

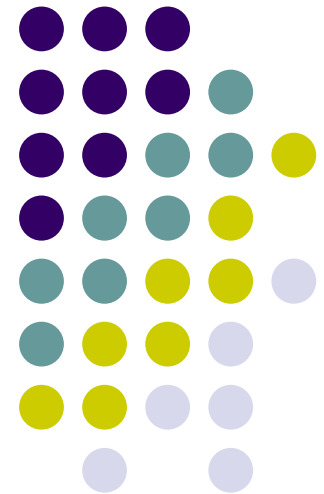
Θεωρία Αποφάσεων

Σ. Λυκοθανάσης, Καθηγητής

Δ. Κοσμόπουλος, Αναπληρωτής Καθηγητής

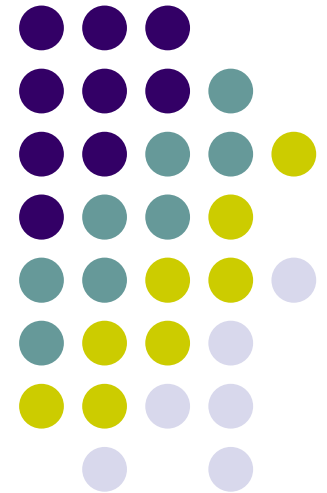
Τμήμα Μηχανικών Η/Υ & Πληροφορικής -
Εργαστήριο Αναγνώρισης Προτύπων

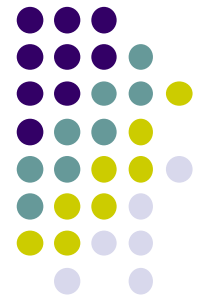
Διευθυντής: Σ. Λυκοθανάσης, Καθηγητής



Κεφάλαιο 4

Μη Παραμετρικές Τεχνικές





Μη παραμετρικές μέθοδοι

- Ούτε η σ.π.π είναι γνωστή, ούτε η διακρίνουσα συνάρτηση

Συμβαίνει συχνά!

Το μόνο που έχουμε είναι τα δεδομένα (με ετικέτες)

Σολομός

Πέρκα

Σολομός

Σολομός



Θέλουμε να εκτιμήσουμε την σ.π.π. από τα δεδομένα

Μη παραμετρικές τεχνικές



- Μη παραμετρικές τεχνικές μπορούν να χρησιμοποιηθούν με αυθαίρετες κατανομές χωρίς καθόλου υποθέσεις για την μορφή τους

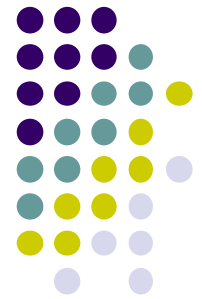
Θα ασχοληθούμε με δυο μεθόδους μη παραμετρικής εκτίμησης:

Παράθυρα Parzen

Εκτιμούν πιθανοφάνεια $p(\mathbf{x}|c_j)$

Μεθοδος των k Πλησιέστερων Γειτόνων (k -Nearest Neighbors)

Μπορούν απ' ευθείας να υπολογίσουν τις εκ των υστέρων πιθανότητες $P(c_j|\mathbf{x})$

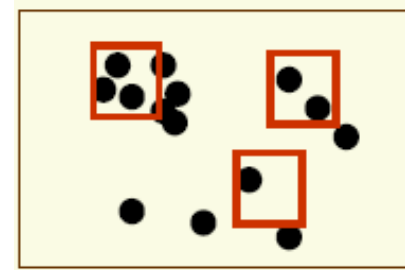


Δυο προσεγγίσεις

$$p(x) \approx \frac{k/n}{V}$$

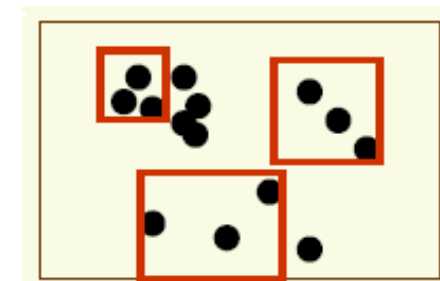
Παράθυρα Parzen

Επίλεξε σταθερό V και υπολόγισε k από τα δεδομένα



Εκτίμηση k Πλησιέστερων Γειτόνων (k-Nearest Neighbors, kNN)

Επίλεξε σταθερό k και υπολόγισε V από τα δεδομένα



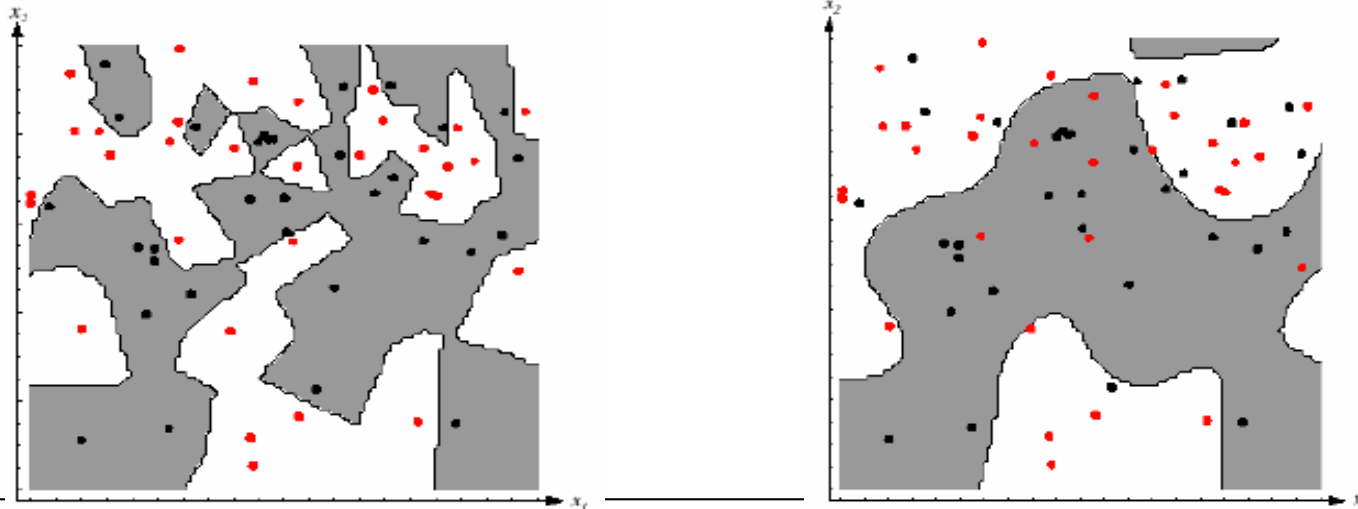
Υπό κατάλληλες συνθήκες και καθώς ο αριθμός των δειγμάτων τείνει προς το άπειρο και οι δυο εκτιμήσεις μπορεί να αποδειχθεί ότι συγκλίνουν στην πραγματική $p(x)$

Ταξινόμηση με Χρήση Παραθύρων Parzen



**Πολύ μικρό παράθυρο →
Πολύ μικρή διαμέριση του
χώρου χαρακτηριστικών,
πράγμα μη επιθυμητό!**

**Μεγαλύτερο παράθυρο
→ Υψηλότερο λάθος κατά την
εκπαίδευση, αλλά καλύτερη
απόδοση γενίκευσης! Καλύτερη
απόδοση γενίκευσης:
Επιθυμητή ιδιότητα.**



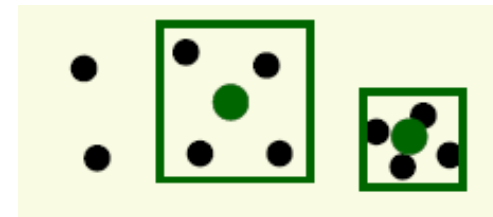
**Στην πράξη, αυτό που θα θέλαμε είναι παράθυρα μικρού πλάτους
στις περιοχές με υψηλή πυκνότητα δεδομένων, και παράθυρα
μεγάλου πλάτους στις περιοχές όπου τα δεδομένα είναι αραιά! Πώς
μπορεί να επιτευχθεί αυτό...;**

k -Πλησιέστεροι Γείτονες (k -NN)



- Η προσέγγιση k -NN μοιάζει με καλή λύση για το πρόβλημα της επιλογής του «καταλληλότερου» μεγέθους του παραθύρου
 - Υποθέτουμε ότι ο όγκος του κελιού (cell) είναι συνάρτηση των δεδομένων εκπαίδευσης
 - «Κεντράρουμε» ένα κελί στο x και το αφήνουμε να μεγαλώσει μέχρις ότου να περιλαμβάνει k δείγματα
 - Τα k αυτά δείγματα ονομάζονται οι k πλησιέστεροι γείτονες του x

■ 2 δυνατές περιπτώσεις:

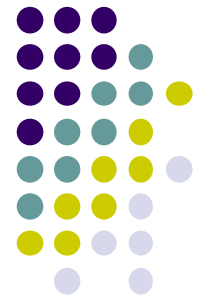


- Η πυκνότητα είναι μεγάλη κοντά στο x : οπότε το κελί θα είναι μικρό και θα παρέχει έχει καλή ανάλυση.
- Η πυκνότητα είναι μικρή κοντά στο x : το κελί θα μεγαλώσει πολύ και θα σταματήσει όταν συναντήσει περιοχές μεγαλύτερης πυκνότητας



Πιο «μαθηματικά»...

- ↳ Επιλέγουμε μια αρχική περιοχή γύρω από το \mathbf{x} όπου θα θέλαμε να υπολογίσουμε την $p(\mathbf{x})$
- ↳ Αυξάνουμε το παράθυρο μέχρι ένα προκαθορισμένο πλήθος k_n δειγμάτων να περιληφθεί εντός του παραθύρου. Αυτοί είναι οι k_n πλησιέστεροι γείτονες του \mathbf{x} .
- ↳ Υπολογίζουμε την πυκνότητα με βάση την τιμή $\frac{k_n/n}{V_n}$
- ↳ Αναγκαίες και ικανές συνθήκες για τη σύγκλιση της $p_n(\mathbf{x})$: $\lim_{n \rightarrow \infty} k_n = \infty$; $\lim_{n \rightarrow \infty} k_n/n = 0$



k -Πλησιέστεροι Γείτονες

Φυσικά τώρα δημιουργείται ένα καινούργιο ερώτημα

- Πώς να διαλέξουμε το k ;

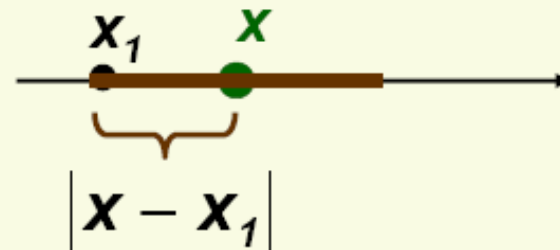
Ένας καλός εμπειρικός κανόνας είναι $k = \sqrt{n}$

- Μπορεί να αποδειχθεί ότι συγκλίνει όταν το n τείνει στο άπειρο
- Όμως, όχι πολύ χρήσιμο στην πράξη

Παράδειγμα 1- D:

Έστω ότι έχουμε ένα δείγμα, δηλαδή $n=1$

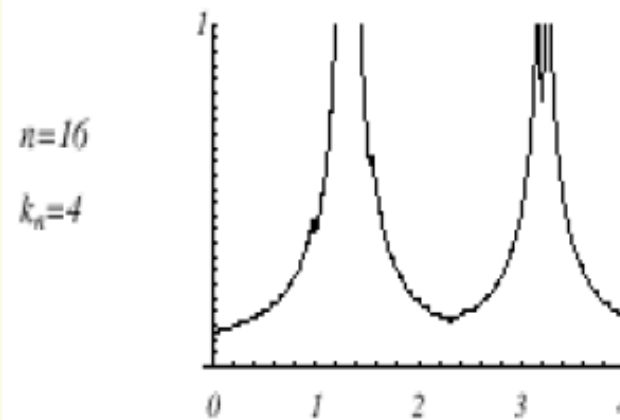
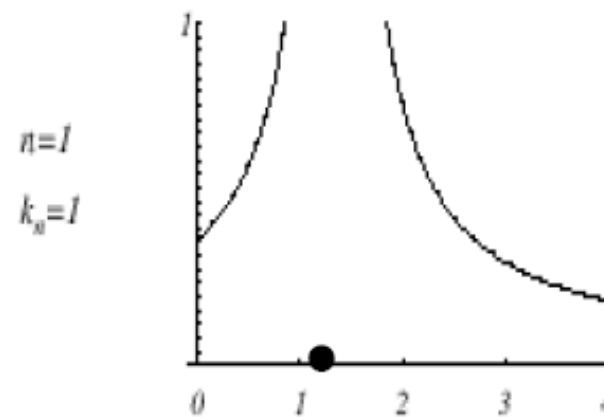
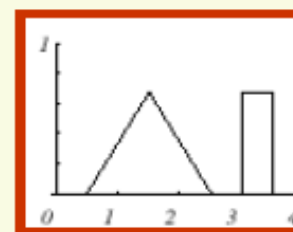
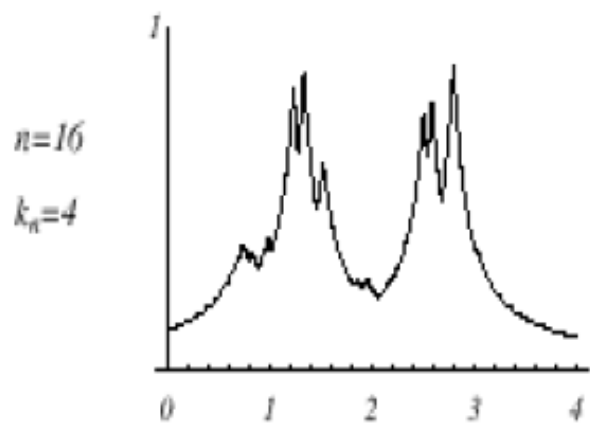
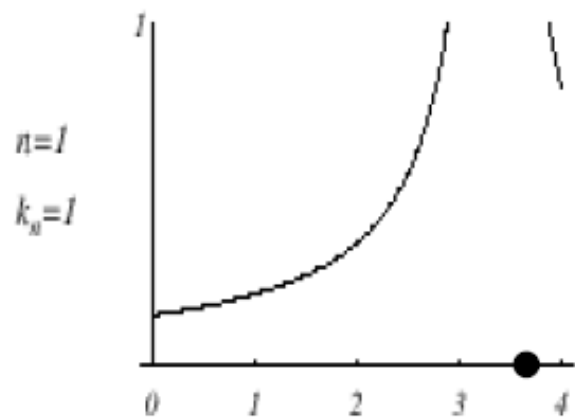
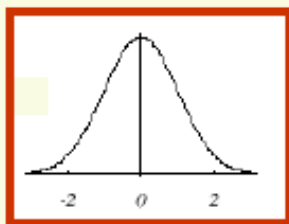
$$p(x) \approx \frac{k/n}{V} = \frac{1}{2|x - x_1|}$$



