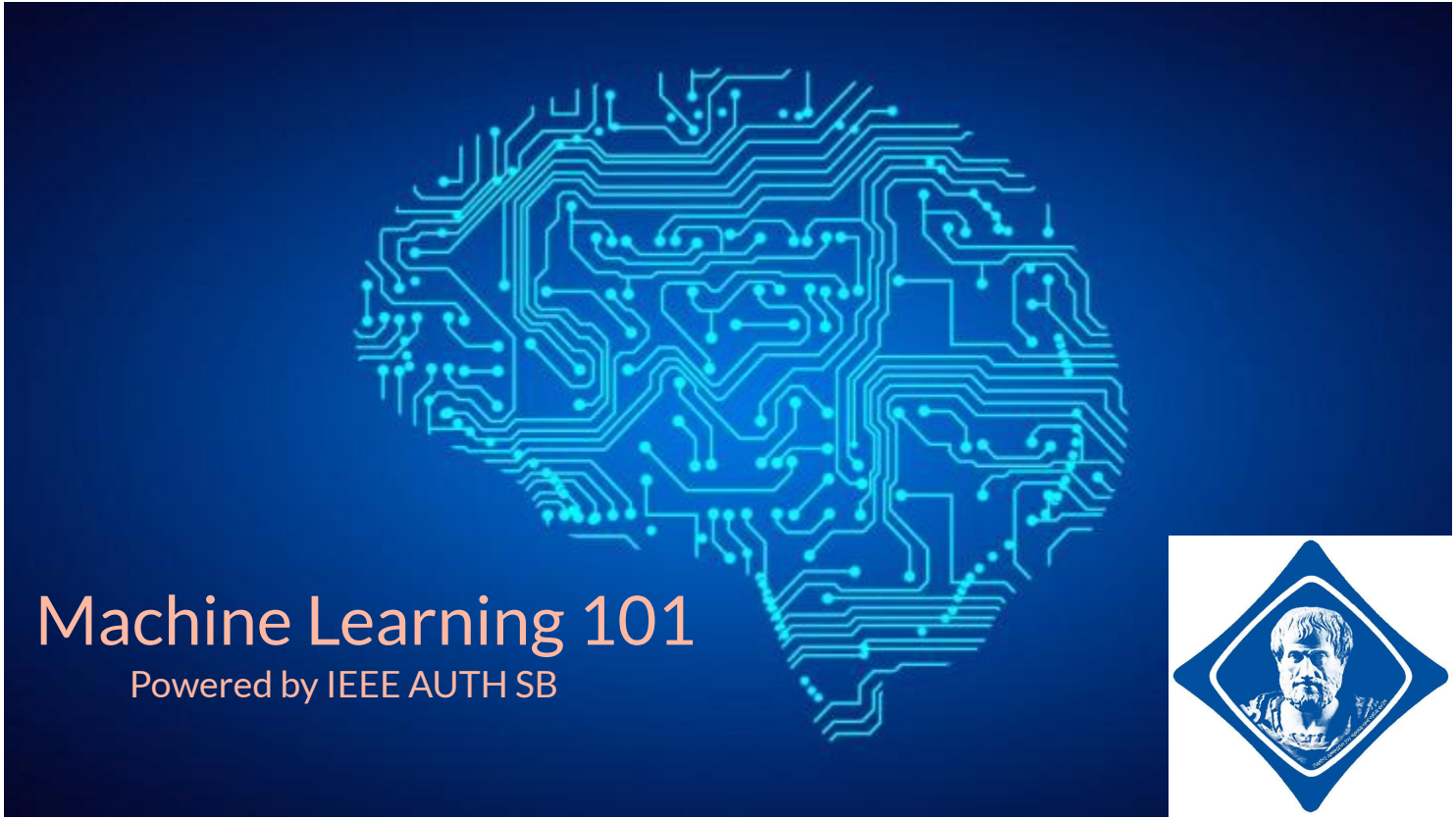

Machine Learning 101
Powered by IEEE AUTH SB

a

Artificial Intelligence

- The science engineering of making intelligent machines, especially intelligent computer programs.

