Μηχανική Μάθηση

Μιχάλης Τίτσιας

Διάλεξη 1ή

Βασικές έννοιες μηχανικής μάθησης

Περιεχόμενα

- Τι είναι μηχανική μάθηση
- Παραδείγματα εφαρμογών μηχανικής μάθησης
- Γενικές τεχνικές μάθησης
- Μάθηση με επίβλεψη
- Μάθηση χωρίς επίβλεψη
- Δεδομένα εκπαίδευσης και δεδομένα ελέγχου
- Η γενική δομή ενός συστήματος μηχανικής μάθησης
- Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

Τι είναι μηχανική μάθηση

Για απλά προβλήματα στην επιστήμη των υπολογιστών (π.χ. εύρεση του συντομότερου μονοπατιού μεταξύ δύο κόμβων ενός δικτύου κτλ), γνωρίζουμε πως να συντάξουμε ένα αλγόριθμο που να τα επιλύει

Για πιο σύνθετα προβλήματα δεν γνωρίζουμε πως να συντάξουμε ένα αλγόριθμο

- Πώς θα γράφαμε ένα πρόγραμμα που να αναγνωρίζει spam emails;
- Να ομαδοποιεί άρθρα στο διαδίκτυο (δες google news);
- Να προτείνει βιβλία, ταίνιες, κτλ σε ένα χρήστη του διαδικτύου (δες Amazon, Netflix);
- Να αναγνωρίζει χειρόγραφους χαρακτήρες, εικόνες, φωνή, φυσική γλώσσα;
- κτλ

Η μηχανική μάθηση ασχολείται με την επιλυση τέτοιων σύνθετων προβλημάτων

Τι είναι μηχανική μάθηση

Ορισμός: Η μηχανική μάθηση ασχολείται με την ανάπτυξη αλγορίθμων που μαθαίνουν (δηλ. βελτιώνουν την επίδοση τους) από δεδομένα ή εμπειρίες

Πιο απλά: μηχανική μάθηση $\Rightarrow$ Αυτοματοποιημένη μάθηση από δεδομένα (learning from data)

(*Οι άνθρωποι μαθαίνουν συνέχως. Πώς μπορούμε να κατασκευάσουμε αυτοματοποιημένες μεθόδους που κάνουν τον ίδιο;*)

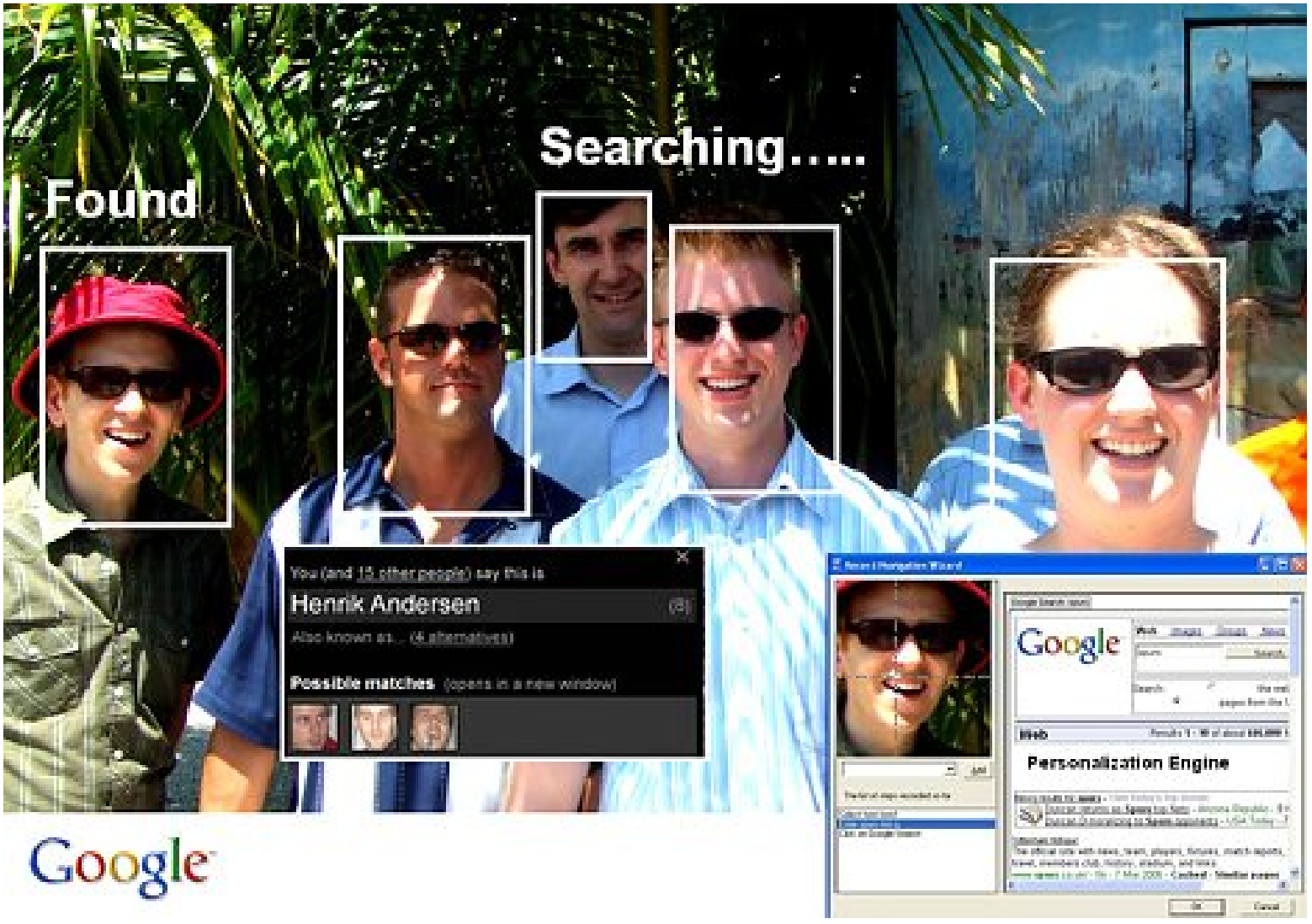
Προεύλεση του τομέα: Τεχνικές μηχανικής μάθησης αρχικά επηρεάστηκαν από την νευροεπιστήμη και τα νευρωνικά δίκτυα Τα τελευταία χρόνια η μηχανική μάθηση έχει έρθει κοντά στην επιστήμη της στατιστικής

- Συχνά λέγεται ότι η μηχανική μάθηση είναι στατιστική ΄κατασκευασμενη΄ από επιστήμονες των υπολογιστών ή της πληροφορικής!

Παραδείγματα εφαρμογών μηχανικής μάθησης

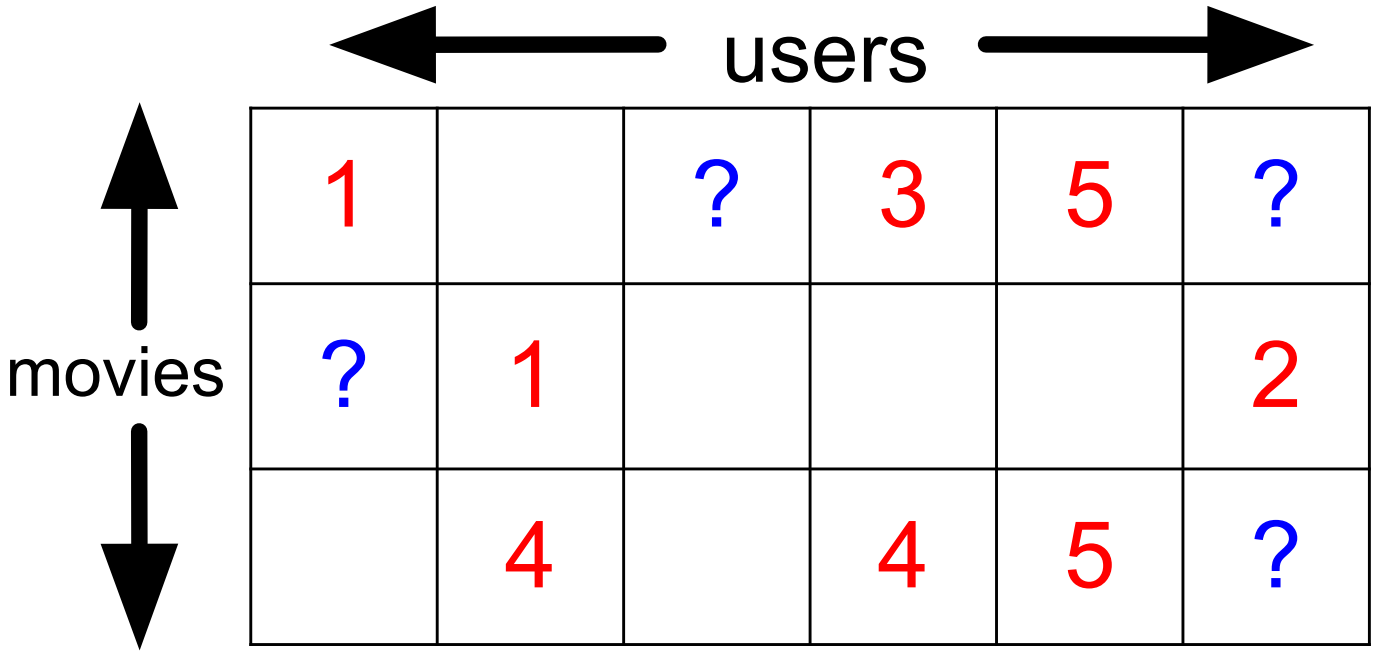
Αναγνώριση χειρόγραφων χαρακτήρων

Παραδείγματα εφαρμογών μηχανικής μάθησης



Αναγνώριση προσώπων σε εικόνες

Παραδείγματα εφαρμογών μηχανικής μάθησης



Recomendation systems

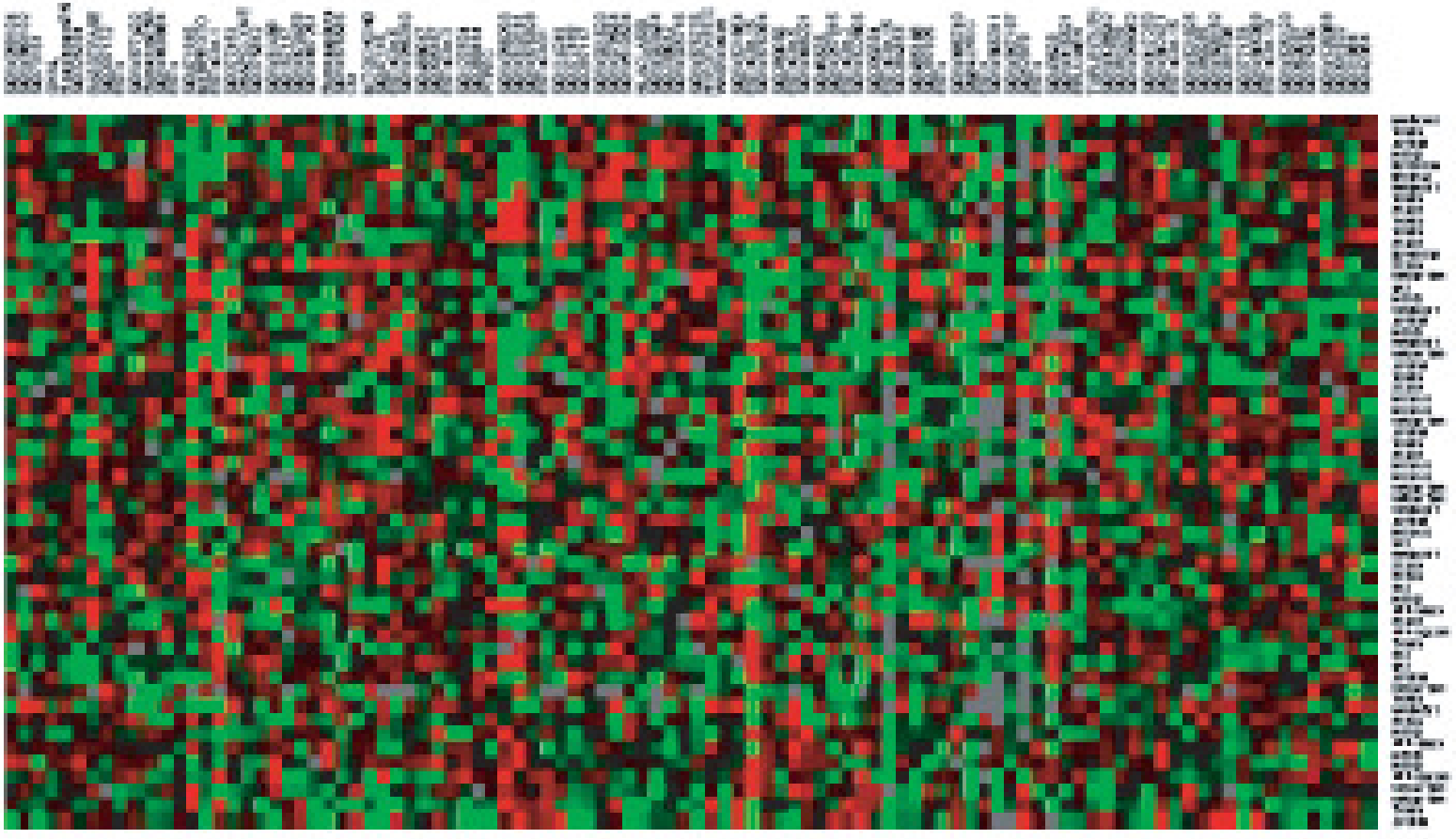
Παραδείγματα εφαρμογών μηχανικής μάθησης

Μάθηση της δεξιότητας παιχτών (π.χ. σε παιχνίδια με δύο παίχτες στο διαδίκτυο) και κατάταξη τους

Πολλές εφαρμογές με σύστηματα που είναι πλέον στην αγορά (π.χ. xbox live κα)

- Μάθηση από δεδόμενα που καθορίζουν το αποτελέσμα ενός παιχνιδιού
- Δυναμική στο χρόνο ενήμερωση της δεξιότητας του κάθε παίχτη
- Πρόβλεψη: Αυτοματοποιημένη σύσταση αντιπάλων σε παίχτες ώστε τα παιχνίδια που προκύπτουν να έχουν ενδιαφέρον για τους παίχτες

Παραδείγματα εφαρμογών μηχανικής μάθησης



Επεξεργάσια δεδομένων στη βιολογία και τη γονιδιωματική επιστήμη

Παραδείγματα εφαρμογών μηχανικής μάθησης

Πάρα πολλές άλλες εφαρμογές

- Επεξεργασία φωνής
- Επεξεργασία φυσικής γλώσσας
- Υπολογιστική όραση
- Αυτόματα συστήματα πλοήγησης: Πώς ένα ρομποτικό όχημα πλοηγείται σε άγνωστο περιβάλλον;
- Δεδομένα στο διαδίκτυο: web click data, placement of adverts, search engines