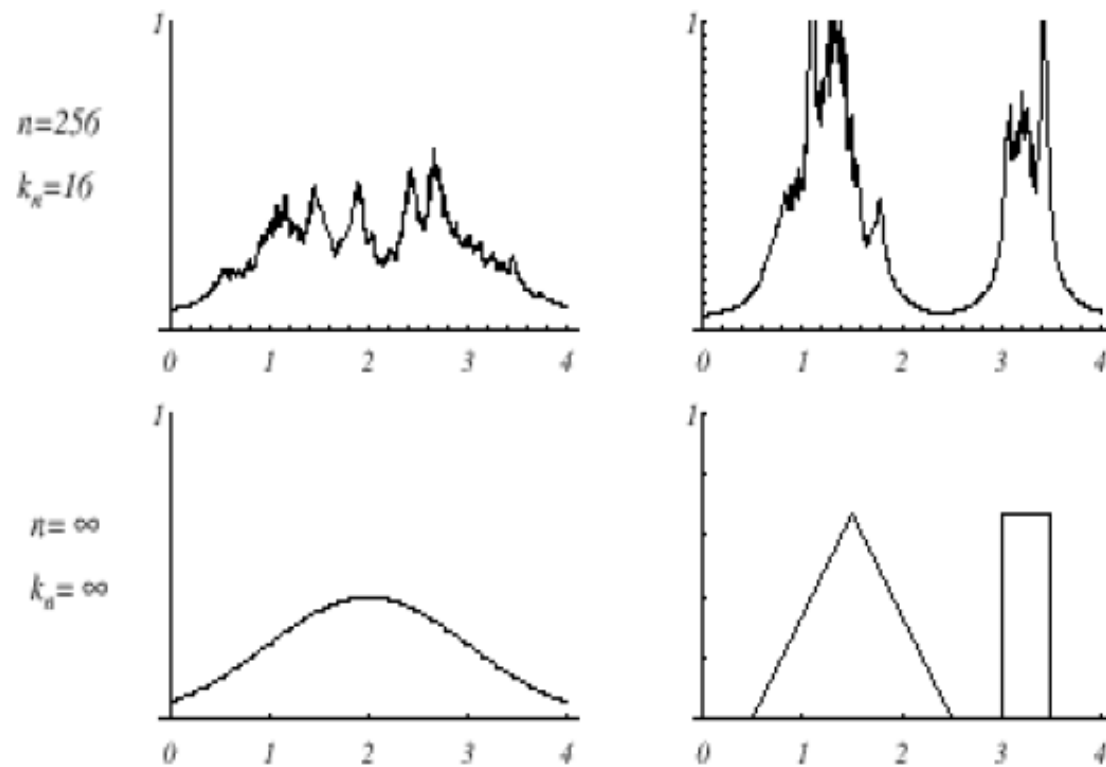
Όμως η $p(x)$ δεν είναι καν σ.π.π

$$\int_{-\infty}^{\infty} \frac{1}{2|x - x_1|} dx = \infty \neq 1$$

Εκτίμηση σ.π.π



Εκτίμηση σ.π.π



Τι παρατηρείτε?

k -Πλησιέστεροι Γείτονες



- Η εκτίμηση της πυκνότητας πιθανότητας $p(x)$ με προσέγγιση KNN, δεν δουλεύει και πολύ καλά στην πράξη, διότι:
 - Δεν πρόκειται καν για σ.π.π.
 - Έχει πολλές ασυνέχειες (έχει πολλές κορυφές, δεν είναι διαφορίσιμη)
 - Ακόμη και για μεγάλες περιοχές όπου δεν παρατηρούνται δείγματα, η τιμή της σ.π.π που εκτιμάται, διαφέρει σημαντικά από το 0 (tails are too heavy).

Θεωρητικά, αν είχαμε άπειρα δείγματα διαθέσιμα, θα μπορούσαμε με την μέθοδο των k πλησιέστερων γειτόνων να δημιουργήσουμε μια σειρά εκτιμήσεων, οι οποίες να συγκλίνουν στην πραγματική σ.π.π. Στην πράξη όμως ο αριθμός των δειγμάτων είναι πάντα περιορισμένος.

k -Πλησιέστεροι Γείτονες



- Δεν είναι όμως λόγος να παρατήσουμε την k -NN προσέγγιση...
- Αντί να προσεγγίζουμε την $p(x)$, ο k -NN μπορεί να χρησιμοποιηθεί για να εκτιμήσει τις εκ των υστέρων πιθανότητες $P(c_i|x)$

Δεν μας χρειάζεται η $p(x)$, αν μπορούμε να έχουμε μια καλή εκτίμηση των $P(c_i|x)$



k -Πλησιέστεροι Γείτονες

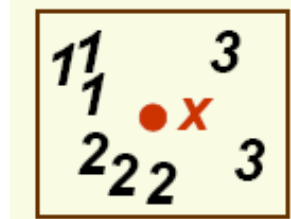
- Πώς μπορούμε να εκτιμήσουμε $P(c_i|x)$ από ένα σύνολο n δειγμάτων?

Ανακαλούμε ότι: $p(x) \approx \frac{k/n}{V}$

Τοποθετούμε ένα κελί με όγκο V γύρω από το x ώστε να περιλάβουμε k δείγματα

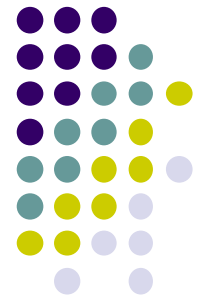
Αν k_i δείγματα μεταξύ των k έχουν ετικέτα c_i

$$p(c_i, x) \approx \frac{k_i/n}{V}$$



Οπότε:

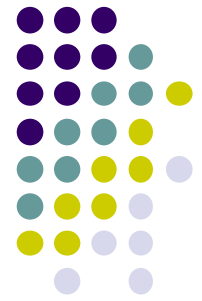
$$p(c_i | x) = \frac{p(x, c_i)}{p(x)} = \frac{p(c_i, x)}{\sum_{j=1}^m p(x, c_j)} \approx \frac{k_i/n}{V \sum_{j=1}^m \frac{k_j/n}{V}} = \frac{k_i}{\sum_{j=1}^m k_j} = \frac{k_i}{k}$$



k -Πλησιέστεροι Γείτονες

Οι εκ των υστέρων πιθανότητες για κάθε κλάση ω_i , σε κάθε μικρή περιοχή του x είναι το ποσοστό των δειγμάτων εντός της περιοχής που έχουν ετικέτα ω_i .

- ↳ n : συνολικός αριθμός προτύπων όλων των κλάσεων
- ↳ k_i : αριθμός προτύπων της κλάσης i στην περιοχή γύρω από το x
- ↳ k : συνολικός αριθμός προτύπων όλων των κλάσεων στην περιοχή γύρω από το x



k -Πλησιέστεροι Γείτονες

- Πολύ απλός κανόνας και συμφωνεί με την «διαισθητική» μας εκτίμηση.
- Αν θεωρήσουμε συνάρτηση κόστους την μηδέν-ένα (ταξινομητής MAP) επιλεγούμε την κλάση, η οποία έχει τα περισσότερα δείγματα στο κελί.
- Όταν δηλαδή δίνεται ένα νέο δείγμα x (χωρίς ετικέτα), βρίσκουμε τα k κοντινότερα δείγματα (μεταξύ των δειγμάτων εκπαίδευσης) και αντιστοιχούμε στο x την ετικέτα της πιο συχνά εμφανιζόμενης κλάσης μεταξύ των γειτόνων.

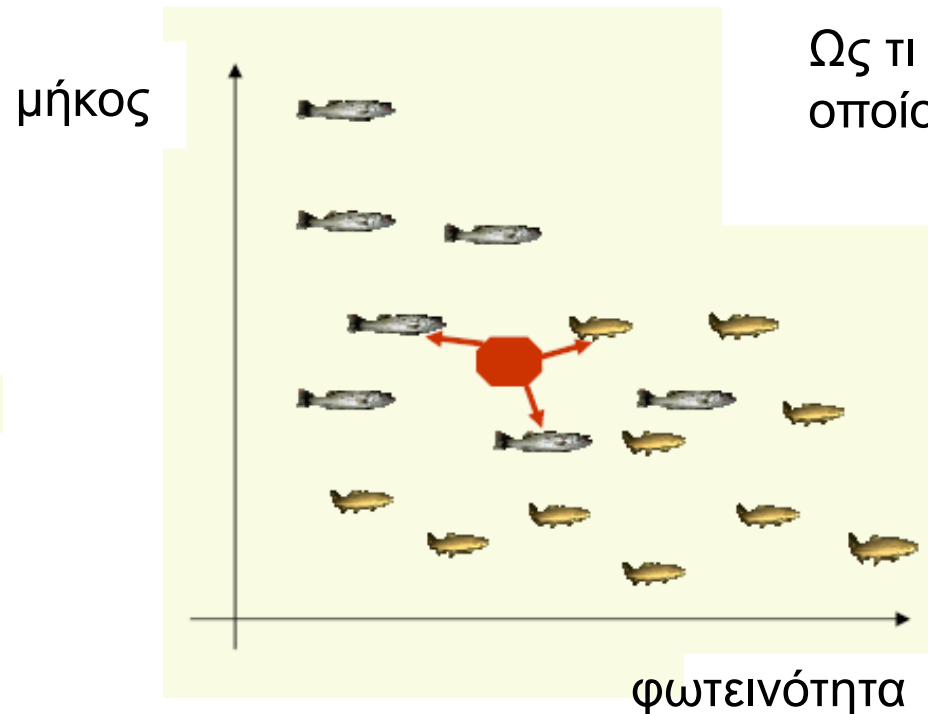
Παράδειγμα



- Έστω ότι έχουμε 2 χαρακτηριστικά (μήκος και φωτεινότητα) και τα δείγματα εκπαίδευσης που φαίνονται στην εικόνα
- Έστω επίσης $k=3$

Πέρκα: 

Σολομός: 



Ως τι θα κατατάξουμε ένα νέο ψάρι, το οποίο βρίσκεται στο κόκκινο πλαίσιο?

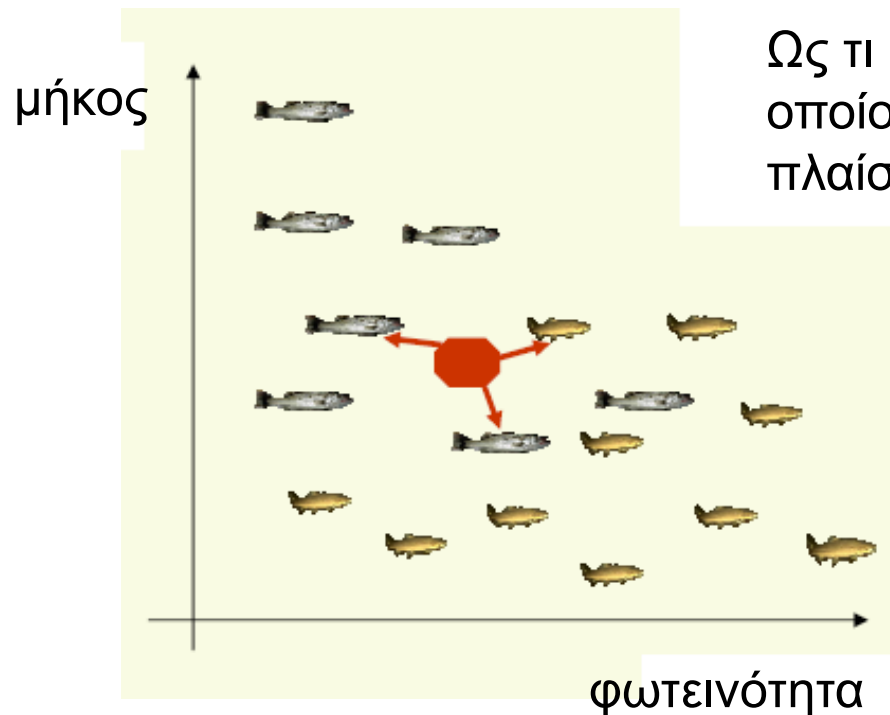
Παράδειγμα



- Έστω ότι έχουμε 2 χαρακτηριστικά (μήκος και φωτεινότητα) και τα δείγματα εκπαίδευσης που φαίνονται στην εικόνα
- Έστω επίσης $k=3$

Πέρκα: 

Σολομός: 



Ως τι θα κατατάξουμε ένα νέο ψάρι, το οποίο βρίσκεται στο πορτοκαλί πλαίσιο?

Κοντινότεροι γείτονες είναι:
2 πέρκες, 1 σολομός

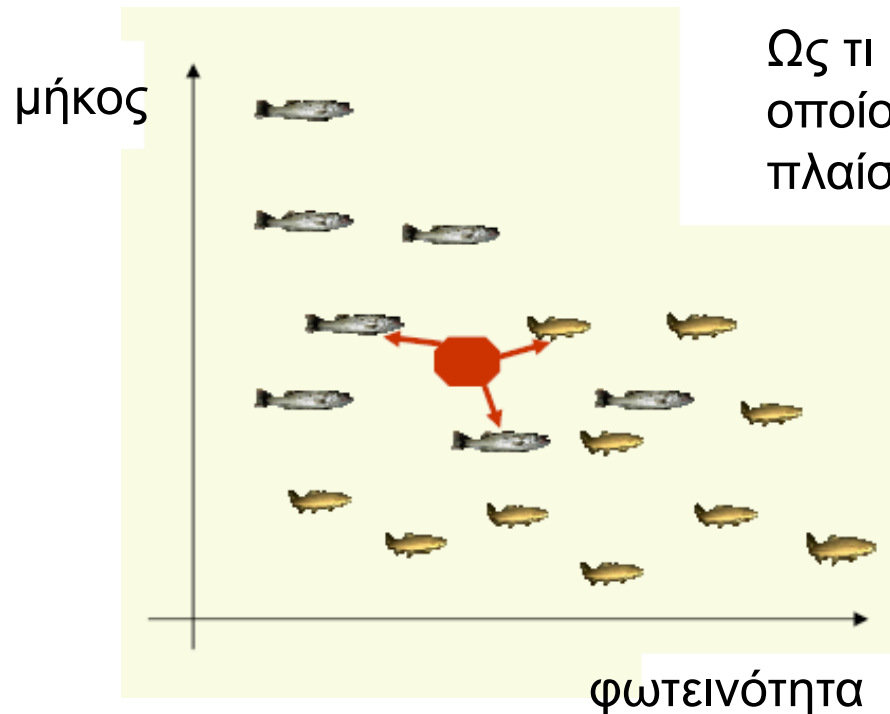
Παράδειγμα



- Έστω ότι έχουμε 2 χαρακτηριστικά (μήκος και φωτεινότητα) και τα δείγματα εκπαίδευσης που φαίνονται στην εικόνα
- Έστω επίσης $k=3$

Πέρκα: 

Σολομός: 

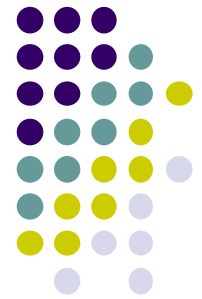


Ως τι θα κατατάξουμε ένα νέο ψάρι, το οποίο βρίσκεται στο πορτοκαλί πλαίσιο?

