جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

ENSIAS

Machine Learning Project

Supervised by: Pr. MAHMOUDI Abdelhak
Presented by:BELKASMI Maryem

2021-2022

Summary

I. Machine learning algorithms

II. Support Vectors Machines SVM

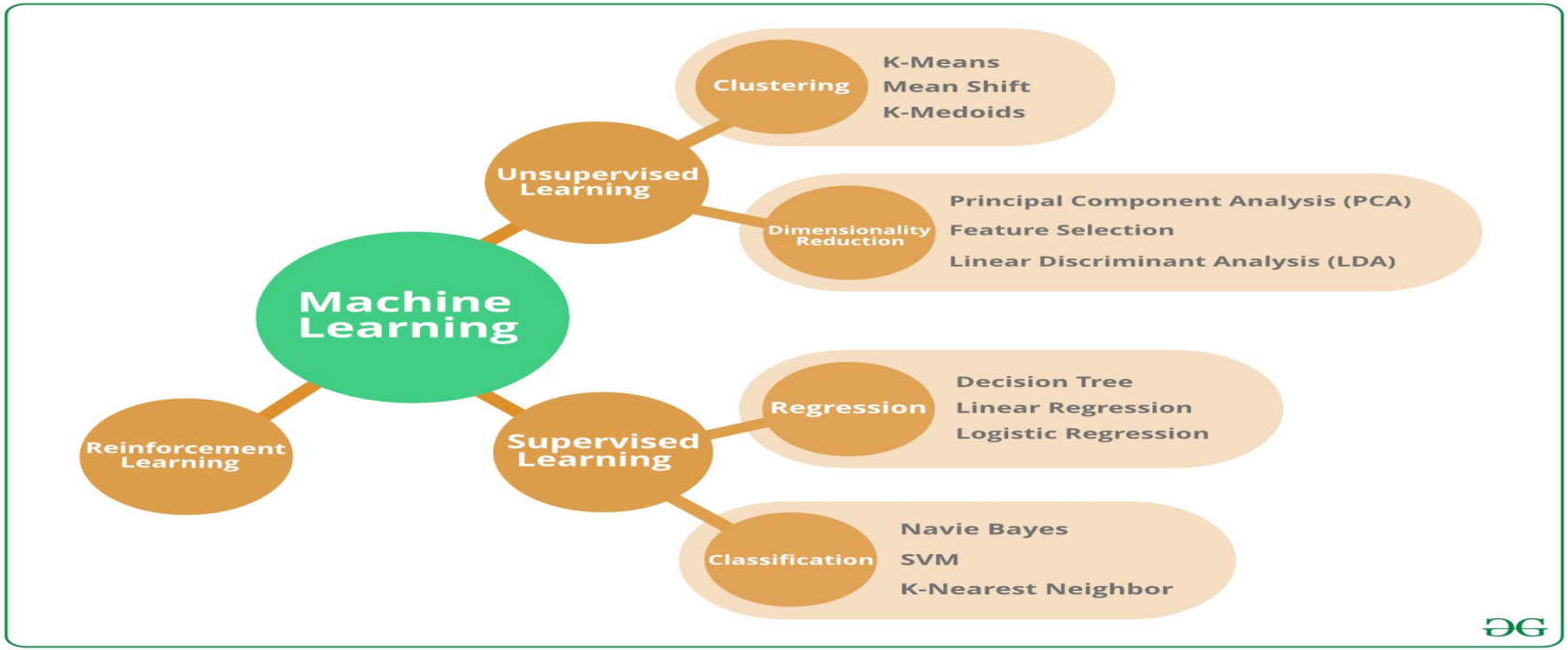
III. Application: E.Coli DNA sequencing