  - John McCarthy, father of AI

- The development of AI started with the intention of creating similar intelligence in machines that we find and regard high in humans.

Machine Learning

- The ability to learn without being explicitly programmed.

- The main elements are statistical analysis and predictive analysis used to spot patterns and find hidden insights based on observed data from previous computations without being programmed on where to look.

ML Categories

- Supervised Learning.
- Unsupervised Learning.
- Reinforcement Learning.

Supervised Learning

- The group of algorithms that require dataset which consists of example input-output pairs. Each pair consists of data sample used to make prediction and expected outcome called label.
- Word “supervised” comes from a fact that labels need to be assigned to data by the human supervisor.
- Then they are used for making predictions on unknown data, that was not a part of training dataset.

Supervised Learning

- **Classification**
  Process of assigning category to input data sample. Example usages: predicting whether a person is ill or not, detecting fraudulent transactions, face classifier.
- **Regression**
  Process of predicting a continuous, numerical value for input data sample. Example usages: assessing the house price, forecasting grocery store food demand, temperature forecasting.

Unsupervised Learning

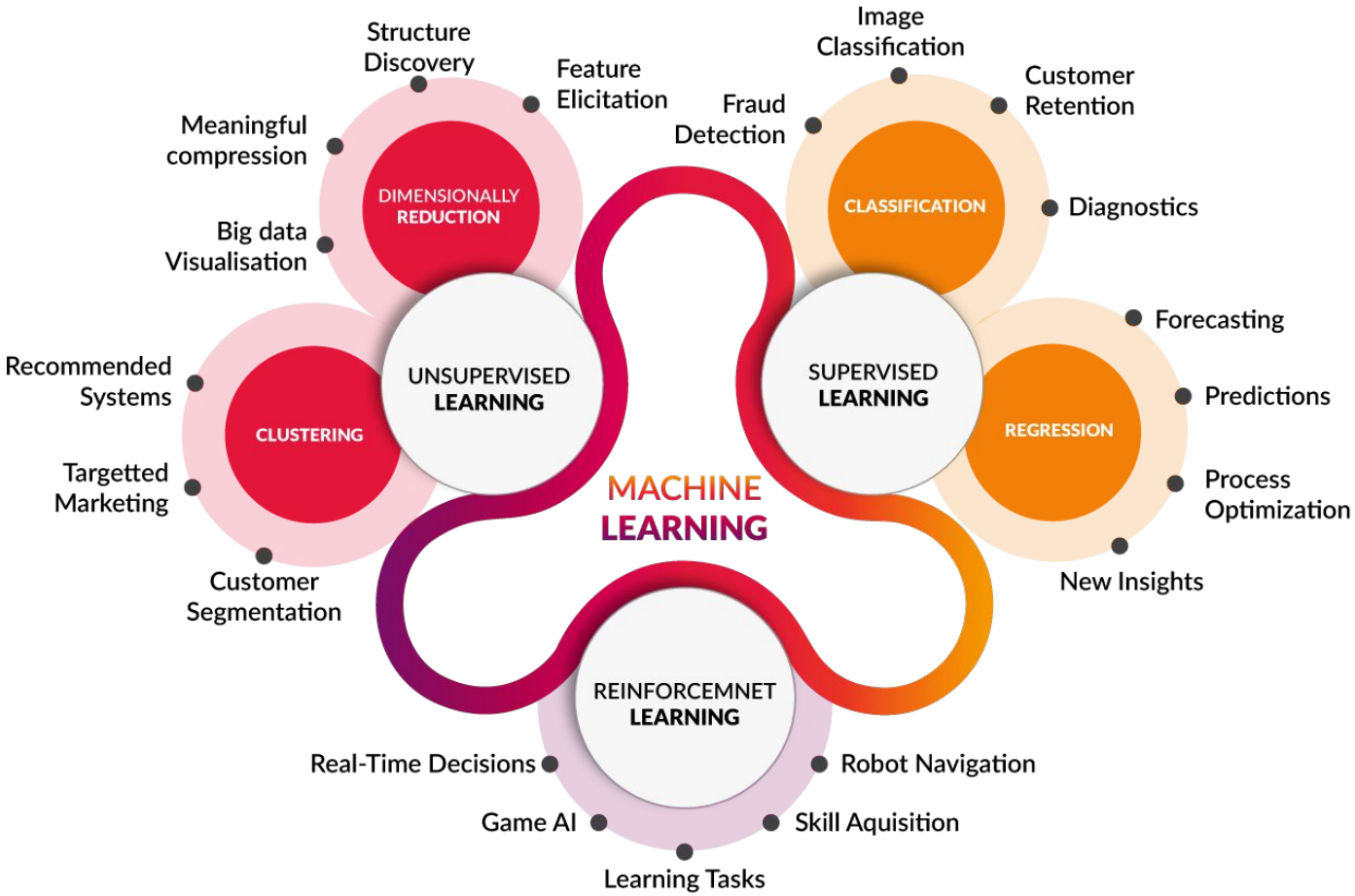
- Group of algorithms that try to draw inferences from non-labeled data (without reference to known or labeled outcomes).
- There are no correct answers.
- Models based on this type of algorithms can be used for discovering unknown data patterns and data structure itself.