Κοντινότεροι γείτονες είναι:
2 πέρκες, 1 σολομός



Πέρκα!



k -Πλησιέστεροι Γείτονες

- Ο k NN κανόνας είναι σίγουρα απλός, αλλά δίνει καλά αποτελέσματα ;
- Υποθέτουμε ότι έχουμε άπειρα δείγματα
- Εξ ορισμού το βέλτιστο σφάλμα δίνεται κατά Bayes, έστω E^* .
- Ο k NN κανόνας δίνει μεγαλύτερο σφάλμα από το E^*
- Ακόμη όμως και για $k=1$ καθώς $n \rightarrow \infty$, μπορεί να αποδειχθεί ότι το σφάλμα είναι μικρότερο του $2E^*$.
- Όσο αυξάνουμε το k , τόσο το άνω όριο του ποσοστού λάθους γίνεται καλύτερο, δηλαδή το ποσοστό λάθους (καθώς $n \rightarrow \infty$) είναι μικρότερος του cE^* , όπου το c μικραίνει όσο μεγαλώνει το k .
- Αν έχουμε πάρα πολλά δείγματα ο k NN κανόνας θα δουλεύει πολύ καλά!

Ταξινομητής Πλησιέστερου Γείτονα



- Ο πιο απλός ταξινομητής KNN είναι αυτός με $k=1$! Αυτός ο ταξινομητής αντιστοιχίζει το x στην τάξη του πλησιέστερου γείτονά του. Είναι καλός αυτός ο ταξινομητής...?

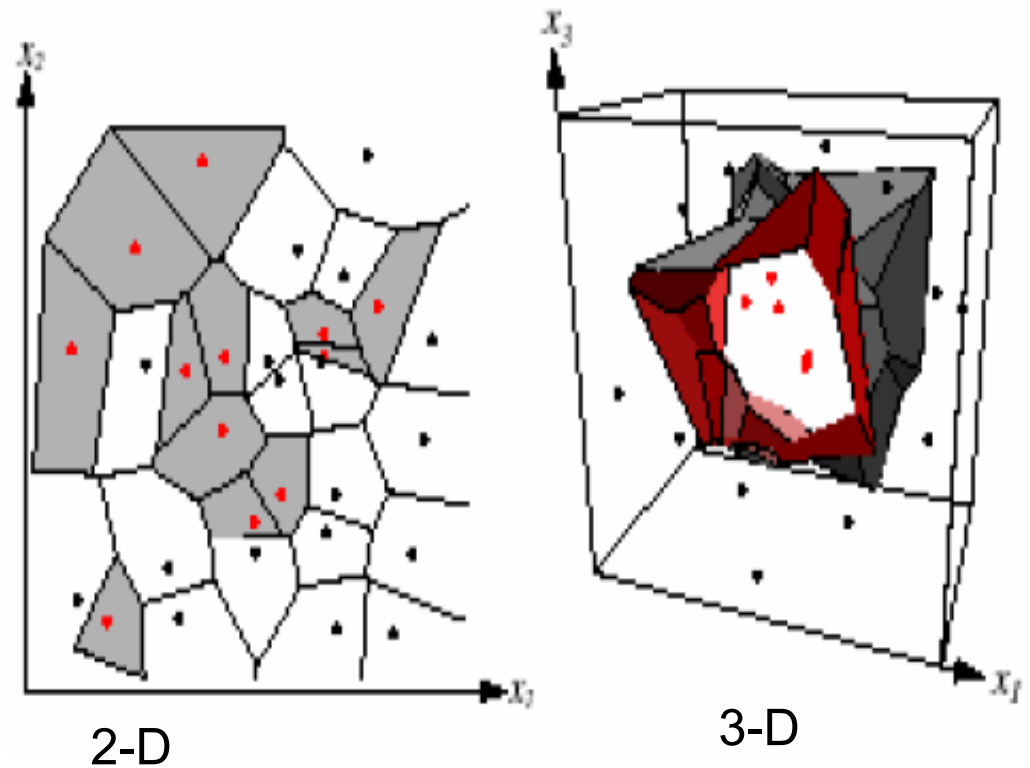
Ταξινομητής Πλησιέστερου Γείτονα



- Ο πιο απλός ταξινομητής KNN είναι αυτός με $k=1$! Αυτός ο ταξινομητής αντιστοιχίζει το x στην τάξη του πλησιέστερου γείτονά του. Είναι καλός αυτός ο ταξινομητής...?

- Ο NN ταξινομητής οδηγεί στο διαχωρισμό του χώρου ως ενός μωσαϊκού Voronoi, όπου κάθε κελί παίρνει την ετικέτα της κλάσης την οποία περιέχει.

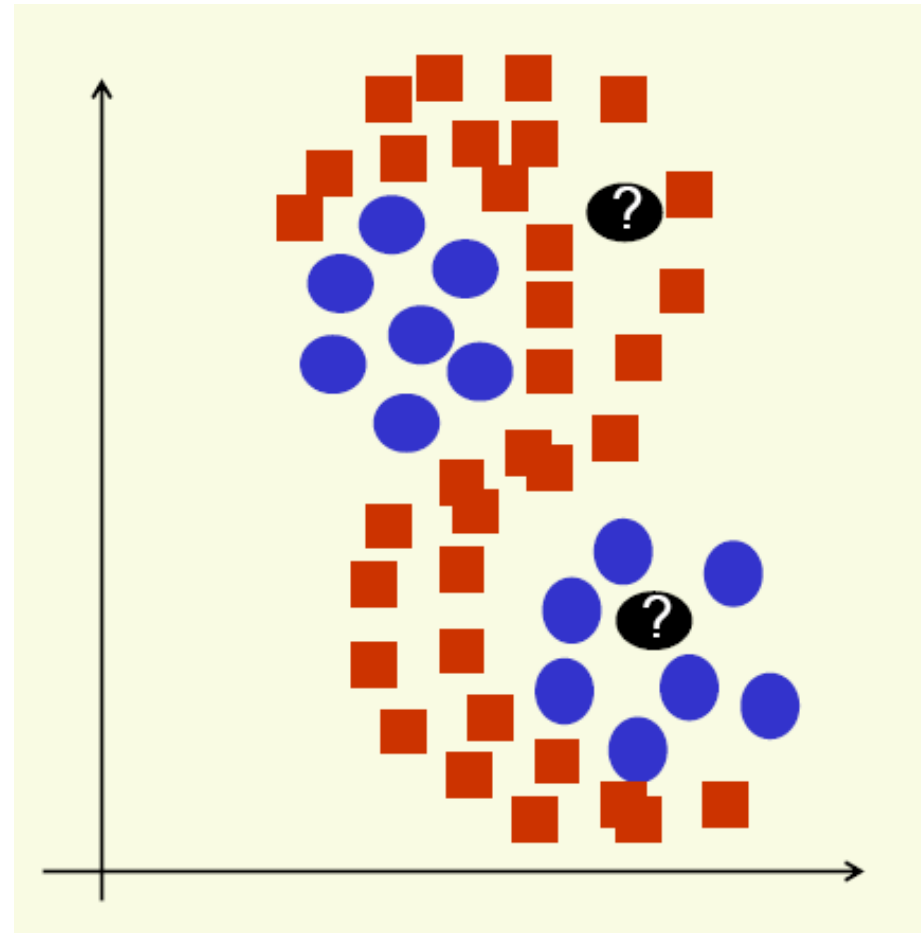
- Δεδομένου απείρου αριθμού δειγμάτων εκπαίδευσης, η πιθανότητα λάθους ταξινόμησης έχει ως πάνω όριο, το διπλάσιο της πιθανότητας σφάλματος του Μπεϋζιανού ταξινομητή (ισχύει για μικρές πιθανότητες λάθους).



Multi-modal σ.π.π.



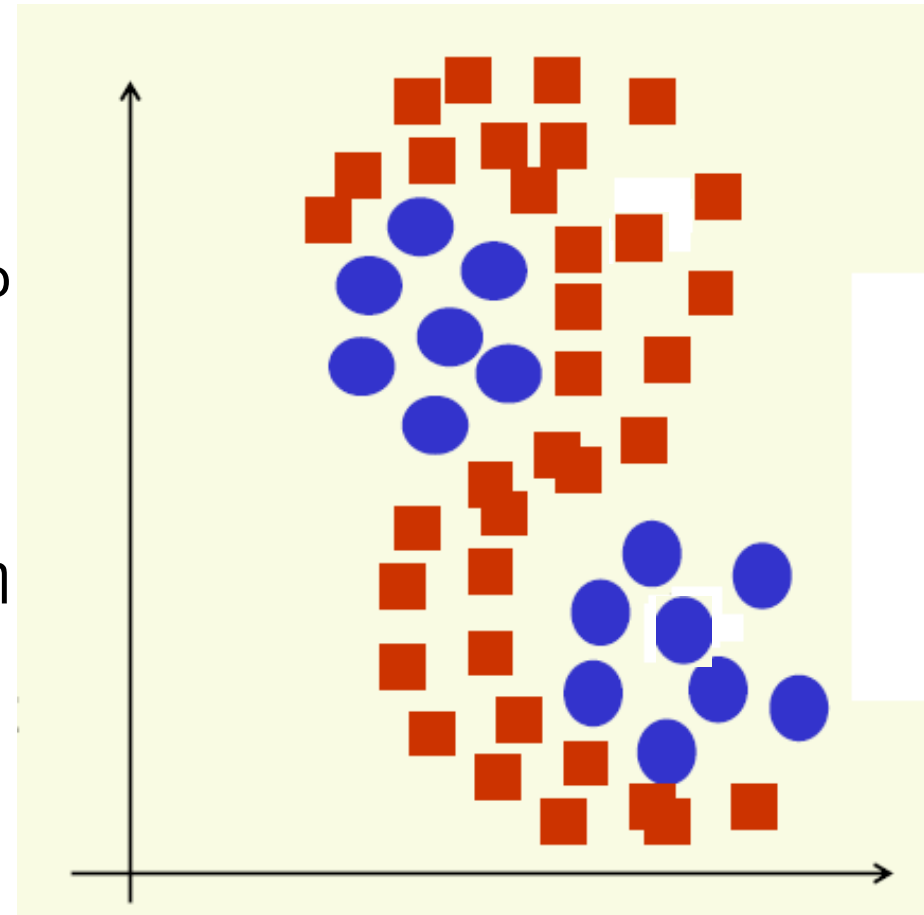
- Οι περισσότερες παραμετρικές τεχνικές εκτίμησης της σ.π.π. δεν θα έδιναν καλά αποτελέσματα για αυτό το πρόβλημα ταξινόμησης σε δυο κλάσεις:
- Ο KNN θα δουλέψει καλά, με την προϋπόθεση ότι υπάρχουν πολλά δείγματα εκπαίδευσης



Multi-modal σ.π.π.



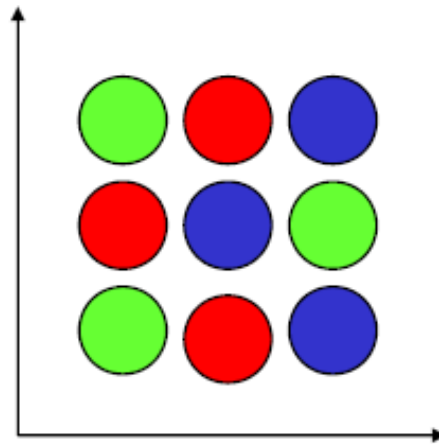
- Οι περισσότερες παραμετρικές τεχνικές εκτίμησης της σ.π.π. δεν θα έδιναν καλά αποτελέσματα για αυτό το πρόβλημα ταξινόμησης σε δυο κλάσεις:
- Ο KNN θα δουλέψει καλά, με την προϋπόθεση ότι υπάρχουν πολλά δείγματα εκπαίδευσης



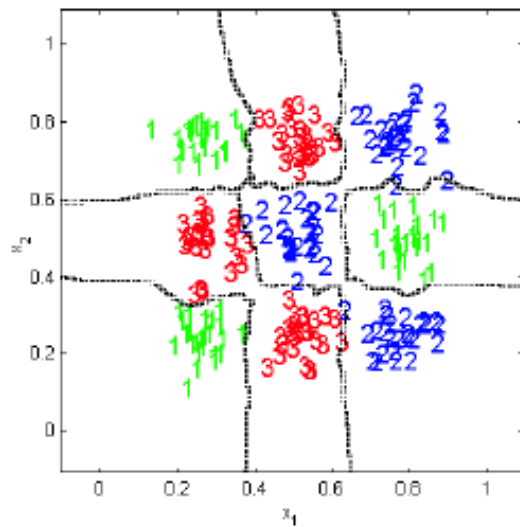
Παράδειγμα



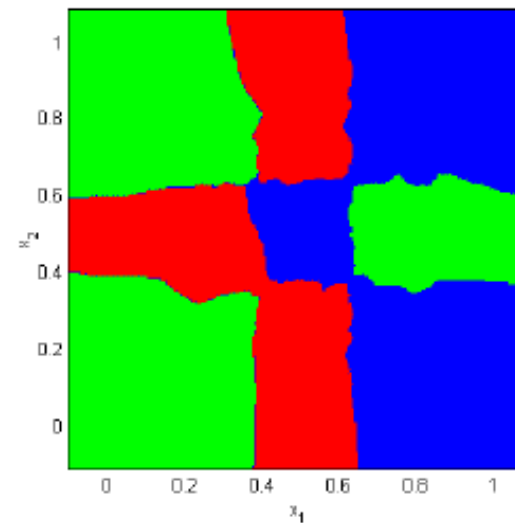
Για $k=5$:



Όρια απόφασης



Περιοχές απόφασης



Επιλογή του k



- Θεωρητικά, αν είναι διαθέσιμος άπειρος αριθμός δειγμάτων, όσο μεγαλύτερο είναι το k τόσο το καλύτερο (το σφάλμα προσεγγίζει περισσότερο το σφάλμα του Bayes)
- Το πρόβλημα όμως είναι ότι για να ισχύει αυτό όλοι οι k γείτονες πρέπει να είναι κοντά στο x
 - Δυνατό όταν ?
 - Αδύνατο ?