IV. Density-Based Spatial Clustering of Applications with Noise DBSCAN

V. Application: Contact Tracing

VI. Conclusion

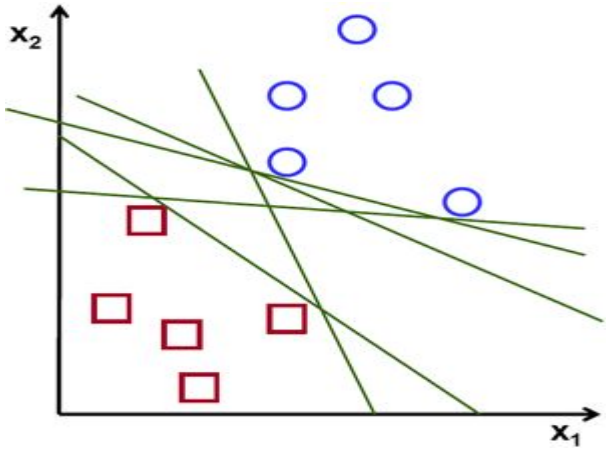
Introduction

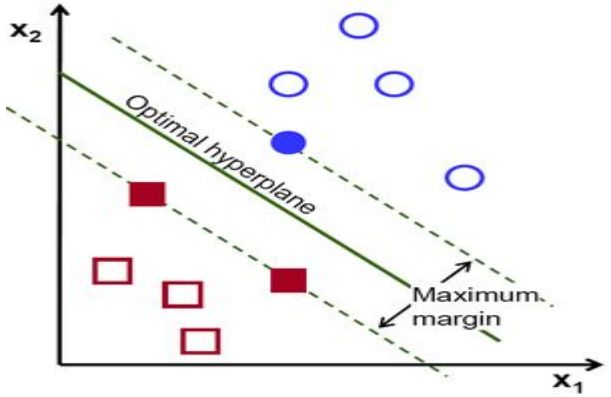


Support Vectors Machines SVM

Definition

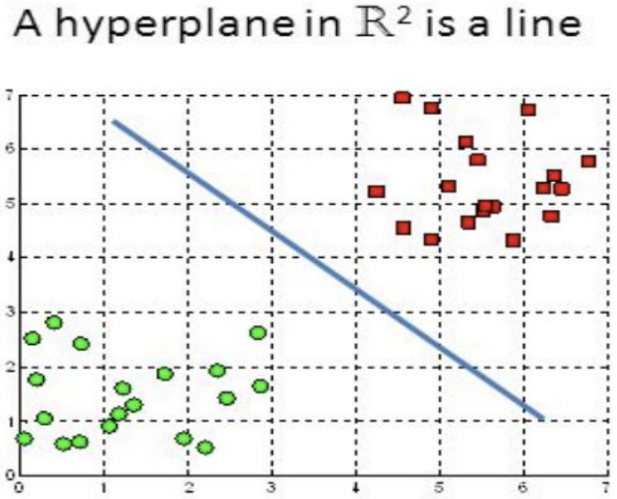
Support vector machine algorithm aims to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

