type II error = FN



- Lets say positive test result means 'is pregnant'

Performance measures from confusion matrix

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

- Accuracy = TP+TN / (TP+TN+FP+FN) i.e. 1 – error
- Precision = TP/(TP+FP)
- Recall = TP/(TP+FN)
- The closer the score to 1, the better the classifier model
- **High recall, low precision:** This means that most of the positive examples are correctly recognized (low FN) but there are a lot of false positives.
- **Low recall, high precision:** This shows that we miss a lot of positive examples (high FN) but those we predict as positive are indeed positive (low FP)

Activity 2.0: Confusion matrix

Sepal length	Sepal width	Petal length	Petal width	Actual Class	Predicted Class
4.4	3.2	1.3	0.2	1	1
5	3.5	1.6	0.6	1	0
5.1	3.8	1.9	0.4	1	0
4.8	3	1.4	0.3	0	0
5.1	3.8	1.6	0.2	0	0
4.6	3.2	1.4	0.2	0	0
5.3	3.7	1.5	0.2	1	1
5	3.3	1.4	0.2	1	1
7	3.2	4.7	1.4	0	0
6.4	3.2	4.5	1.5	1	1
6.9	3.1	4.9	1.5	1	1
5.5	2.3	4	1.3	0	1
6.5	2.8	4.6	1.5	0	1
5.7	2.8	4.5	1.3	0	0
6.3	3.3	4.7	1.6	0	0
4.9	2.4	3.3	1	1	1
6.6	2.9	4.6	1.3	1	1
5.2	2.7	3.9	1.4	0	1
5	2	3.5	1.4	0	0
5.1	2.3	3.6	1.2	0	1

- Let positive be 1, and negative is 0;
- Given the Iris dataset as in the table, quantify TP, TN, FP and FN
- Determine the accuracy, precision and recall

Activity 2.0 Answer: Confusion matrix

	Predicted: 0	Predicted: 1
Actual: 0	7 (TN)	4 (FP)
Actual: 1	2 (FN)	7 (TP)

	Predicted: 1	Predicted: 0
Actual: 1	7 (TP)	2 (FN)
Actual: 0	4 (FP)	7 (TN)

Let positive be 1, and negative is 0;

- **TP=7, TN=7, FN=2, FP=4**
- Accuracy = TP+TN / (TP+TN+FP+FN)

  =7+7/ (7+7+2+4) = 0.7
- Recall = TP/(TP+FN) = 7/(7+2)=0.78
- Precision = TP/(TP+FP)= 7/(7+4) = 0.64

How real world use ML?

https://towardsdatascience.com/real-world-examples-of-applied-machine-learning-from-ai-conference-4d4678700c6

Real world examples

- Uber COTA (Customer Obsession Ticket Assistant)

http://w4nderlu.st/assets/javascript/ViewerJS/#http://w4nderlu.st/content/3-publications/cota-improving-the-speed-and-accuracy-of-customer-support-through-ranking-and-deep-networks/cota-o-reilly-ai-conference-2018.pdf

- AirBNB ; categorizing listing photos

https://medium.com/airbnb-engineering/categorizing-listing-photos-at-airbnb-f9483f3ab7e3

- ZocDoc ; finding in network physicians from insurance cards

https://www.zocdoc.com/about/blog/tech/making-sense-of-insurance-cards-using-deep-learning/

- Spotify ; How does it know your kind of music?

https://medium.com/s/story/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe

Computational Intelligence

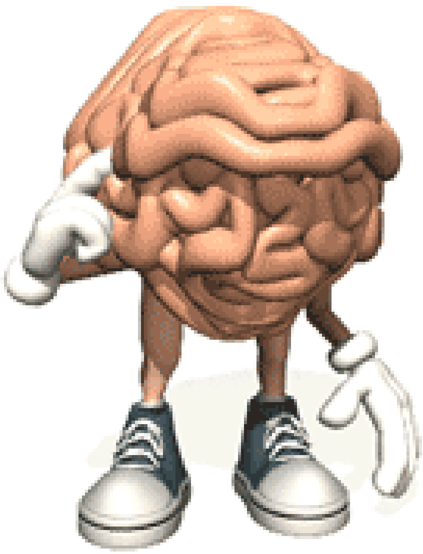
What methods machines use to learn?

Outline

- What is AI?
- Now, come Computational Intelligence (CI)
- Scope of CI
- Importance of CI
- CI Techniques
  - Artificial Neural Network
  - Genetic Algorithm
  - Fuzzy Logic

What is AI?