Επιλογή του k



- Θεωρητικά, αν είναι διαθέσιμος άπειρος αριθμός δειγμάτων, όσο μεγαλύτερο είναι το k τόσο το καλύτερο (το σφάλμα προσεγγίζει περισσότερο το σφάλμα του Bayes)
- Το πρόβλημα όμως είναι ότι για να ισχύει αυτό όλοι οι k γείτονες πρέπει να είναι κοντά στο x
 - Δυνατό όταν είναι διαθέσιμος άπειρος αριθμός δειγμάτων
 - Αδύνατο στην πράξη, επειδή ο αριθμός των δειγμάτων είναι πεπερασμένος

Επιλογή του k



- Στη πράξη

1. Το k πρέπει να είναι μεγάλο ώστε να ελαχιστοποιείται το ποσοστό λάθους

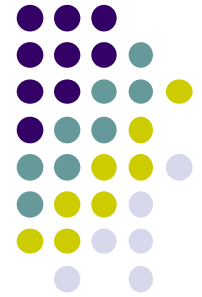
- Πολύ μικρό k θα οδηγήσει σε όρια απόφασης με ?
- Επιπλέον, μεγάλο k δίνει εκτίμηση για την ακρίβεια της ταξινόμησης

2. Το k πρέπει να είναι αρκετά μικρό ώστε να λαμβάνονται υπόψη μόνο κοντινά σημεία.

- Πολύ μεγάλο k θα οδηγήσει σε ? όρια απόφασης
- Επιπλέον αυξάνει το υπολογιστικό κόστος

Δεν είναι καθόλου απλό να βρει κανείς τον κατάλληλο συμβιβασμό!

Επιλογή του k



- Στη πράξη

1. Το k πρέπει να είναι μεγάλο ώστε να ελαχιστοποιείται ο ρυθμός λάθους

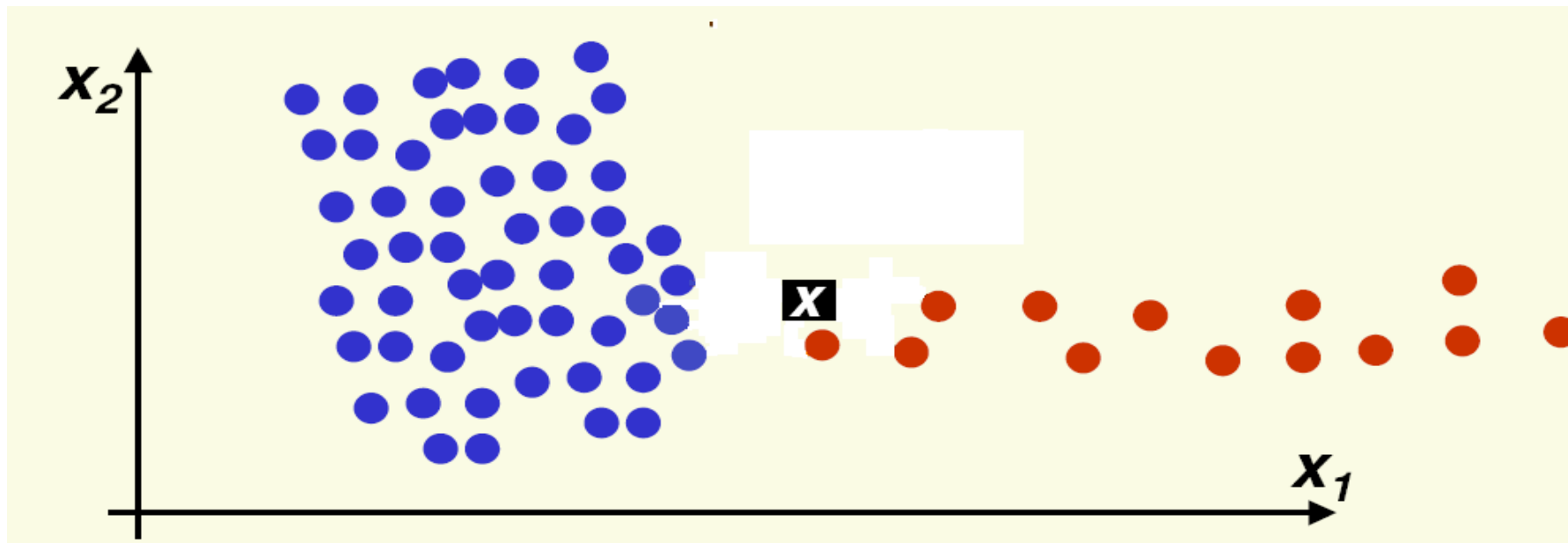
- Πολύ μικρό k θα οδηγήσει σε όρια απόφασης με θόρυβο
- Επιπλέον, μεγάλο k δίνει εκτίμηση για την ακρίβεια της ταξινόμησης

2. Το k πρέπει να είναι αρκετά μικρό ώστε να λαμβάνονται υπόψη μόνο κοντινά σημεία.

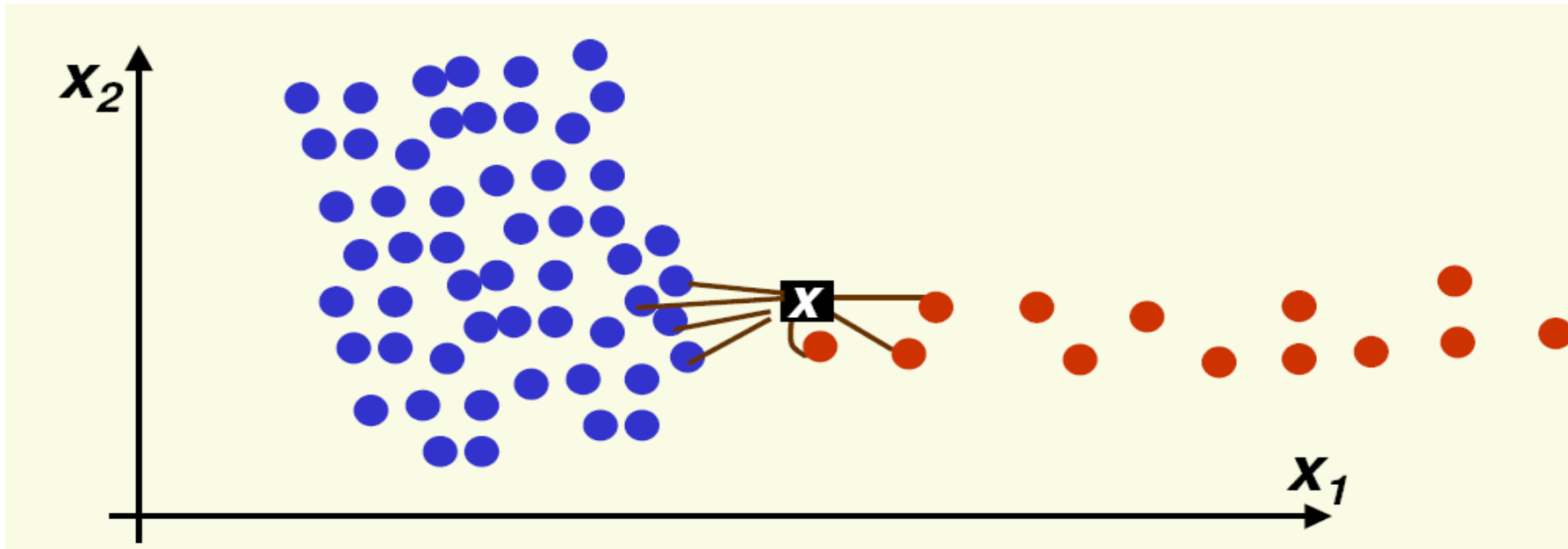
- Πολύ μεγάλο k θα οδηγήσει σε «oversmoothed» όρια απόφασης
- Επιπλέον αυξάνει το υπολογιστικό κόστος

Δεν είναι καθόλου απλό να βρει κανείς τον κατάλληλο συμβιβασμό!

Επιλογή του k



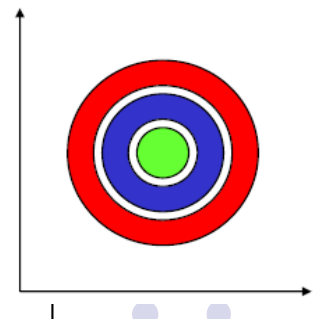
Επιλογή του k



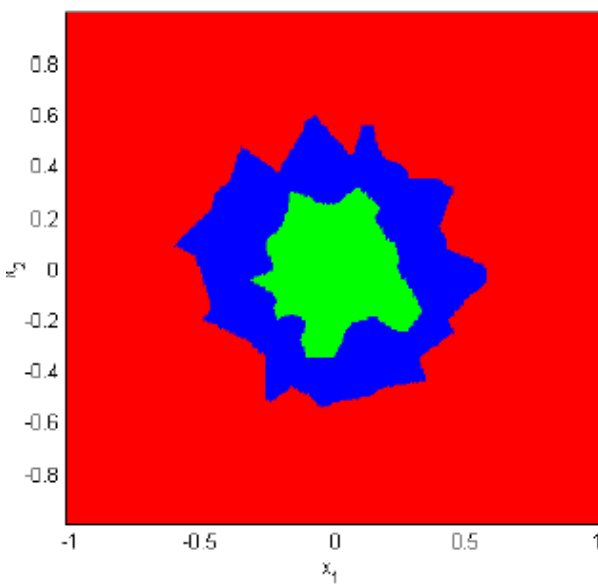
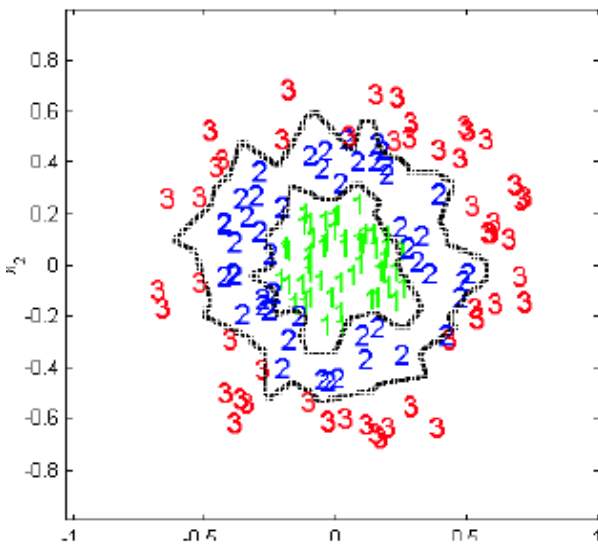
Για $k=1, 2$ και 3 το σημείο x ταξινομείται σωστά στη κόκκινη κλάση

Για μεγαλύτερα k η ταξινόμηση είναι λανθασμένη, μπλε κλάση

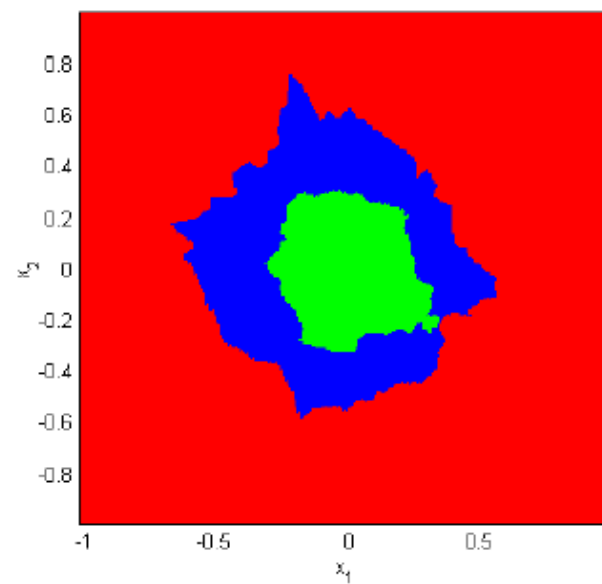
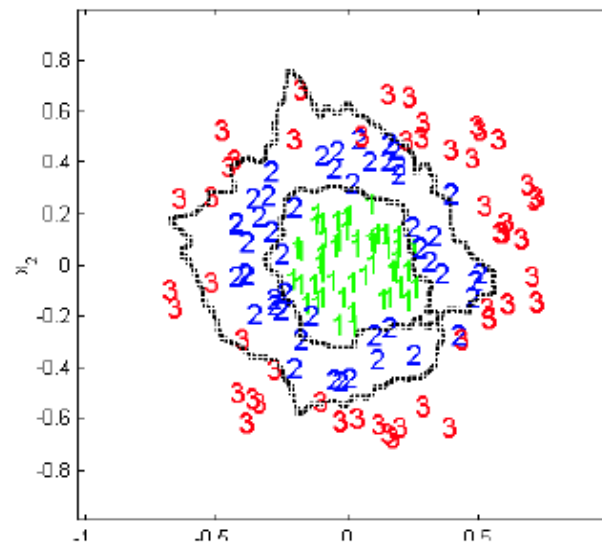
Παράδειγμα



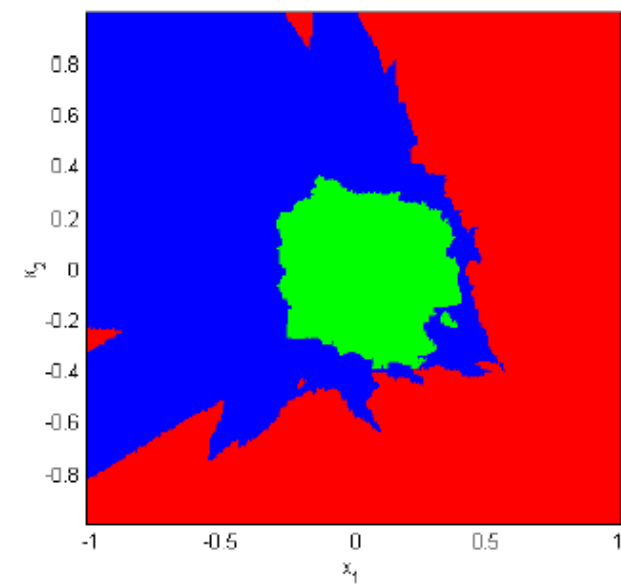
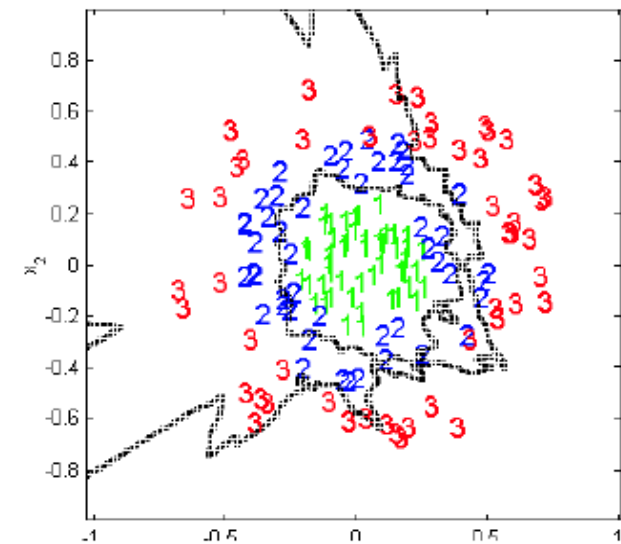
1-NN



5-NN



20-NN

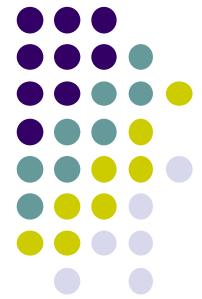


Υπολογιστική Πολυπλοκότητα



- Ο βασικός kNN αλγόριθμος αποθηκεύει όλα τα δείγματα. Υποθέτουμε ότι έχουμε n δείγματα καθε ένα με διάσταση d :
 - $O(d)$ για να υπολογιστεί η απόσταση ως προς ένα δείγμα
 - $O(nd)$ για να βρεθεί ένας κοντινότερος γείτονας
 - $O(knd)$ για να βρεθούν k κοντινότεροι γείτονες
 - Οπότε η πολυπλοκότητα είναι $O(knd)$
- Απαγορευτική για μεγάλο αριθμό δειγμάτων...
- Όμως χρειαζόμαστε μεγάλο αριθμό δειγμάτων για να δώσει καλά αποτελέσματα ο KNN!