Γενικές τεχνικές μάθησης

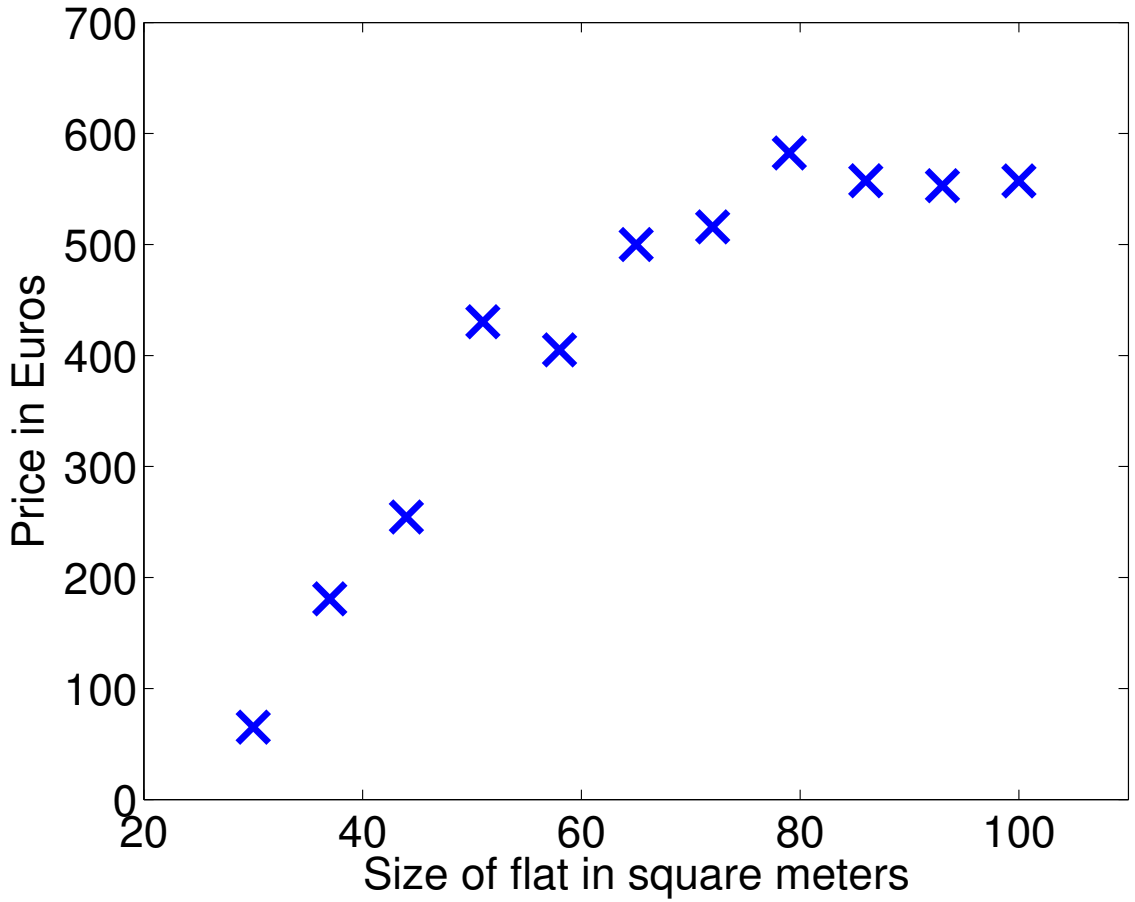
Μάθηση με επίβλεψη (supervised learning): Μάθηση ενός συστήματος που προβλέπει μια μεταβλητή εξόδου (output) δοθέντος μιας μεταβλητής εισόδου (input)

Μάθηση χωρίς επίβλεψη (unsupervised learning): Εύρεση δομής (structure/pattern) στα δεδομένα

Ενισχυτική μάθηση (reinforcement learning): Μάθηση με εξερεύνηση σε άγνωστο περιβάλλον

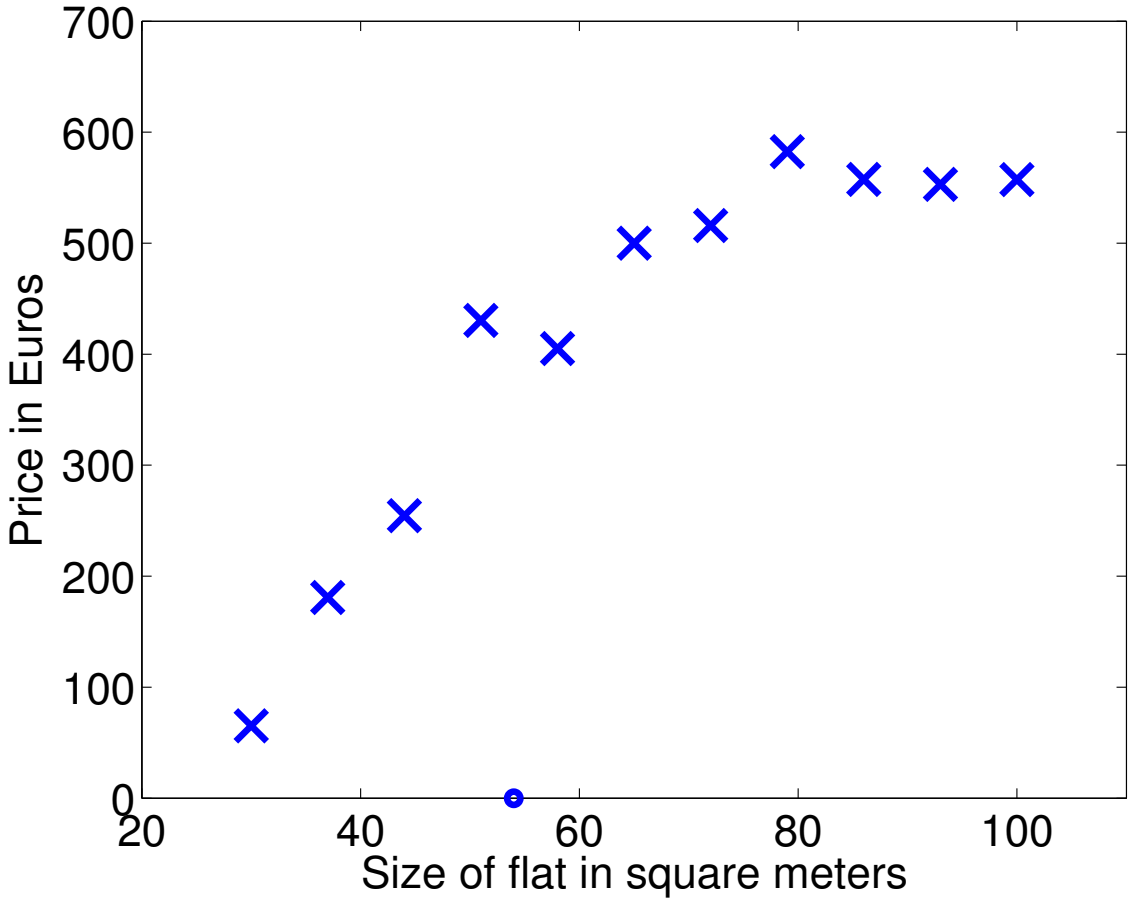
- χρησιμοποιείται στην ρομποτική

Μάθηση με επίβλεψη



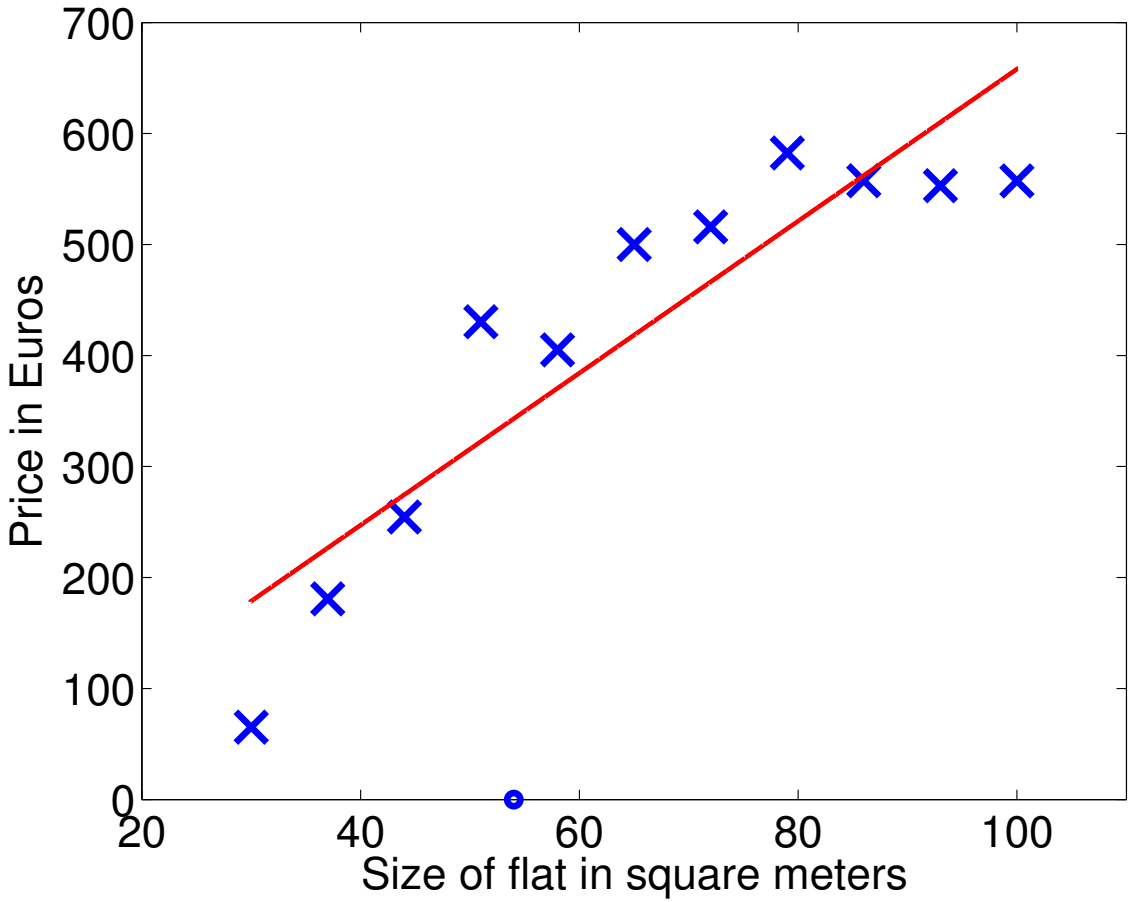
Δεδομένα τιμών ενοικίου διαμερισμάτων σε σχέση με το μέγεθος σε τετραγωνικά μέτρα

Μάθηση με επίβλεψη



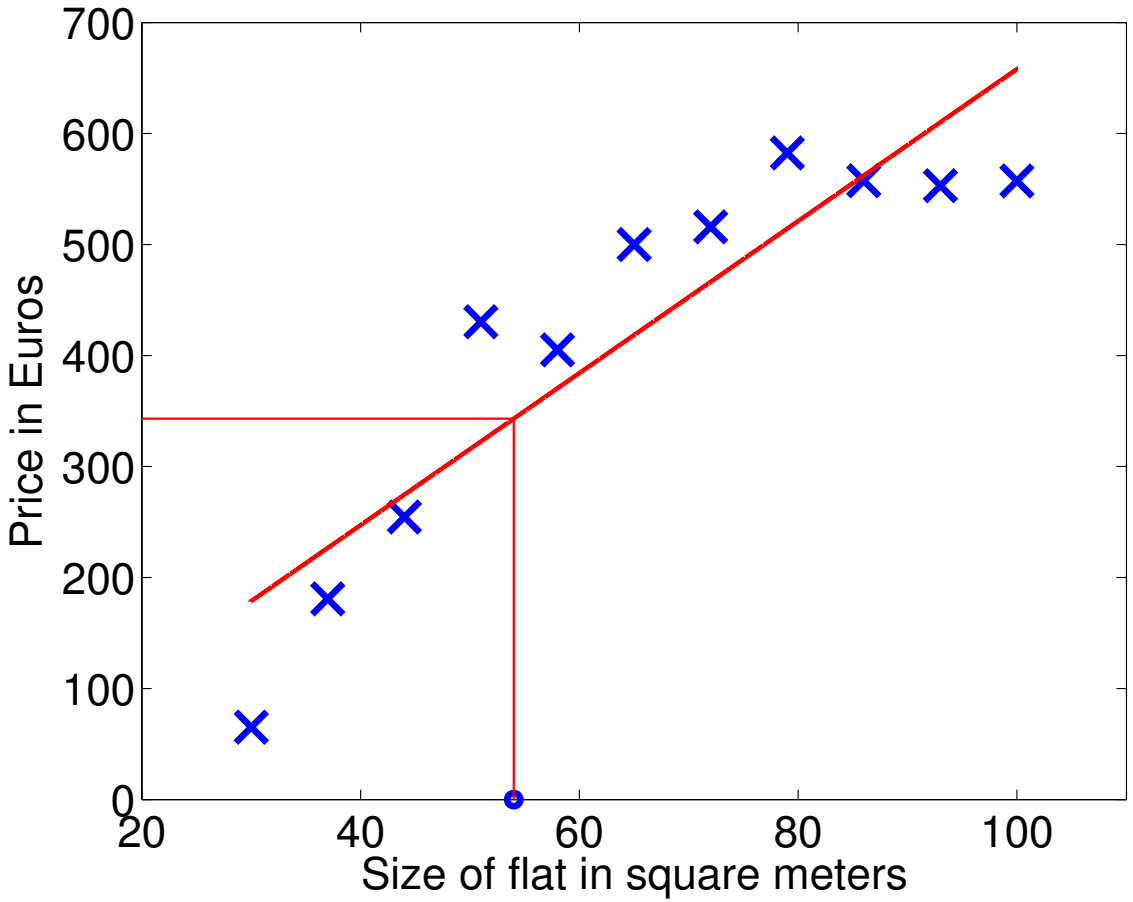
Πώς θα μπορούσαμε να προβλέψουμε την τιμή του ενοικίου για ένα διαμέρισμα με $54m^2$ (μπλε κουκκίδα);

Μάθηση με επίβλεψη



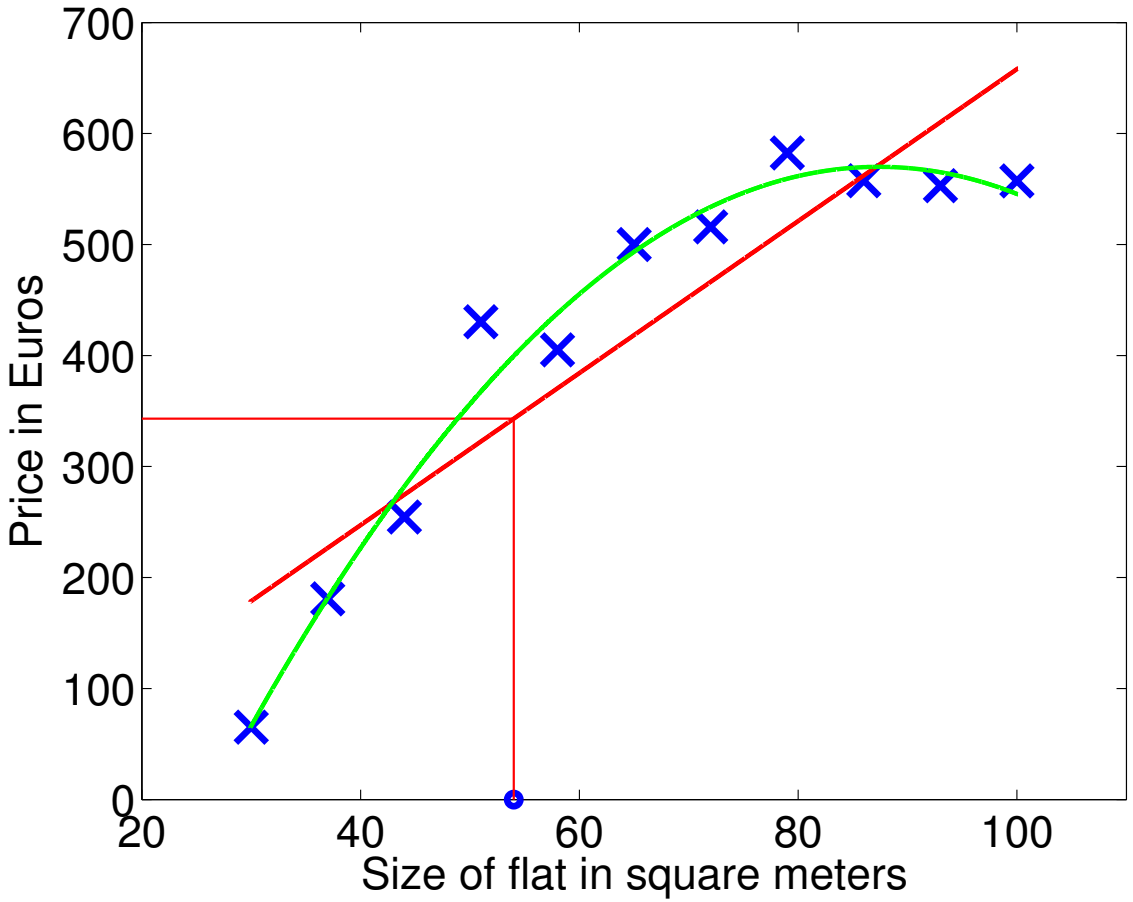
θα μπορούσαμε να ταιριάξουμε μια ευθεία γραμμή στα αρχικά δεδομένα

