TETRIS DAY-15

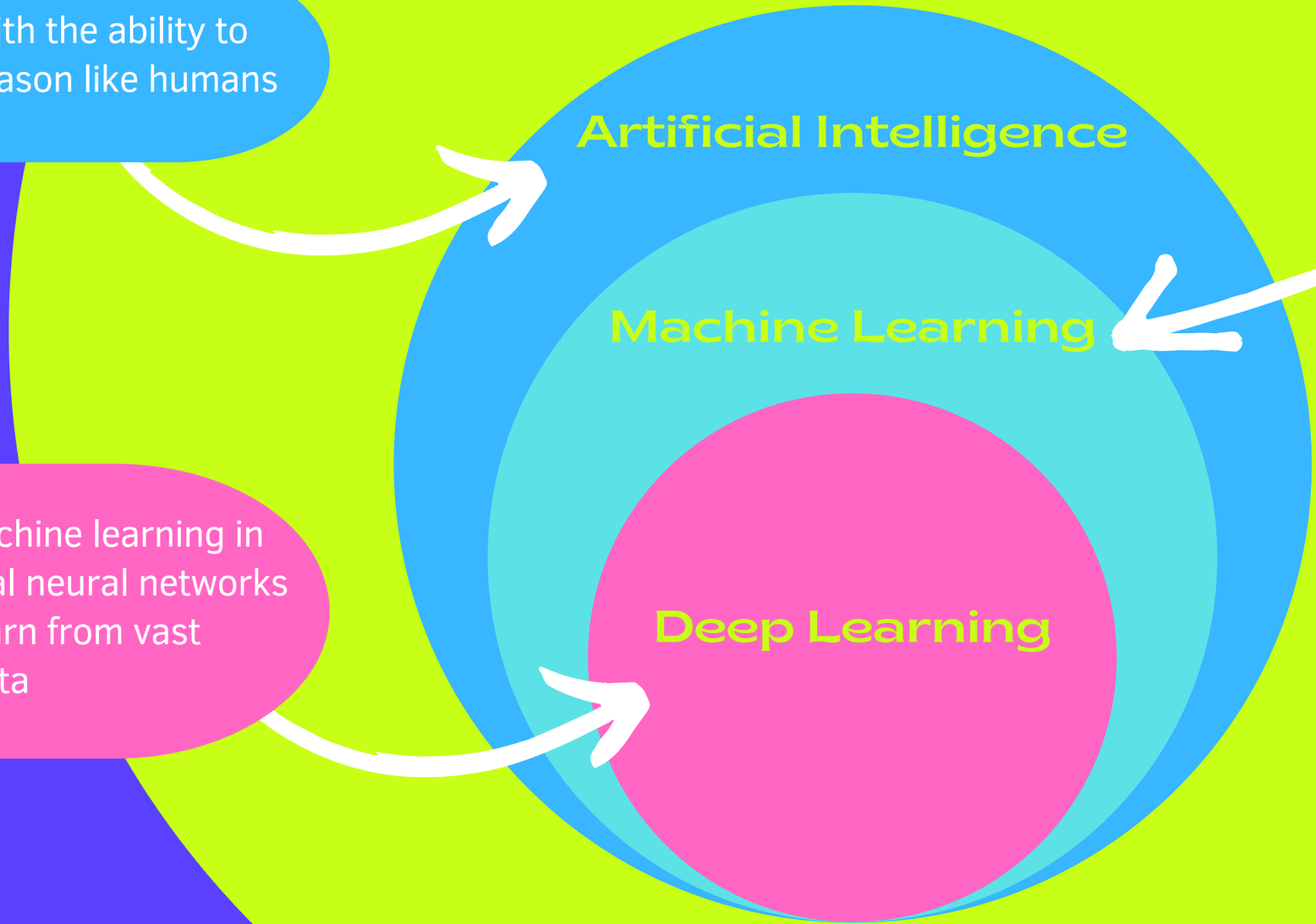
Introduction to Machine Learning

22 Juli 2022

Author: Eduardus Tjitrahardja

Machine Learning is a branch of artificial intelligence (AI)

