# Εισαγωγή στην Ταξινόμηση

Μηχανική Μάθηση

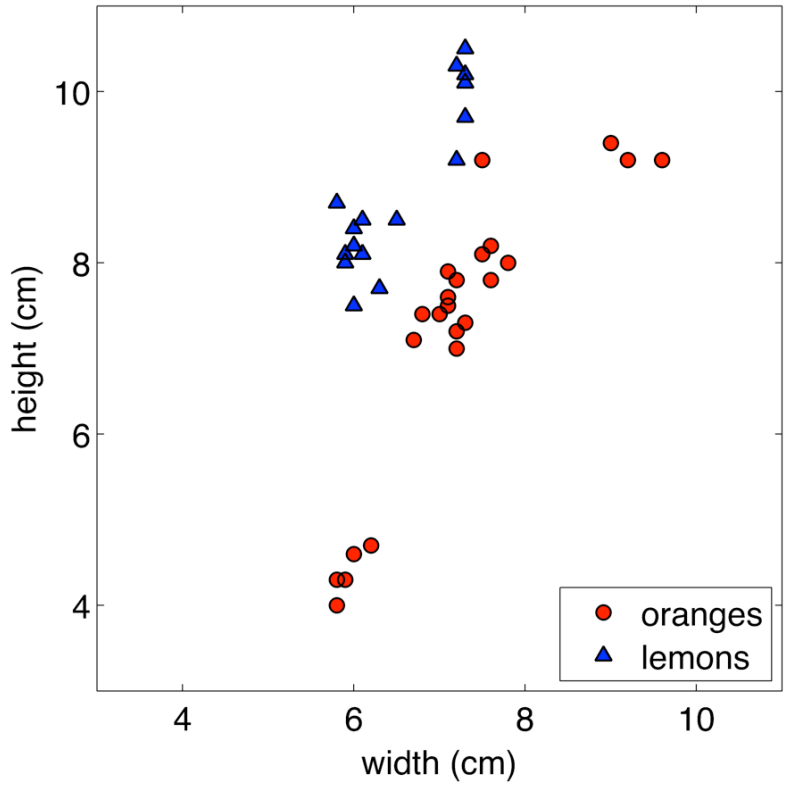
ΔΠΜΣ Επιστήμη Δεδομένων και Μηχανική Μάθηση

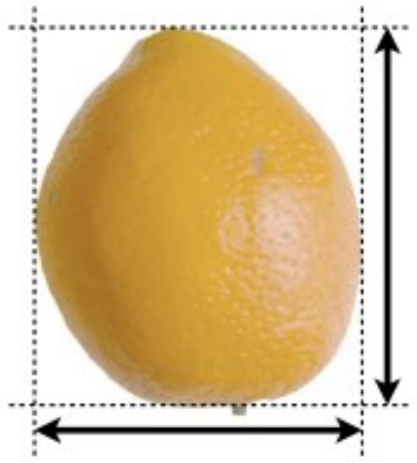
Γιώργος Αλεξανδρίδης – gealexandri@islab.ntua.gr

# Κατηγορίες ταξινομητών

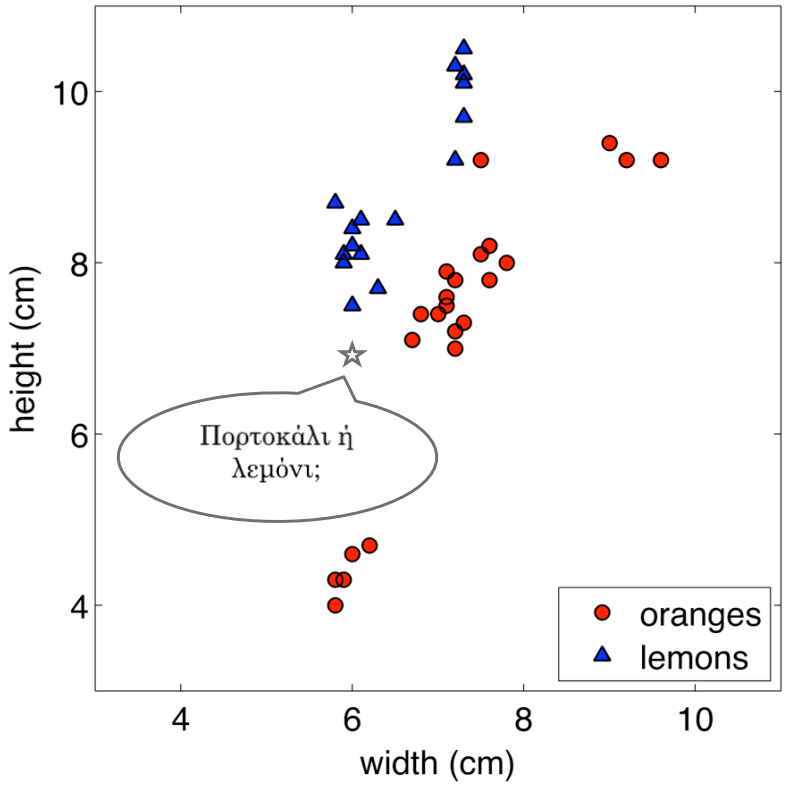
1. **Διαμεριστικοί** (divisive)
   - Προσπαθούν να *τεμαχίσουν* το χώρο των δεδομένων σε *μη-επικαλυπτόμενες* υπό-περιοχές εντός των οποίων βρίσκονται δεδομένα μιας κλάσης
   - Ταξινομητές πλησιέστερων γειτόνων, δέντρα αποφάσεων, (πολυεπίπεδα) νευρωνικά δίκτυα πρόσθιας τροφοδότησης, μηχανές διανυσμάτων υποστήριξης, …
2. **Παραγωγικοί** (generative)
   - Στατιστική θεώρηση των δεδομένων
   - Προσπαθούν να μάθουν την *υποκείμενη κατανομή* (underlying distribution) που παράγει τα δεδομένα
   - Αφελείς μπεϋζιανοί ταξινομητές, γκαουσιανά μοντέλα μίξης, κρυφά μαρκοβιανά μοντέλα,…

# Πρόβλημα ταξινόμησης: Πορτοκάλια ή Λεμόνια;



# Ταξινόμηση και Επαγωγή (Induction)



# Εκμάθηση Μέσω Παραδειγμάτων

Instance-based Learning

# Εκμάθηση μέσω Παραδειγμάτων

- **Μη-παραμετρικά μοντέλα** (*non-parametric models*)
- Πρόκειται για απλά μοντέλα τα οποία χρησιμοποιούνται για την προσέγγιση προβλημάτων συνεχών τιμών (παλινδρόμηση) ή διακριτών τιμών (ταξινόμηση)
  - Ταξινομητές ***k*-πλησιέστερων γειτόνων** (*k-nearest neighbors classifier* – *k*NN), **δίκτυα ακτινικών συναρτήσεων βάσης** (*radial basis function* – RBF), ..
- Η διαδικασία της μάθησης είναι ισοδύναμη με την αποθήκευση των **δεδομένων** που χρησιμοποιούνται για την **εκπαίδευση** του μοντέλου (*training data*)
- Τα νέα στιγμιότυπα ταξινομούνται χρησιμοποιώντας «ομοειδή» στιγμιότυπα από το σύνολο εκπαίδευσης
  - Τα νέα στιγμιότυπα αναφέρονται επίσης και ως **στιγμιότυπα δοκιμής** (*test instances*) ή **δεδομένα δοκιμής** (test data)
- Κωδικοποιούν λογικές **υποκείμενες υποθέσεις** (*underlying assumptions*)
  - Οι κλάσεις (έξοδος) μεταβάλλονται «**ομαλά**» (*smooth*) συναρτήσει της εισόδου
  - Τα δεδομένα καταλαμβάνουν έναν υπο-χώρο του αρχικού χώρου μεγάλων διαστάσεων

