Supervised versus Unsupervised Learning

Ogwal-Awio Kenneth

Lira University

Lecture objectives

- By the end of this lecture students should be able to:
  - ✓Comprehend the difference between supervised, and unsupervised learning
  - ✓Explain the problem peculiar to each method of machine learning
  - ✓Identify the algorithms for each method of machine learning
  - ✓Explain the uses and applications of the methods of machine learning

What is supervised learning?

- Supervised learning is a subdomain of machine learning in which an algorithm is employed to learn the mapping function from the input variable (x) to the output variable (y).

  i.e. **y = f(x)**

- It is about learning a mapping function from inputs to outputs.
- Supervised learning algorithms utilize datasets to make predictions.
  - The utilized datasets are called training datasets.
- It is based on labelled data

Labelled dataset

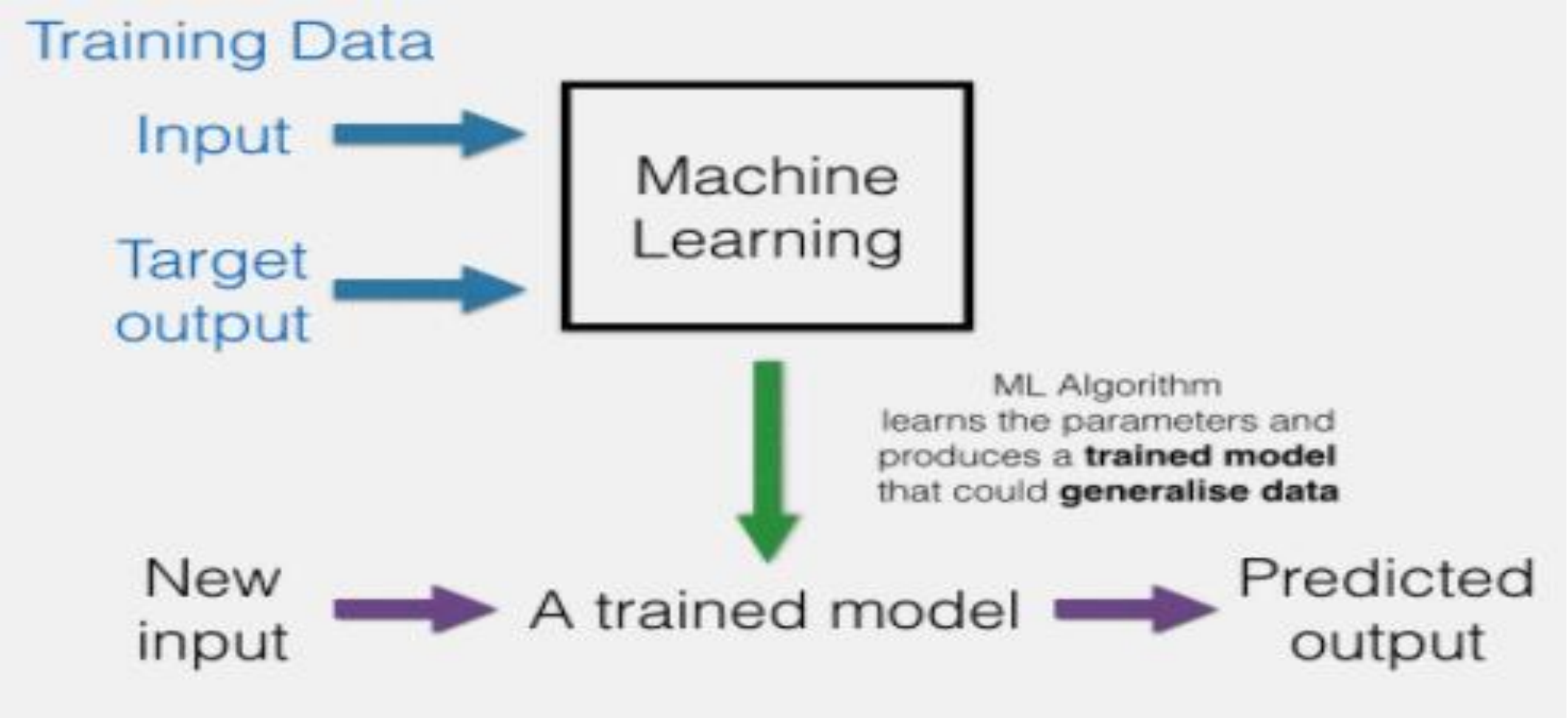
No	SIZE	COLOUR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Rounded shape with a depression at the top	Tomato
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Rounded shape without a depression at the top	Orange

- Here, size, colour and shape form the input training dataset, while fruit name form the labels, also called classes.
- A supervised learning algorithm infers a function from this given labeled training data.

Supervised learning…

- The training data consists of a set of training examples.
  - ➢ Each example is a pair consisting of an input object and a desired output value
    - ○ The input data is also called input vector, and output value is also called a label, category, or supervisory signal.
- A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for determining the classes of new inputs.
- The input vector must be sufficient to allow, and too small or too large.

Learning process in supervised learning



Supervised learning

- The algorithm should correctly determine the class labels for new instances of inputs.
  - ➢ In Machine learning, an algorithm that is capable of learning a classification predictive model is called a classification algorithm.
    - This requires the learning algorithm to generalize from the training data to unseen situations in a “reasonable” way.
- Examples of the common classification algorithms include K-Nearest Neighbors (K-NN) algorithm, Decision trees (DT) algorithm, Logistic regression (LR) algorithm, and Naïve Bayes (NB) algorithm.

Supervised learning problems

- A supervised learning problem can be a:
  - ✓Classification problem
  - ✓Regression problem