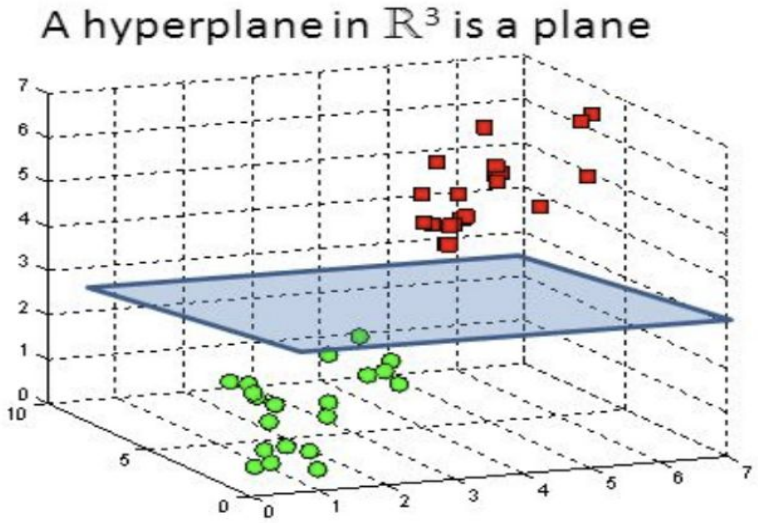




Support Vectors Machines SVM

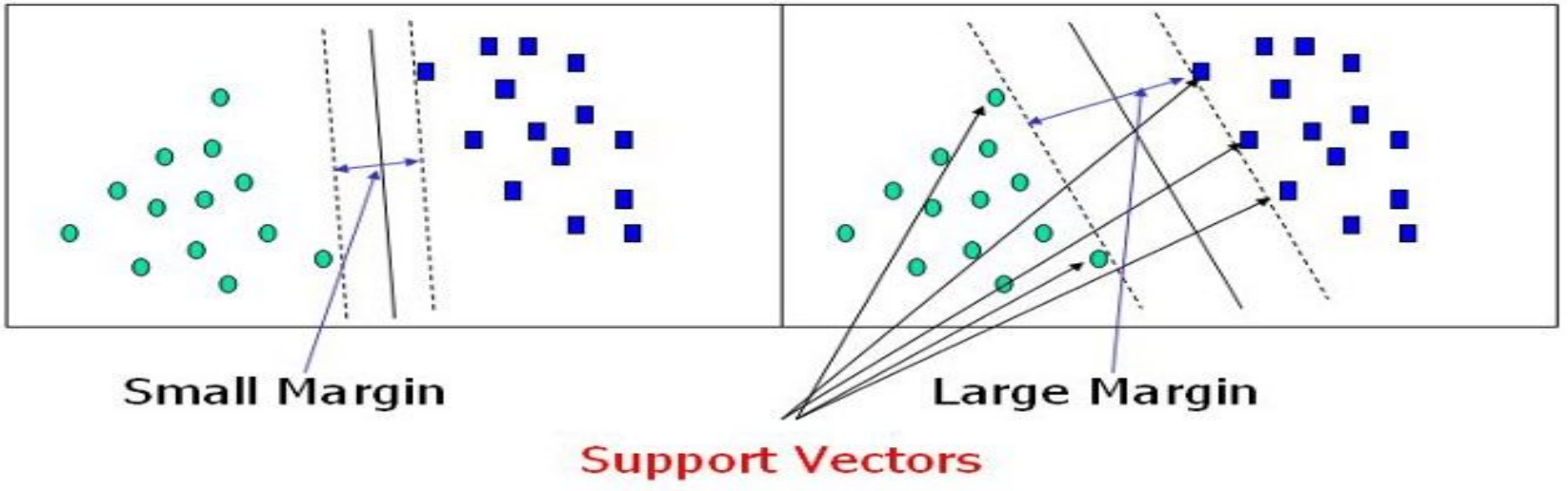
Hyperplanes





Support Vectors Machines SVM

Support Vectors and Margin



E.Coli DNA classifying

E. coli is a facultative anaerobic bacteria found in the intestine of vertebrates2. Indeed, it makes up about 80% of our aerobic intestinal flora.

Density-Based Spatial Clustering of Applications with Noise DBSCAN

- DBSCAN algorithm is based on the intuitive notion of “clusters” and “noise”.

- The key idea is that for each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points.

Why DBSCAN?

Other algorithmes are:

- affected by the presence of noise and outliers

  in the data
- suitable only for compact and well-separated cluster

DBSCAN Parameters

The DBSCAN algorithm basically requires 2 parameters:

- minPts: The minimum number of points (a threshold) clustered together for a region to be considered dense.

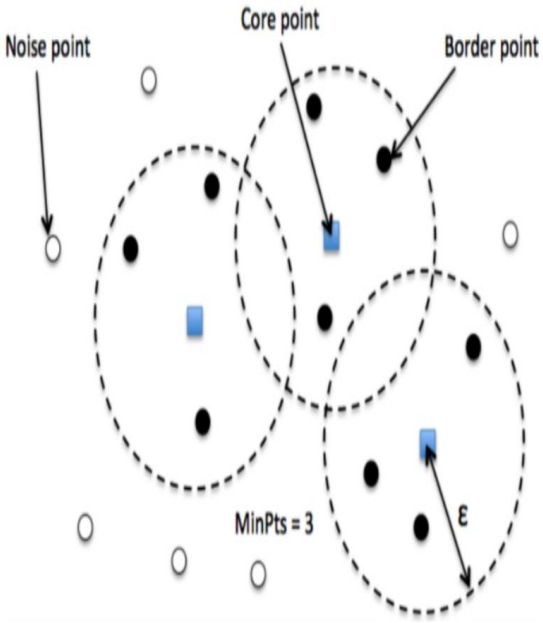
- eps (ε): A distance measure that will be used to locate the points in the neighborhood of any point.