# Ταξινομητής Πλησιέστερου Γείτονα

- Έστω ότι το σύνολο $T$ των δειγμάτων δεδομένων εκπαίδευσης ανήκουν στον Ευκλείδειο χώρο ($x \in \Re^d$)
- **Λειτουργία**
  - Η ετικέτα που λαμβάνει κάθε νέο δείγμα είναι η ετικέτα του πλησιέστερου δείγματος δεδομένων εκπαίδευσης
  - Συνήθως χρησιμοποιείται η Ευκλείδεια απόσταση

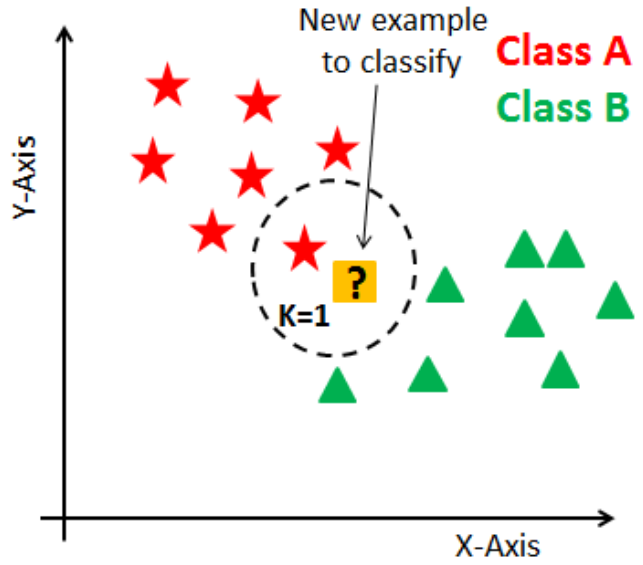
$$\left\|\boldsymbol{x}^{(a)} - \boldsymbol{x}^{(b)}\right\|_2 = \sqrt{\sum_{j=1}^{d}\left(x_j^{(a)} - x_j^{(b)}\right)^2}$$

- **Αλγόριθμος**
  1. Βρες το παράδειγμα $(\boldsymbol{x}^*;\ t^*)$ από τα δεδομένα εκπαίδευσης το οποίο είναι «εγγύτερα» στο νέο στιγμιότυπο $\boldsymbol{x}$

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}^{(i)} \in T}{arg\min}\ \text{distance}(\boldsymbol{x}^{(i)}, \boldsymbol{x})$$

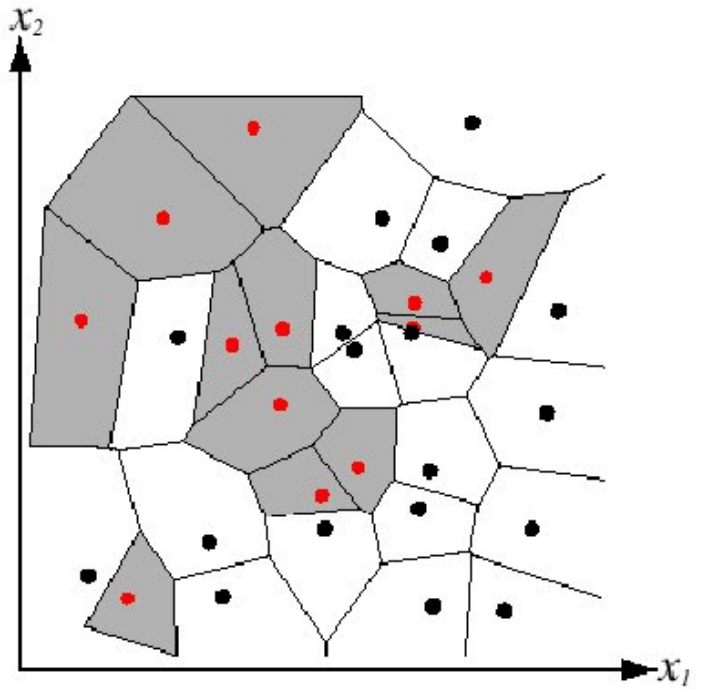
  1. Ανάθεσε στο $\boldsymbol{x}$ την ετικέτα $t^*$: $(\boldsymbol{x};\ t^*)$

- Στην πραγματικότητα, δεν χρειάζεται να υπολογίσουμε την ρίζα
  - Γιατί;

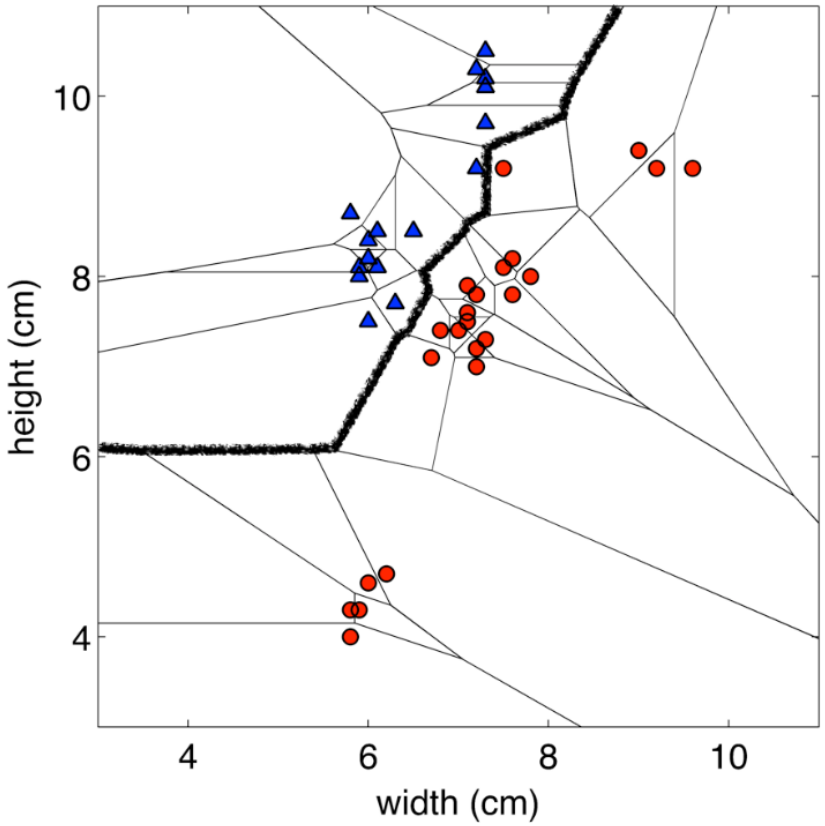


# Όρια απόφασης

- Ο ταξινομητής πλησιέστερου γείτονα δεν υπολογίζει άμεσα όρια απόφασης μεταξύ των κλάσεων
  - Ωστόσο, αυτά προκύπτουν έμμεσα
- Όρια απόφασης: **Διαγράμματα Voronoi**
  - Ο χώρος των δεδομένων χωρίζεται σε διαφορετικές περιοχές ανάλογα με την ετικέτα (κλάση) τους
  - Τα ευθύγραμμα τμήματα αποτελούν τις μεσοκαθέτους μεταξύ δύο «γειτονικών» στιγμιοτύπων των δεδομένων εκπαίδευσης