Μάθηση με επίβλεψη



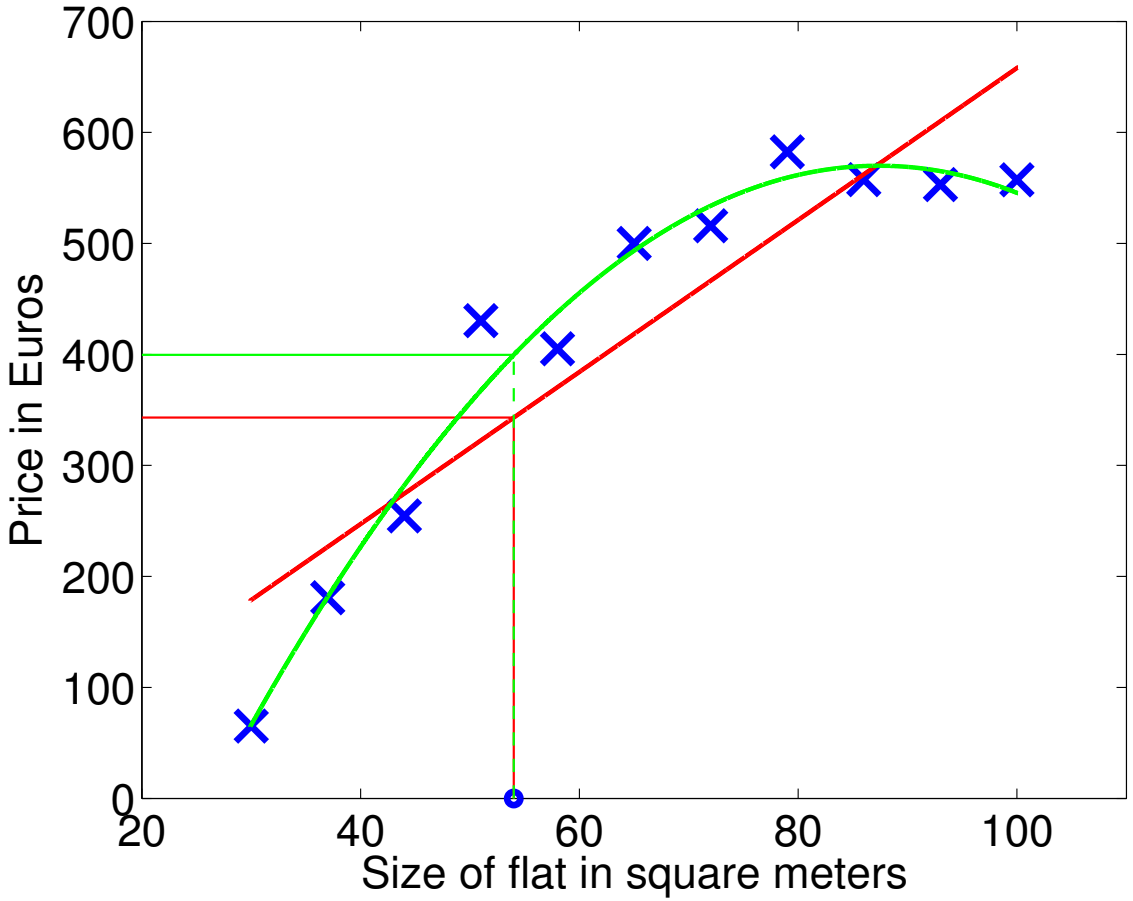
και έπειτα να κάνουμε προβλέψεις με βάση την γραμμή που βρήκαμε

Μάθηση με επίβλεψη



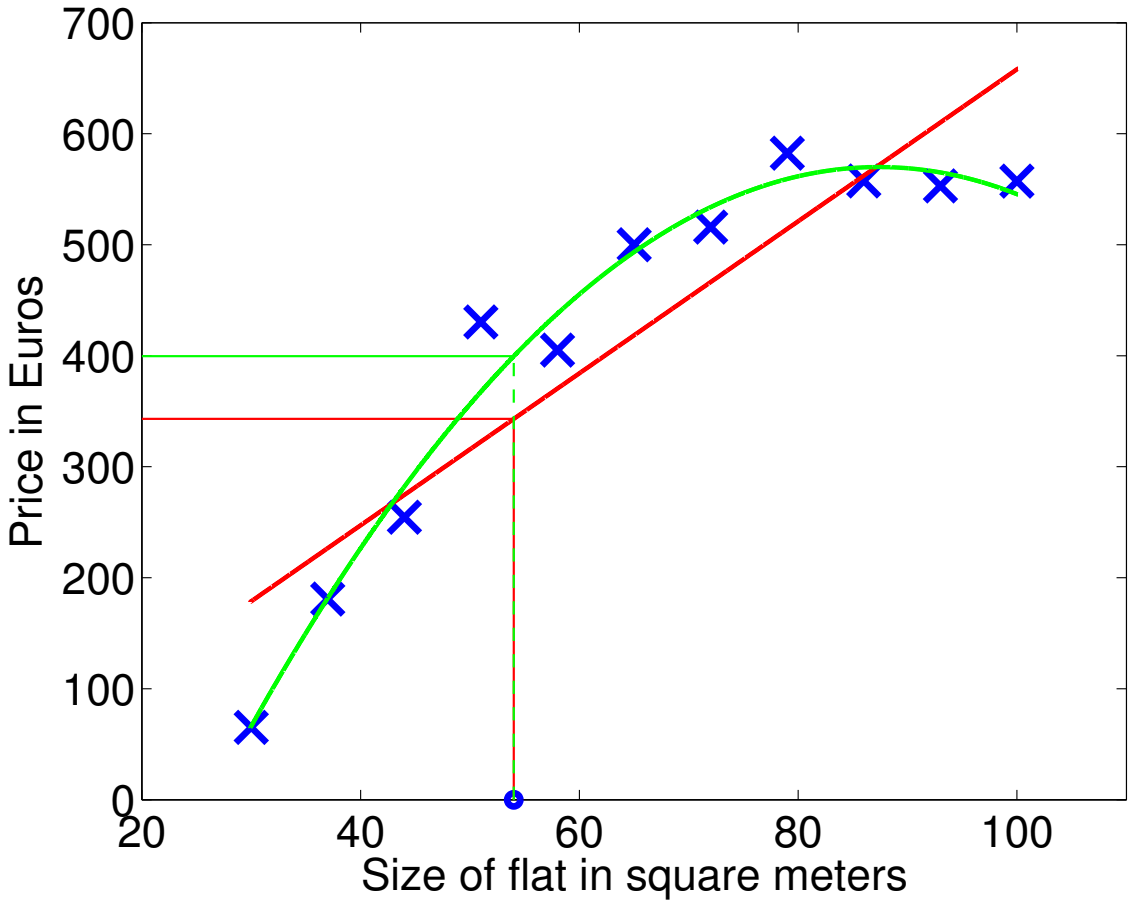
Εναλλακτικά θα μπορούσαμε να ταιριάξουμε μια τετραγωνική συνάρτηση

Μάθηση με επίβλεψη



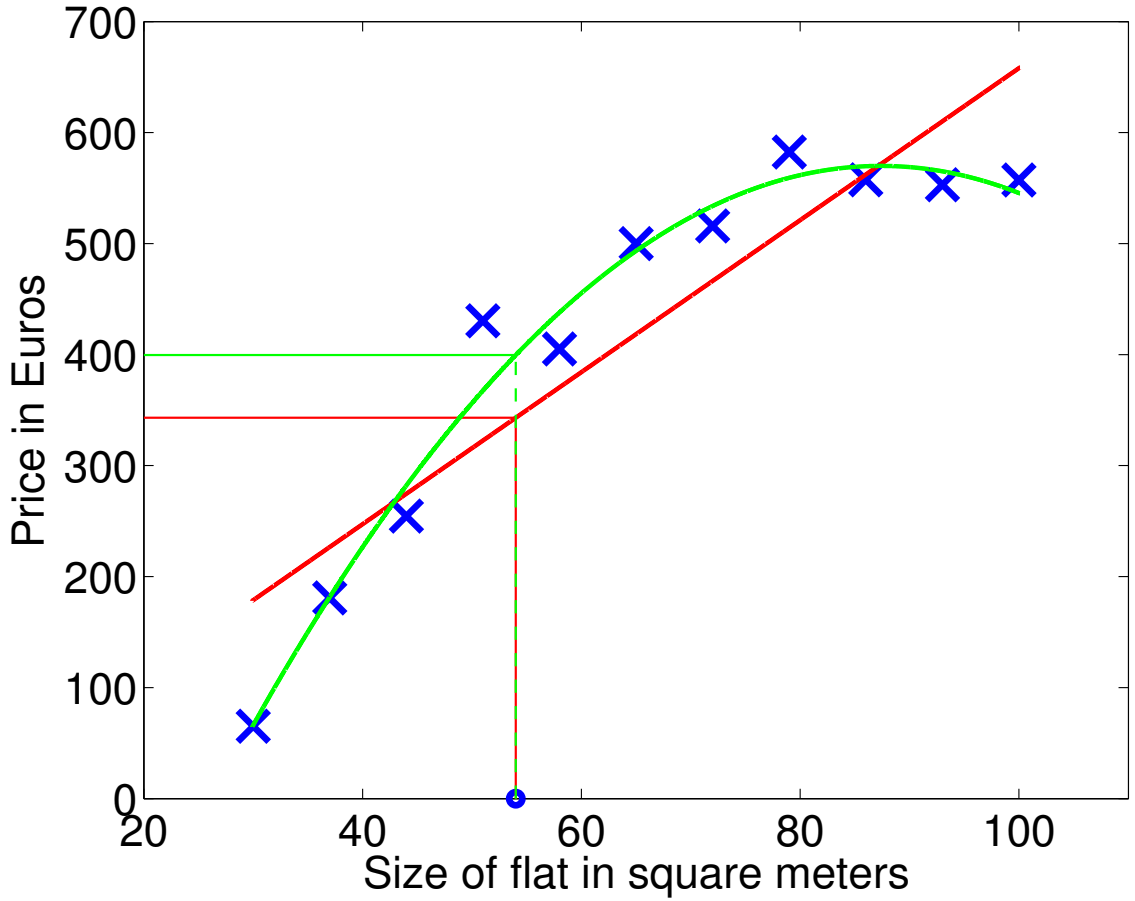
η οποία δίνει διαφορετική πρόβλεψη

Μάθηση με επίβλεψη



Το πρόβλημα αυτό αποτελεί πρόβλημα μάθησης με επίβλεψη γιατί στα αρχικά δεδομένα μας δίνεται αυτό που θέλουμε να προβλέψουμε, δηλ. μας δίνονται οι τιμές των ενοικίων μαζί με τα μεγέθη των διαμερισμάτων

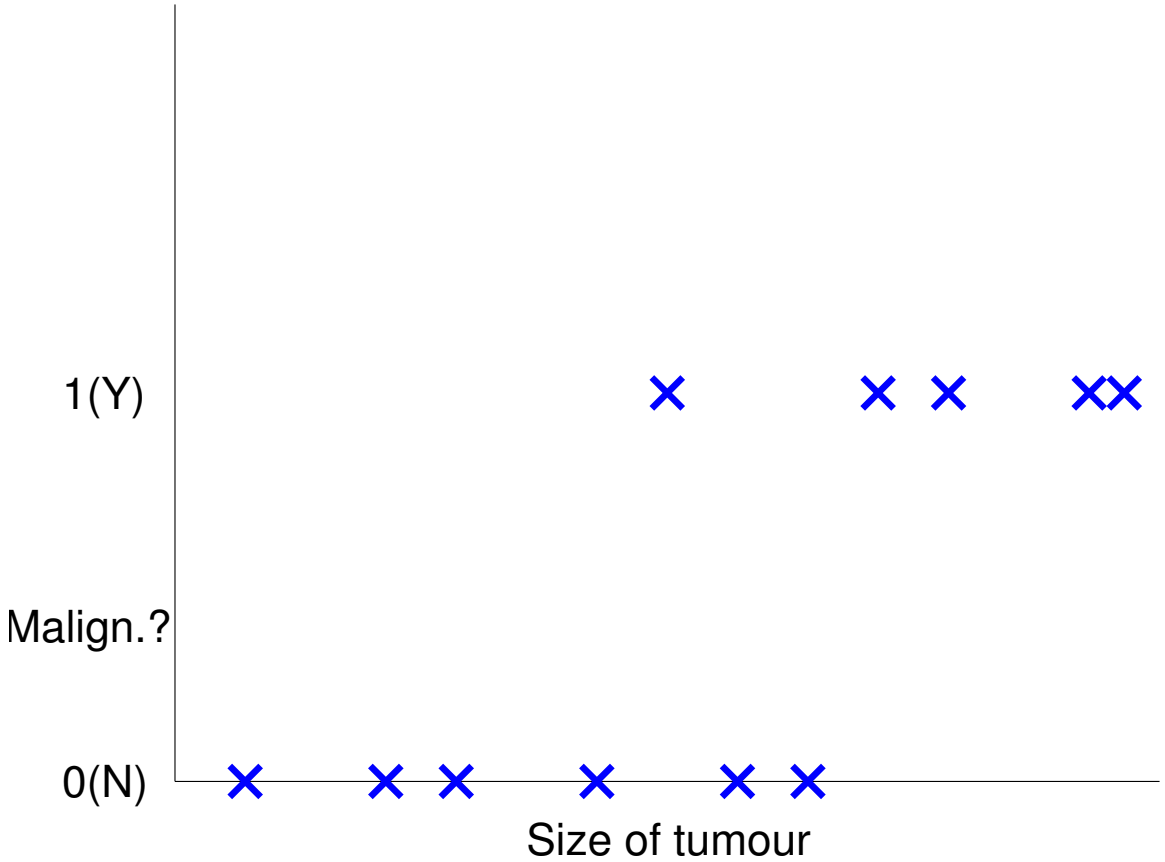
Μάθηση με επίβλεψη



Πιο συγκεκριμένα το πρόβλημα αυτό ονομάζεται παλινδρόμηση (regression)

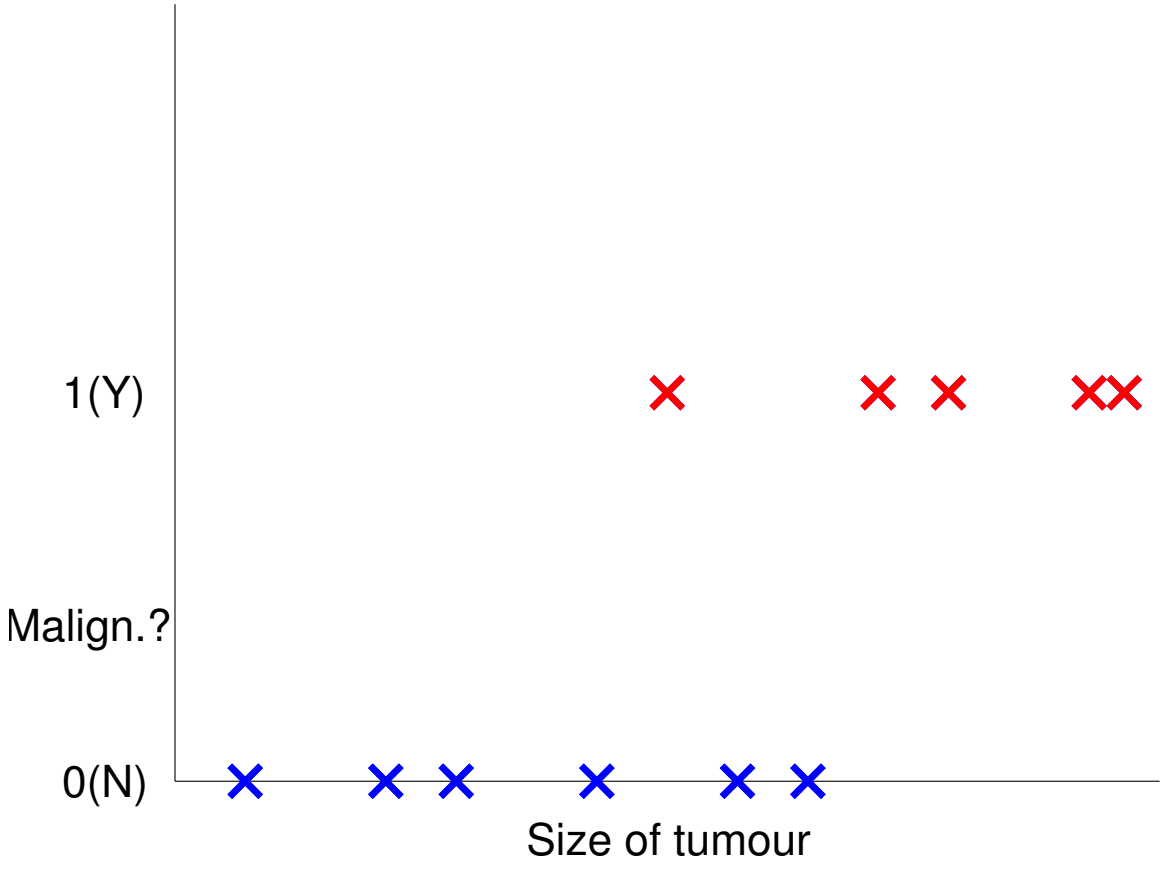
- λόγω του ότι αυτό που θέλουμε να προβλέψουμε παίρνει συνεχείς τιμές