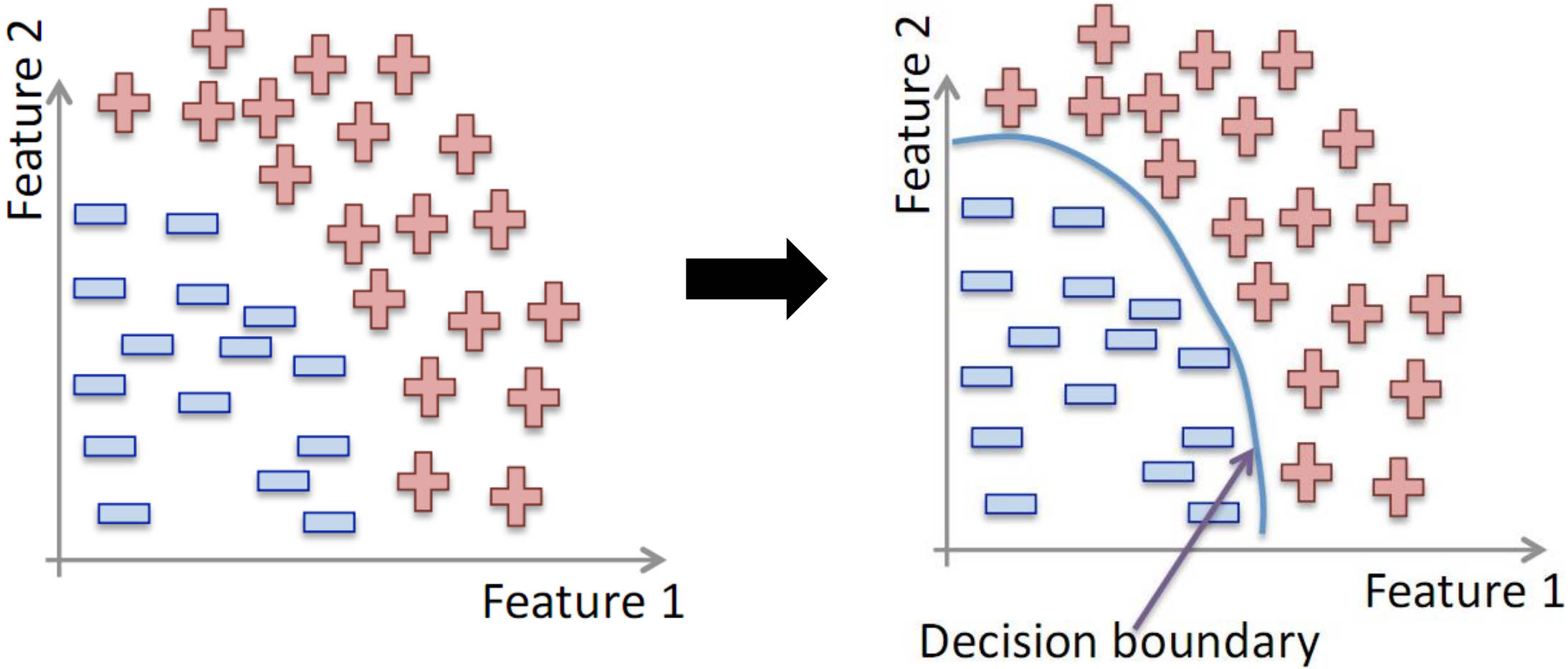
Classification

- Classification algorithms estimate the mapping function (f) from the input variables (x) to discrete or categorical output variables (y).
  - The output variable in classification is categorical (or discrete), i.e. it classification returns categorical output
  - In a way, it solves a yes/no problem
    - i.e. whether something meets its required standards, or not.
- The goal of classification is to predict the target class.
  - Thus, classification predicts a discrete class label output for an example.

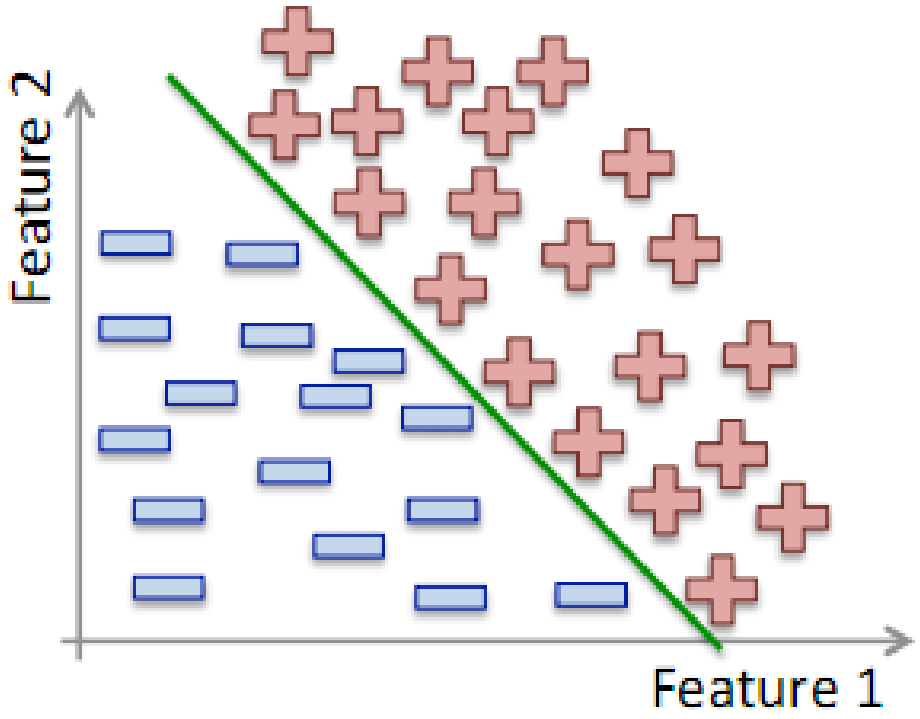
Classification…

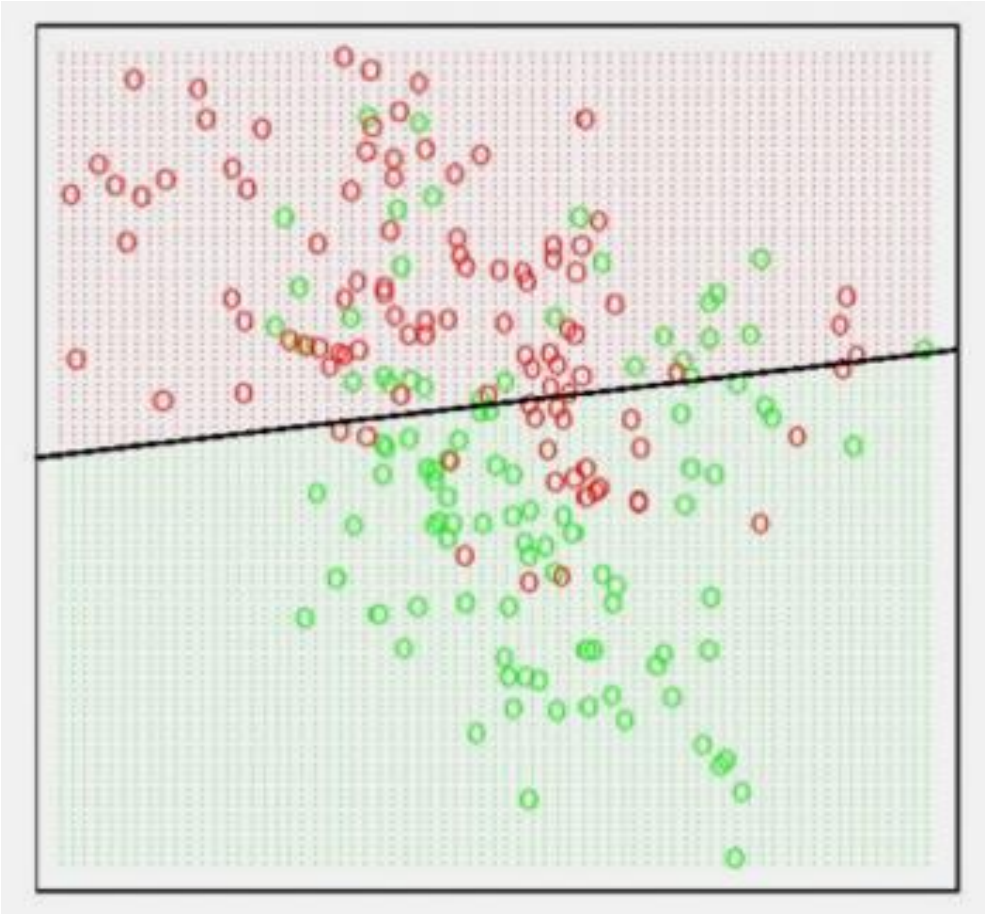
- If the trained model is for predicting any of two target classes, then it is known as binary classification, or a two-class problem.
  - For example basing the student profile to predict whether the student will pass or fail, and predicting whether a customer will buy the new product or not, are both problems which can be addressed with binary classification.
- If we have to predict from more than two target classes, it is known as multi-classification.
  - For example, considering all subject details of a student to predict which subject the student will score more, and identifying the object in an image, are both multi-classification problems. Other application include customer segmentation, audio and image categorization, and text analysis.