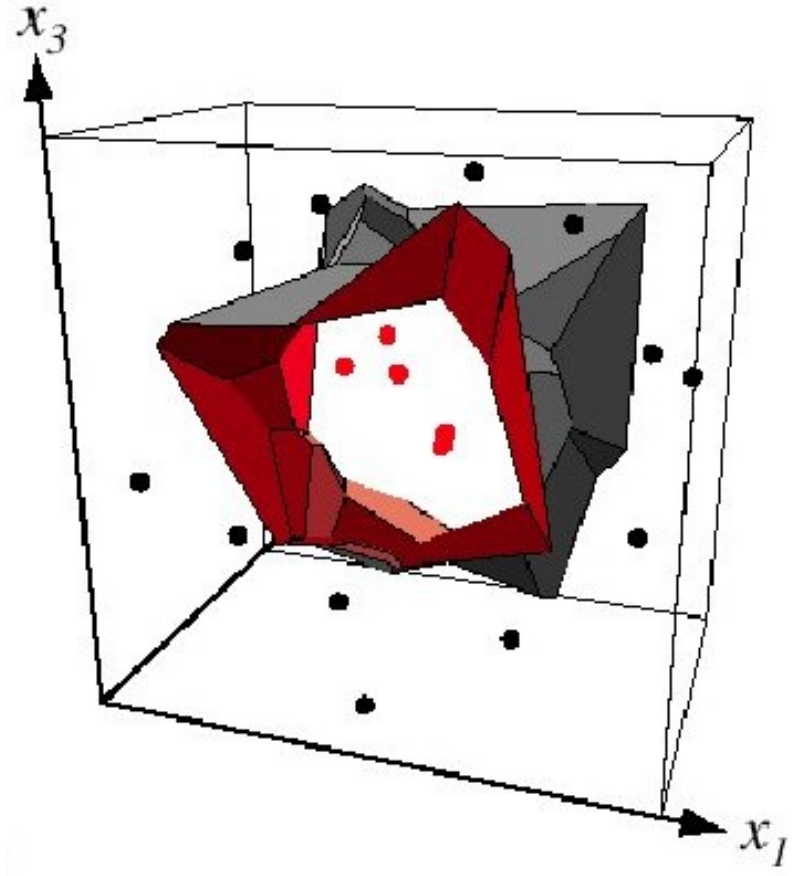


# Πορτοκάλια ή Λεμόνια: Όριο απόφασης




Σύνθετο όριο απόφασης (όχι ευθύγραμμο τμήμα)

# Όριο απόφασης στις 3 διαστάσεις

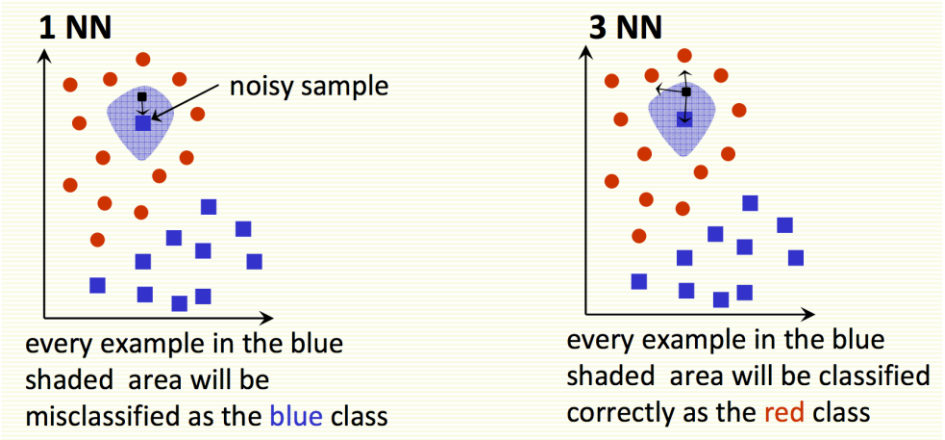


# $k$-πλησιέστεροι γείτονες

- Ο ταξινομητής πλησιέστερου γείτονα είναι «*ευαίσθητος*» στην ύπαρξη *θορύβου* στα δεδομένα

- **Λύση**
  - «*Ομαλοποίηση*» της ταξινόμησης μέσω «*ψηφοφορίας*» των $k$ πλησιέστερων γειτόνων

- **Αλγόριθμος**
  1. Βρες τα $k$ πλησιέστερα παράδειγμα $(\boldsymbol{x}^{(i)}; t^{(i)})$ από τα δεδομένα εκπαίδευσης τα οποία είναι «εγγύτερα» στο νέο στιγμιότυπο $\boldsymbol{x}$
  2. Ανάθεσε στο $\boldsymbol{x}$ την ετικέτα $y$:
     - $\mathrm{y} = \underset{t^{(z)}}{argmax} \sum_{r=1}^{k} \delta(t^{(z)}, t^{(r)})$



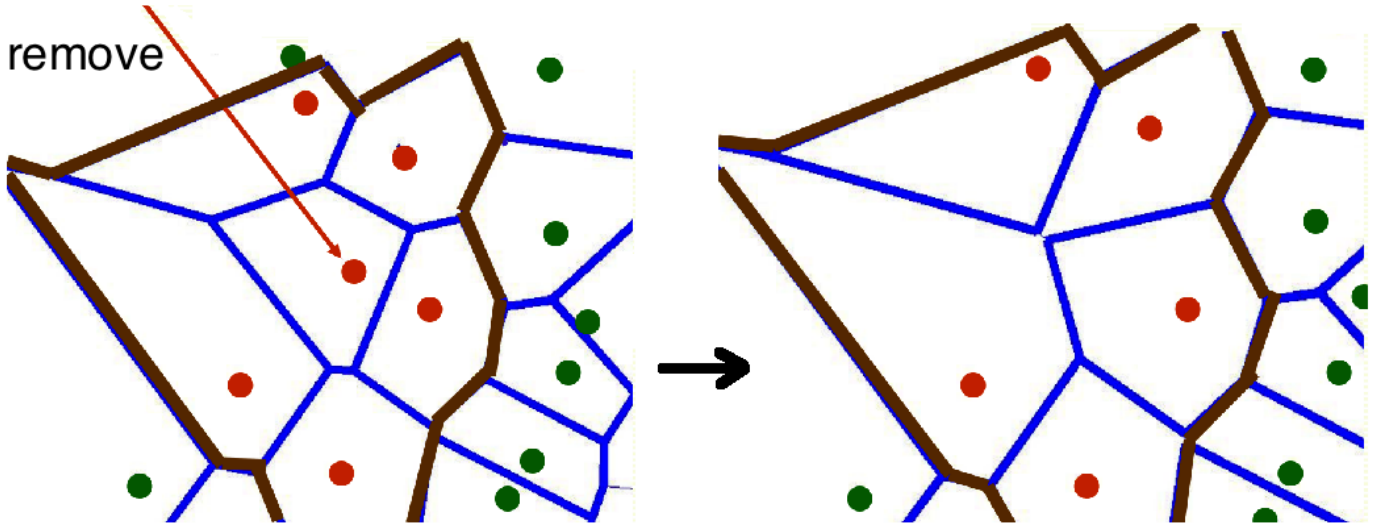
# Προσδιορισμός $k$

- **Πως** βρίσκουμε την **κατάλληλη** τιμή για το $k$;
  - Όσο μεγαλύτερο είναι, τόσο βελτιώνεται η απόδοση του ταξινομητή
  - Ωστόσο αν είναι αρκετά μεγάλο, τότε το εύρος της «γειτονιάς» μεγαλώνει, συμπεριλαμβάνοντας δείγματα τα οποία μπορεί να βρίσκονται πολύ μακριά από το νέο στιγμιότυπο
- Εύρεση $k$ μέσω τεχνικών **διασταυρούμενης επικύρωσης** (*cross-validation*)
  - Θα μιλήσουμε σε επόμενες διαλέξεις και στο εργαστήριο για αυτές
- Εμπειρικός κανόνας
  - $k < \sqrt{N}$, όπου $N$ το πλήθος των παραδειγμάτων εκπαίδευσης