Unsupervised Learning

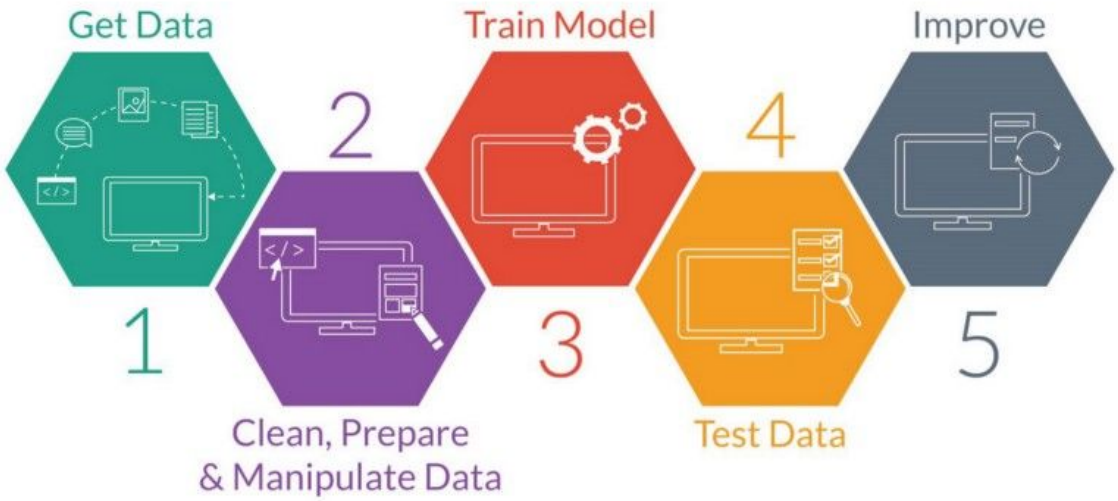
- **Clustering**
  Process of dividing and grouping similar data samples together. Example usages: segmentation of supermarkets or customers, data visualisation.
- **Dimensionality Reduction**
  A process of compressing features into so-called principal values which conveys similar information concisely. Example usages: speeding up other Machine Learning algorithms by reducing numbers of calculations, finding a group of most reliable features in data.

Reinforcement Learning

- Produces so-called agents.
- The agent role is to receive information from the environment and react to it by performing an action. The information is fed to an agent in form of numerical data, called state, which is stored and then used for choosing right action.
- As a result, an agent receives a reward that can be either positive or negative. The reward is a feedback that can be used by an agent to update its parameters.
- Trial and error training process.