Μάθηση με επίβλεψη



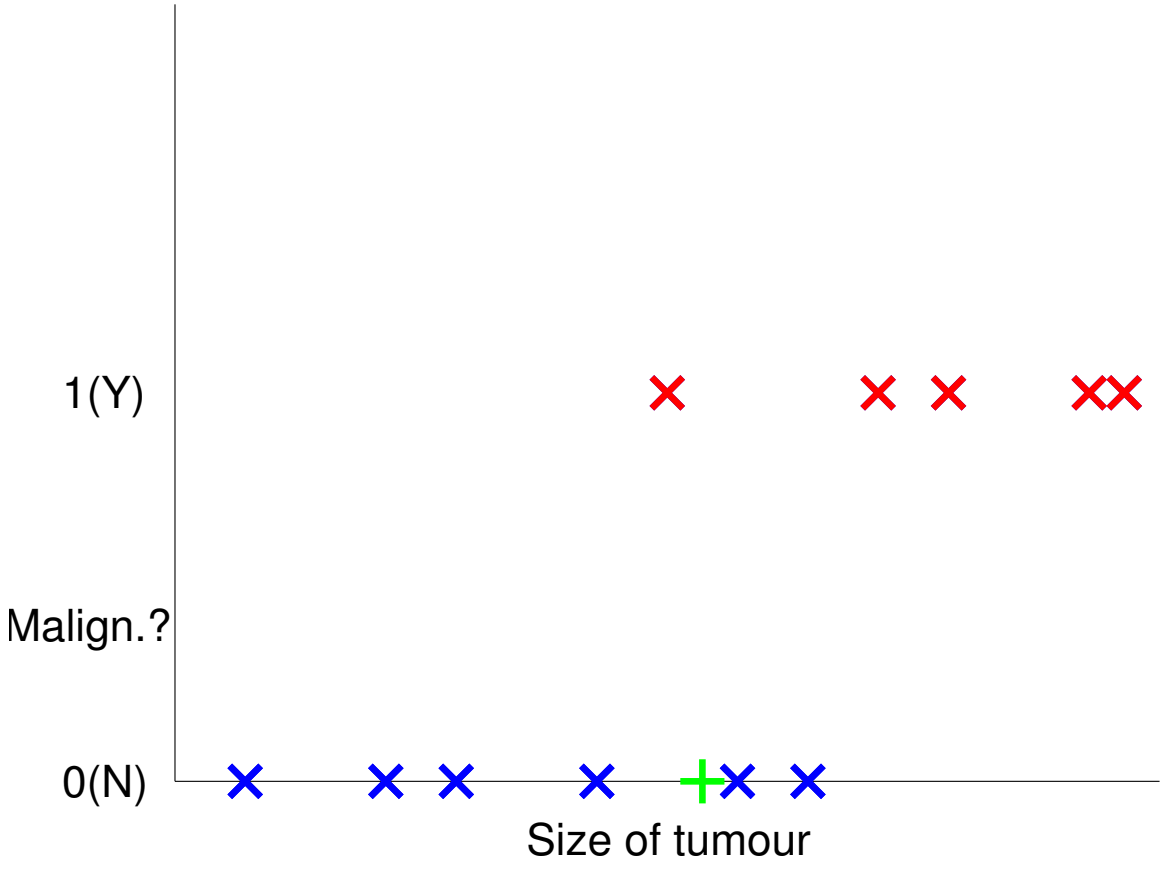
Έστω δεδομένα ασθενών ώστε στον οριζόντιο άξονα (δεδομένα εισόδου) δίνεται το μέγεθος ενός «ενδεχομένως» καρκινικού όγκου και στο κάθετο άξονα (δεδομένα εξόδου) 0 ή 1 για τις δύο περιπτώσεις, δηλ. καλοήθης ή κακοήθης

Μάθηση με επίβλεψη



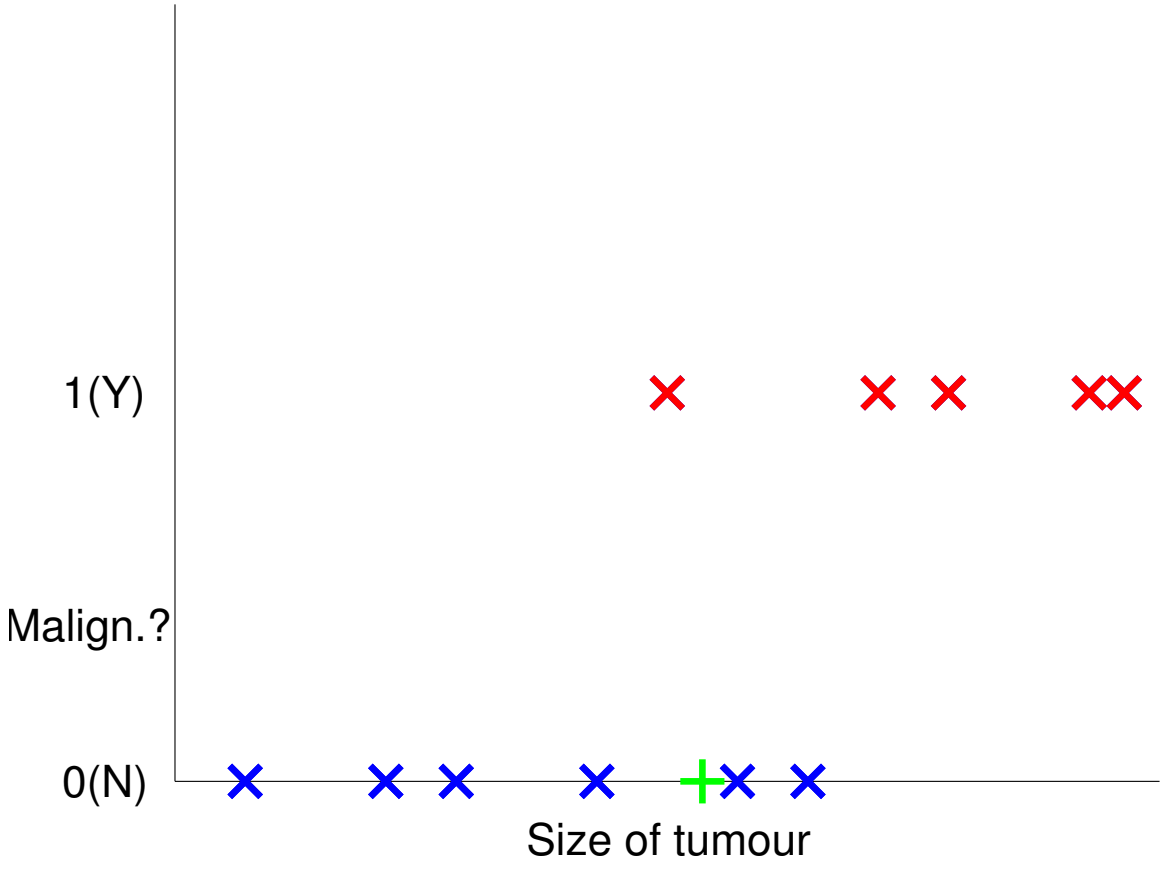
Κάθε δεδομένο ανήκει σε μια κατηγορία, δηλ στην κατηγορία 0 ή την κατηγορία 1

Μάθηση με επίβλεψη



Πρόβλημα μηχανικής μάθησης: Για ένα νέο όγκο πως μπορούμε να εκτιμήσουμε την πιθανότητα να είναι καλοήθης ή κακοήθης;

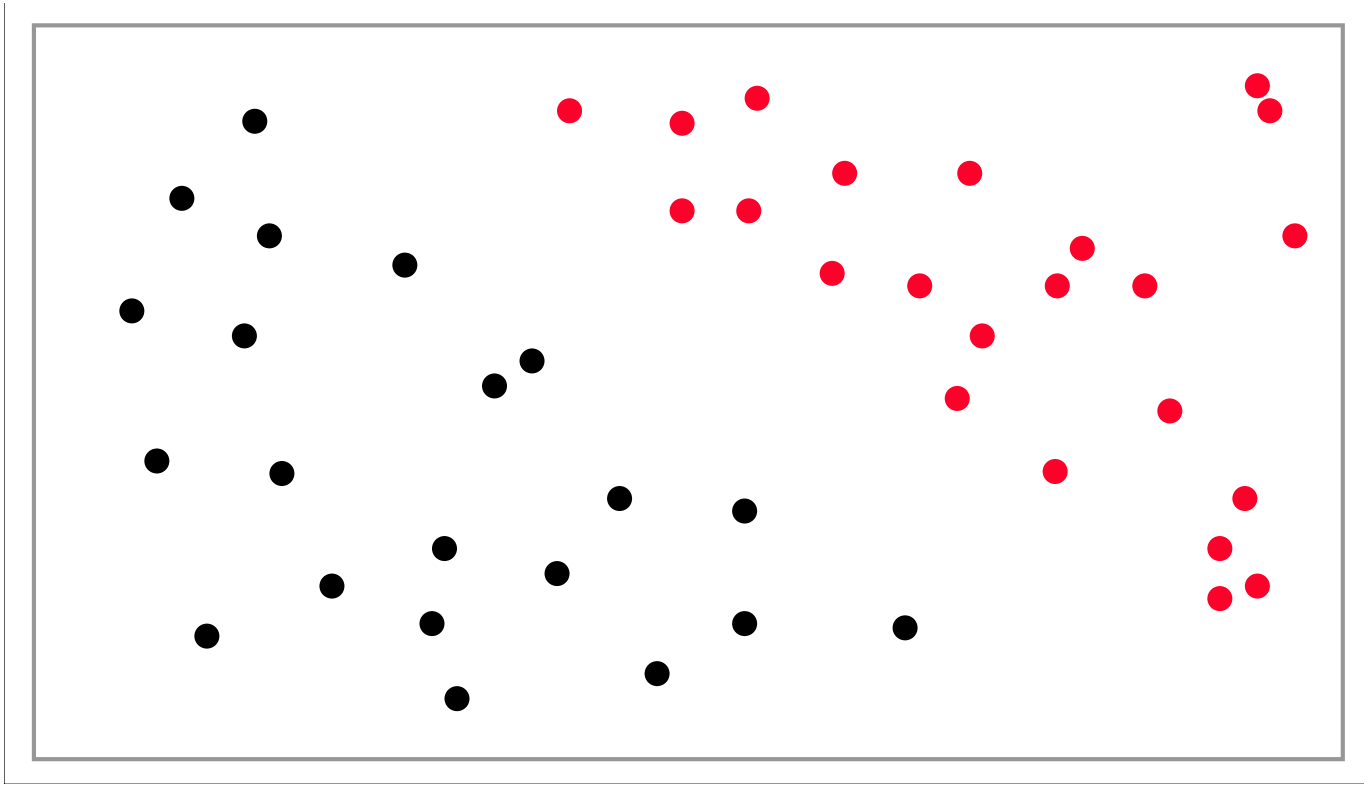
Μάθηση με επίβλεψη



Το πρόβλημα αυτό ονομάζεται κατηγοριοποίηση δεδομένων

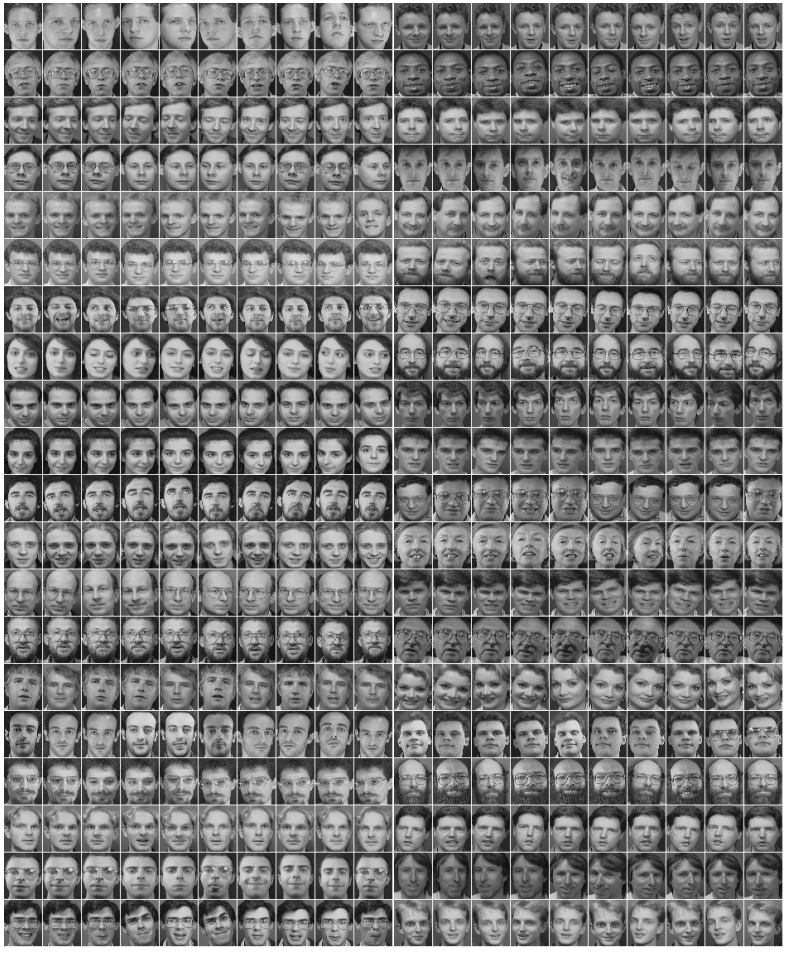
- Τα δεδομένα εξόδου παίρνουν ένα διακριτό σύνολο τιμών
- δηλ. έχουμε ένα σύνολο δυνατών κατηγοριών