# $k$-πλησιέστεροι γείτονες: ζητήματα πολυπλοκότητας

- **Υψηλή πολυπλοκότητα** κατά τη διαδικασία ελέγχου
  - Για να βρούμε έναν πλησιέστερο γείτονα πρέπει να υπολογίσουμε την απόσταση από όλα τα δεδομένα εκπαίδευσης
  - Λύσεις
    - Χρησιμοποίηση υποσυνόλου των διαστάσεων, χρήση αποδοτικών δομών δεδομένων (πχ kd-trees), υπολογισμός προσεγγιστικής απόστασης, αφαίρεση πλεονασματικών δεδομένων, …
- **Υψηλές απαιτήσεις αποθήκευσης**
  - Πρέπει να αποθηκεύσουμε στη μνήμη όλα τα δεδομένα εκπαίδευσης
  - Λύσεις
    - Αφαίρεση πλεονασματικών δεδομένων
- **Δεδομένα πολλών διαστάσεων** («Κατάρα διαστατικότητας»)
  - Το πλήθος των απαιτούμενων δεδομένων εκπαίδευσης αυξάνει εκθετικά όσο αυξάνουν οι διαστάσεις
  - Επίσης αυξάνει και το υπολογιστικό κόστος
  - Λύσεις
    - Εφαρμογή τεχνικών μείωσης διαστάσεων και επιλογής χαρακτηριστικών

# Αφαίρεση πλεονασματικών δεδομένων




Αν όλοι οι γείτονες έχουν την ίδια κλάση, μπορούμε να αφαιρέσουμε το δείγμα δεδομένων

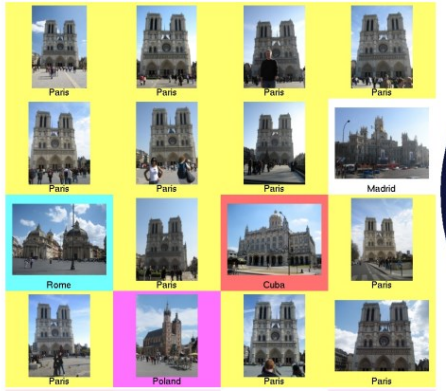
# Εφαρμογή Ταξινόμησης Πλησιέστερων Γειτόνων

- *Σε ποιο μέρος τραβήχτηκε η ακόλουθη φωτογραφία;*
  - James Hays, Alexei A. Efros. im2gps: estimating geographic information from a single image. CVPR'08. Project page: http://graphics.cs.cmu.edu/projects/im2gps/]



# Που τραβήχθηκε η ακόλουθη φωτογραφία;

- **Δεδομένα εκπαίδευσης**
  - 6 εκ. εικόνες από το Flickr που περιέχουν μεταδεδομένα τοποθεσίας
  - Αρκετά πυκνή δειγματοληψία σε όλη την υφήλιο
- Αναπαράσταση κάθε φωτογραφίας με συγκεκριμένα, περιγραφικά χαρακτηριστικά
- Πρόβλεψη τοποθεσίας για νέες φωτογραφίες μέσω ταξινόμησης $k$-πλησιέστερων γειτόνων
  - Οι ερευνητές βρήκαν ότι το βέλτιστο $k$ ήταν 120




# $k$-πλησιέστεροι γείτονες: Συμπεράσματα

- Σχηματίζουν περίπλοκα όρια απόφασης, τα οποία προσαρμόζονται στην πυκνότητα των δεδομένων εκπαίδευσης
- Σε περιπτώσεις που τα δεδομένα εκπαίδευσης είναι πολλά, η μέθοδος των $k$-πλησιέστερων γειτόνων λειτουργεί ικανοποιητικά
- **Ζητήματα**
  1. Ευαισθησία στο θόρυβο
  2. Ευαισθησία στο εύρος τιμών των χαρακτηριστικών των δεδομένων
  3. Η έννοια της απόστασης μεταξύ στιγμιοτύπων δεδομένων χάνει τη σημασία της όσο οι διαστάσεις μεγαλώνουν
  4. Γραμμική υπολογιστική πολυπλοκότητα συναρτήσει του πλήθους των δεδομένων εκπαίδευσης (ψευδο-πολυωνυμικός αλγόριθμος)

# Αφελής Μπεϋζιανός Ταξινομητής

Naïve Bayesian Classifier

# Αφελείς Μπεϋζιανοί Ταξινομητές

- **Naïve Bayesian Classifiers** (NBC)
  - Οικογένεια *πιθανοτικών ταξινομητών* οι οποίοι βασίζονται στο **Θεώρημα του Bayes**
  - Υποθέτουν ισχυρή ανεξαρτησία μεταξύ των χαρακτηριστικών των δεδομένων
    - Εξου και ο χαρακτηρισμός *αφελείς* (naïve)
- **Thomas Bayes** (1701-1761)
  - Βρετανός στατιστικολόγος, φιλόσοφος και ιερωμένος της Πρεσβυτεριανής Εκκλησίας
- **Θεώρημα Bayes** ή **Κανόνας Bayes**
  - Έστω $A_1, A_2, \cdots, A_n$ **αμοιβαίως αποκλειόμενα** (ανεξάρτητα) ενδεχόμενα που καθορίζουν δειγματικό χώρο $S$. Έστω $B$ ένα οποιοδήποτε ενδεχόμενο του χώρου, τέτοιο ώστε $P(B) > 0$. Τότε

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

# Θεώρημα Bayes: Απλή μορφή

- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, όπου $A, B$ ενδεχόμενα με $P(B) > 0$
- Το θεώρημα του Bayes
  - επιτρέπει την *ενημέρωση* της πιθανότητας εμφάνισης ενός ενδεχομένου, **ενσωματώνοντας νέα πληροφορία**
  - Ενσωματώνει τις *εκ των προτέρων πιθανότητες* (prior probabilities) για να δημιουργήσει εκ των *υστέρων πιθανότητες* (posterior probabilities)
- $P(A|B)$*: εκ των υστέρων πιθανότητα* εμφάνισης ενδεχομένου Α, δεδομένου ότι το ενδεχόμενο Β έχει συμβεί
- $P(A)$*: εκ των προτέρων πιθανότητα* εμφάνισης ενδεχομένου Α, πριν την πραγματοποίηση νέας παρατήρησης
- $P(B|A)$: *πιθανοφάνεια* (likelihood) ενδεχομένου $A$
  - πιθανότητα να συμβεί το $B$ ενώ το $A$ έχει ήδη συμβεί
- $P$(Β): πιθανότητα εμφάνισης ενδεχομένου Β
  - Καλείται και *«απόδειξη»* (evidence)