Classification

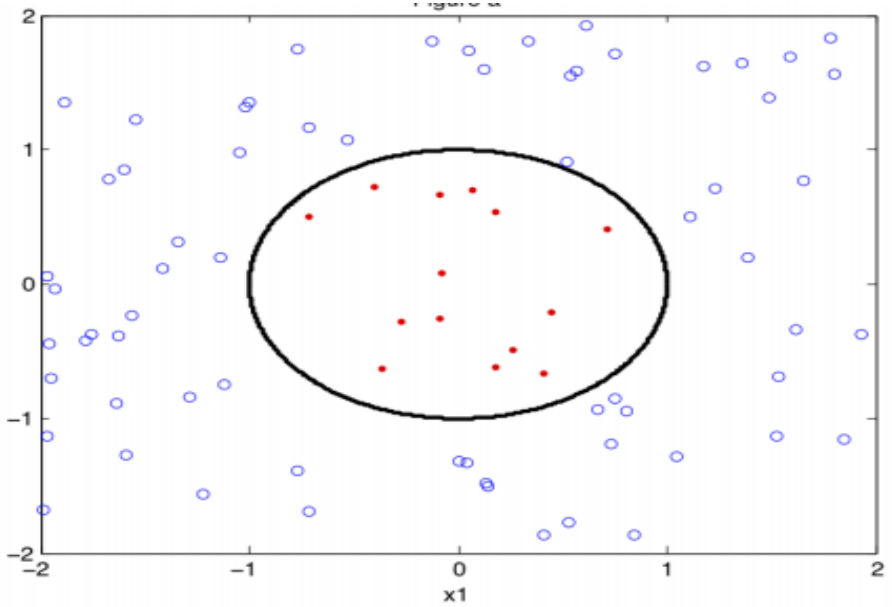


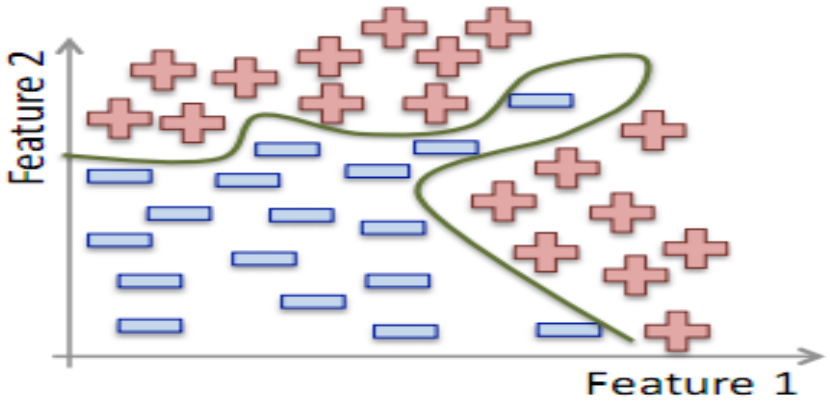
More examples – Linear classification…

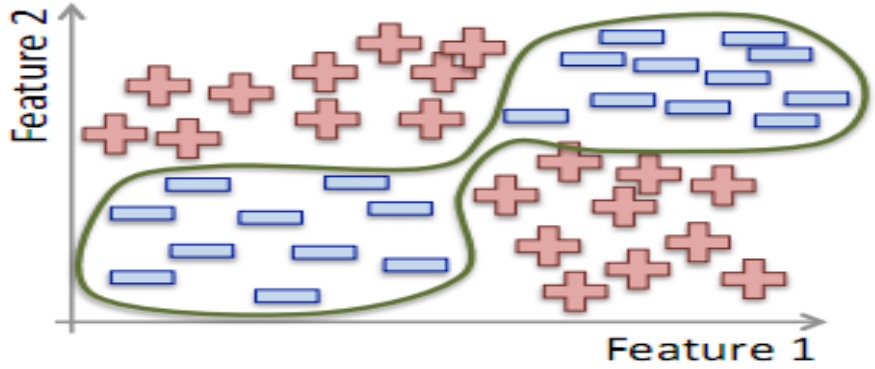




More examples - Non-linear classification…







Applications of classification

- Determining where email is a spam or not
- Determining where face is for male or female
- Attack detection in computer network flow
- Determining whether fruit is banana or orange
- Approving credit as yes or no
- Etc, etc, …

Regression

- Regression algorithms attempt to estimate the mapping function (f) from the input variables (x) to numerical or continuous output variables (y).
- In this case, y is a real value
  - A real values here can be an integer or a floating point value.
- Regression prediction problems are usually about quantities or sizes.
  - It predicts a discrete or a continuous numerical value.
- A regression output is represented by a quantity determined based on the inputs of the model rather than being confined to a set of possible labels.

Regression algorithms

- Common regression algorithms include;
  - ✓Linear regression (LR)
  - ✓Support Vector Regression (SVR)
  - ✓Regression trees

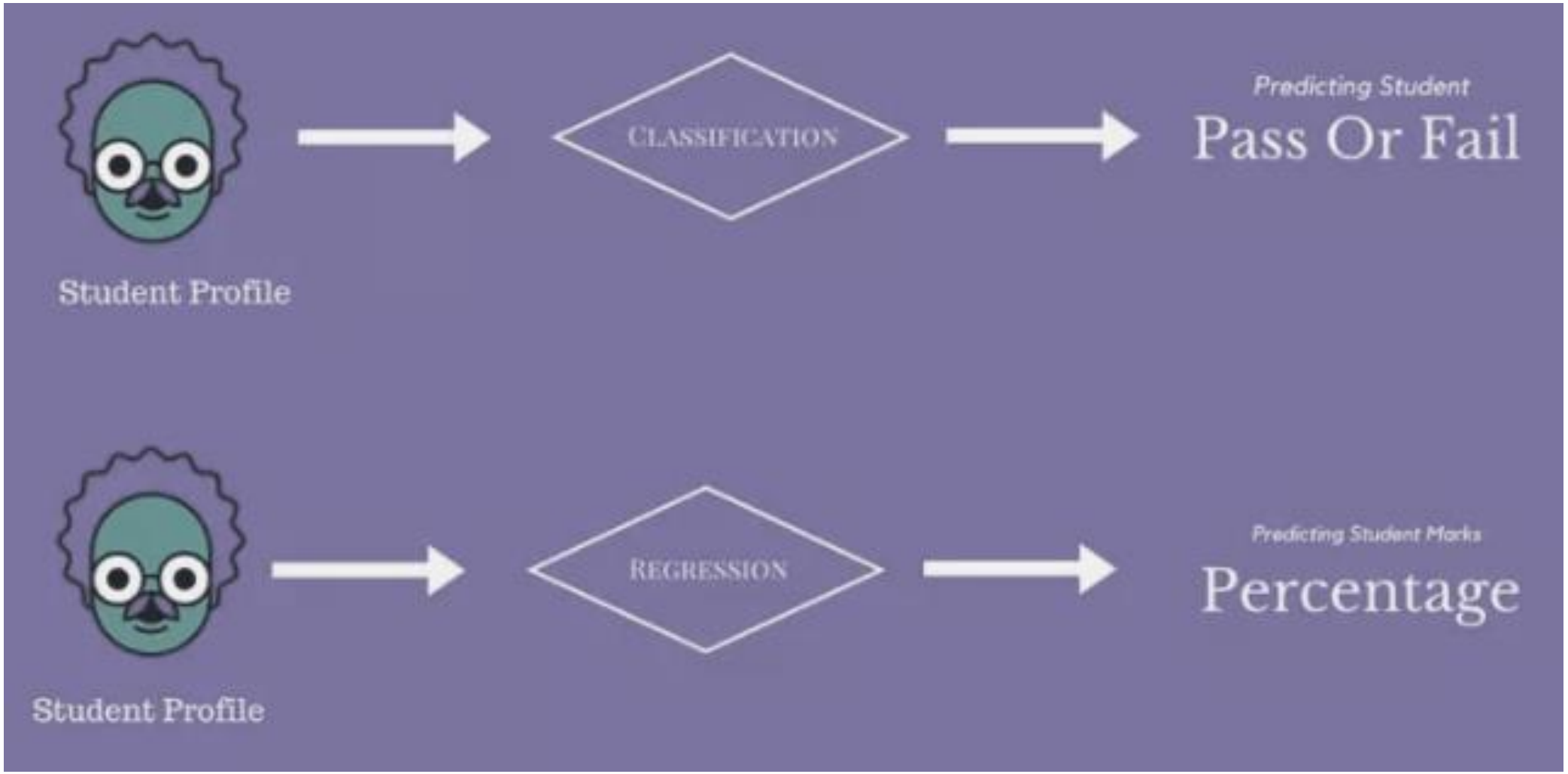
❑Other algorithms can be used for both classification and regression

- ➢E.g. Decision Tress and Random Forest.