Μείωση Υπολογιστικού Φόρτου



➤ Υπολογισμός μερικών αποστάσεων (partial distance)

↳ Χρήση ενός υποσυνόλου r από τις d διαστάσεις, για τον υπολογισμό της απόστασης του δείγματος προς ταξινόμηση από τα δείγματα εκπαίδευσης:

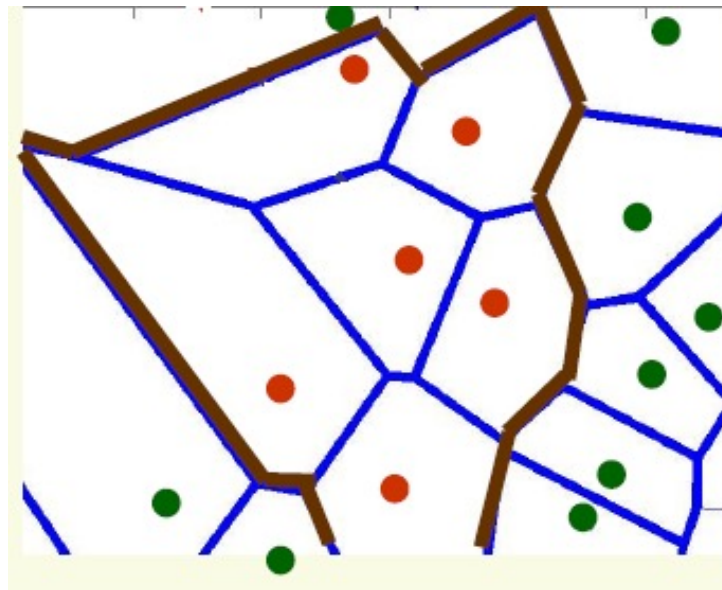
$$D_r(\mathbf{a}, \mathbf{b}) = \left(\sum_{k=1}^r (a_k - b_k)^2 \right)^{1/2}$$

- :↳ Δημιουργία δένδρων όπου τα δείγματα εκπαίδευσης (πρότυπα) συνδέονται επιλεκτικά έτσι ώστε για την ταξινόμηση νέου δείγματος, να απαιτείται ο υπολογισμός της απόστασής του από ορισμένα κομβικά πρότυπα (entry or root) και τα συνδεδεμένα πρότυπα αυτών.
- ↳ Διαγραφή προτύπων που περιβάλλονται από πρότυπα της ίδιας κλάσης.
1. Κατασκεύασε το διάγραμμα Voronoi των αρχικών προτύπων
 2. Για κάθε πρότυπο, αν κάποιος γείτονάς του δεν ανήκει στην ίδια κλάση με αυτό, τσέκαρέ το.
 3. Διέγραψε τα μή τσεκαρισμένα πρότυπα και κατασκεύασε το νέο διάγραμμα Voronoi

Editing 1-NN



- Ένα δείγμα είναι άχρηστο και μπορεί να αφαιρεθεί αν όλοι οι Voronoi γείτονες έχουν την ίδια κλάση

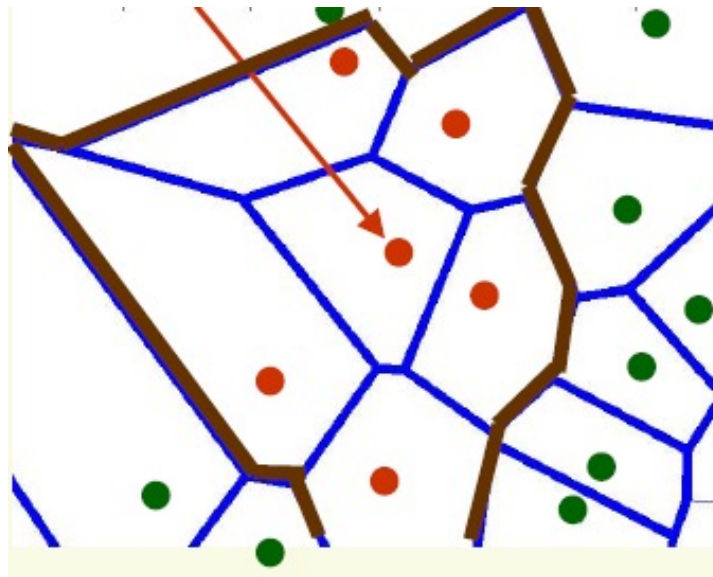


- ✓ Μικραίνει ο αριθμός των δειγμάτων
- ✓ Τα όρια απόφασης παραμένουν τα ίδια

Editing 1-NN



- Ένα δείγμα είναι άχρηστο και μπορεί να αφαιρεθεί αν όλοι οι Voronoi γείτονες έχουν την ίδια κλάση

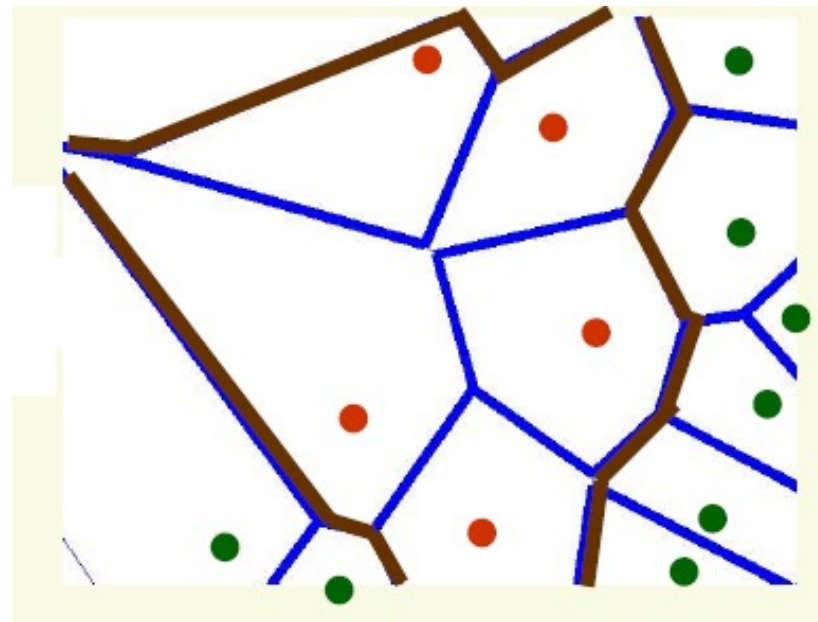


- ✓ Μικραίνει ο αριθμός των δειγμάτων
- ✓ Τα όρια απόφασης παραμένουν τα ίδια

Editing 1-NN



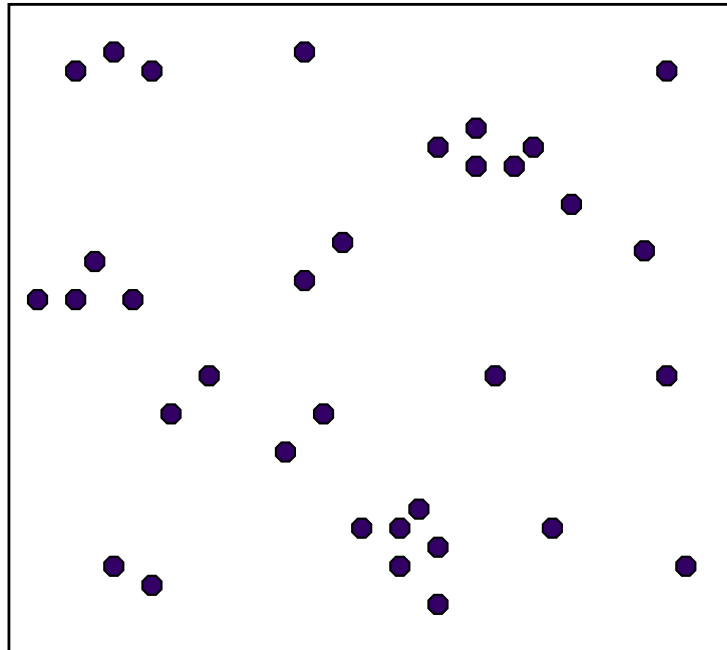
- Ένα δείγμα είναι άχρηστο και μπορεί να αφαιρεθεί αν όλοι οι Voronoϊ γείτονες έχουν την ίδια κλάση



- ✓ Μικραίνει ο αριθμός των δειγμάτων
- ✓ Τα όρια απόφασης παραμένουν τα ίδια



KD-Tree

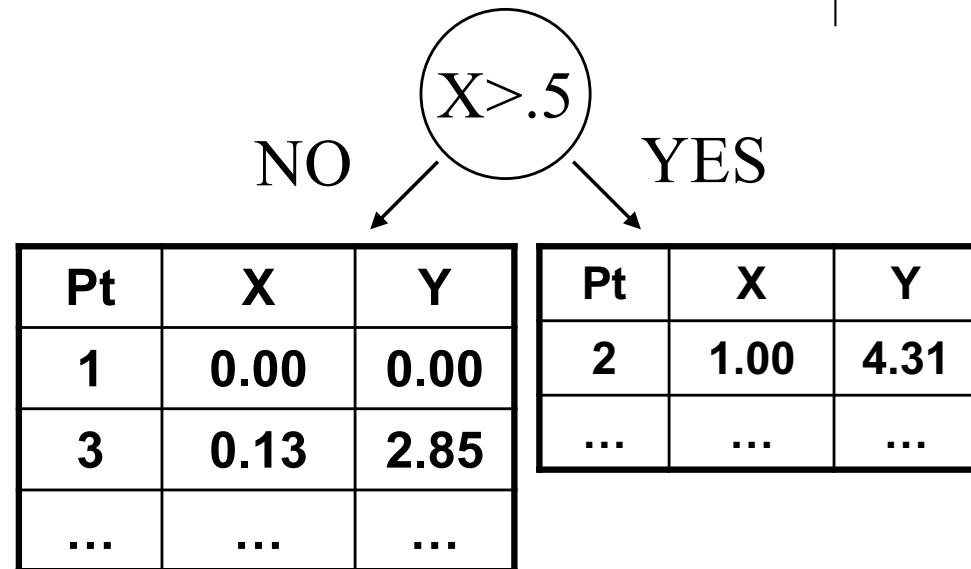
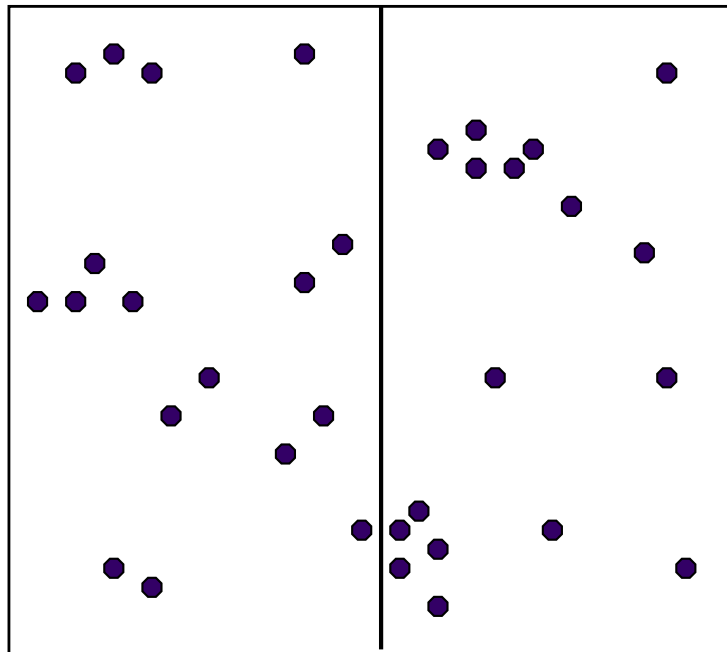


Pt	X	Y
1	0.00	0.00
2	1.00	4.31
3	0.13	2.85
...

Ξεκινάμε με μια λίστα με n σημεία.



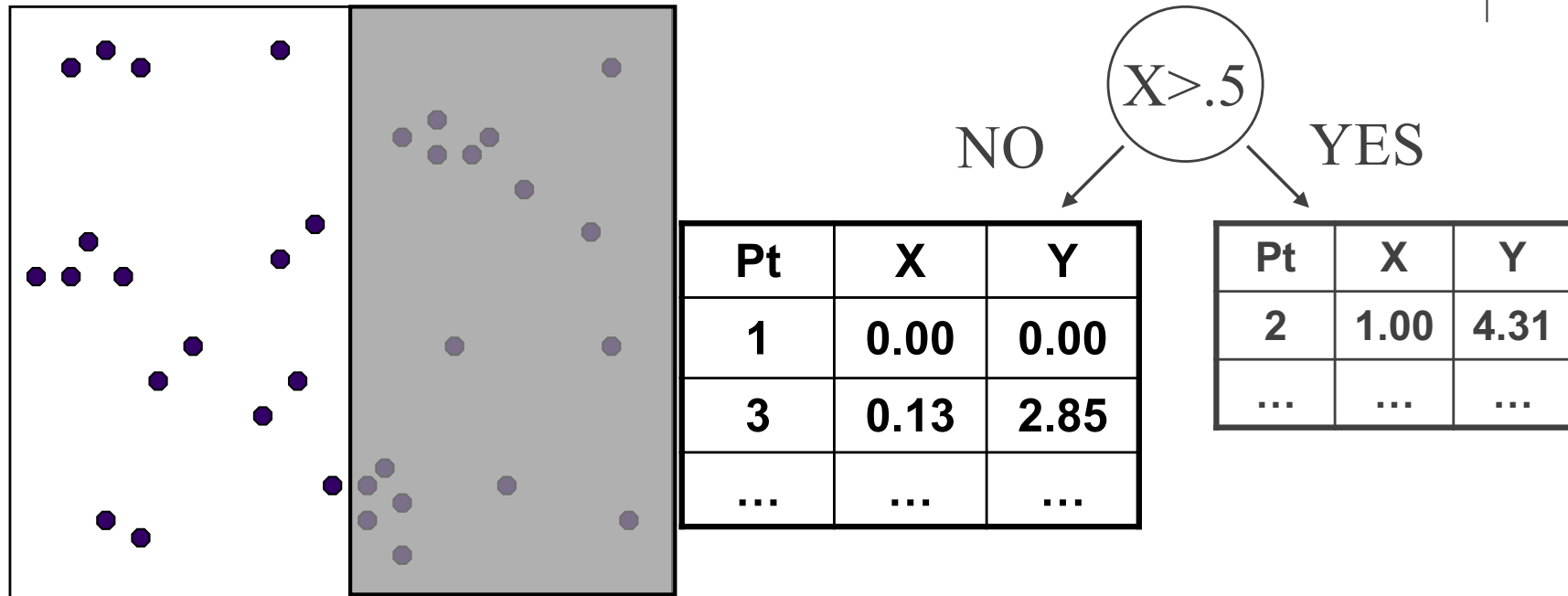
KD-Tree



Μπορούμε να χωρίσουμε τα σημεία σε δυο ομάδες επιλέγοντας μια διάσταση X και μια τιμή V και διαχωρίζοντας τα σημεία σε $X > V$ και $X \leq V$.