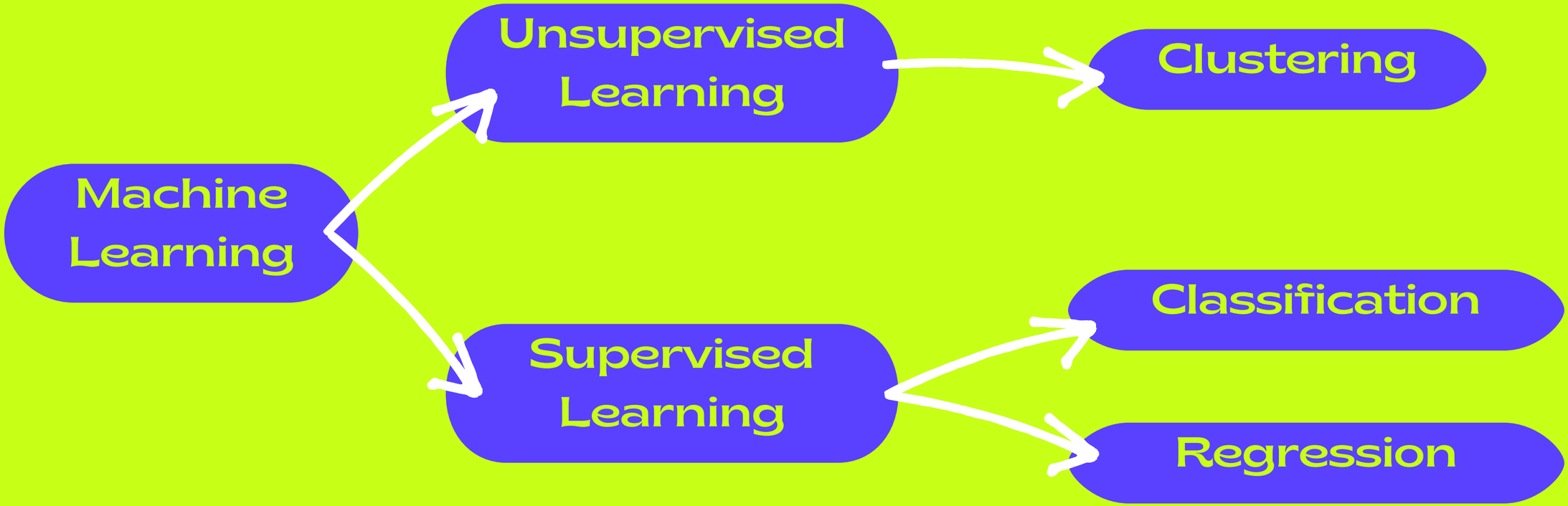
Programs with the ability to learn and reason like humans



Algorithms with the ability to learn without being explicitly programmed

Subset of machine learning in which artificial neural networks adapt and learn from vast amount of data

Machine Learning Techniques



TETRIS

What is Supervised and Unsupervised Learning?

Supervised Learning

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process.

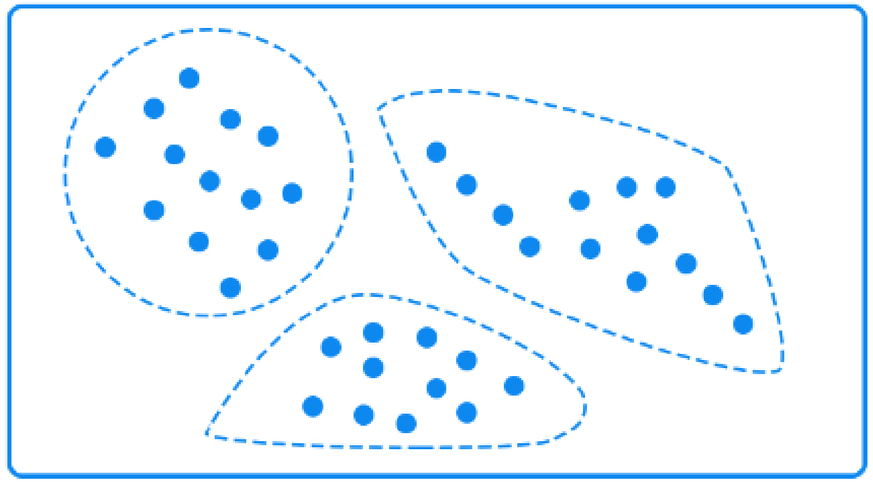
Unsupervised Learning

Unsupervised learning is where you have unlabeled data (or no target variable) in the dataset.
The goal of Unsupervised Learning Algorithms is to find some structure in the dataset.