MACHINE LEARNING
UNSUPERVISED LEARNING
SUPERVISED LEARNING
REINFORCEMNET LEARNING
DIMENSIONALLY REDUCTION
CLUSTERING
CLASSIFICATION
REGRESSION
Structure Discovery
Feature Elicitation
Meaningful compression
Big data Visualisation
Recommended Systems
Targetted Marketing
Customer Segmentation
Image Classification
Customer Retention
Fraud Detection
Diagnostics
Forecasting
Predictions
Process Optimization
New Insights
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

The Process


Get Data
Train Model
Improve
2
4
1
3
5
Clean, Prepare
& Manipulate Data
Test Data

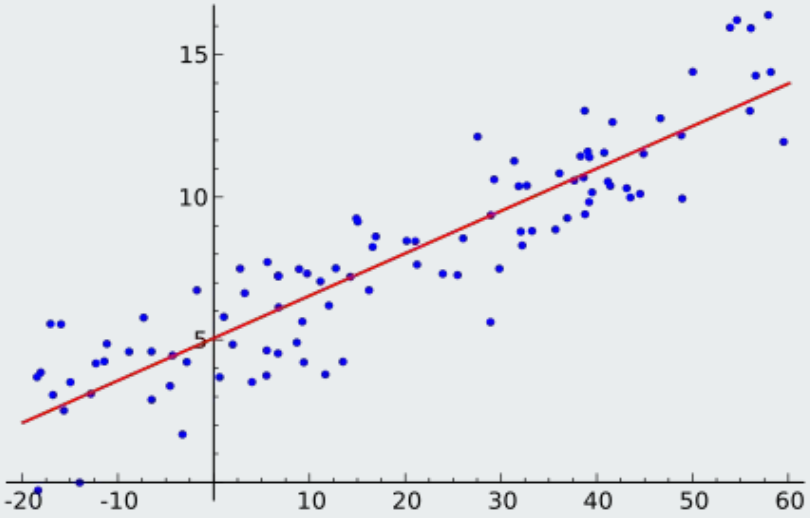
Questions?

Ένα απλό πρόβλημα

Έστω ότι θέλουμε να προβλέψουμε το κόστος ασφάλισης υγείας διαφόρων πελατών, ανάλογα με χαρακτηριστικά όπως η ηλικία, το φύλο, ο αριθμός των παιδιών κλπ.

Πώς?

Linear Regression



Ορισμός

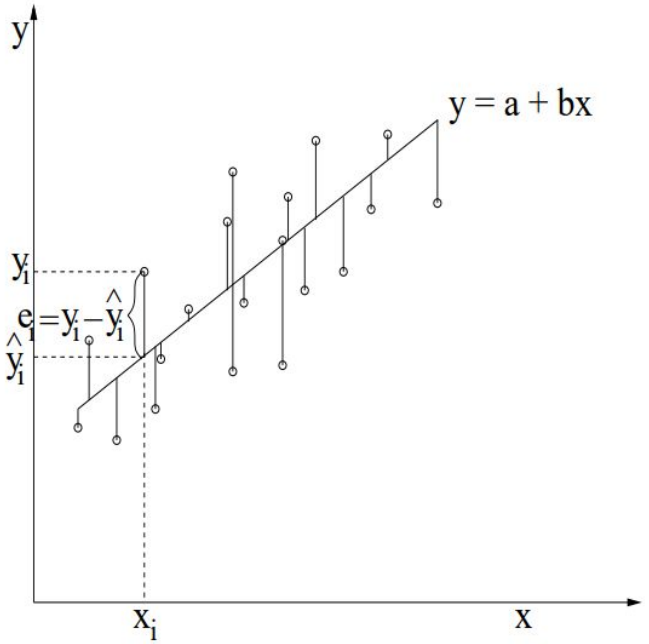
- Στην γραμμική παλινδρόμηση υποθέτουμε ότι μια εξαρτημένη μεταβλητή είναι γραμμικός συνδυασμός των ανεξάρτητων
- Διακρίνεται σε απλή και πολλαπλή
- Απλή λέγεται όταν έχουμε μόνο μία ανεξάρτητη μεταβλητή
- Πολλαπλή λέγεται όταν έχουμε τουλάχιστον δύο ανεξάρτητες μεταβλητές και ουσιαστικά αποτελεί γενίκευση της απλής