# Θεώρημα Bayes: Παράδειγμα

Έστω ότι έχουμε κατάστημα ηλεκτρολογικού υλικού και προμηθευόμαστε λαμπτήρες από *τρεις* κατασκευαστές: τον **A**, τον **B** και τον **C**. Πιο συγκεκριμένα ο **A** μας προμηθεύει το *80%* των λαμπτήρων που πουλάμε, ο **B** το *15%* και ο **C** το υπόλοιπο *5%*. Επίσης, οι κατασκευαστές μας έχουν ενημερώσει ο μεν **A** ότι το *4%* των λαμπτήρων του είναι ελλαττωματικό, ο **B** το *6%* και ο **C** το *9%*. Δεδομένου ότι ένας πελάτης μας επιστρέφει **έναν λαμπτήρα** πίσω ως **ελλαττωματικό**, **ποια** είναι η **πιθανότητα** να έχει κατασκευαστεί από τον **A**;

**Λύση**

$P(A) = 0{,}8, P(B) = 0{,}15, P(C) = 0{,}05$

$P(E|A) = 0{,}04, P(E|B) = 0{,}06, P(E|C) = 0{,}09$

Εφαρμογή θεωρήματος Bayes

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C)} = \frac{0{,}04 \cdot 0{,}8}{0{,}04 \cdot 0{,}8 + 0{,}06 \cdot 0{,}15 + 0{,}09 \cdot 0{,}05} \approx 0{,}7033$$

Άρα η πιθανότητα ο ελλαττωματικός λαμπτήρας να έχει κατασκευαστεί από τον Α είναι **περίπου 70,33%**

# Αφελής Μπεϋζιανός Ταξινομητής: Μοντέλο

- **Αφελής υπόθεση** (naïve assumption)
  - Όλα τα $n$ χαρακτηριστικά του νέου δείγματος δεδομένων $\boldsymbol{x}$ είναι ανεξάρτητα μεταξύ τους
  - $\boldsymbol{x} = \{x_1, x_{2,} \cdots, x_n\}$
  - $P(x_i |\, x_1, x_{2,} \cdots, x_{i-1,} x_{i+1,} \cdots, x_n, C_k) = P(x_i|C_k)$, $C_k$: Κλάση δείγματος
  - $P(\boldsymbol{x}|C_k) = \prod_{i=1}^{n} P(x_i|C_k)$
- **Εφαρμογή** Θεωρήματος Bayes
  - $P(C_k|\boldsymbol{x}) = \frac{P(C_k)P(\boldsymbol{x}|C_k)}{P(\boldsymbol{x})} \propto P(C_k)P(\boldsymbol{x}|C_k) \Longrightarrow P(C_k|\boldsymbol{x}) \propto P(C_k)\prod_{i=1}^{n} P(x_i|C_k)$
  - $P(\boldsymbol{x})$: Πιθανότητα εμφάνισης συγκεκριμένου δείγματος δεδομένων, σταθερά
    - *«Απόδειξη»* (evidence)
- **Ταξινόμηση** νέου δείγματος δεδομένων $\boldsymbol{x}$
  - $\hat{y} = \underset{k \in 1,\cdots K}{\operatorname{argmax}} P(C_k)\prod_{i=1}^{n} P(x_i|C_k)$

# Αφελής Μπεϋζιανός Ταξινομητής: Εκπαίδευση Μοντέλου

- **Ταξινομητής**: $\hat{y} = \underset{k \in 1,\cdots K}{\operatorname{argmax}} P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$
- **Δεδομένα εκπαίδευσης** $\{X, \boldsymbol{y}\}$
  - $m$ στιγμιότυπα (instances) $n$ χαρακτηριστικών το κάθε ένα, μαζί με την αντίστοιχη ετικέτα τους
- **Παράμετροι μοντέλου**
  1. $P(C_k)$: Υπολογισμός κατανομής κλάσεων στο σύνολο δεδομένων εκπαίδευσης
  2. $P(x_i|C_k)$ : Υπολογισμός πιθανότητας εμφάνισης κάθε χαρακτηριστικού σε κάθε κλάση, στο σύνολο δεδομένων εκπαίδευσης
- Με την προσθήκη νέων δεδομένων εκπαίδευσης, οι τιμές των παραμέτρων του μοντέλου *ενημερώνονται*

# Αφελής Μπεϋζιανός Ταξινομητής: Παράδειγμα

**Παρατηρήσεις βροχόπτωσης**

| **Θερμοκρασία** | **Υγρασία** | **Βροχή** |
|---|---|---|
| Κρύο | Υψηλή | Ναι |
| Κρύο | Χαμηλή | Όχι |
| Μέση | Χαμηλή | Ναι |
| Μέση | Μέτρια | Όχι |
| Ζέστη | Μέτρια | Όχι |
| Ζέστη | Υψηλή | Όχι |

Αν σήμερα η *Θερμοκρασία* είναι **Μέση** και η *Υγρασία* **Υψηλή**, θα βρέξει ή όχι;

**Λύση**

- $P(B) = \frac{2}{6}$, $P(OB) = \frac{4}{6}$
- $P(M|B) = \frac{1}{2}$, $P(M|OB) = \frac{1}{4}$
- $P(Y|B) = \frac{1}{2}$, $P(Y|OB) = \frac{1}{4}$
- $P(B|M,Y) \propto P(B)\ P(M|B)P(Y|B) = \frac{2}{6} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{12}$
- $P(OB|M,Y) \propto P(OB)\ P(M|OB)P(Y|OB) = \frac{4}{6} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{24}$
- Επειδή $P(B|M,Y) > P(OB|M,Y)$ ο αφελής μπεϋζιανός ταξινομητής προβλέπει ότι <u>θα βρέξει</u> σήμερα

# Συμπεράσματα

- **Περιορισμοί**
  - Ελλιπή δεδομένα εκπαίδευσης
    - Αν η τιμή κάποιου χαρακτηριστικού *δεν υπάρχει καθόλου* στα δεδομένα εκπαίδευσης, τότε *δεν μπορεί* να προσδιοριστεί σε ποια κλάση ανήκει το δείγμα.
  - Συνεχή χαρακτηριστικά
    - Πρέπει να βρεθεί η *υποκείμενη κατανομή* τους (πχ κανονική, Laplace, Poisson)
  - Μη-ανεξαρτησία μεταξύ των χαρακτηριστικών των δεδομένων
    - Εφαρμογή *τεχνικών απεικόνισης* των χαρακτηριστικών σε νέο χώρο, όπου η ανεξαρτησία τους είναι διασφαλισμένη ως ένα βαθμό (π.χ. ανάλυση κυρίων συνιστωσών)
- **Εφαρμογές**
  - Φίλτρα ενοχλητικής αλληλογραφίας
  - Εκτιμητές ρίσκου για χορήγηση δανείων/έκδοση πιστωτικών καρτών
  - …