Applications of regression

- Sample applications include:
  - ✓Prediction of prices of houses when provided with a dataset about the houses.
  - ✓Predicting whether it will rain, and if it will, how much rain we will get.
  - ✓Predicting the likelihood of increase in customers' demand.
  - ✓Knowledge extraction
  - ✓Compression
  - ✓Outlier detection e.g. exceptions that are not covered by the rule

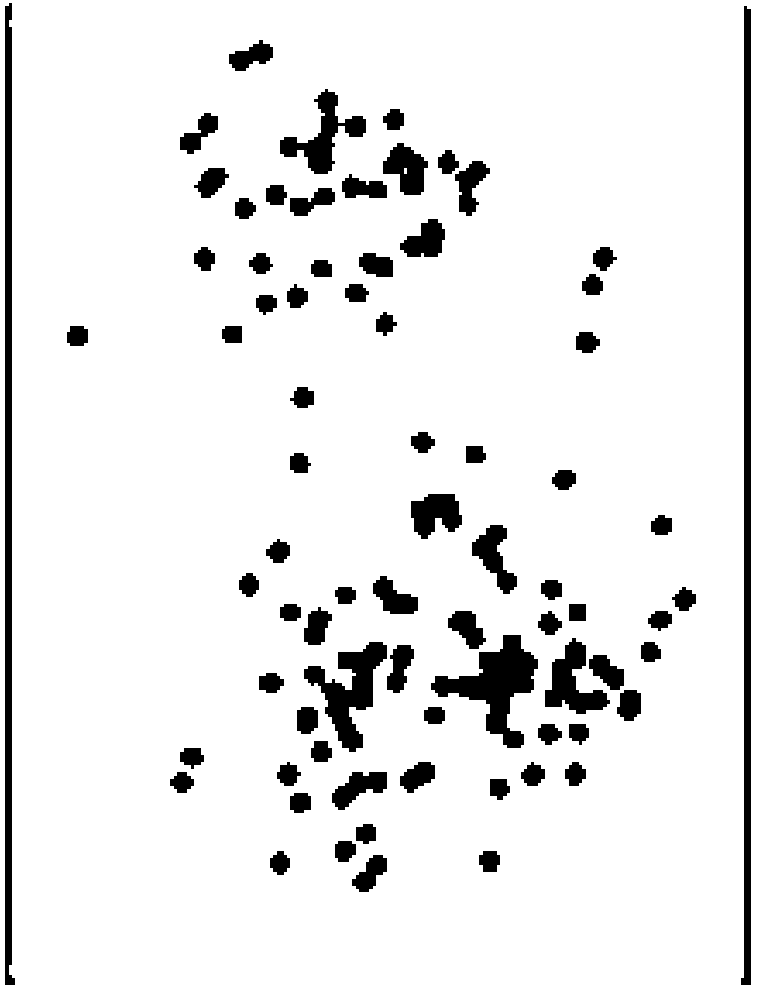
Classification vs Regression



Unsupervised learning

What is unsupervised learning?

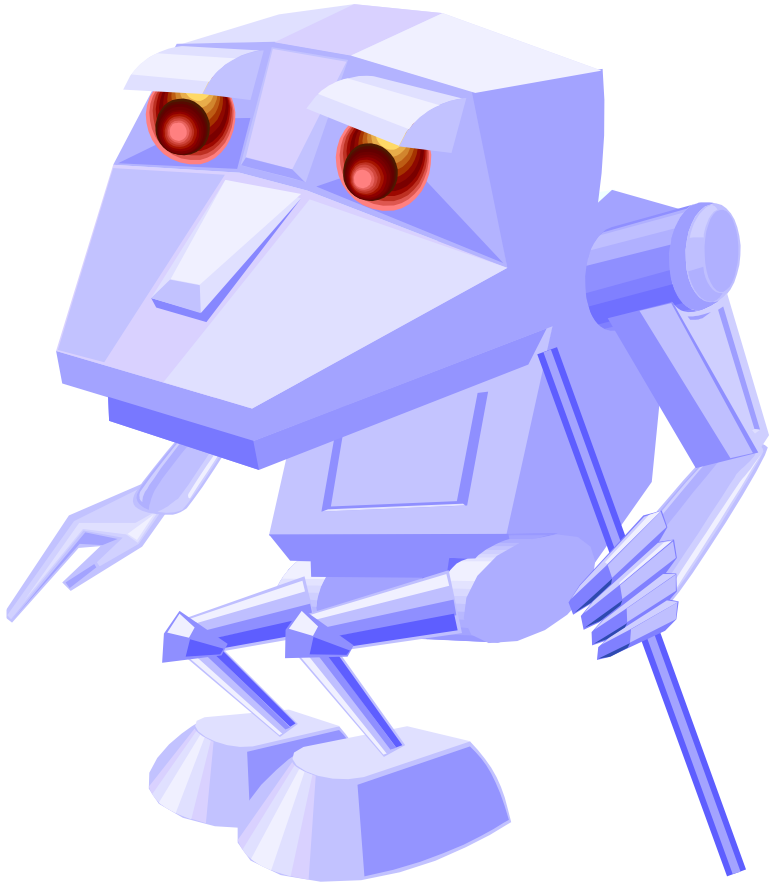
- Unsupervised Learning is the acquisition of knowledge on useful structure in a dataset which does not have labeled classes, optimization criterion, feedback signal, or any other information beyond the raw data
  - ➢ It is also referred to as self-organisation, or modelling probability density of inputs.
- It learns from unlabeled data to discover patterns in the data, and perform cohesive grouping and associations based on frequent co-occurrence.



Unsupervised learning…

- Unsupervised learning identifies commonalities in the example data and predicts based on the presence or absence of commonalities.
  - ➢It is a prelude to discovery of underlying properties of the dataset.
- It takes as training examples the set of attributes/features alone.
- The purpose of unsupervised learning is to attempt to find natural partitions in the training set.
  - ➢The core approach is by clustering of inputs to generate the output.