Απλή Γραμμική Παλινδρόμηση

$$\hat{y}_i = a + bx_i$$

Συνάρτηση κόστους:

$$\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$



Συνάρτηση κόστους

- Προκειμένου να αξιολογήσουμε πόσο ικανό είναι ένα μοντέλο να κάνει προβλέψεις, χρησιμοποιούμε μια συνάρτηση κόστους
- Μοντέλο πρόβλεψης στην πολλαπλή γραμμική:

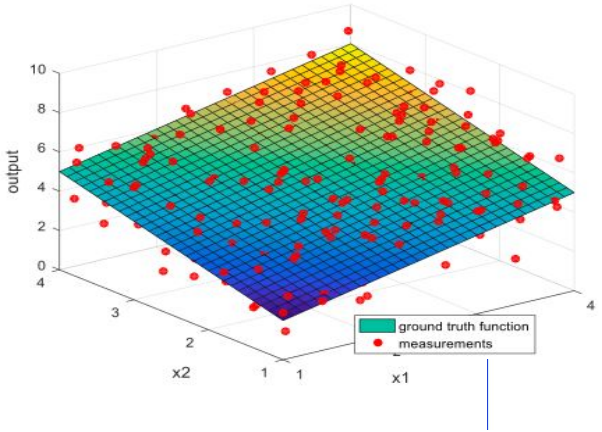
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

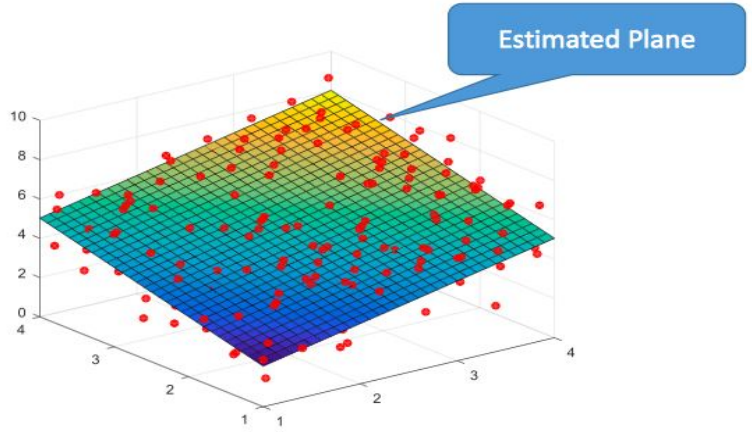
- Συνάρτηση κόστους:

$$S(\beta) = \sum_{i=1}^{m} \left| y_i - \sum_{j=0}^{n} X_{ij} \beta_j \right|^2$$

Γραμμική παλινδρόμηση πολλών μεταβλητών

- **Παραδείγματα:**





27

Στο παραπάνω διάγραμμα διασποράς οι κόκκινες τελείες αντιπροσωπεύουν τις παρατηρήσεις του dataset. Το επίπεδο που περνά ανάμεσα από αυτές είναι το μοντέλο πρόβλεψης. Η συνάρτηση κόστους υπολογίζει την κατακόρυφη απόσταση της καθε παρατήρησης από το επίπεδο και αθροίζει τα τετράγωνα των αποστάσεων αυτών.