- Ability to interact with the real world
- Reasoning and Planning
- Learning and Adaptation

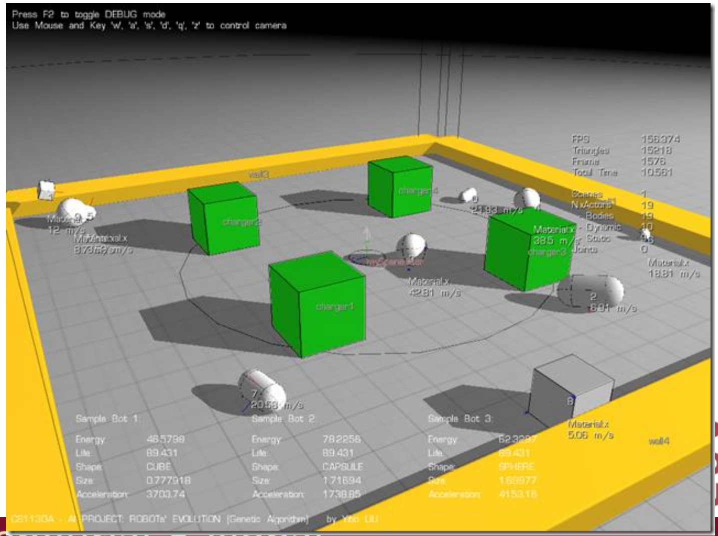


Now, come Computational Intelligence (CI)

- A methodology involving computing that exhibits an ability to learn and/or to deal with new situations, such that the system is perceived to possess one or more attributes of reason, such as generalization, discovery, association and abstraction (Russel et al, 1998).
- CI is the field of computing that draws from the successes of natural systems to develop new ways of solving computational problems in the real world.

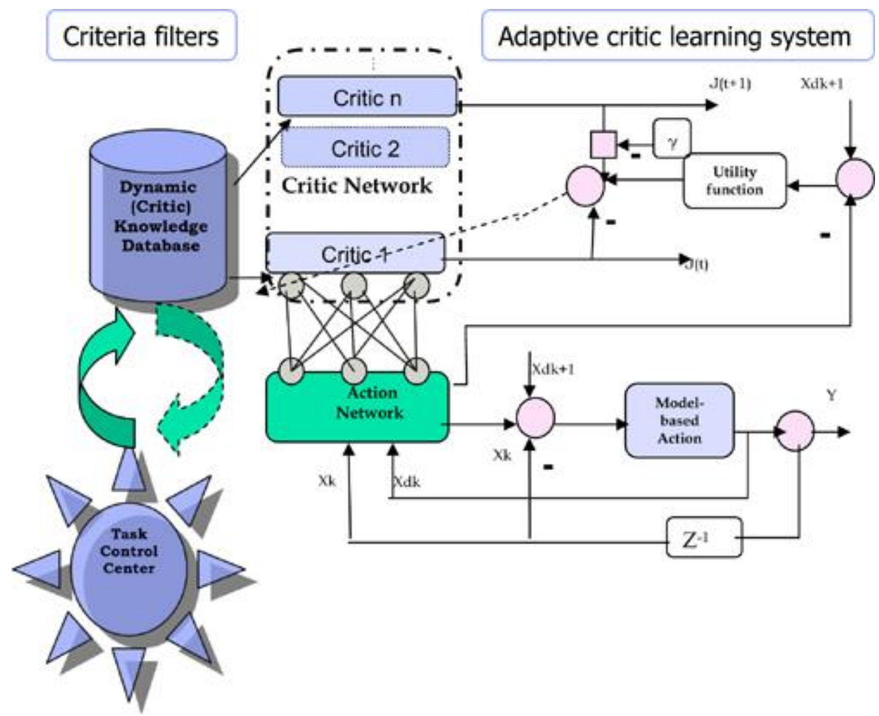
Scope of CI

- Practical adaptation concepts, paradigms, algorithms and implementations that enable or facilitate appropriate actions (intelligent behavior) in complex and changing environments.



Scope of CI

- Fuzzy, imprecise or imperfect data
- No available mathematical algorithm
- Optimal solution unknown
- Rapid prototyping required
- Only domain experts available
- Robust system required
- Learning capability



More recently:

- DNA Computing
- Quantum Computing

Importance of CI

- Solve the problems in realistic situation (not ideal) such as fuzzy, imperfect or incomplete, only domain experts available.
- Complementary approaches with mathematical based algorithm solutions
- Encourage new technology development especially in Quantum and DNA computing.

CI Techniques

- Artificial Neural Network
- Genetic Algorithm
- Fuzzy Logic