DBSCAN output

Core: This is a point that has at least m points within distance n from itself.

Border: This is a point that has at least one Core point at a distance n

Noise: This is a point that is neither a Core nor a Border.
And it has less than m points within distance n from itself.

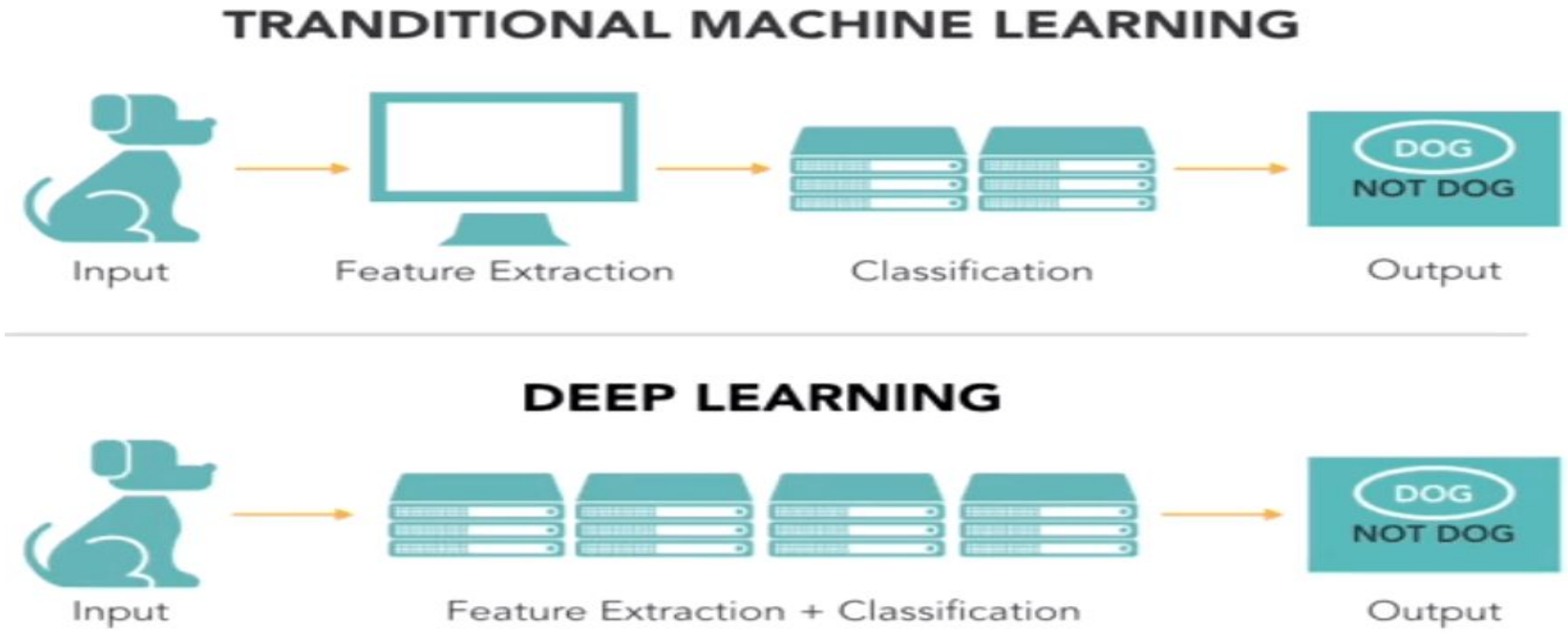


Contact Tracing

The process used to identify those who come into contact with people who have tested positive for contagious diseases, like Covid19.

During a pandemic, performing contact tracing correctly can help reduce the number of people to get infected or speed up the process of treating infected people. Doing so can help save many lives.

Conclusion


TRANDITIONAL MACHINE LEARNING
DOG
NOT DOG
Input
Feature Extraction
Classification
Output
DEEP LEARNING
DOG
NOT DOG
Input
Feature Extraction + Classification
Output

Refernces

- https://towardsdatascience.com/contact-tracing-using-less-than-30-lines-of-python-code-6c5175f5385f
- https://github.com/akshkshay/DNA-Sequences-Classification/blob/master/DNA_Sequences_Classification.ipynb
- https://dataanalyticspost.com/Lexique/svm/
- https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/