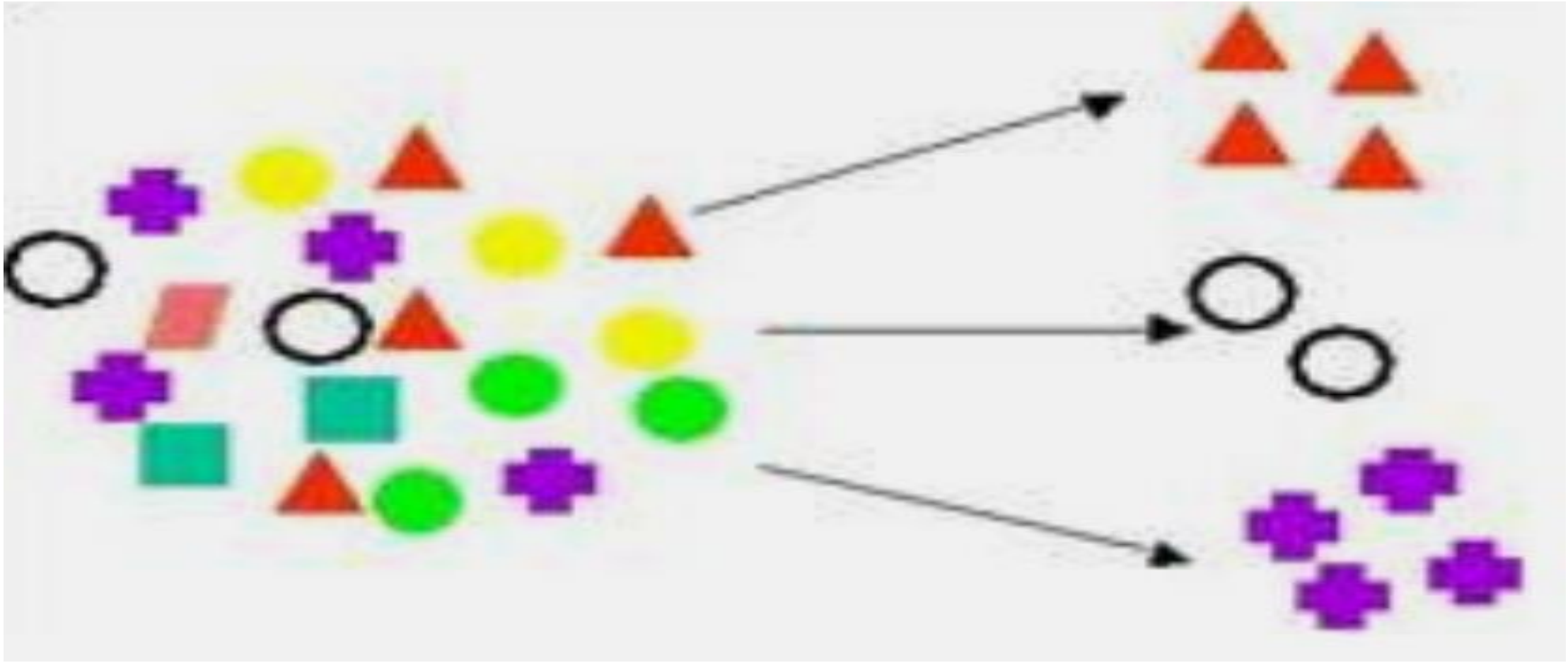
Unsupervised learning



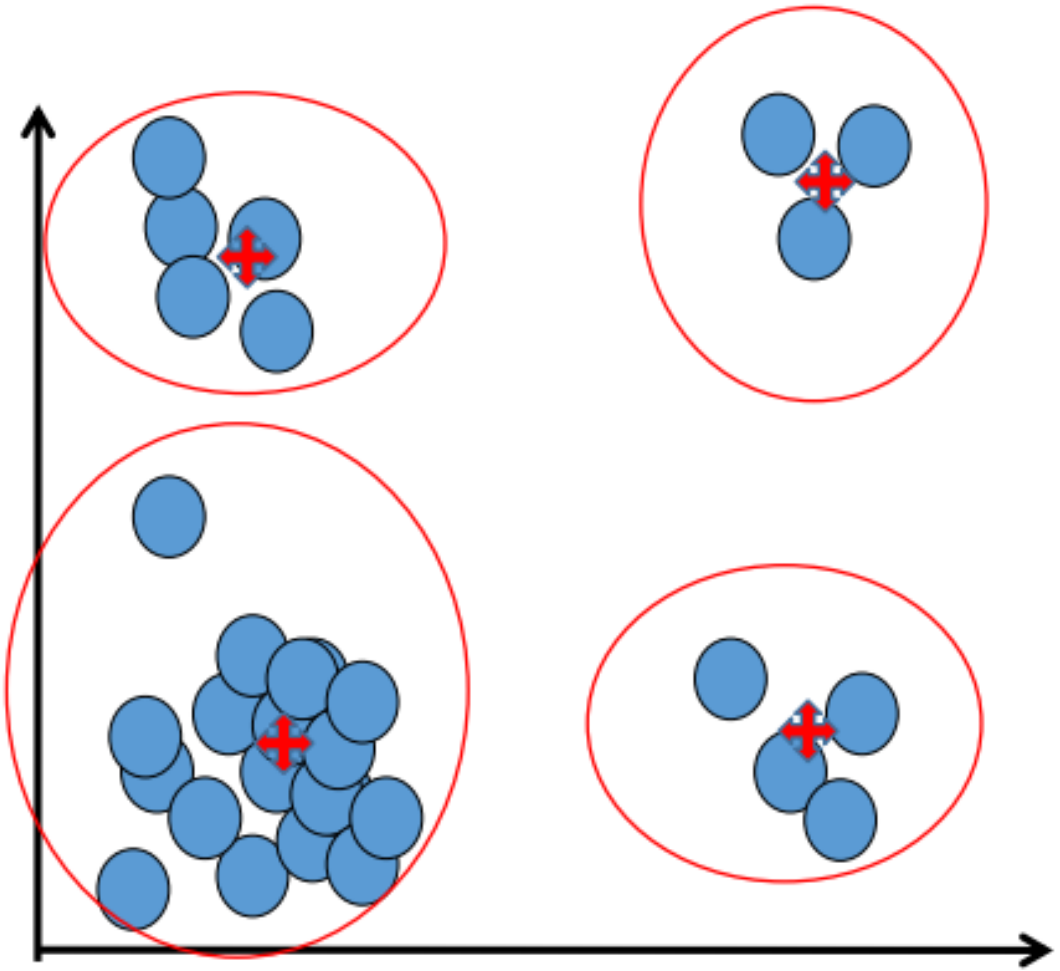
Clustering

- In clustering, the data has no target attribute, but we want to explore the data to find any intrinsic structures in them.
  - ➢ A cluster is a collection of data items which are similar between them and dissimilar to data items of other clusters
- The goal of clustering is to
  - ✓Group data points that are similar or close to each other, and
  - ✓Identify such groupings without basing on labels.