# Γραμμική Ταξινόμηση

Linear Classification

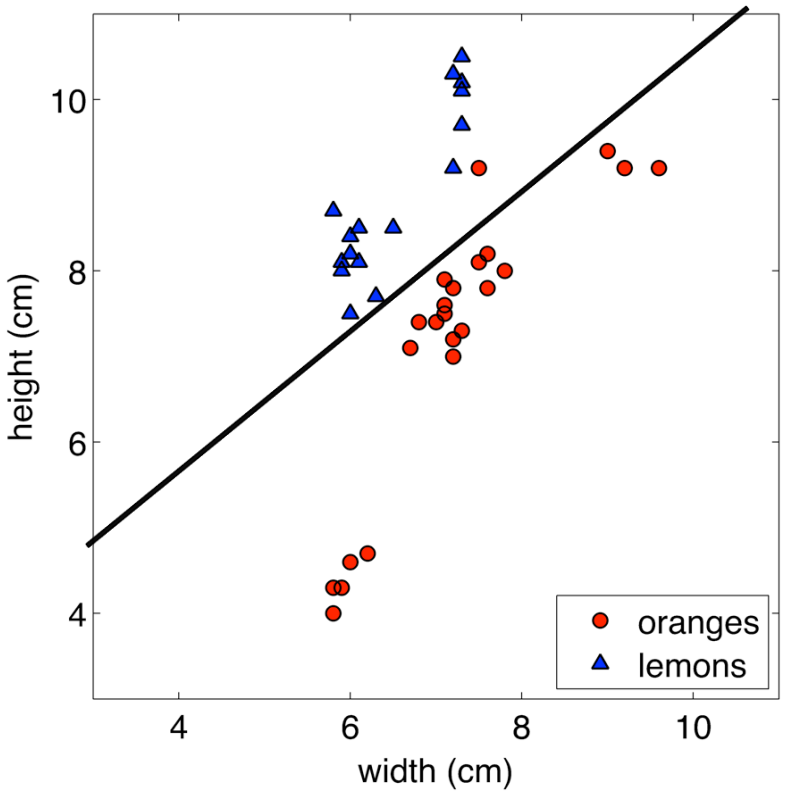
# Πρόβλημα ταξινόμησης: Πορτοκάλια ή Λεμόνια;

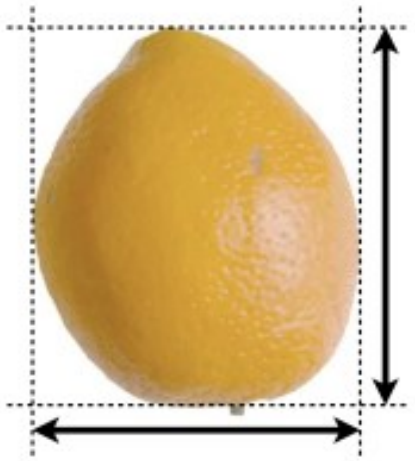
Πρόβλημα δυαδικής ταξινόμησης

Μπορούμε να βρούμε ένα γραμμικό όριο απόφασης που να χωρίζει τα χαρακτηριστικά (χώρο της εισόδου) σε υποπεριοχές όπου να κυριαρχούν δείγματα μόνο της μιας κλάσης

$$y = \text{sgn}(w_0 + w_1 x_1 + w_2 x_2)$$

sgn(): συνάρτηση προσήμου
$w_0, w_1, w_2$: παράμετροι μοντέλου



- Αν τις γνωρίζω, περιγράφω το χώρο και δεν χρειάζεται να αποθηκεύω στη μνήμη όλα τα δείγματα!

# Συνάρτηση Απόφασης

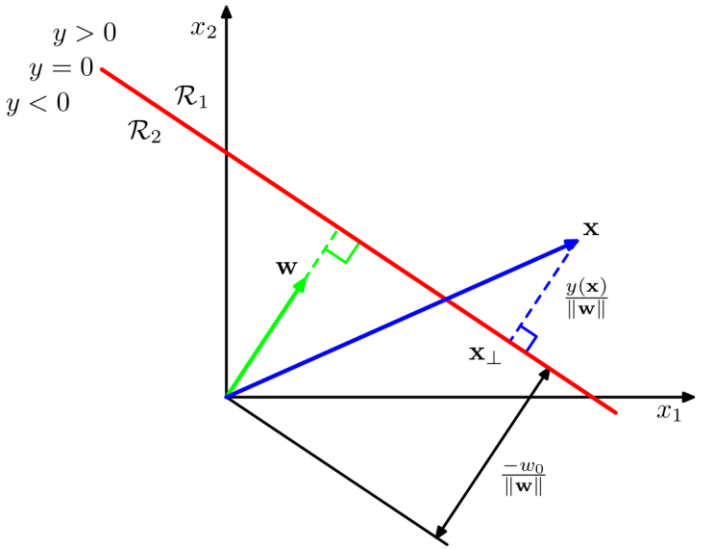
- Στόχος
  - Ανάθεση διανύσματος $\boldsymbol{x}$ στην κλάση $c_i$
- Γραμμική σχέση μεταξύ διανύσματος εισόδου και ταξινόμησης
  - $y(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0$
  - $\boldsymbol{w}$: **Διάνυσμα βαρών** (*weight vector*)
  - $w_0$: **Πόλωση** (*bias*)
- Παράδειγμα δυαδικής ταξινόμησης
  - Ταξινόμηση στην κλάση $c_1$ αν $y(\boldsymbol{x}) \geq 0$ ή στην $c_0$ αν $y(\boldsymbol{x}) < 0$
  - Δημιουργία **επιπέδου** ή **ορίου απόφασης** (*decision surface/boundary*)
- Έστω ότι έχουμε δύο διανύσματα $\boldsymbol{x}_1$ και $\boldsymbol{x}_2$ τα οποία ταξινομούνται πάνω στο όριο απόφασης: $y(\boldsymbol{x}_1) = y(\boldsymbol{x}_2) = 0 \Rightarrow \boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$
  - Διάνυσμα βαρών κάθετο στο όριο απόφασης: $\frac{\boldsymbol{w}^T \boldsymbol{x}}{\|\boldsymbol{w}\|} = -\frac{w_0}{\|\boldsymbol{w}\|}$

# Όριο απόφασης

- Οποιοδήποτε σημείο του χώρου μπορεί να εκφραστεί συναρτήσει της προβολής του πάνω στο επίπεδο διαχωρισμού

$$\boldsymbol{x} = \boldsymbol{x}_{\perp} + r\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$$

- Προκύπτει εύκολα πως $r = \frac{y(\boldsymbol{x})}{\|\boldsymbol{w}\|}$
  - Άσκηση για το σπίτι!
- Απόσταση επιπέδου διαχωρισμού από την αρχή των αξόνων οφείλεται στην πόλωση (παράγοντας $-\frac{w_0}{\|\boldsymbol{w}\|}$)



# Γραμμική ταξινόμηση πολλαπλών κλάσεων

- Αν έχουμε $k$ κλάσεις, ορίζουμε $k$ διαφορετικές γραμμικές συναρτήσεις απόφασης
  - $y_k(\boldsymbol{x}) = \boldsymbol{w}_k^T \boldsymbol{x} + w_{k0}$
- Λειτουργία ταξινόμησης
  - $j = \arg\max_{i \in k} y_i(\boldsymbol{x})$
- Όριο απόφασης μεταξύ κλάσεων $C_m$ και $C_n$ καθορίζεται από τη συνθήκη

$$y_m(\boldsymbol{x}) = y_n(\boldsymbol{x}) \Rightarrow (\boldsymbol{w_m} - \boldsymbol{w_n})^{\boldsymbol{T}} \boldsymbol{x} + (w_{m0} - w_{n0}) = 0$$

  - Ίδια μορφή με το όριο απόφασης της δυαδικής ταξινόμησης
  - Ισχύουν οι αντίστοιχες ιδιότητες