Πίσω στο πρόβλημα - Dataset

Οι παρατηρήσεις που χρησιμοποιούμε οργανώνονται σε ένα αρχείο .csv

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Έστω λοιπόν ότι:
X_1: age
X_2: sex
X_3: bmi
X_4: children
X_5: smoker
Y: charges

Time for some code

You will need:

✓ numpy
✓ pandas
✓ scikit-learn

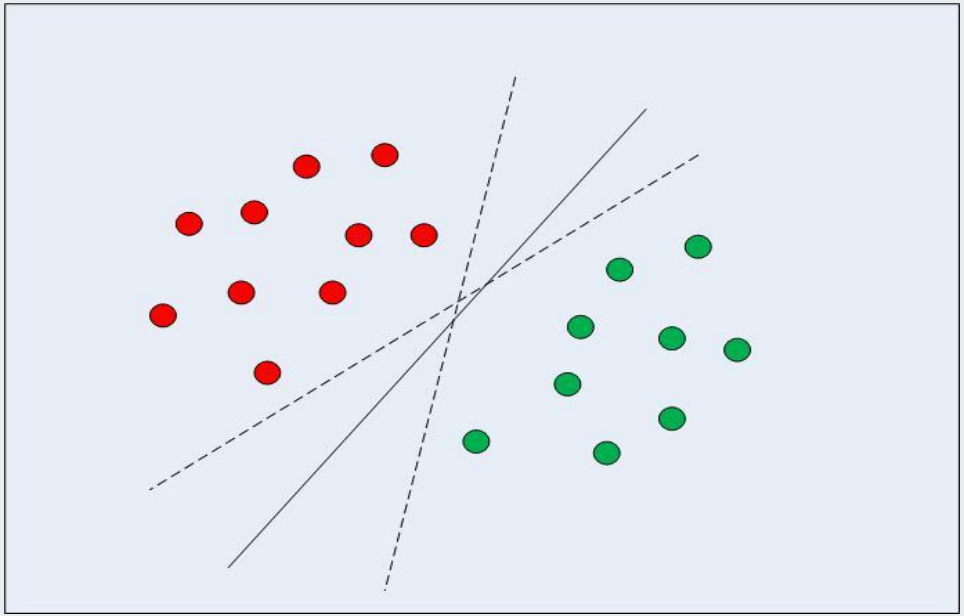
...και άλλο απλό πρόβλημα

Έστω τώρα ότι θέλουμε να μαντέψουμε αν ένας ασθενής θα εμφανιστεί σε ένα ραντεβού.

Τι γίνεται όμως όταν το πρόβλημα έχει σαν έξοδο διακριτές τιμές (παρών ή απών)?

	Gender	Age	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap	SMS_received	No-show
0	F	62	0	1	0	0	0	0	No
1	M	56	0	0	0	0	0	0	No
2	F	62	0	0	0	0	0	0	No
3	F	8	0	0	0	0	0	0	No
4	F	56	0	1	1	0	0	0	No

Classification - SVM



Support Vector Machines

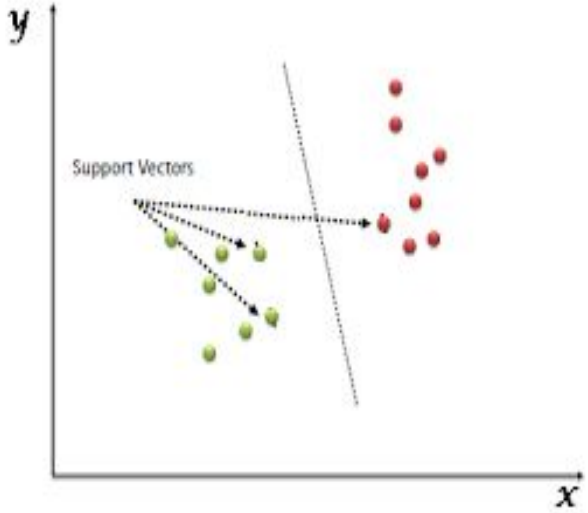
Είναι supervised machine learning αλγόριθμος που χρησιμοποιείται κυρίως για classification

- Κάθε εγγραφή του Dataset αναπαρίσταται σαν σημείο στον N-διάστατο χώρο (όπου N είναι ο αριθμός των features της κάθε εγγραφής).
- Η τιμή του κάθε feature είναι η συντεταγμένη της εκάστοτε εγγραφής.
- Το classification γίνεται διαχωρίζοντας τα σημεία με το “ καλύτερο “ δυνατό υπερεπίπεδο (decision boundary).