Example of clustering

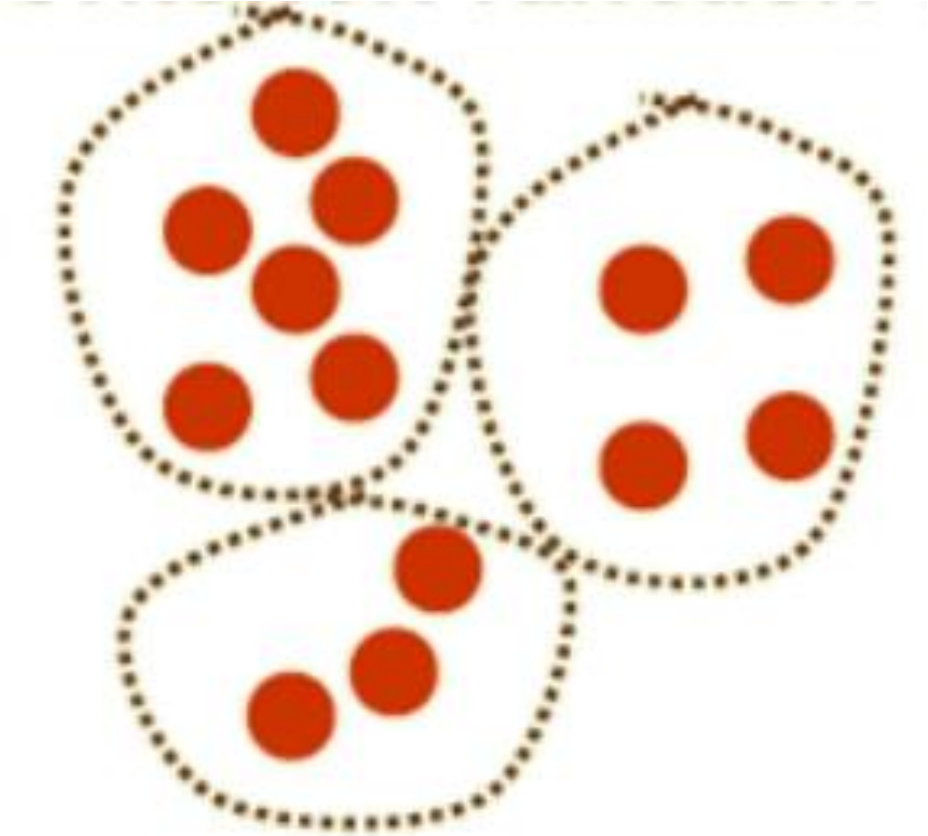


Cluster centroids

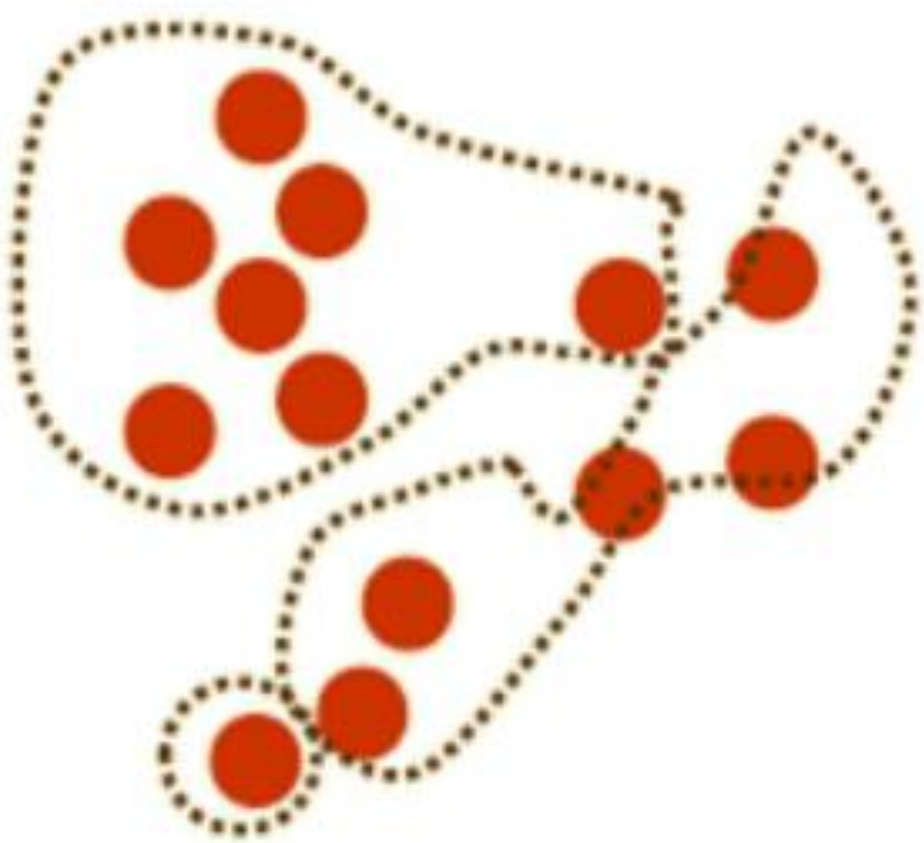


- Each cluster is represented by its central point known as a centroid
  - A cluster boundary is determined by the farthest data point in the cluster.

Which is a better way of clustering?



Good clustering

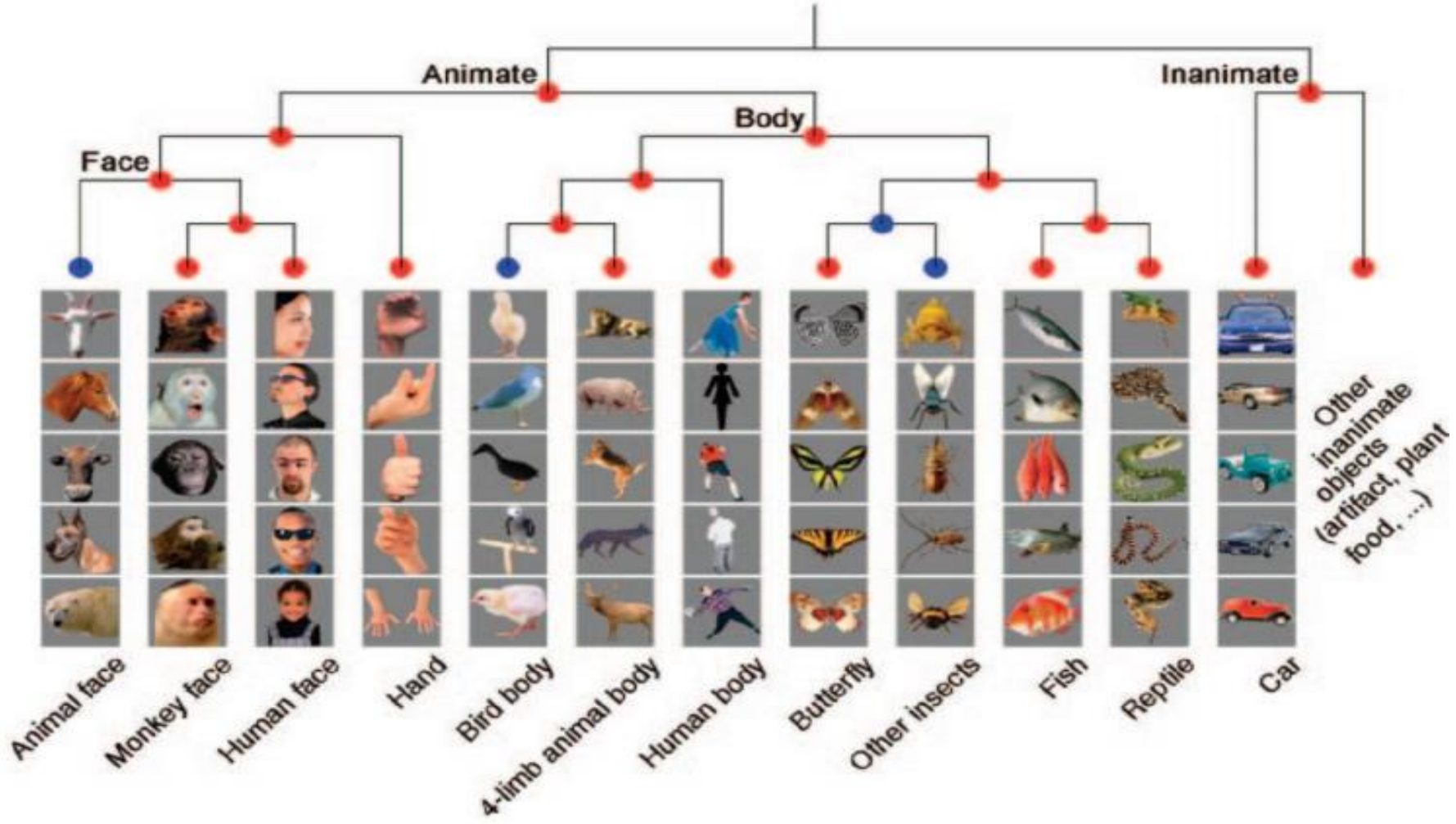


Bad clustering

Types of clustering

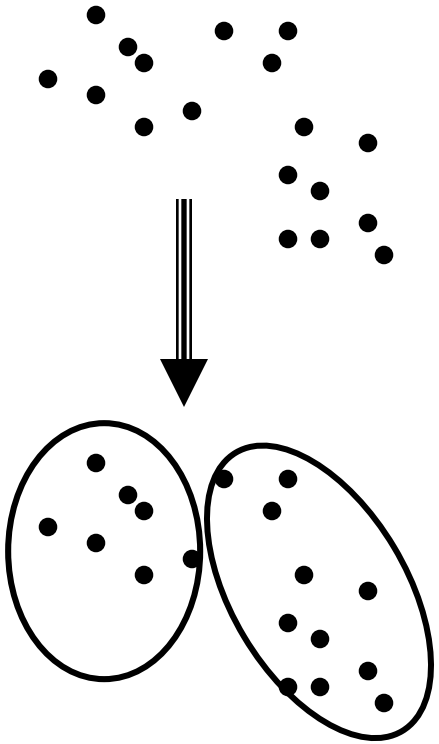
- Common types of clustering include:
  - ✓Exclusive clustering – also known as hard clustering, is where a given data point in n-dimensional space only belongs to one cluster. E.g. K-means method.
  - ✓Overlapping clustering – also called soft clustering, is where a given data point can belong to more than one cluster. E.g. Fuzzy C-means and Fuzzy K-means.
  - ✓Hierarchical clustering – where a hierarchy of clusters is built using top-down (divisive clustering) or bottom-up (agglomerative clustering) approach.
  - ✓Flat clustering – a simple clustering technique where no hierarchy is present
  - ✓Spectral clustering – where the nodes in a graph are linked based on similarity weights
  - ✓Model-based clustering – also called probabilistic clustering, is where data is modelled using a standard statistical method after partitioning the training set into clusters with each cluster composed of values from different clusters, in order to find a model that best fits the data.