# Μέθοδος Ελαχίστων Τετραγώνων

- Χρησιμοποιείται για τον προσδιορισμό των χαρακτηριστικών των γραμμικών συναρτήσεων απόφασης
  - Δηλαδή των διανυσμάτων των βαρών και της πόλωσης
- Έστω ότι έχουμε πρόβλημα ταξινόμησης $k$ κλάσεων και επίσης
  1. $T = \{\boldsymbol{x}_i, t_i\}, i = 1, \dots, N$ το σύνολο των δεδομένων εκπαίδευσης
  2. $y_n(\boldsymbol{x}) = \boldsymbol{w}_n^T \boldsymbol{x} + w_{n0}, i = 1, \dots, k$ το γραμμικό μοντέλο για την κλάση $C_n$
- Εκφράζουμε τα $k$ γραμμικά μοντέλα υπό τη μορφή πίνακα $y(\boldsymbol{x}) = \widetilde{\boldsymbol{W}}^T \widetilde{\boldsymbol{x}}$, όπου
  - $\widetilde{\boldsymbol{W}}$ επαυξημένος πίνακας βαρών, του οποίου η $n$-οστή στήλη είναι το διάνυσμα $(w_{n0}, \boldsymbol{w}_n^T)^{\mathrm{T}}$
  - $\widetilde{\boldsymbol{x}} = (1, \boldsymbol{x}^T)^{\mathrm{T}}$ επαυξημένο διάνυσμα εισόδου
  - Ενσωματώνουμε την πόλωση στα βάρη
- $\widetilde{\boldsymbol{X}}$: πίνακας δεδομένων, του οποίο η $i$-οστή γραμμή είναι το διάνυσμα $\widetilde{\boldsymbol{x}}_i^T$
- $\mathbf{T}$: πίνακας ετικετών, του οποίου η $i$-οστή γραμμή είναι το διάνυσμα $\boldsymbol{t}_n^T$

# Μέθοδος Ελαχίστων Τετραγώνων (συνέχεια)

- **Αντικειμενική συνάρτηση** (*objective function*) μεθόδου ελαχίστων τετραγώνων
  - Επίσης γνωστή και ως **συνάρτηση σφάλματος/απώλειας** (*error/loss function*)

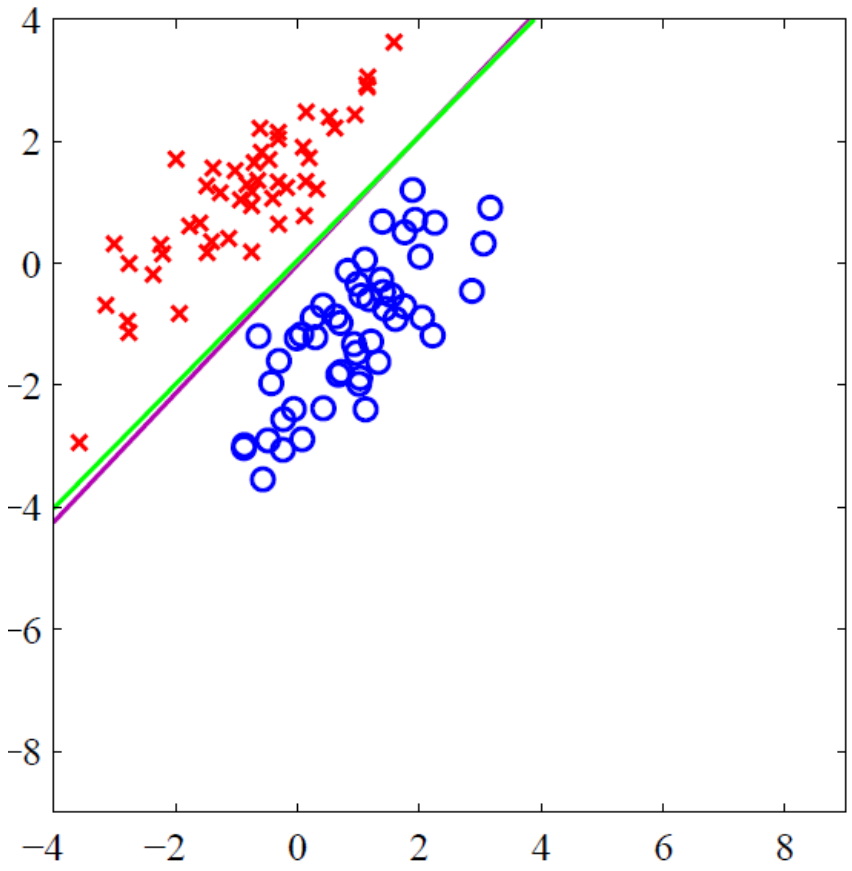
$$E(\widetilde{\boldsymbol{W}}) = \frac{1}{2}\mathrm{Tr}\left\{(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{W}} - \boldsymbol{T})^{T}(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{W}} - \boldsymbol{T})\right\}$$

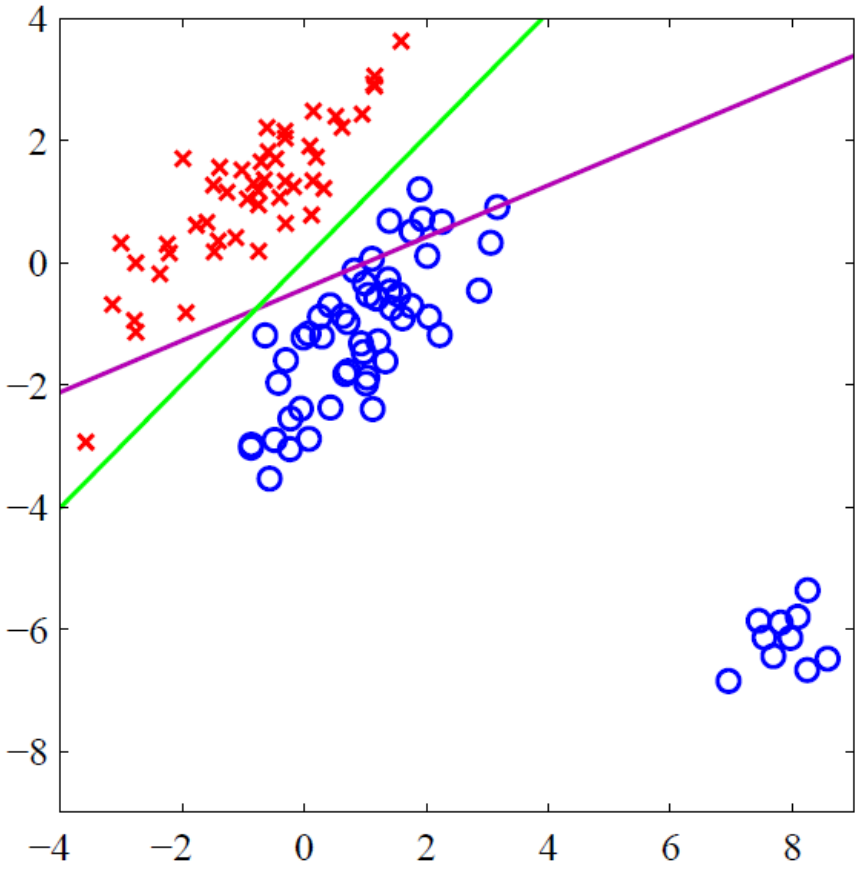
- Εύρεση ελαχίστου συνάρτησης $E(\widetilde{\boldsymbol{W}})$
  - $\frac{\partial E(\widetilde{\boldsymbol{W}})}{\partial \widetilde{\boldsymbol{W}}} = 0 \Longrightarrow \cdots \Longrightarrow \widetilde{\boldsymbol{W}} = (\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{T} = \widetilde{\boldsymbol{X}}^{+}\boldsymbol{T}$
  - $\widetilde{\boldsymbol{X}}^{+}$: ψευδοαντίστροφος του $\widetilde{\boldsymbol{X}}$
  - Οι πράξεις άσκηση για το σπίτι!
- Τελικά $y(\boldsymbol{x}) = \widetilde{\boldsymbol{W}}^{T}\widetilde{\boldsymbol{x}} = \mathbf{T}^{\mathrm{T}}(\widetilde{\boldsymbol{X}}^{+}\boldsymbol{T})^{\mathrm{T}}\widetilde{\boldsymbol{x}}$
  - Το διάνυσμα των βαρών προσδιορίστηκε πλήρως ντετερμινιστικά από τα δεδομένα εκπαίδευσης