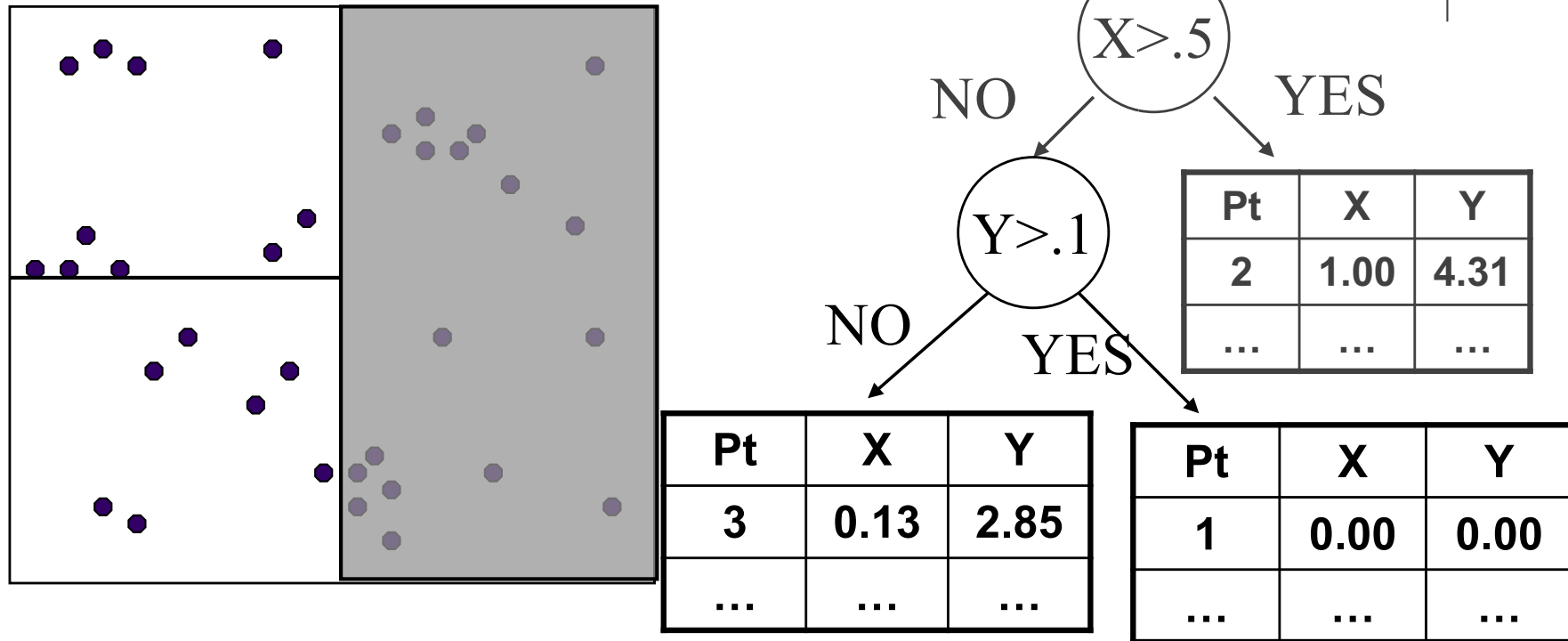
KD-Tree



Μετά μπορούμε να θεωρήσουμε την κάθε ομάδα ξεχωριστά και να ξαναχωρίσουμε τα σημεία (στην ίδια/διαφορετική διάσταση).



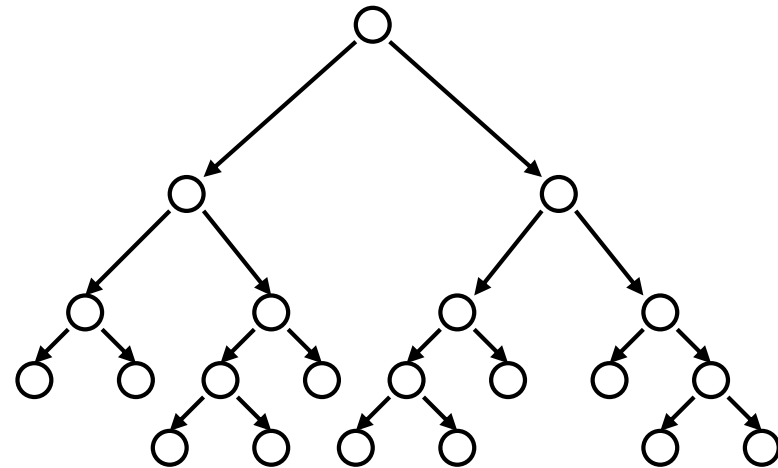
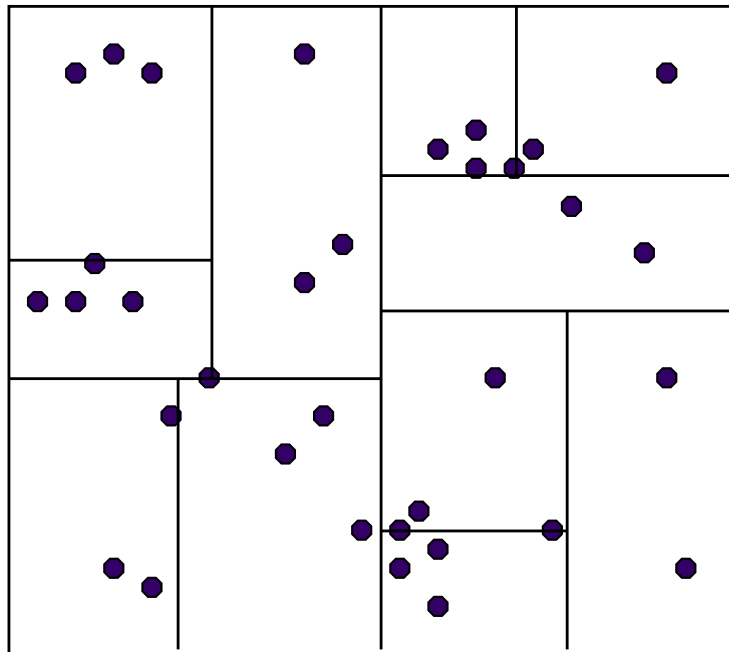
KD-Tree



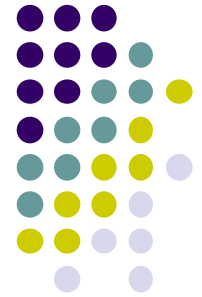
Μετά μπορούμε να θεωρήσουμε την κάθε ομάδα ξεχωριστά και να ξαναχωρίσουμε τα σημεία (στην ίδια/διαφορετική διάσταση).



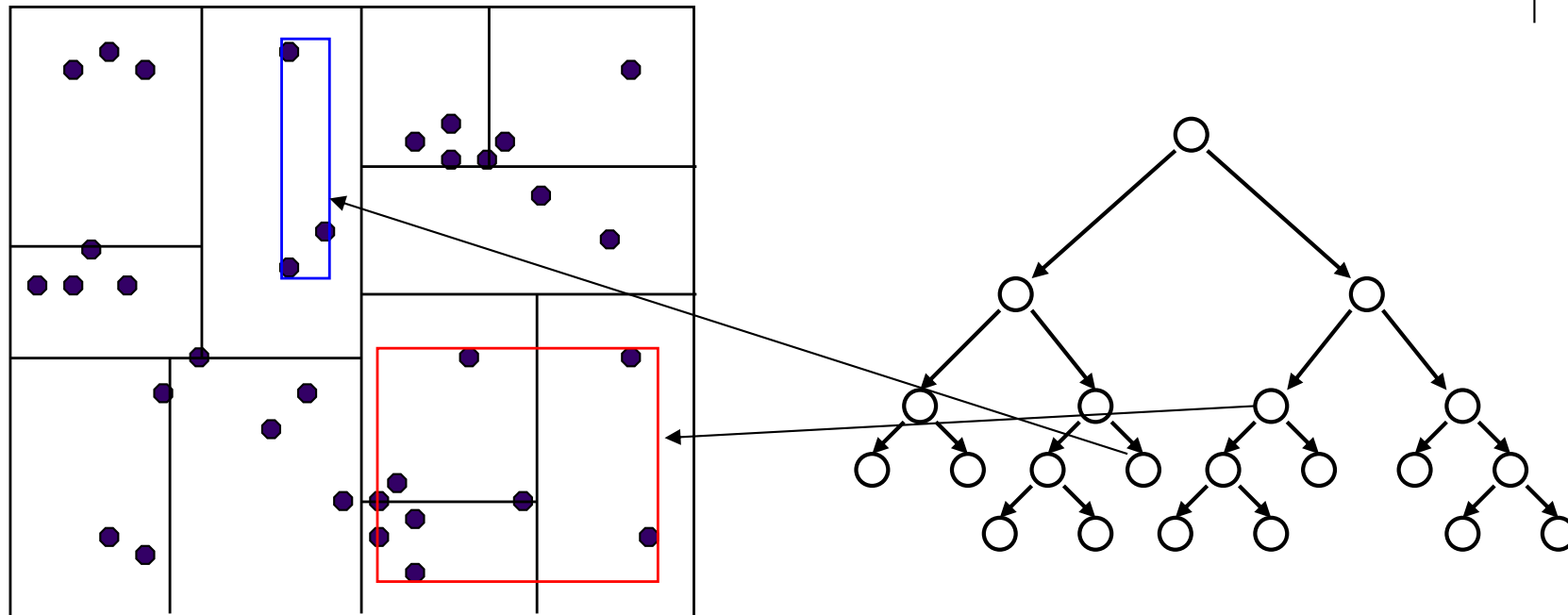
KD-Tree



Συνεχίζουμε να διαχωρίζουμε τα δεδομένα, δημιουργώντας μια δομή δέντρου. Κάθε κόμβος δίχως παιδιά (leaf node), περιέχει μια λίστα με σημεία.



KD-Tree



Κρατάμε πληροφορία σε κάθε κόμβο:

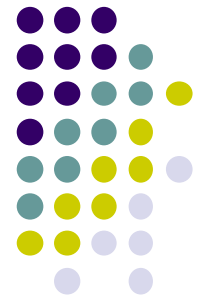
- Τη διάσταση κατά την οποία χωρίζουμε
- Το κατώφλι διαχωρισμού
- Τα (σφιχτά) όρια των σημείων σε αυτό ή κάτω από αυτόν τον κόμβο (bounding box).

KD-Tree

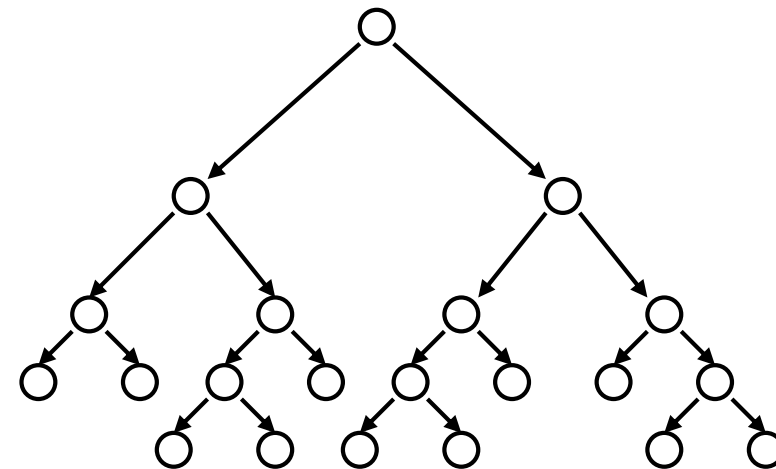
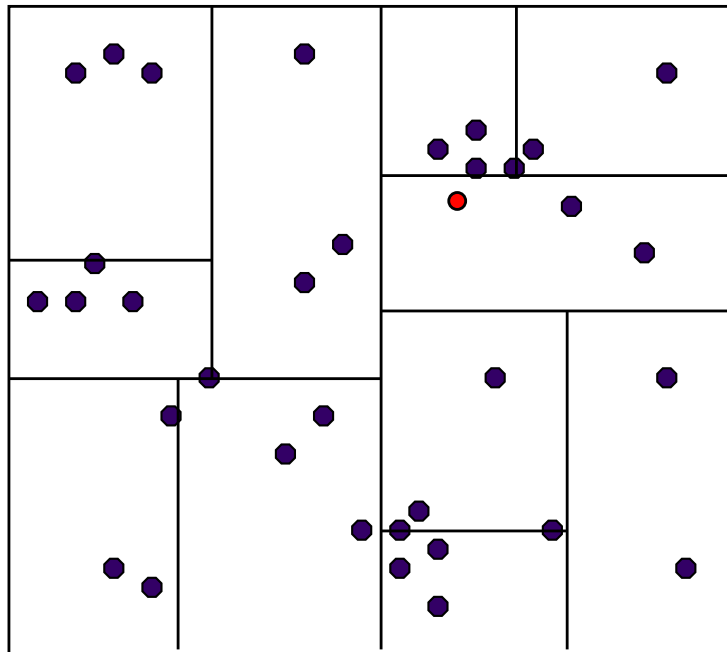


Χρησιμοποιούμε ευρετικούς κανόνες για να αποφασίσουμε για τον διαχωρισμό:

- Σε ποια διάσταση γίνεται ο διαχωρισμός? Πιο «πλατύ»
- Ποια τιμή χρησιμοποιείται για τον διαχωρισμό? Median των τιμών αυτής της διάστασης.
- Πότε σταματάμε? Όταν είναι λιγότερα από m σημεία ή το κελί έχει κάποιο ελάχιστο πλάτος.