TETRIS

Unsupervised Learning

Clustering



Ex:

K-Means

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

The similarity of those groups can be determined using their euclidean distance.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

TETRIS

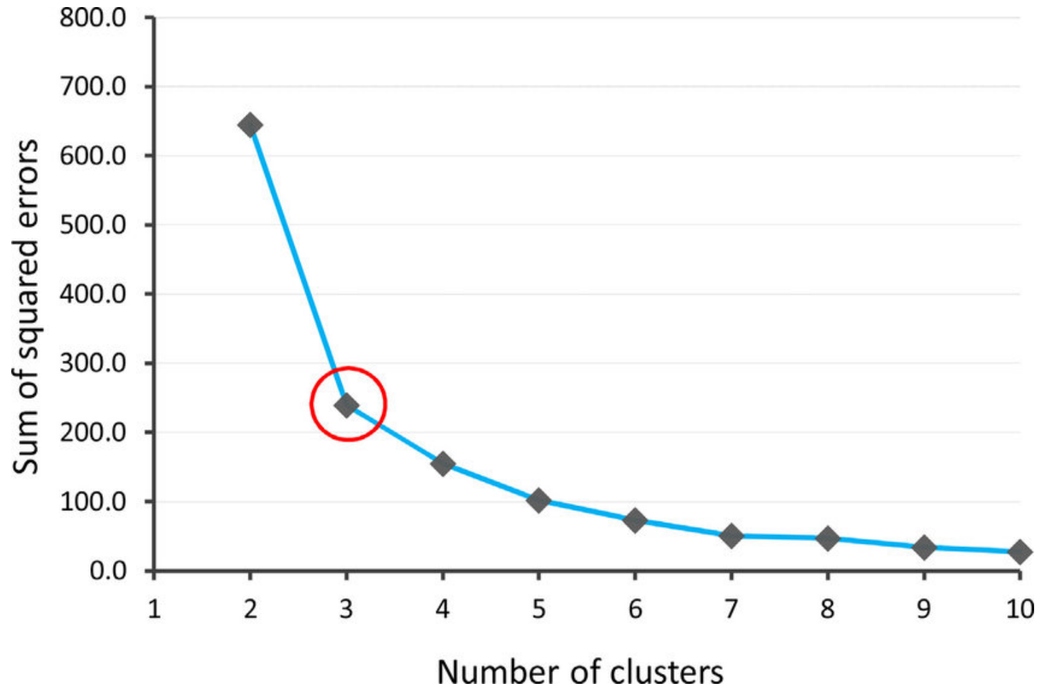
Clustering

Elbow Method

Elbow is one of the most well-known methods to determine the best k for K-Means model and boost your model performance.

It picks up a range of values and takes the best among them. It calculates the sum of the square (**SSE**) of the points and calculates the average distance.

We call it Elbow Method because the graph plotted in this case resembles an elbow. The value of k falling on the joint part is what we consider to be the optimum value.



SSE is the sum of the squared **centroid**'s distance from its cluster elements' mean.
Centroid is the centre point of the cluster.

Supervised Learning

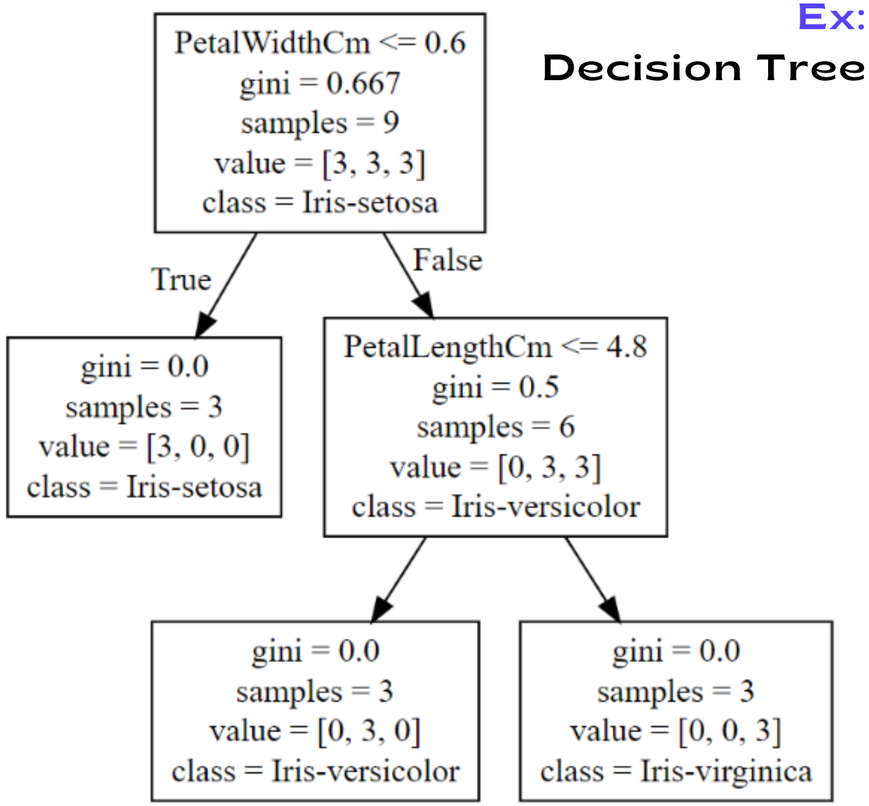
Classification

Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. The targets/labels are given in the dataset.