Μάθηση με επίβλεψη



Τα δεδονένα εισόδου θα μπορούσαν να είναι δισδιάστατα

Μάθηση με επίβλεψη



ή πολυδιάστατα

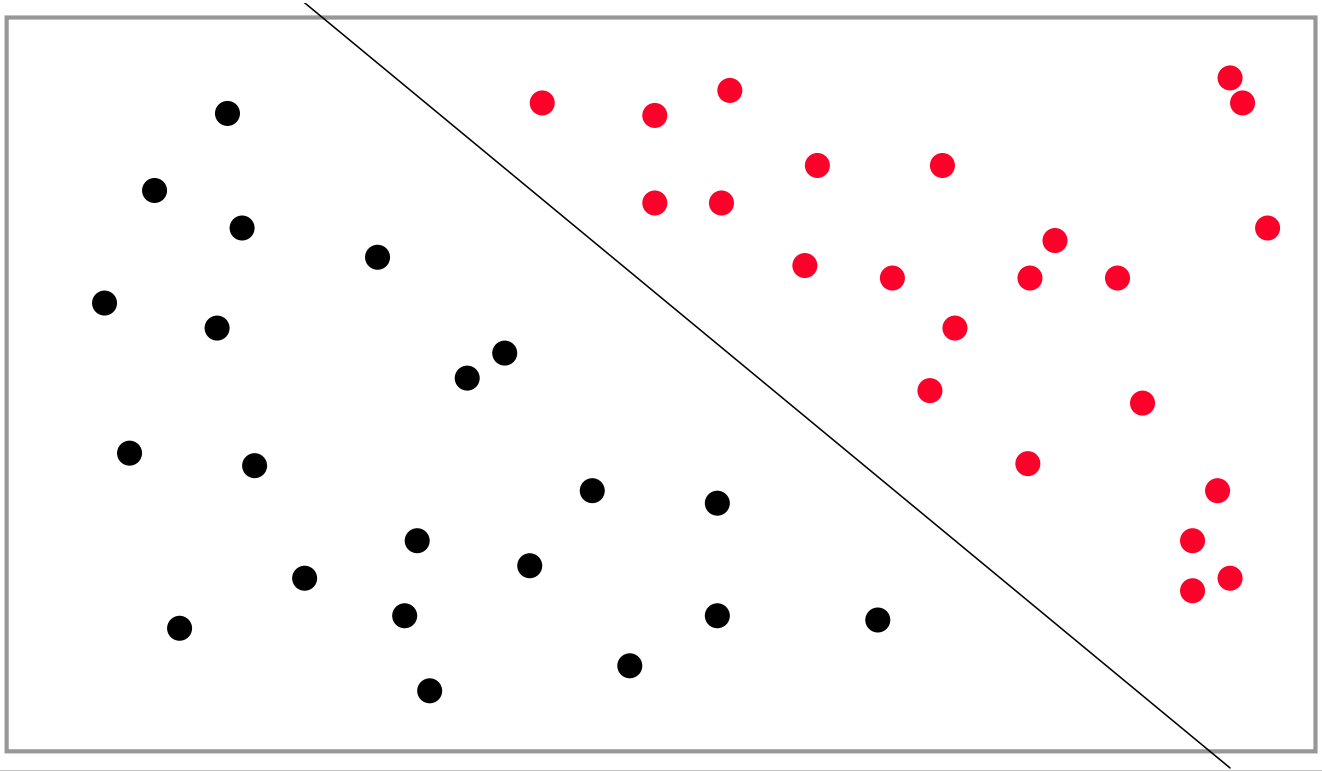
(όπως στο παράδειγμα που θα θέλαμε να κατηγοριοποιούμε εικόνες ανάλογα με το φύλο)

Μάθηση με επίβλεψη

Στην πιο γενική περίπτωση τα δεδομένα είναι πολυδιάστατα και έχουμε πολλές κατηγορίες

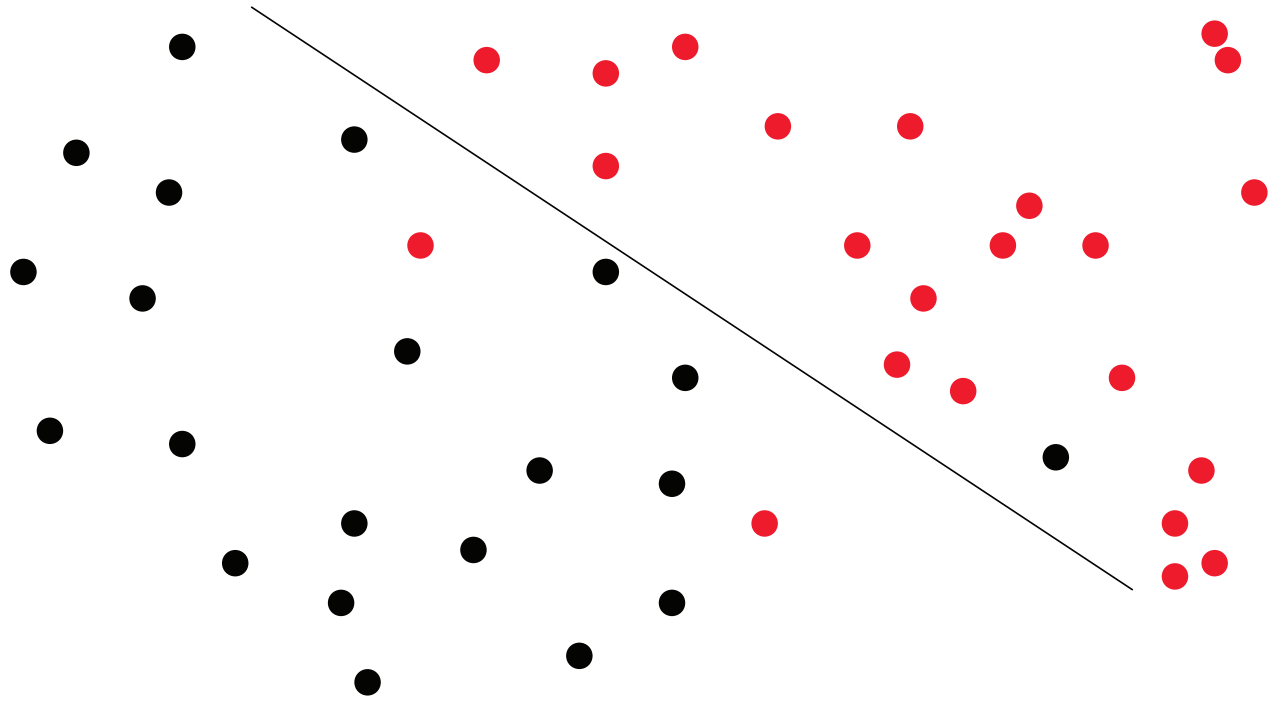
(όπως στο παράδειγμα που θα θέλαμε να κατηγοριοποιούμε εικόνες σε ένα από τα 10 αριθμητικά ψηφία)

Μάθηση με επίβλεψη



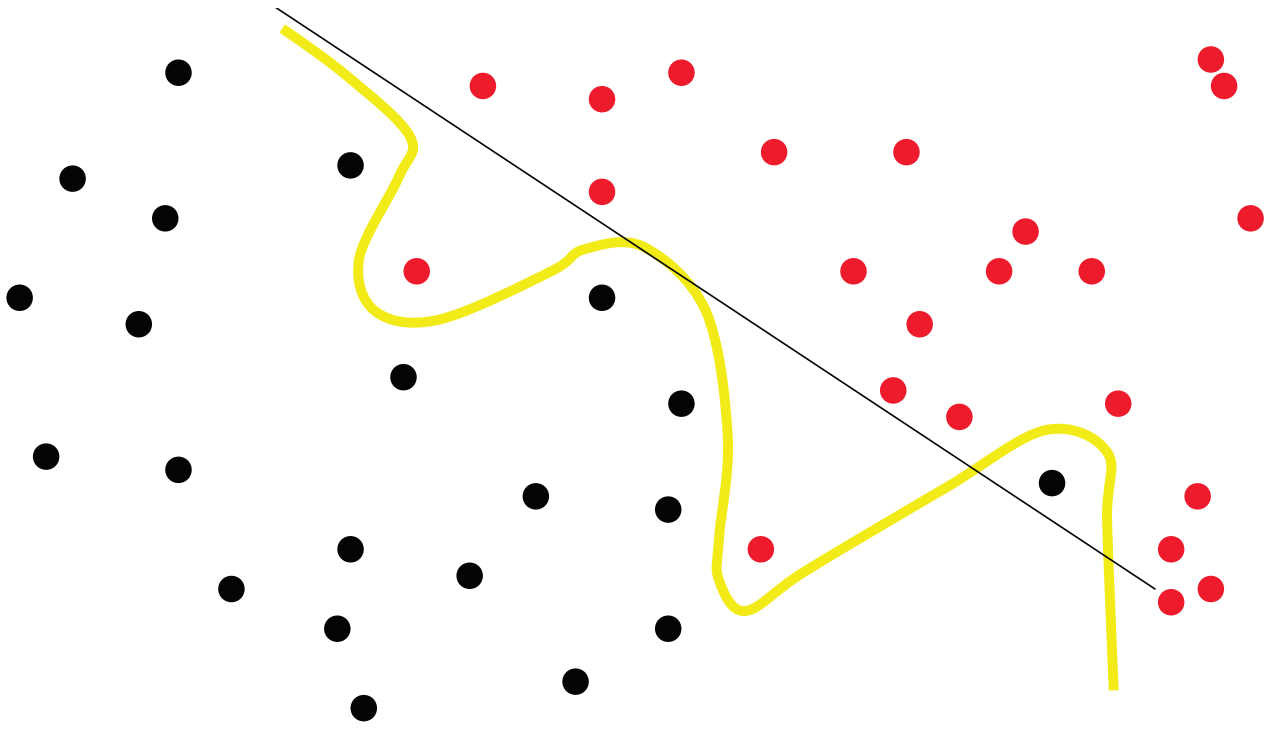
Μιας μέθοδος μηχανικής επιδιώκει τον διαχωρισμό των δύο κατηγοριών

Μάθηση με επίβλεψη



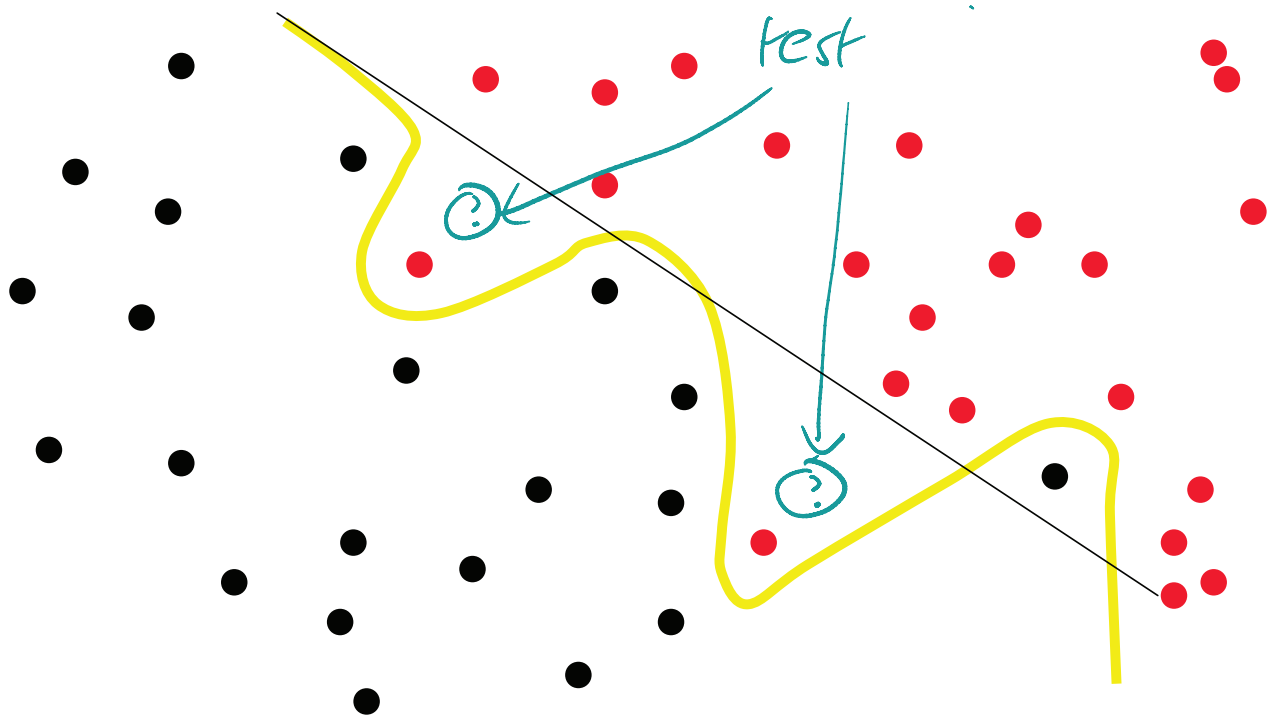
Όπως και στο πρόβλημα παλινδρόμησης το πρόβλημα μπορεί να είναι σύνθετο και να υπάρχουν διαφορετικές λύσεις από τις οποίες θα θέλαμε να βρούμε την καλύτερη

Μάθηση με επίβλεψη



Ποιο μοντέλο είναι καλύτερο; Η **μαύρη** ή η κίτρινη γραμμή;

Μάθηση με επίβλεψη

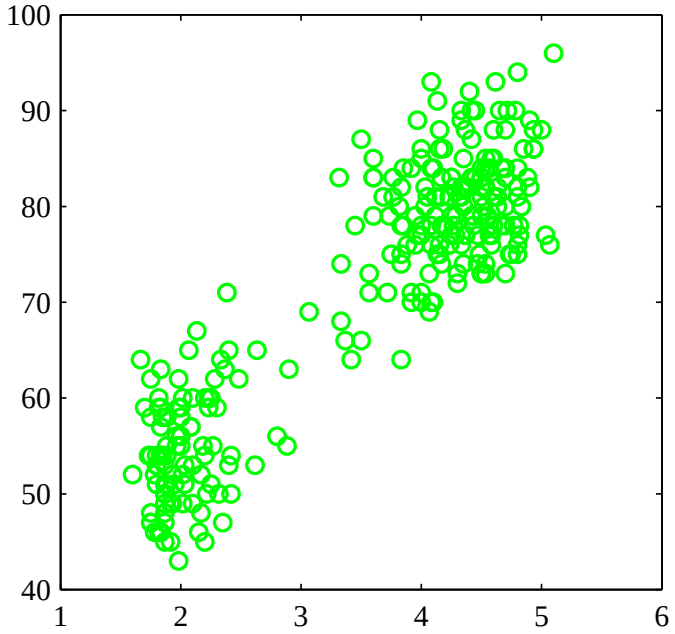


Η επίδοση σε άγνωστα δεδομένα είναι αυτή που μετράει!