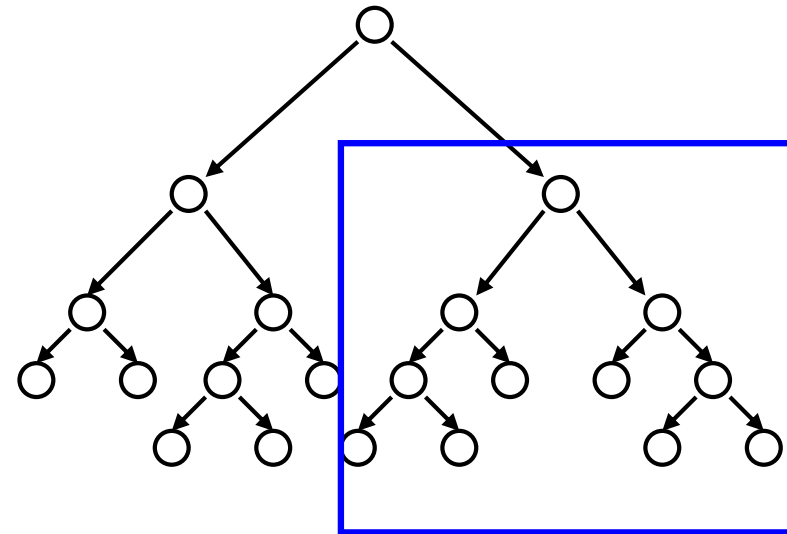
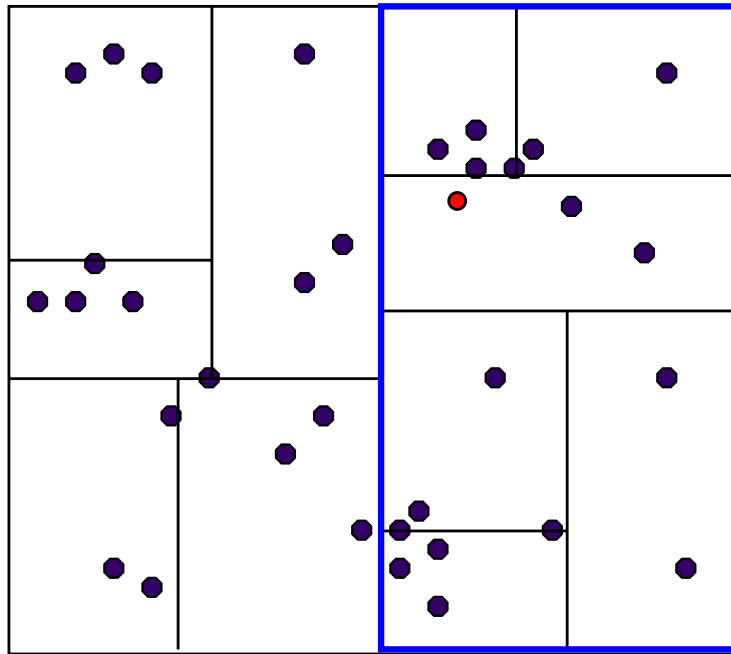
Nearest Neighbor και KD-Trees



Διατρέχουμε το δέντρο ψάχνοντας τον κοντινότερο γείτονα για το δεδομένο σημείο.



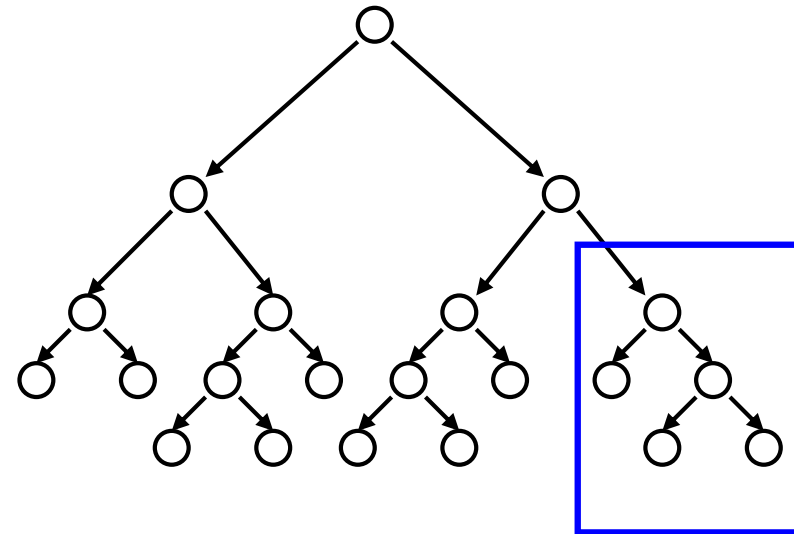
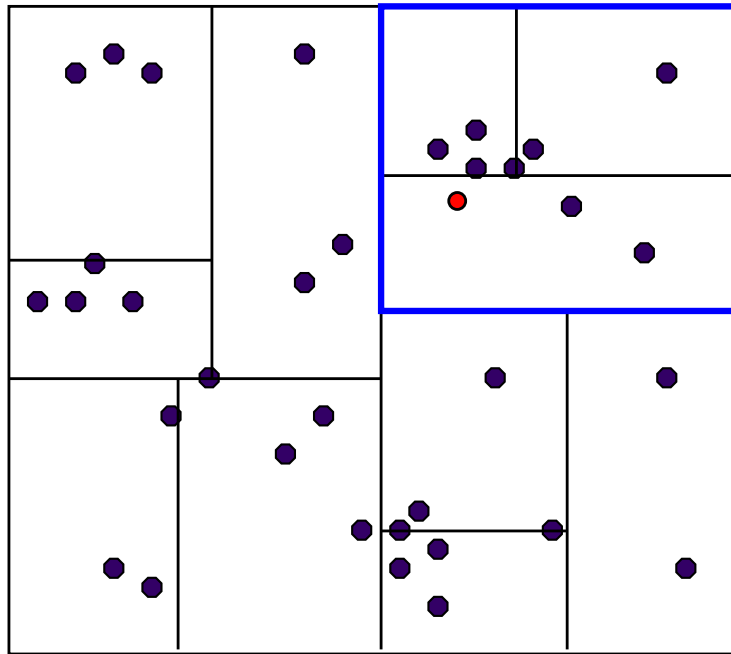
Nearest Neighbor with KD Trees



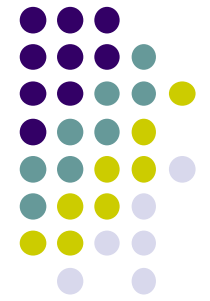
Εξετάζουμε αρχικά τα κοντινότερα σημεία: Εξετάζουμε τον κλάδο ο οποίος είναι πιο κοντά στο δεδομένο σημείο.



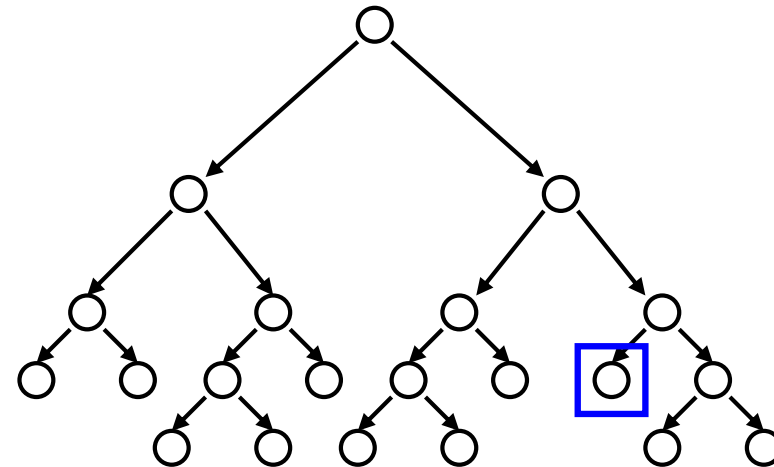
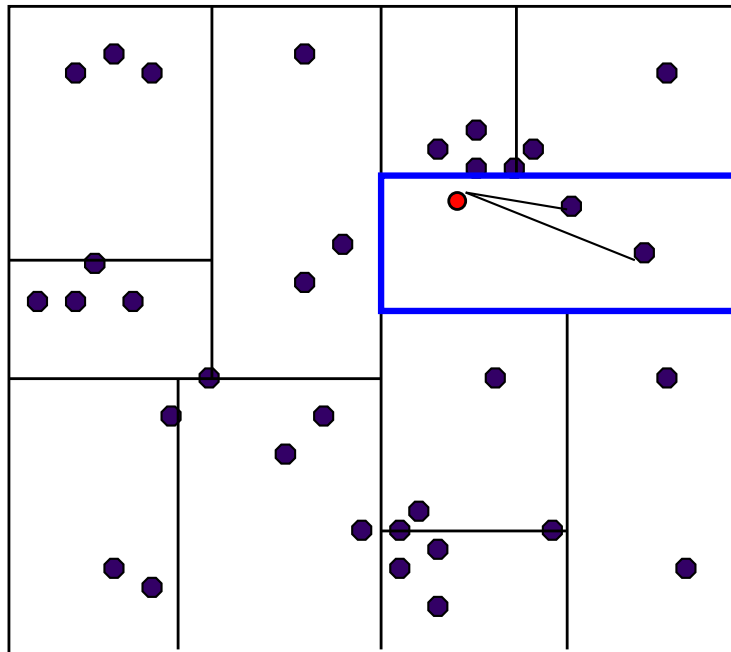
Nearest Neighbor και KD-Trees



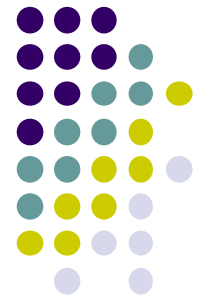
Εξετάζουμε αρχικά τα κοντινότερα σημεία: Εξετάζουμε τον κλάδο ο οποίος είναι πιο κοντά στο δεδομένο σημείο.



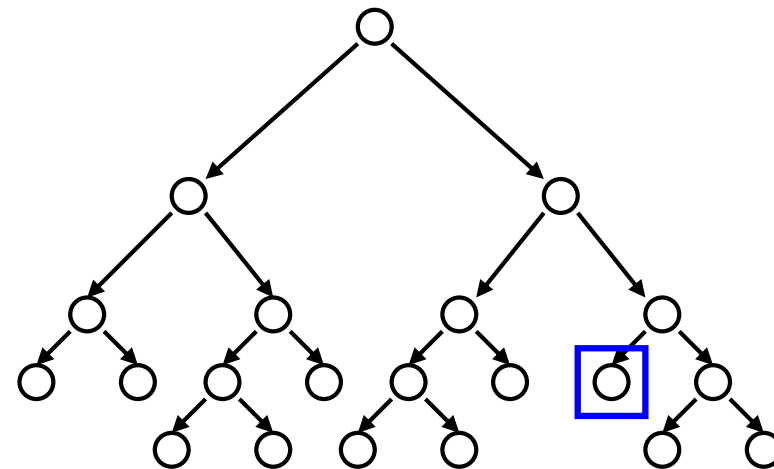
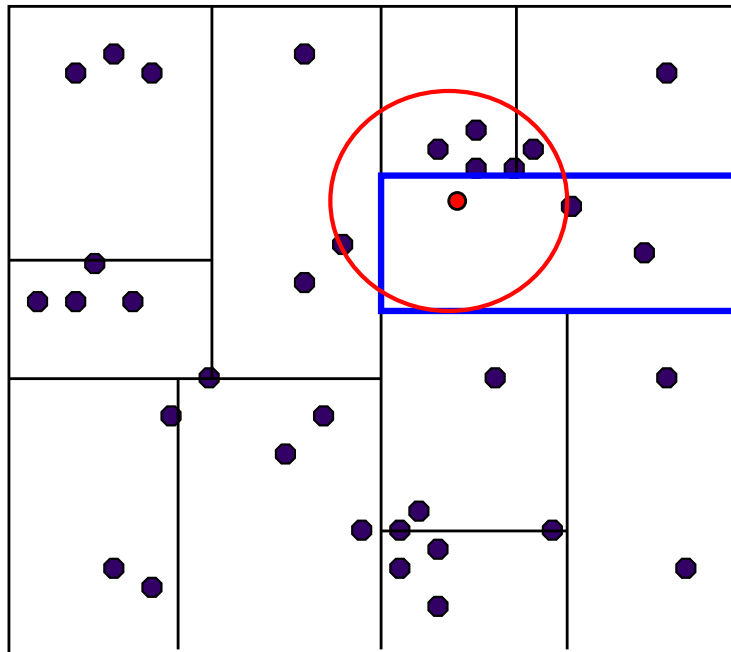
Nearest Neighbor και KD-Trees



Όταν φτάσουμε σε φύλλο υπολογίζουμε την απόσταση του δεδομένου σημείου προς κάθε σημείο του κόμβου-φύλλου.



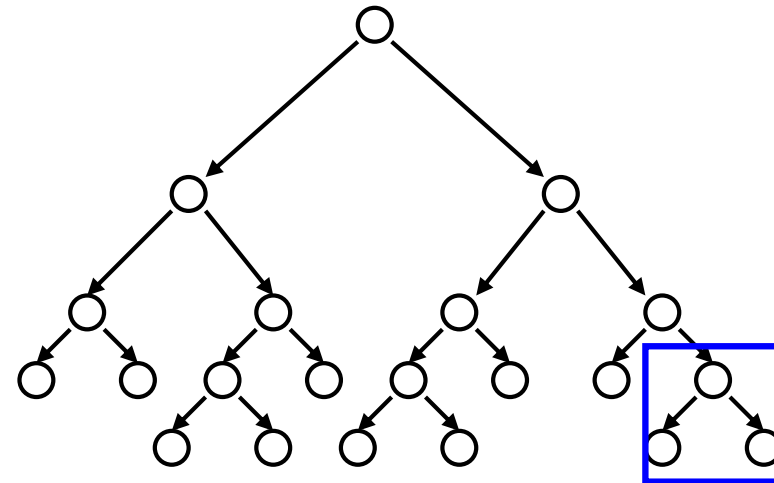
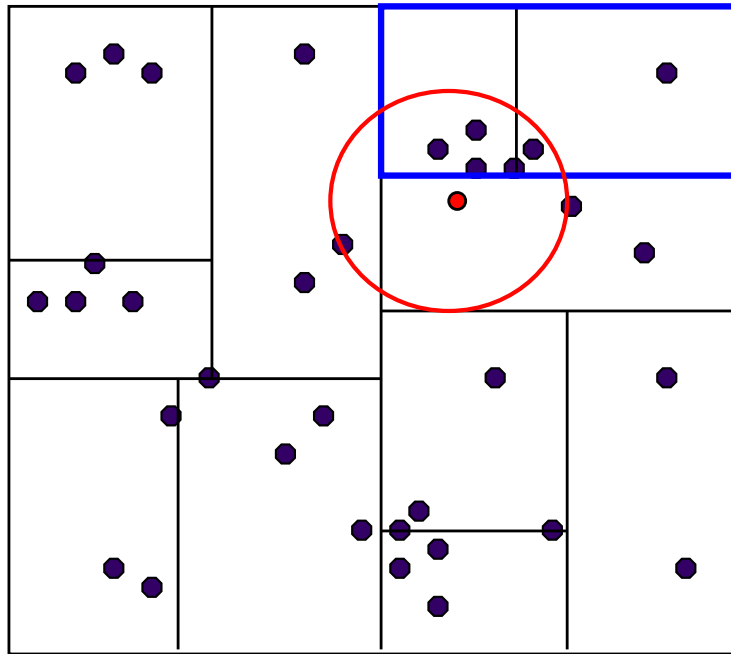
Nearest Neighbor και KD-Trees



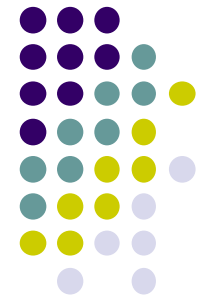
Όταν φτάσουμε σε φύλλο υπολογίζουμε την απόσταση του δεδομένου σημείου προς κάθε σημείο του κόμβου-φύλλου.