Example – Hierarchical clustering

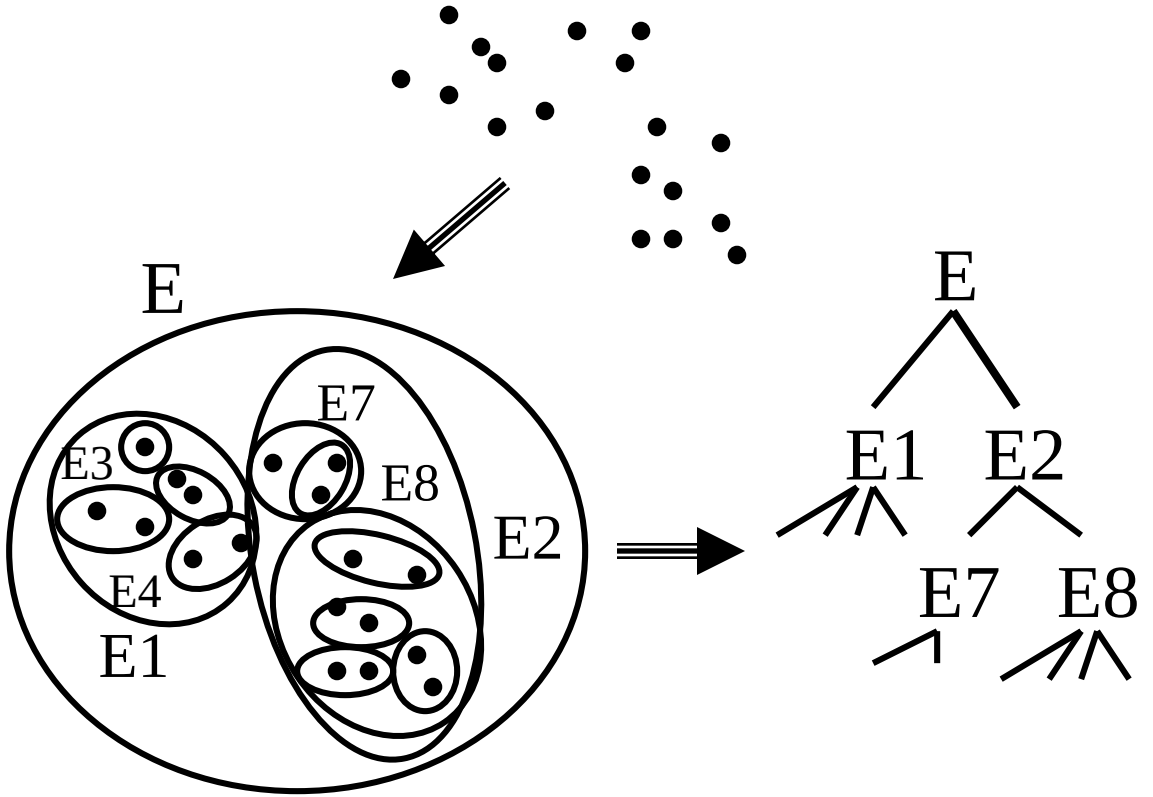


Flat Clustering vs Hierarchical Clustering

Flat Clustering



Hierarchical Clustering



Important elements in clustering

- Similarity function
  - Also called proximity measure function, it is a real-valued function that quantifies the similarity between two objects.
- Stopping criteria.
  - Since an clustering algorithm computes successive approximations to the solution of a linear system, a practical test is needed to determine when to stop the iteration.
- Cluster quality
  - There are a number of metrics used to quantify the quality of clustering, and so we need to determine which is more suitable for that application.

When to use clustering

- Clustering can be used in the following situations:
  - ✓When the data has high dimensionality
  - ✓When the dataset is large
  - ✓When the dataset consists of different kinds of attributes, such as numerical, categorical, or binary data.
  - ✓When there is need to discover clusters with arbitrary shapes.
  - ✓When the dataset contains noisy, missing or erroneous data.
  - ✓When the results should be interpretable, comprehensible and usable.

Applications of clustering

- ✓Climatology
  - Study of climates
- ✓Bioinformatics
  - Processing, storage and retrieval of data related to life.
- ✓Data mining
  - Finding insights from data
- ✓Image processing
  - Object and character recognition
- ✓Medical imaging
  - Finding patterns from images related to life.

Applications of clustering…

- ✓Pattern recognition
  - Finding and classifying data based on knowledge already gained from patterns and/or their representation.
- ✓Spatial data analysis
  - Analysis of space
- ✓Etc, etc

Challenges with Unsupervised Learning

- Data mining deals with **large databases**
- Scalability with respect to number of instance
  - ➢Use a random sample (possible bias)
- Dealing with mixed data
  - ➢Many algorithms only make sense for numeric data
- High dimensional problems
  - ➢Can the algorithm handle many attributes?
  - ➢How do we interpret a cluster in high dimensions?
- Shape of clusters

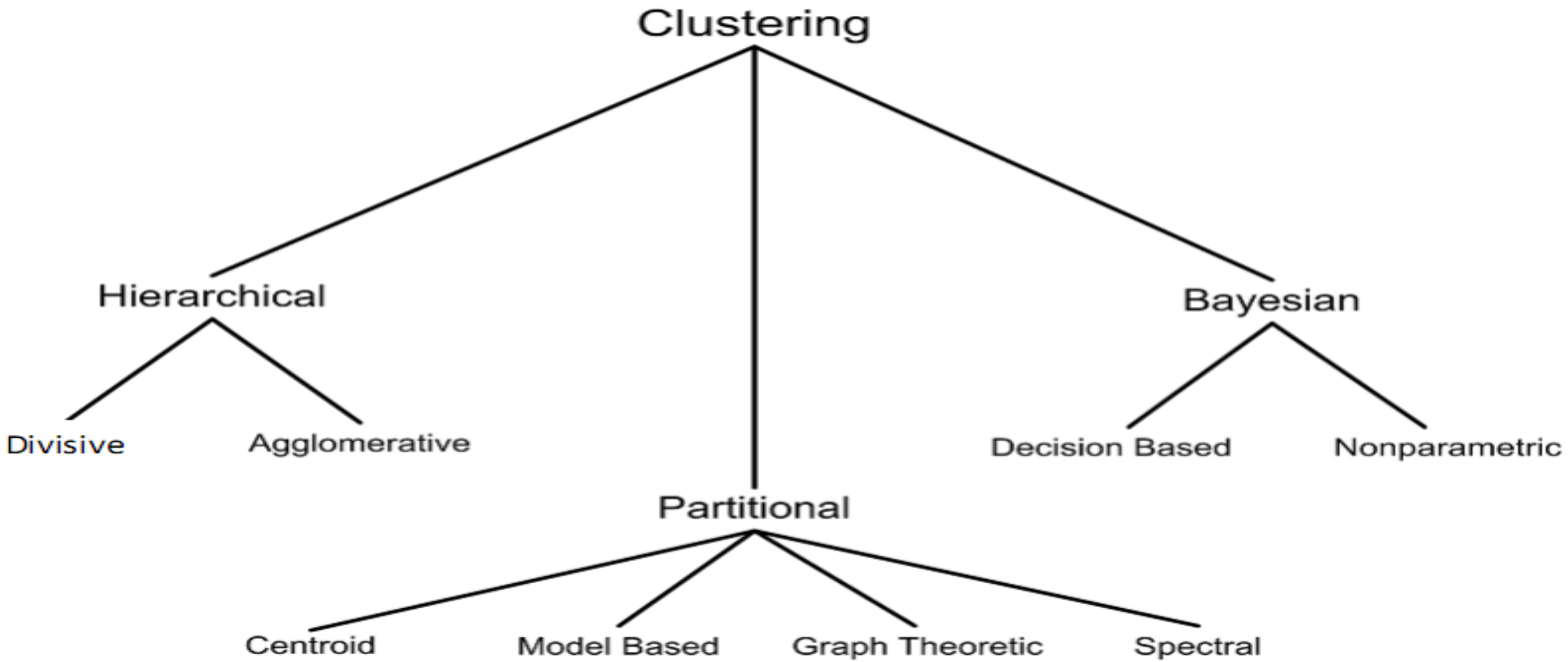
Challenges with Clustering

- Minimum domain knowledge (e.g., knowing the number of clusters)
- Noisy data
- Insensitivity to instance order
- Interpretability and usability

Applications of unsupervised learning

- Understanding and visualization
- Anomaly detection
- Information retrieval
- Data compression