Μάθηση χωρίς επίβλεψη

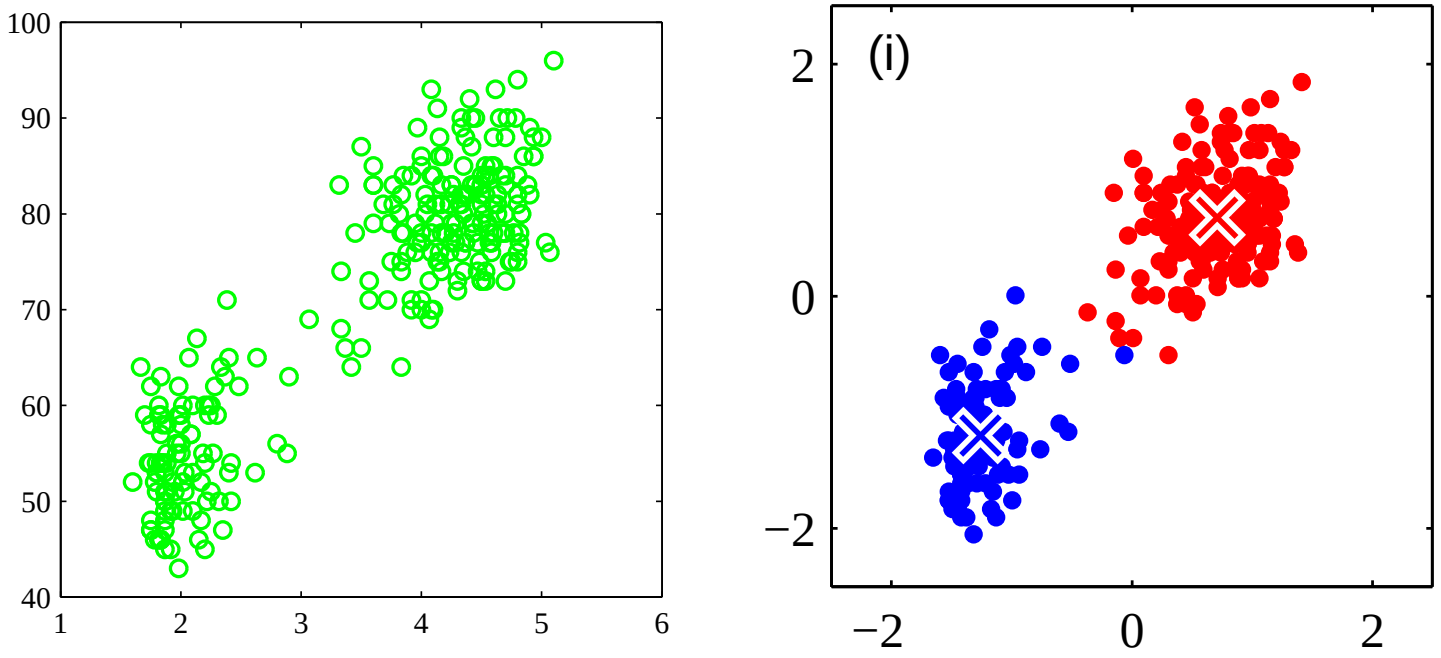
Στην μάθηση χωρίς επίβλεψη μας δίνονται μόνο δεδομένα εισόδου και θα θέλαμε να ανακαλύψουμε δομή

Μάθηση χωρίς επίβλεψη



Υπάρχει κάποια δομή στα δεδομένα

Μάθηση χωρίς επίβλεψη



Ομαδοποίηση (clustering): Διαχωρισμός των δεδομένων σε ομάδες

Μάθηση χωρίς επίβλεψη

Blind Source Separation.... Demo

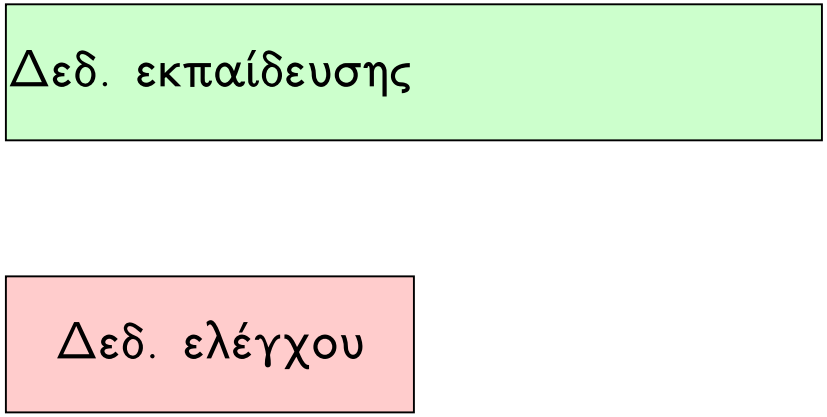
Δεδομένα εκπαίδευσης και δεδομένα ελέγχου

1. Μάθηση με επίβλεψη (supervised learning)
2. Μάθηση χωρίς επίβλεψη (unsupervised learning)

Και στις δύο παραπάνω τεχνικές μάθησης ενδιαφερόμαστε για μεθόδους που γενικεύουν καλά σε άγνωστα δεδομένα

Υπό αυτή την έννοια τα δεδομένα χωρίζονται σε δεδομένα εκπαίδευσης (training data) και δεδομένα ελέγχου (test data)

Δεδομένα εκπαίδευσης και δεδομένα ελέγχου



- Το σύστημα μηχανικής μάθησης εκπαιδεύεται χρησιμοποιώντας τα δεδομένα εκπαίδευσης
- Η επίδοση του συστήματος μετράται μέσω των δεδομένων ελέγχου

Η επίδοση στα δεδομένα ελέγχου δεν πρέπει σε καμιά περίπτωση να χρησιμοποιείται κατά την εκπαίδευση, διότι σε μια τέτοια περίπτωση δεν θα έχουμε ένα ανεξάρτητο κριτήριο μέτρησης της επίδοσης

Η γενική δομή ενός συστήματος μηχανικής μάθησης

Ένα σύστημα μηχανική μάθησης αποτελείται από

Δεδομένα:

- Συλλογή και προεπεξεργασία δεδομένων (feature selection/extraction)

Μοντέλο ή υπόθεση:

- Π.χ. η γραμμική ή η τετραγωνική συνάρτηση στο πρόβλημα παλινδρόμησης
- Εξαρτάται από άγνωστους παραμέτρους

Αλγόριθμοι εκπαίδευσης:

- Συναρτήσεις κόστους βάσει των οποίων μαθαίνουμε τις άγνωστες παραμέτρους του μοντέλου
- Αλγόριθμοι βελτιστοποίησης

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

Ένα σημαντικό θέμα αφορά την επιλογή μοντέλου (model selection) ώστε να επιτυγχάνουμε την

- αποφυγή των φαινομένων υπερεκπαίδευσης (overfitting) και υποεκπαίδευσης (underfitting)

Εν τέλει η επιτυχία της μεθόδου κρίνεται από την ικανότητα γενίκευσης (δηλ. από την επίδοση στα δεδομένα ελέγχου)

Ας κατανοήσουμε τις παραπάνω έννοιες με ένα παράδειγμα

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

- Έστω ότι έχουμε τα ακόλουθα δεδομένα εκπαίδευσης

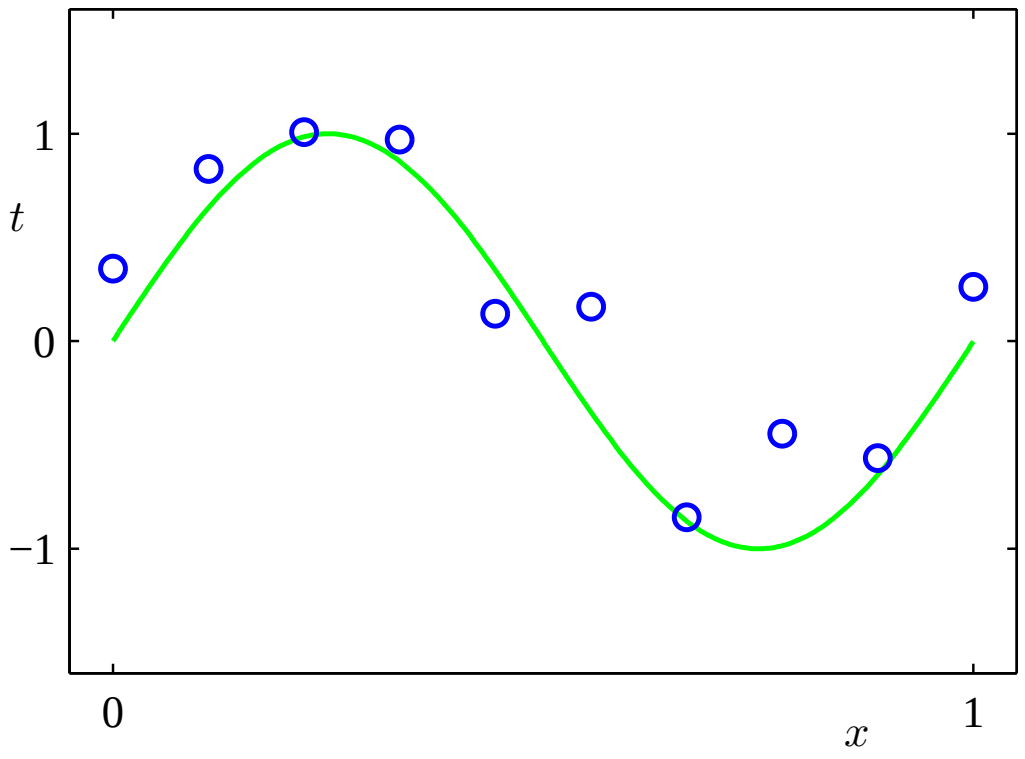
$$\mathcal{D} = \{\mathbf{x}_n, t_n\}_{n=1}^{N}$$

όπου κάθε $\mathbf{x}_n$ είναι ένα δεδομένο εισόδου και t_n το αντίστοιχο δεδομένο εξόδου

- Πρόβλημα μάθησης: Κατασκευή ενός συστήματος που να μαθαίνει να προβλέπει την έξοδο t_* για κάθε άγνωστο δεδομένο εισόδου $\mathbf{x}_*$
- Αν $t \in \mathbb{R} \Rightarrow$ πρόβλημα παλινδρόμησης (regression)
- Αν $t \in \{1, \ldots, K\} \Rightarrow$ πρόβλημα κατηγοριοποίησης (classification)

Ας αναλύσουμε τα παραπάνω με ένα παράδειγμα προβλήματος παλινδρόμησης

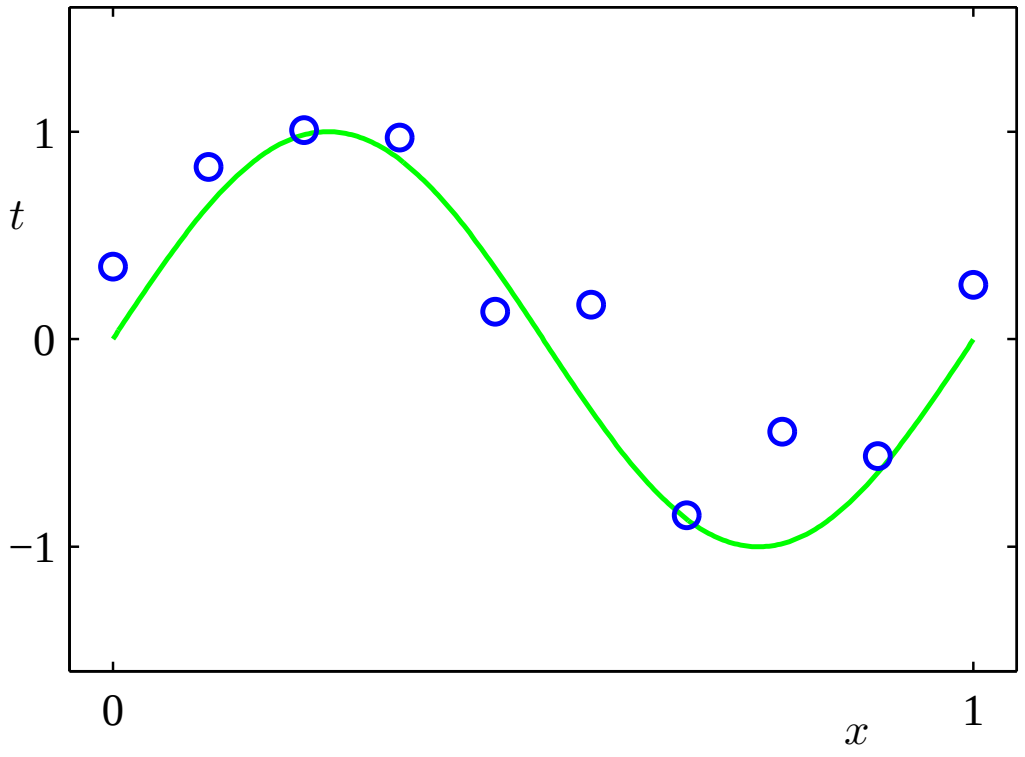
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Έστω 10 τεχνητώς κατασκευασμένα δεδομένα $(x_n, t_n)_{n=1}^{N}$ τα οποία φαίνονται στο σχήμα

Η πράσινη γραμμή δείχνει την πραγματική συνάρτηση που έχει παράγει τα δεδομένα την οποία θα θέλαμε να προσεγγίσουμε εκπαιδεύοντας ένα σύστημα μηχανικής μάθησης

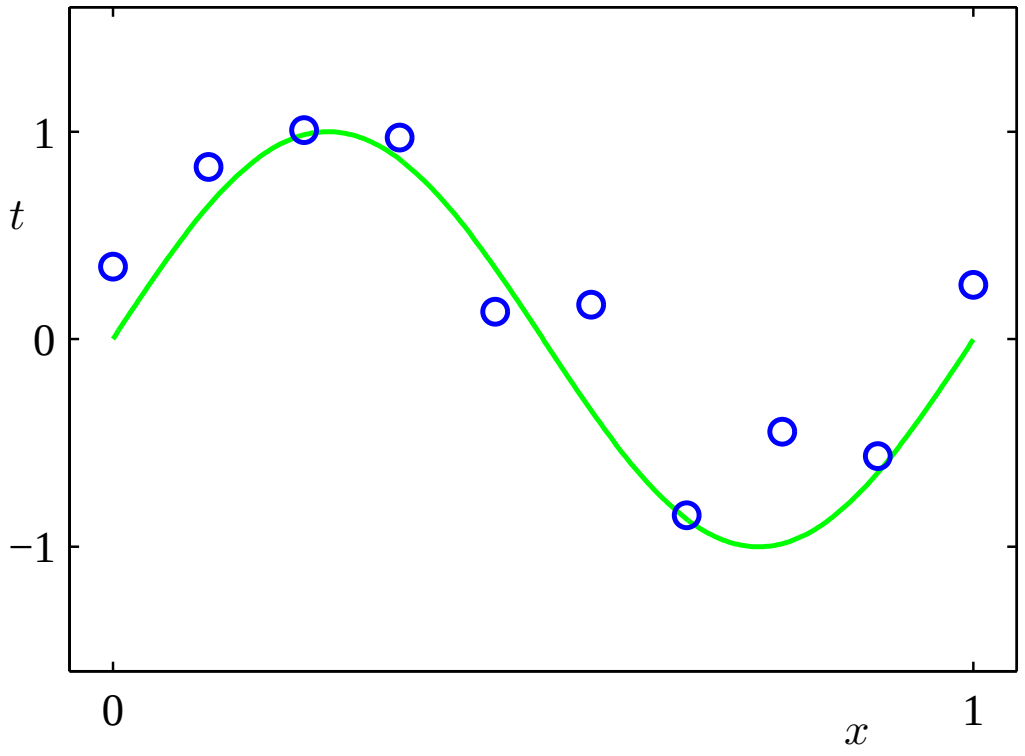
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Σε αυτό το απλό πρόβλημα δεν απαιτείται προεπεξεργασία των δεδομένων εισόδου

Οπότε ας προχωρήσουμε στην κατασκευή ενός μοντέλου

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

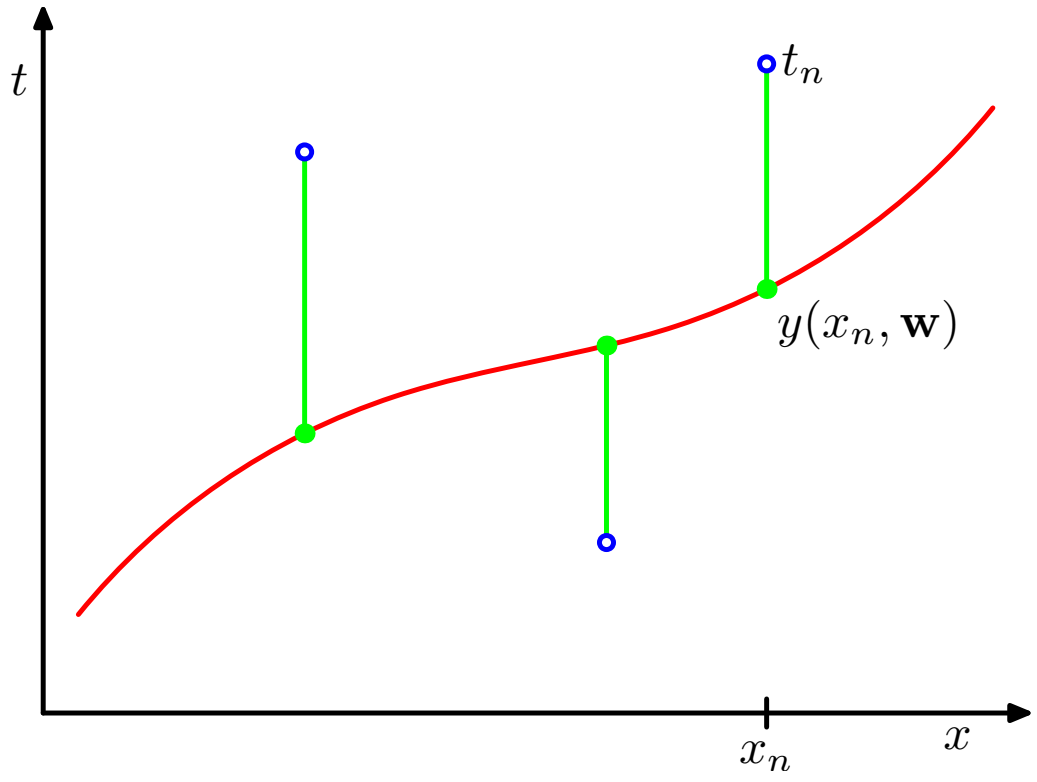


Επιλέγουμε ένα πολυώνυμο M-βαθμού ως μοντέλο

$$y(x, \mathbf{w}) = w_0 + w_1 x + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