Τρόπος Λειτουργίας

Σκοπός του αλγορίθμου είναι να κατηγοριοποιήσει τα δεδομένα σε μία από τις δύο κλάσεις.

- Ότι είναι δεξιά του υπερεπιπέδου “ανήκει” στην “κόκκινη” κλάση και αριστερά στην “πράσινη” κλάση.
- Το υπερεπίπεδο δημιουργείται με βάση τα support vectors.



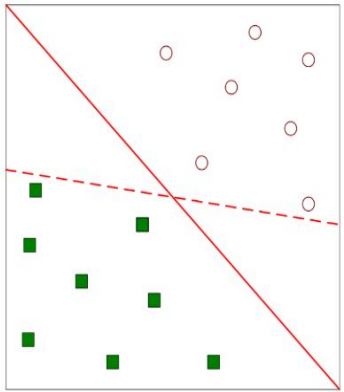
Τρόπος Λειτουργίας (2)

Στην αρχή αναφέραμε ότι ο SVM βρίσκει το ‘καλύτερο’ δυνατό υπερεπίπεδο διαχωρισμού.

- Τι εννοούμε ‘καλύτερο’?
- Πως προσδίδουμε σε ένα επίπεδο μία αξιολόγηση (π.χ. καλό)?
- Πως είμαστε σίγουροι ότι για δεδομένο dataset υπάρχει το ‘καλύτερο’ υπερεπίπεδο?

Δύο πιθανά υπερεπίπεδα που διαχωρίζουν τις κλάσεις φαίνονται δίπλα.

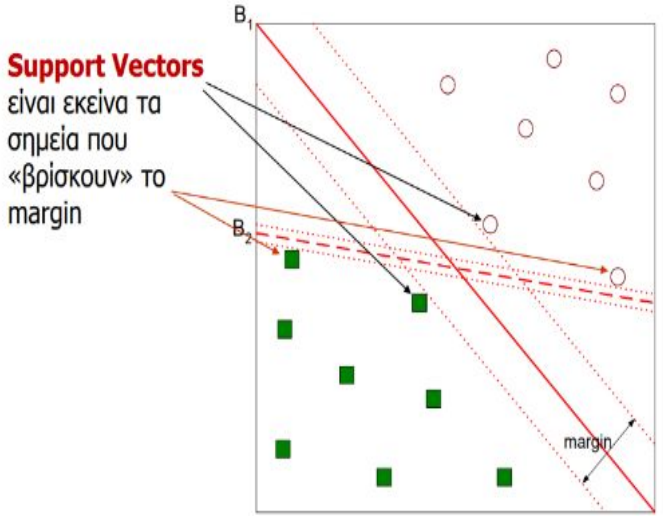
Ποιο από τα δύο θα διαλέγατε?



Τρόπος Λειτουργίας (3)

Προκειμένου να μπορέσουμε να αξιολογήσουμε ένα επίπεδο χρειαζόμαστε μία μετρική αξιολόγησης.

- Για αυτό χρησιμοποιείται η έννοια του **Margin** (Περιθώριο) ως:
  η ελάχιστη απόσταση ενός σημείου από το διαχωριστικό επίπεδο
- Όλα τα σημεία που επιτυγχάνουν την ελάχιστη αυτή απόσταση ονομάζονται **Support Vectors**



Questions?

Useful Links

- Python: https://docs.python.org/3/tutorial/
- Machine Learning info for beginners: https://medium.com/machine-learning-for-humans/why-machine-learning-matters-6164faf1df12
- Find more datasets here: https://www.kaggle.com/
- An SVM example: https://stackabuse.com/implementing-svm-and-kernel-svm-with-pythons-scikit-learn/

Thank you!