What makes classification and clustering different? Classification has a label from each data in the dataset and Clustering doesn't has a label at all.

Ex:

Decision Tree



TETRIS

Classification - Decision Tree

Gini Impurity

is a function that determines how well a decision tree was split.

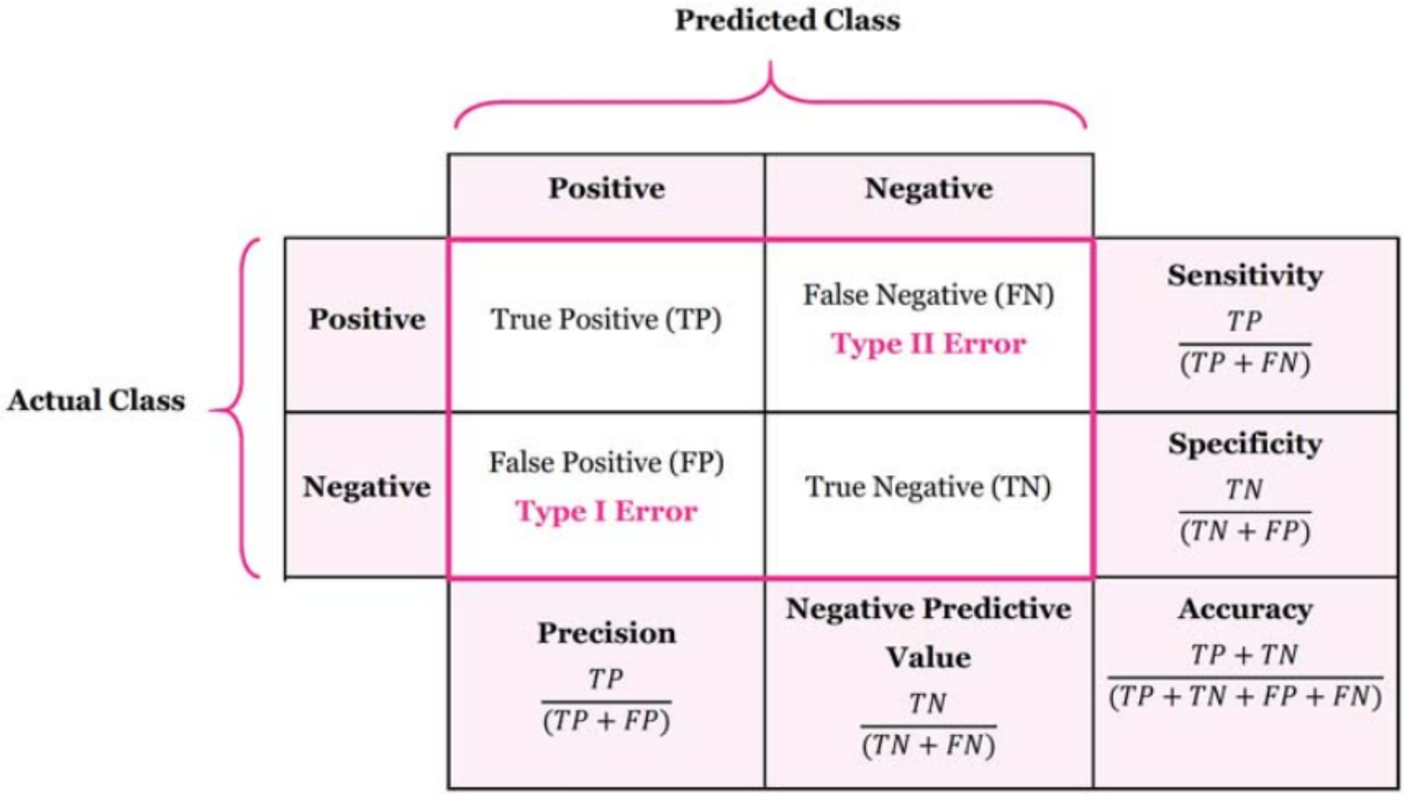
If the result is approcing:
1 -> Very Random
0 -> Very Uniform

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Where:
p is the probality of the label

MSGR

Confidence Matrix



Evaluate The Model

to evaluate you model,
you can use the confidence matrix on the left to calculate your accuracy, recall, precisition, etc.

TETRIS

Thank You!

-TeamOne