Χρειαζόμαστε κάποιο αλγόριθμο εκπαίδευσης του μοντέλου, δηλ. εύρεσης κατάλληλων τιμών για τις παραμέτρους $\mathbf{w}$

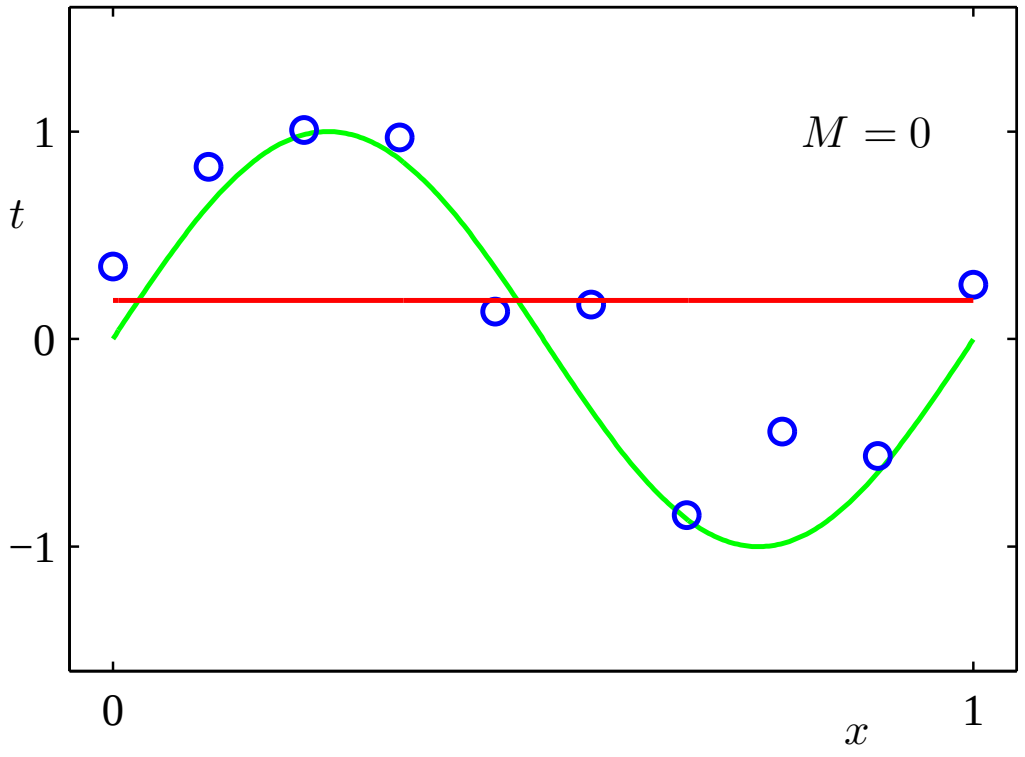
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Ο αλγόριθμος που επιλέγουμε θα βασιστεί στην ελαχιστοποίηση της ακόλουθης συνάρτησης (μέθοδος ελαχίστων τετραγώνων):

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} (y(x_n, \mathbf{w}) - t_n)^2$$

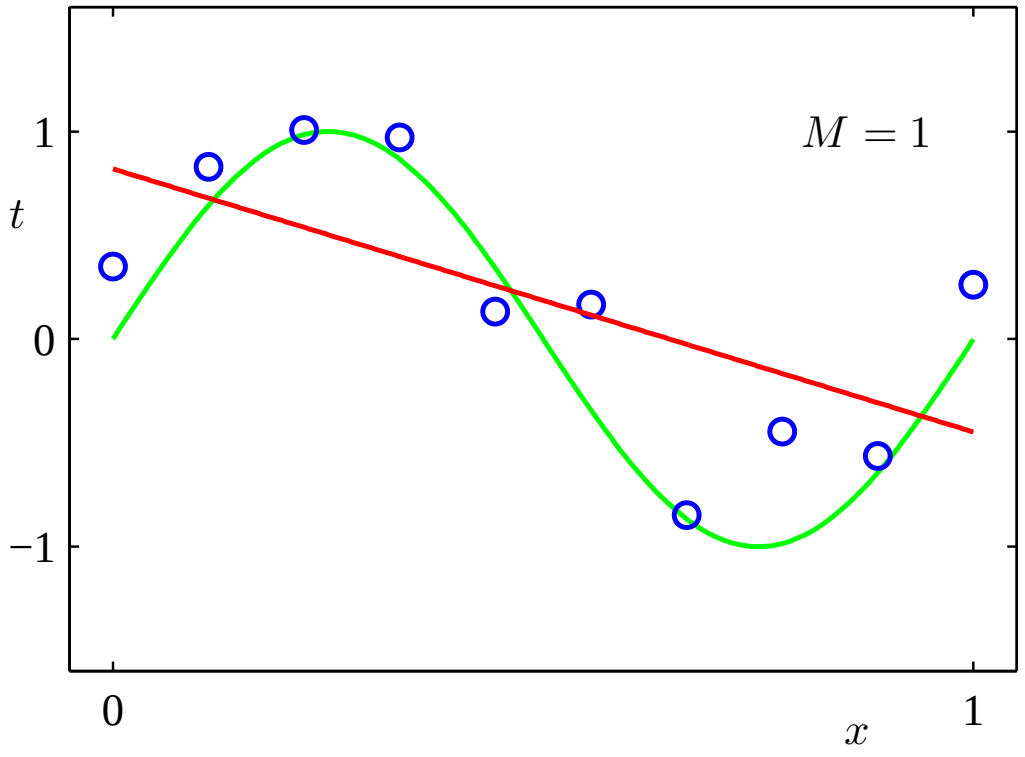
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Το εκπαιδευμένο 0-βαθμού πολυώνυμο

$$y(x, \mathbf{w}^*) = w_0^*$$

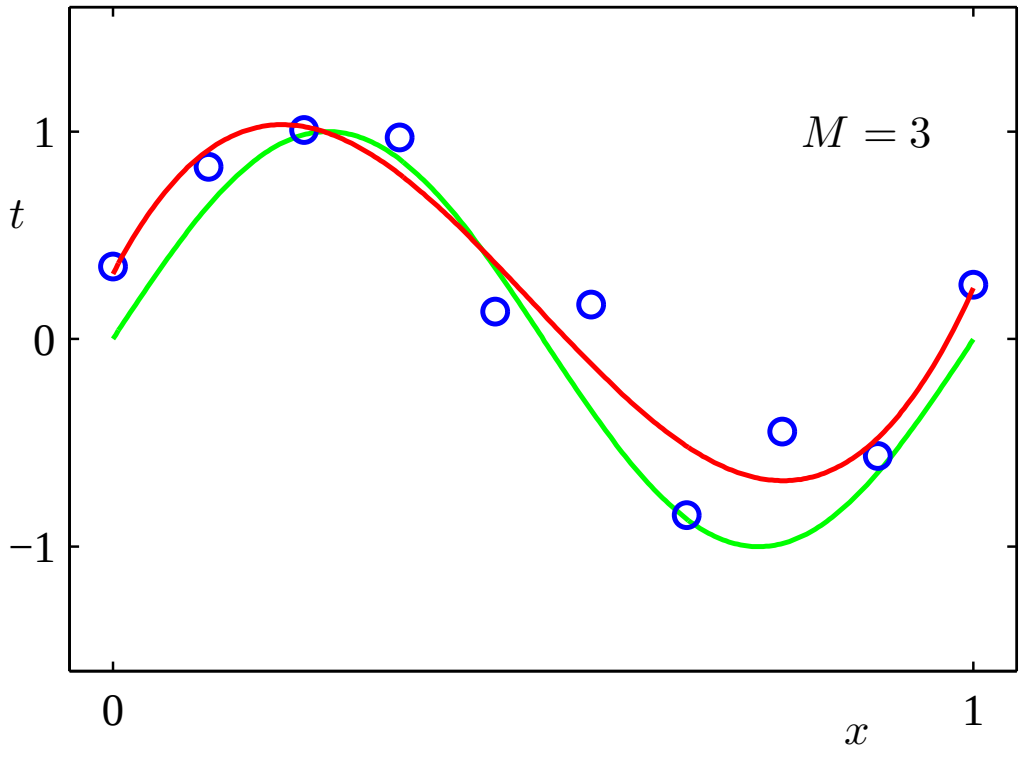
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Το εκπαιδευμένο 1-βαθμού πολυώνυμο

$$y(x, \mathbf{w}^*) = w_0^* + w_1^* x$$

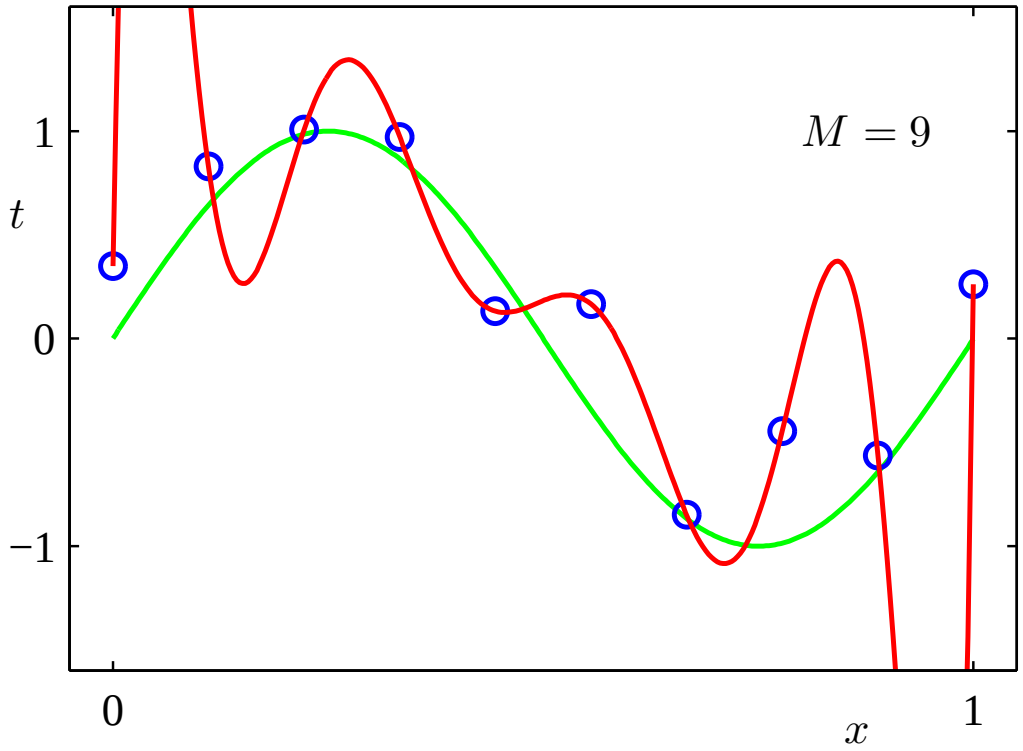
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Το εκπαιδευμένο 3-βαθμού πολυώνυμο

$$y(x, \mathbf{w}^*) = w_0^* + w_1^* x + w_2^* x^2 + w_3^* x^3$$

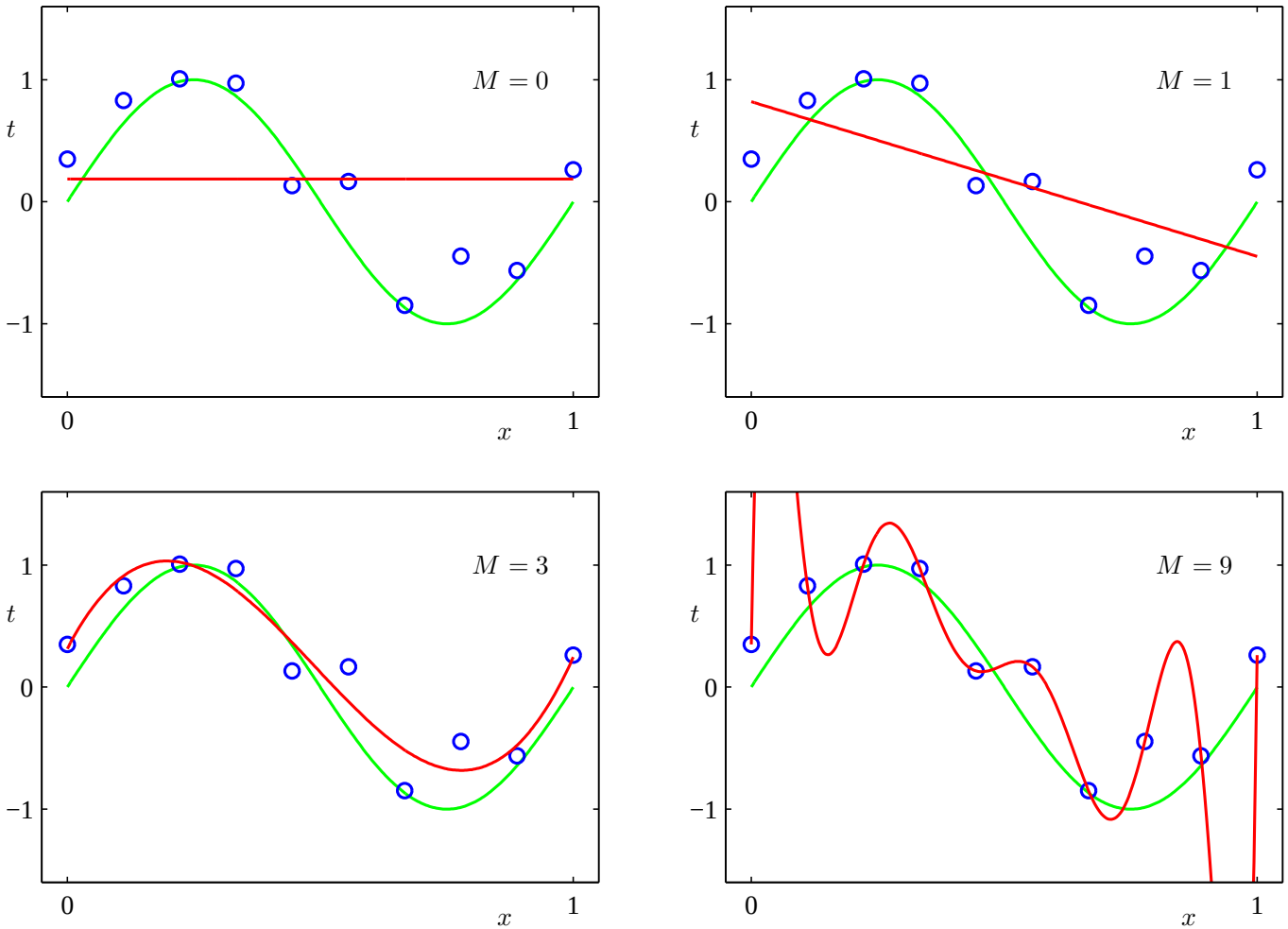
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Το εκπαιδευμένο 9-βαθμού πολυώνυμο

$$y(x, \mathbf{w}^*) = \sum_{j=0}^{9} w_j^* x^j$$

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

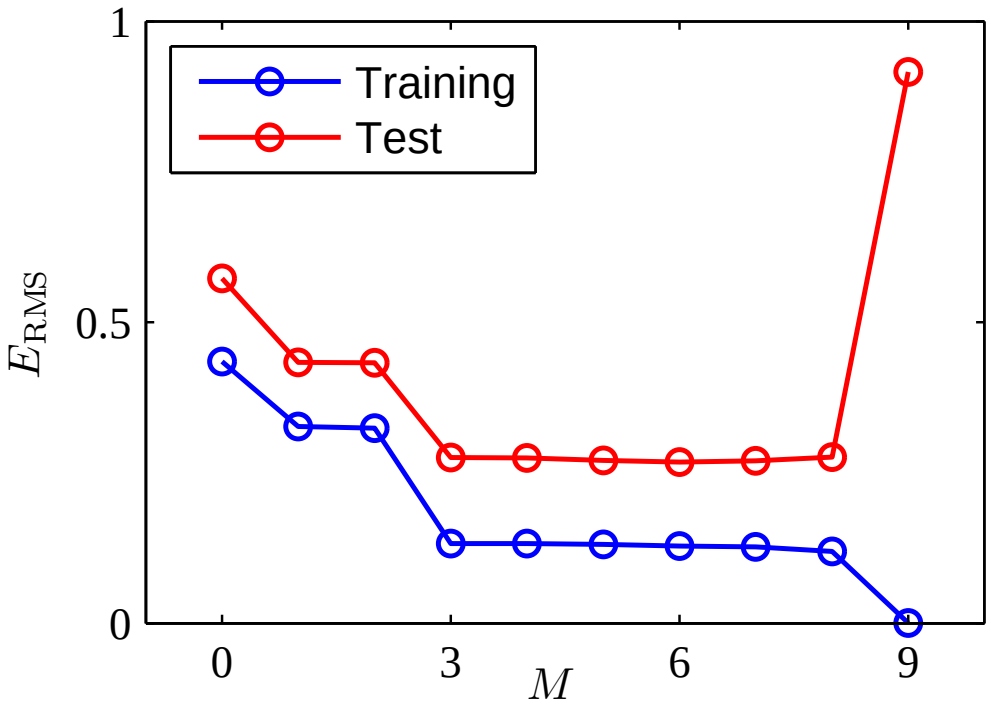


Το $M = 9$ μοντέλο είναι υπερεκπαιδευμένο (overfitted)