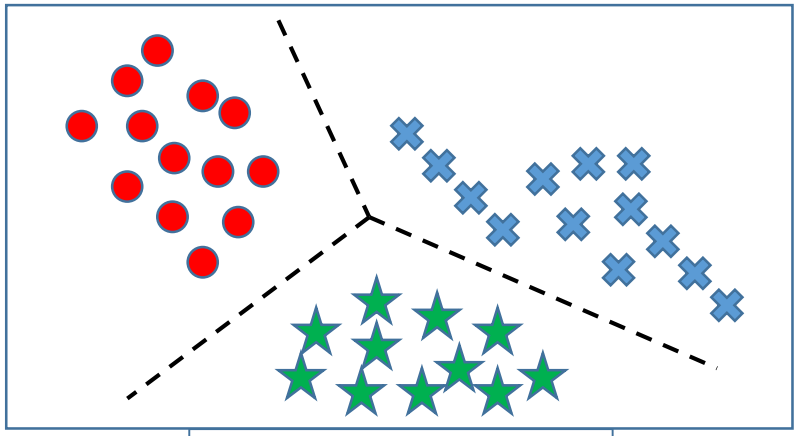
Clustering in a nutshell



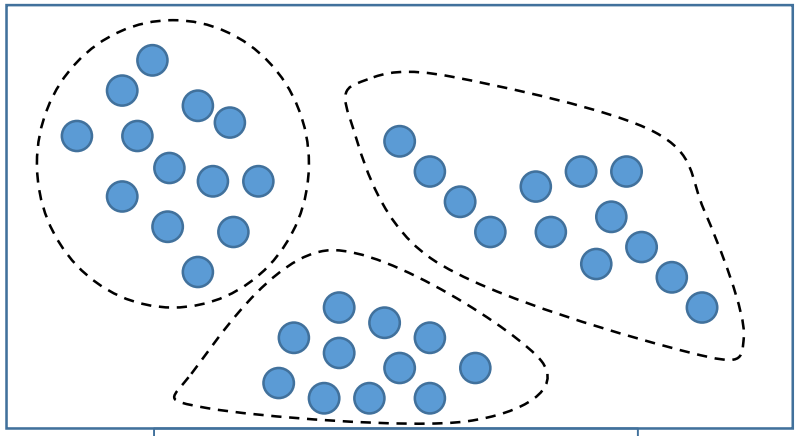
Clustering in a nutshell…

- Hierarchical algorithms – find successive clusters using previously established clusters. They include:
  - Agglomerative algorithms – begin with each element as a separate cluster and merge them into successive larger clusters, until all clusters are merged into a single cluster (root cluster).
  - Divisive algorithms – begin with the whole set and proceed to recursively divide it into successive smaller clusters, until only singleton clusters of individual data points remain.
- Partitional algorithms – determine all clusters at once, but can also be used as devise algorithms in the hierarchical clustering.
- Bayesian algorithms – generate a posteriori distribution over the collection of all partitions of the data.

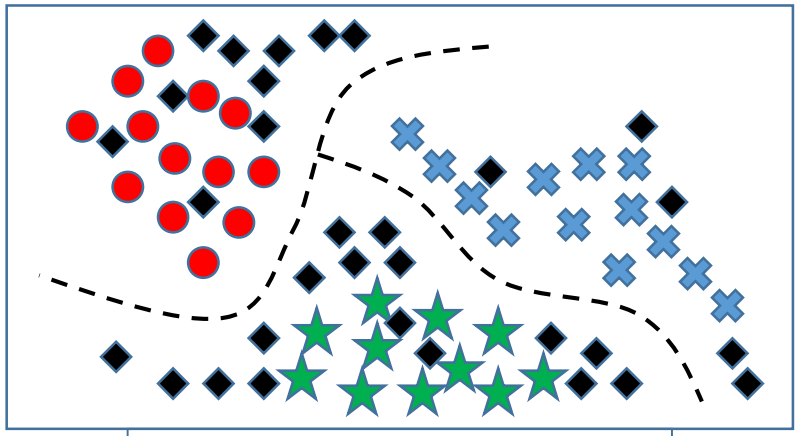
Summary of ML algorithms



Supervised learning

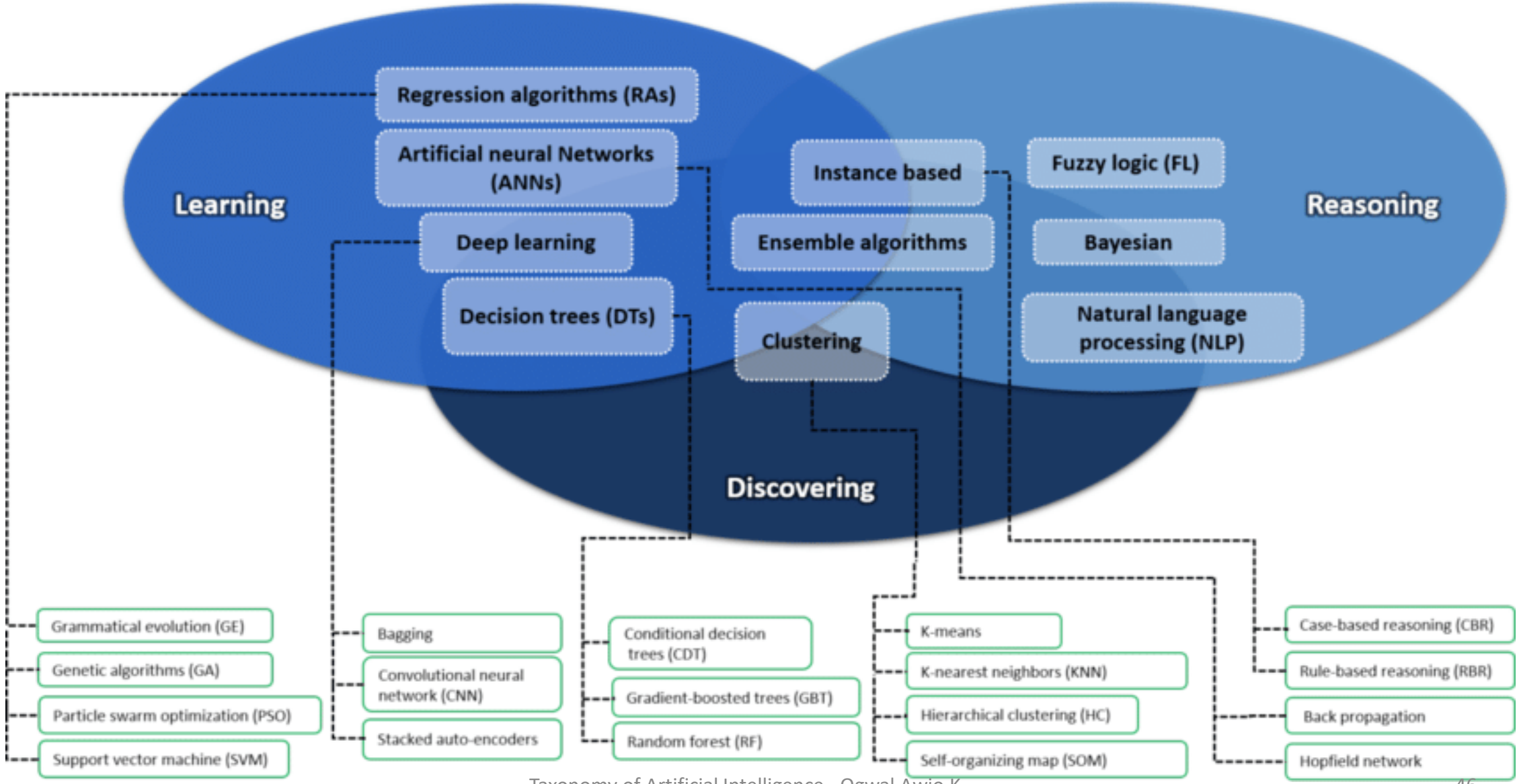


Unsupervised learning



Semi-supervised learning

A taxonomy of some of the best known AI methods



Reading Assignment

- Read about use of the Measures of evaluating a machine learning product based on the Confusion Matrix. Submission to [akogwal@gmail.com] is optional, as it will not attract any mark.

Big thanks