# Μέθοδος Ελαχίστων Τετραγώνων: Χαρακτηριστικά

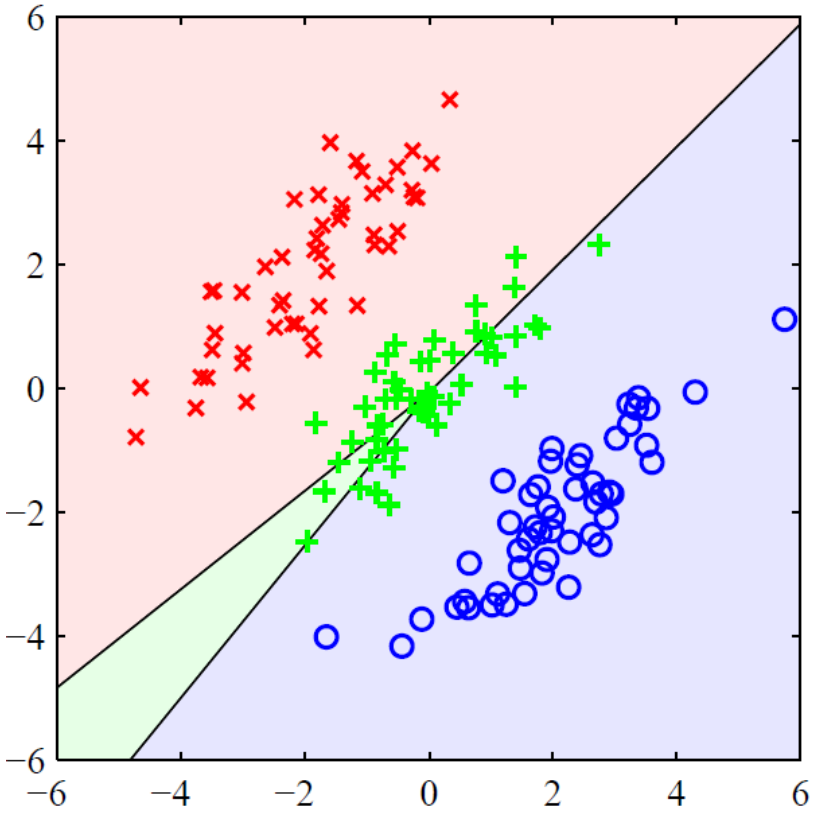
- Πλεονεκτήματα
  - Προσδιορισμός βαρών μέσω ακριβούς λύσης κλειστής μορφής
- Μειονεκτήματα
  - Πολύ μεγάλη ευαισθησία σε **έκτοπες τιμές** (*outliers*)
    - Δείγματα αρκετά μακριά από ομοειδή τους έχουν μεγάλη επίδραση στον καθορισμό του ορίου απόφασης εξαιτίας της βελτιστοποίησης του τετραγωνικού σφάλματος
  - Υποθέτει ότι τα χαρακτηριστικά των δεδομένων υπακούν στην κανονική κατανομή
    - Δεν μοντελοποιεί σωστά χαρακτηριστικά που ακολουθούν άλλες πιθανοτικές κατανομές

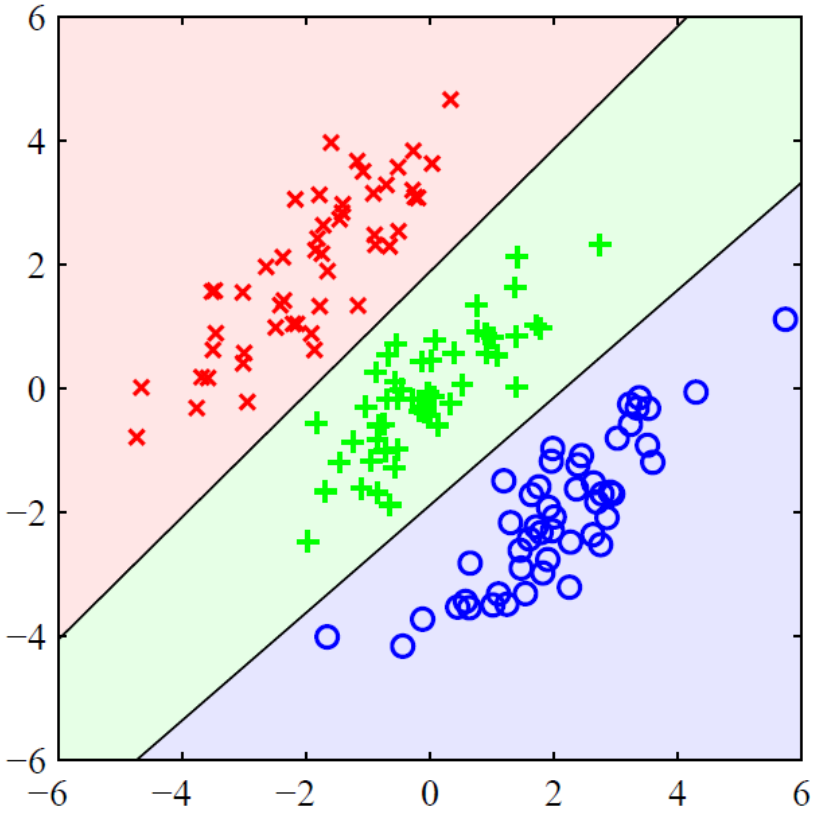
# Μέθοδος Ελαχίστων Τετραγώνων: Έκτοπες Τιμές



# Μέθοδος Ελαχίστων Τετραγώνων: Μη-γκαουσιανά χαρακτηριστικά



# Βιβλιογραφία


1. Ταξινομητές Πλησιέστερων Γειτόνων (Εκμάθηση μέσω Παραδειγμάτων)
   - M. Kubat – Εισαγωγή στη Μηχανική Μάθηση
     - Κεφάλαιο 3
2. Αφελείς Μπεϋζιανοί Ταξινομητές
   - M. Kubat – Εισαγωγή στη Μηχανική Μάθηση
     - Κεφάλαιο 2 – Ενότητες 2.1-2.4
3. Γραμμική Ταξινόμηση
   - C. M. Bishop – Αναγνώριση Προτύπων και Μηχανική Μάθηση
     - Κεφάλαιο 4 – Ενότητα 4.1 – Υποενότητες 4.1.1-4.1.3