Τα $M = 0, 1$ μοντέλα είναι υποεκπαιδευμένα (underfitted)

Το $M = 3$ μοντέλο είναι το καλύτερο

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

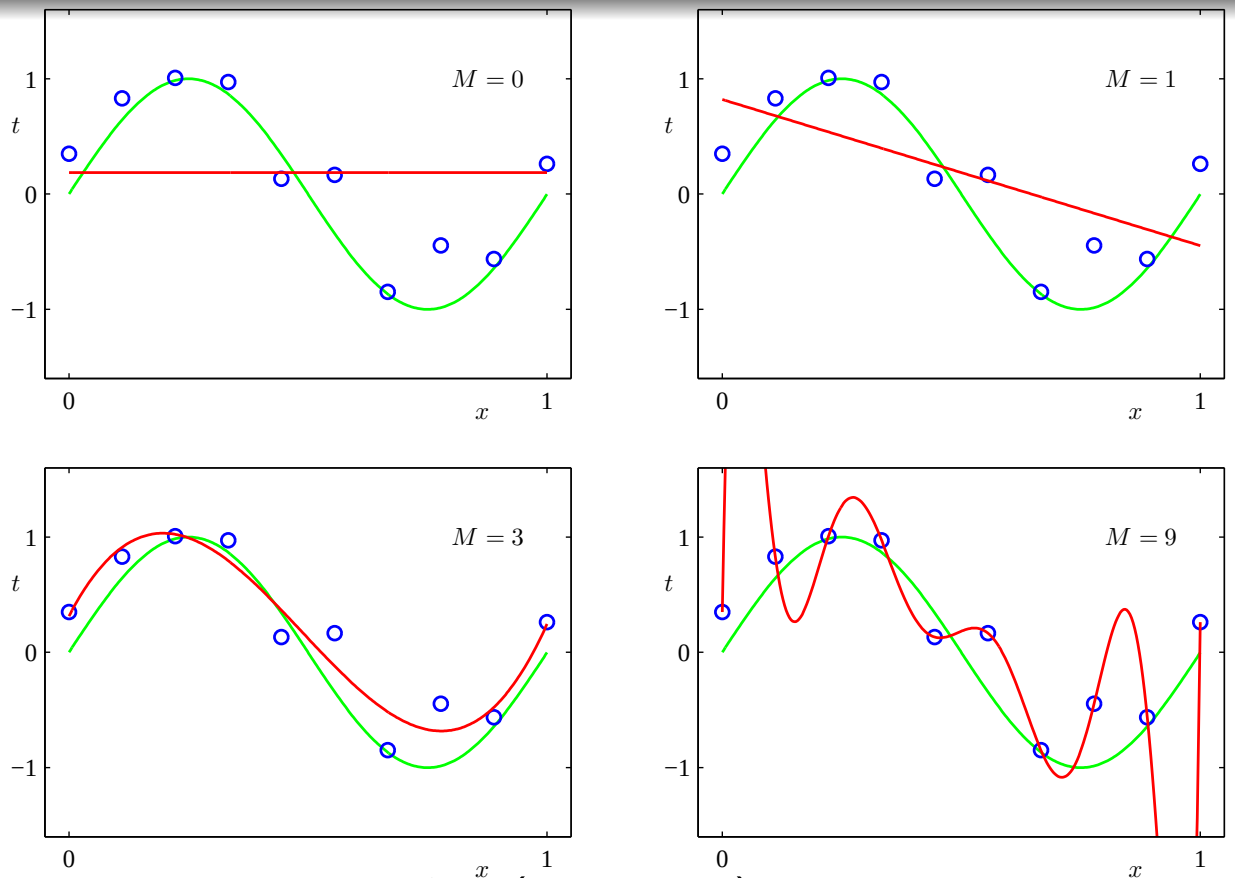


Επίδοση στα δεδομένα ελέγχου: Μέσο-σφάλμα (root-mean-square-error)

$$\sqrt{\frac{\sum_{x_*}(y(x_*, \mathbf{w}^*) - t_*)^2}{\text{Αριθμός δεδομένων ελέγχου}}}$$

(τα υποεκπαιδευμένα και υπερεκπαιδευμένα μοντέλα δεν έχουν καλή επίδοση)

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Τα υποεκπαιδευμένα μοντέλα ($M = 0, 1$) δεν είναι αρκετά ευέλικτα

- $\Rightarrow$ δεν υπάρχει καμιά ελπίδα να βελτιωθεί η επίδοση τους στο συγκεκριμένο πρόβλημα

Αντιθέτως το υπερεκπαιδευμένο μοντέλο ($M = 9$) θα μπορούσε να έχει πιο γενική χρήση

- $\Rightarrow$ αν υπήρχε τρόπος να περιορίσουμε την ευελιξία του ανάλογα με τις ανάγκες του προβλήματος

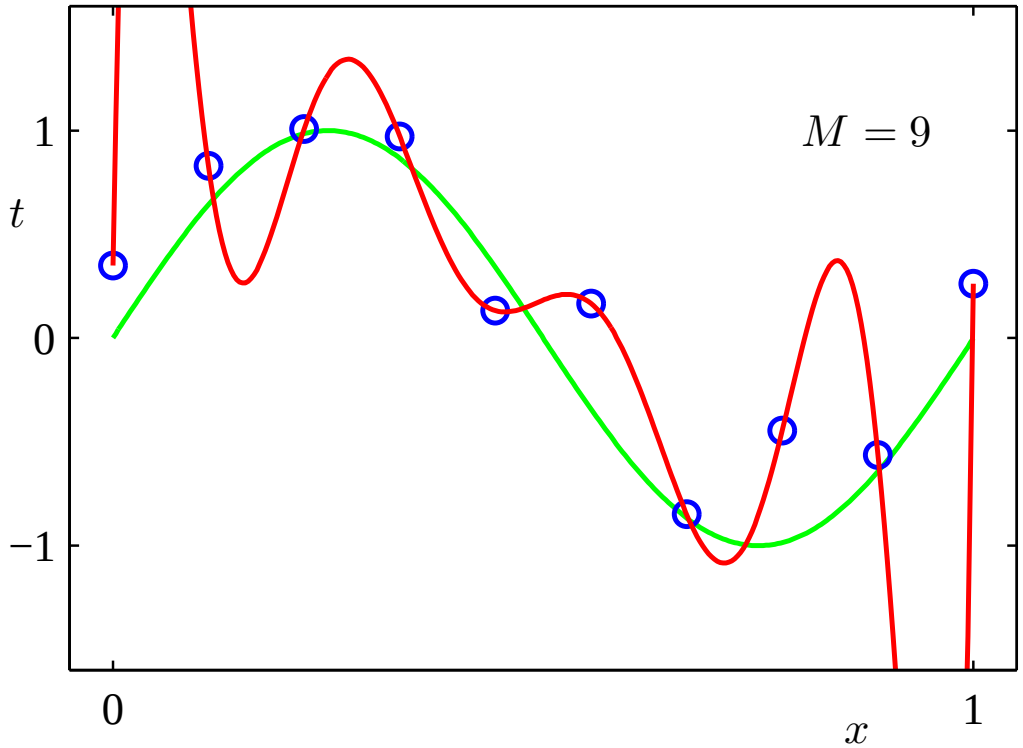
Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

Μια πολύ σημαντική ιδεά στην κατασκευή συστημάτων μηχανικής μάθησης βασίζεται στο εξής σκεπτικό

- **Πραγματικά προβλήματα είναι σύνθετα**
  - $\Rightarrow$ σχεδόν ποτέ δεν γνωρίζουμε ή μπορούμε να μαντέψουμε το ιδανικό μοντέλο
- Οπότε μια καλή πολιτική είναι η χρήση **πολύ ευέλικτων μοντέλων** (ως default!)
  - $\Rightarrow$ που ενδεχομένως θα μπορούσαν να επιλύσουν και τα πιο σύνθετα προβλήματα
- Έπειτα θα θέλαμε κατά περίπτωση να **προσάρμοζουμε/περιορίζουμε την ευελιξία** των μοντέλων αυτών
  - $\Rightarrow$ ώστε να αποφεύγεται η υπερεκπαίδευση

Η παράπανω ιδέα ονομάζεται **κανονικοποίηση (regularization)**

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση



Επιστρέφοντας στην παλινδρόμηση με πολυώνυμα είναι εύλογο να ισχυριστούμε ότι το $M = 9$ μοντέλο αντιμετωπίζει το πρόβλημα της υπερεκπαίδευσης λόγω της συνάρτησης κόστους

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} (y(x_n, \mathbf{w}) - t_n)^2$$

η οποία δεν μπορεί να αποτρέψει την υπερεκπαίδευση και οδηγεί σε ακραίες τιμές των παραμέτρων (δες επόμενη διαφάνεια)

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

	$M=0$	$M=1$	$M=3$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

Οι τιμές των παραμέτρων των υπερεκπαιδευμένων μοντέλων είναι πολύ μεγάλες/ακραίες (σε απόλυτη τιμή)

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

Κανονικοποίηση (regularization) των παραμέτρων **w**

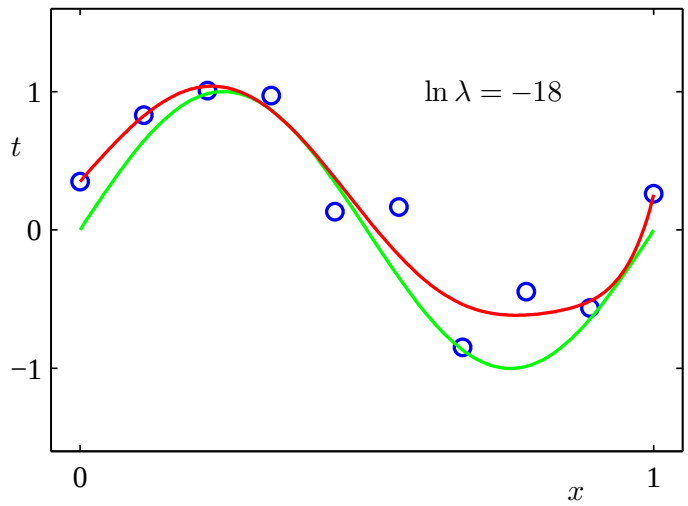
- Θα θέλαμε μια νέα συνάρτηση κόστους που να αποτρέπει μεγάλες τιμές των παραμέτρων

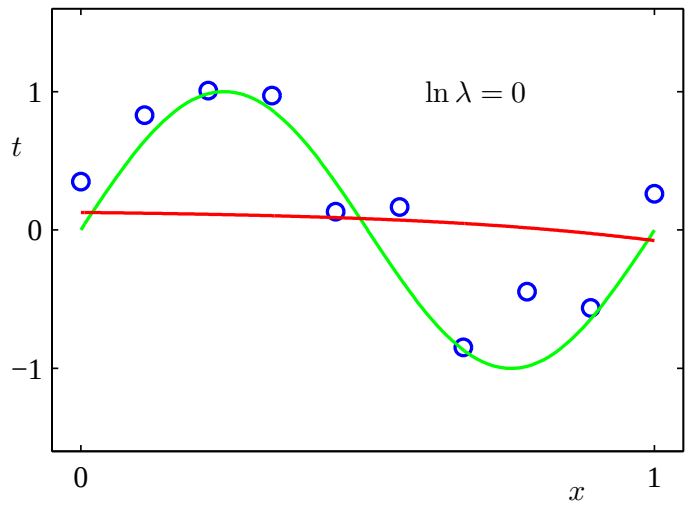
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 + \lambda\frac{||\mathbf{w}||^2}{2}$$

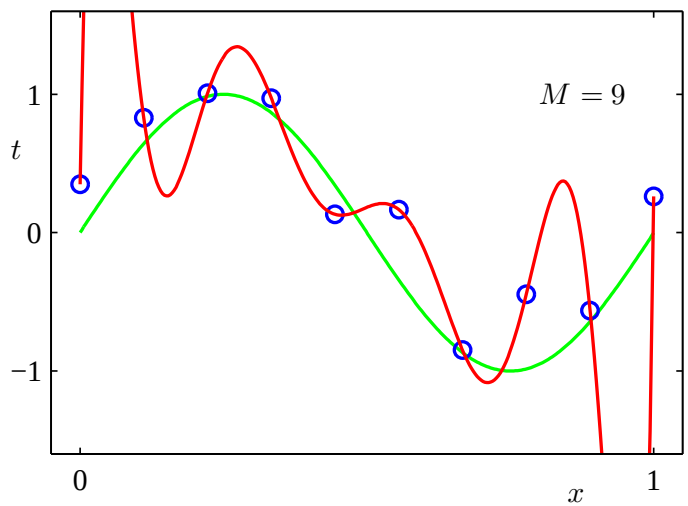
  όπου $||\mathbf{w}|| = w_0^2 + w_1^2 + \ldots + w_M^2$ και $\lambda > 0$

- Ο όρος $\lambda\frac{||\mathbf{w}||^2}{2}$ 'τιμωρεί' μεγάλες τιμές των παραμέτρων
- λ ονομάζεται παράμετρος κανονικοποίησης

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση







Το $M = 9$ μοντέλο εκπαιδευμένο για διαφορετικές τιμές του λ. Για κάποια τιμή του λ το μοντέλο φαίνεται ιδανικό!

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

Επιλογή του λ επηρεάζει τις τιμές των παραμέτρων. Πώς θα μπορούσαμε να επιλέγουμε με αυτόματο τρόπο το λ;

Υπερεκπαίδευση, υποεκπαίδευση, κανονικοποίηση

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 + \lambda\frac{||\mathbf{w}||^2}{2}$$

Πολλές αναπάντητες ερωτήσεις:

- Ποια είναι η ερμηνεία πίσω από την χρήση της $E(\mathbf{w})$; (θα μπορούσε η $E(\mathbf{w})$ να είχε άλλη μορφή;)
- Αν τα δεδομένα εξόδου παίρνουν διακριτές τιμές, θα ήταν η $E(\mathbf{w})$ κατάλληλη;
- Πώς επιλέγουμε την τιμή του λ;

Η θεωρία πιθανότητων και η στατιστική αποτελούν το θεωρητικό υπόβαθρο βάσει του οποίου μπορούμε να απαντήσουμε στα παραπάνω ερωτήματα και να κατασκευάζουμε αλγορίθμους μηχανικής μάθησης

Επίλογος

- Διάβασμα για το σπίτι: section 1.0 και 1.1 από το βιβλίο του Bishop
- Επόμενο μάθημα: Η τεχνική cross validation, εισαγωγή στην θεωρία πιθανοτήτων και στα πιθανοτικά μοντέλα