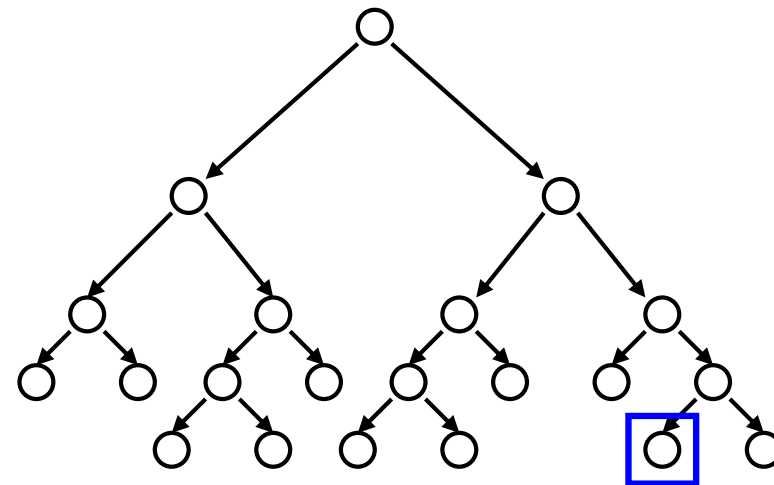
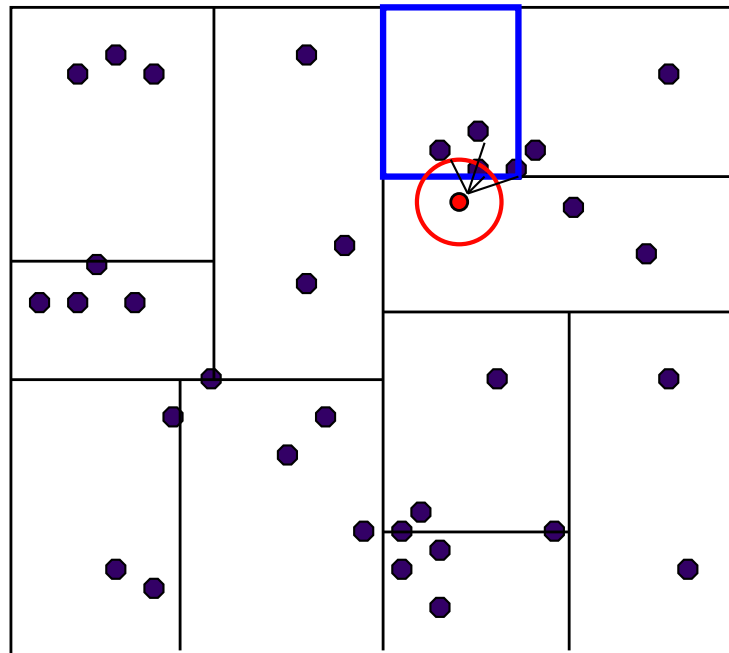
Nearest Neighbor και KD-Trees



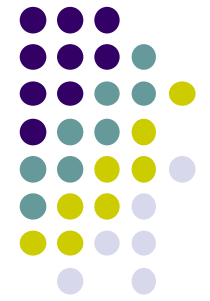
Μετά μπορούμε να γυρίσουμε προς τα πίσω και να δοκιμάσουμε τον άλλο κλάδο σε κάθε κόμβο τον οποίο επισκεφτήκαμε.



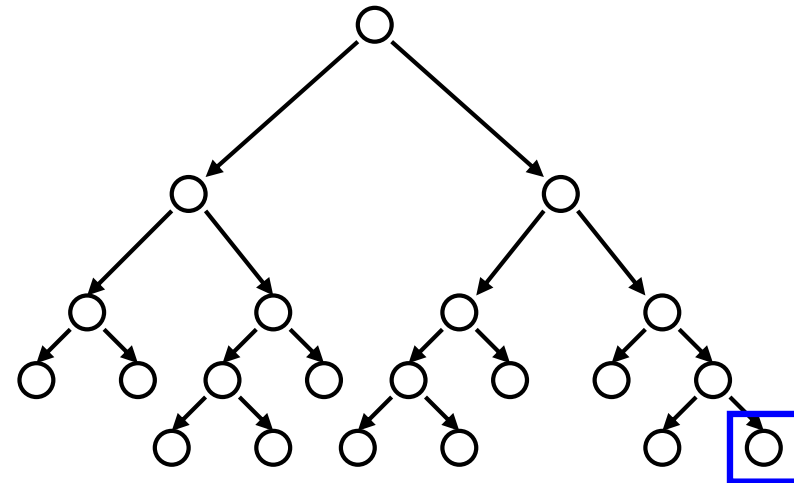
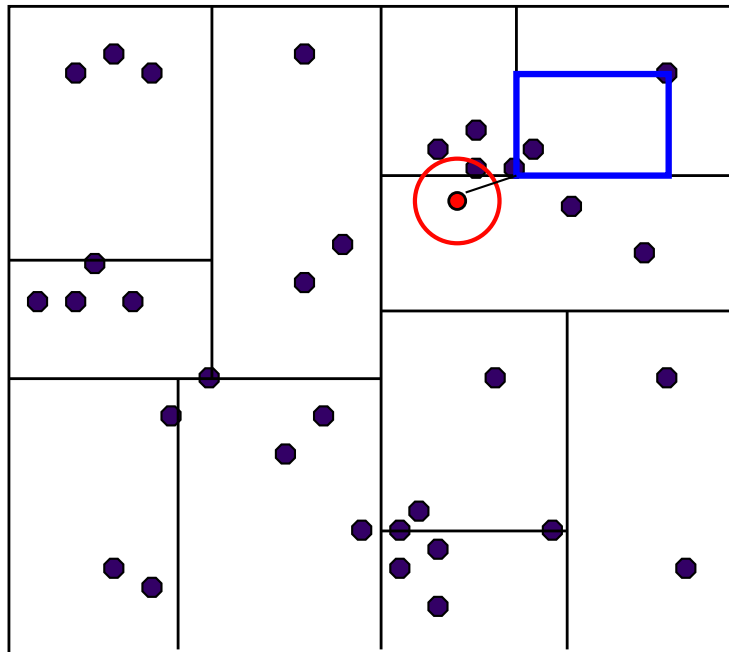
Nearest Neighbor και KD-Trees



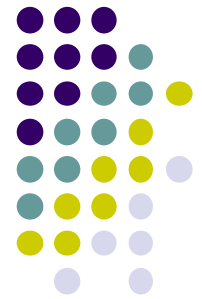
Κάθε φορά που βρίσκουμε νέο κοντινότερο κόμβο μπορούμε να προσαρμόζουμε τα όρια απόστασης.



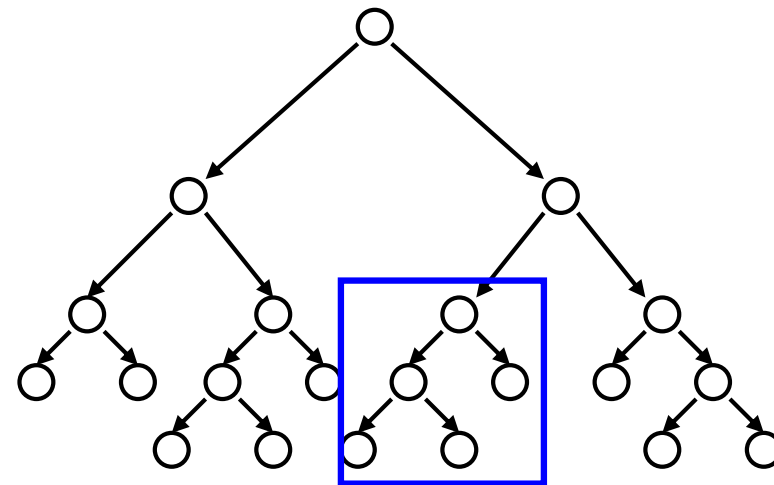
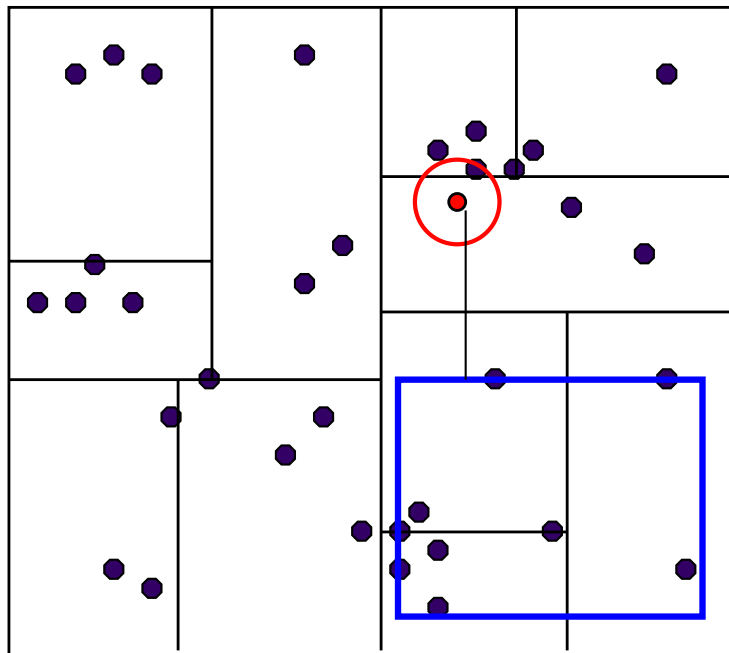
Nearest Neighbor και KD-Trees



Χρησιμοποιώντας τα όρια απόστασης και τα όρια των δεδομένων κάτω από κάθε κόμβο, μπορούμε να «κλαδέψουμε μέρη του δέντρου τα οποία ΔΕΝ θα μπορούσαν να περιέχουν τον κοντινότερο γείτονα.



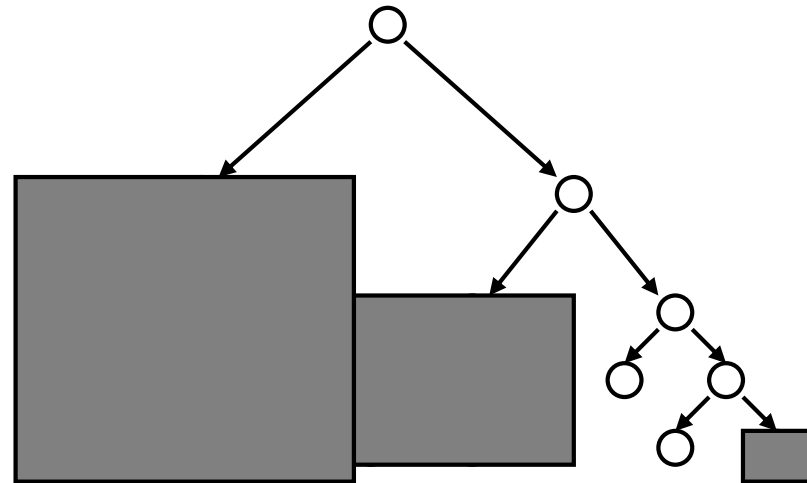
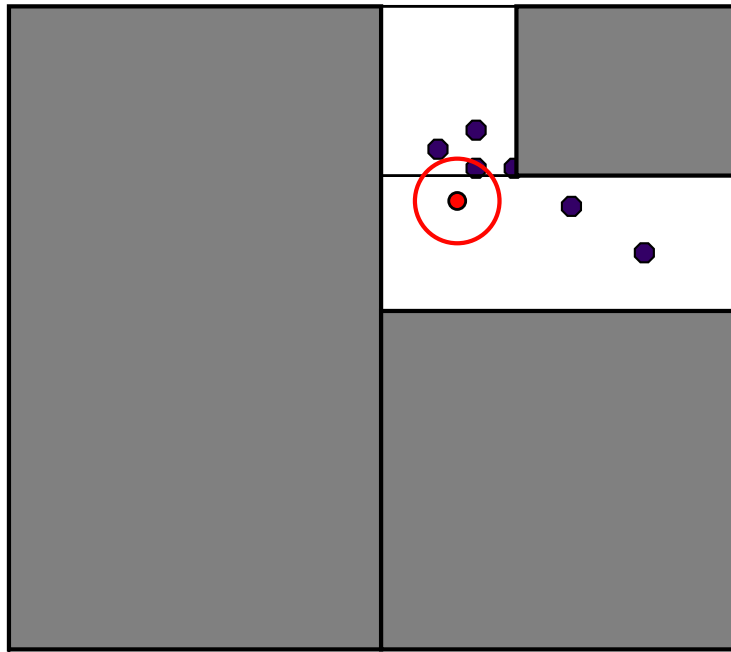
Nearest Neighbor και KD-Trees



Χρησιμοποιώντας την τρέχουσα ελάχιστη απόσταση και την απόσταση του bounding box κάτω από κάθε κόμβο, μπορούμε να κλαδέψουμε τους κόμβους που ΔΕΝ θα μπορούσαν να περιέχουν τον κοντινότερο γείτονα.



Nearest Neighbor και KD-Trees



Χρησιμοποιώντας την τρέχουσα ελάχιστη απόσταση και την απόσταση του bounding box κάτω από κάθε κόμβο, μπορούμε να κλαδέψουμε τους κόμβους που ΔΕΝ θα μπορούσαν να περιέχουν τον κοντινότερο γείτονα.

Επιλογή Μέτρου



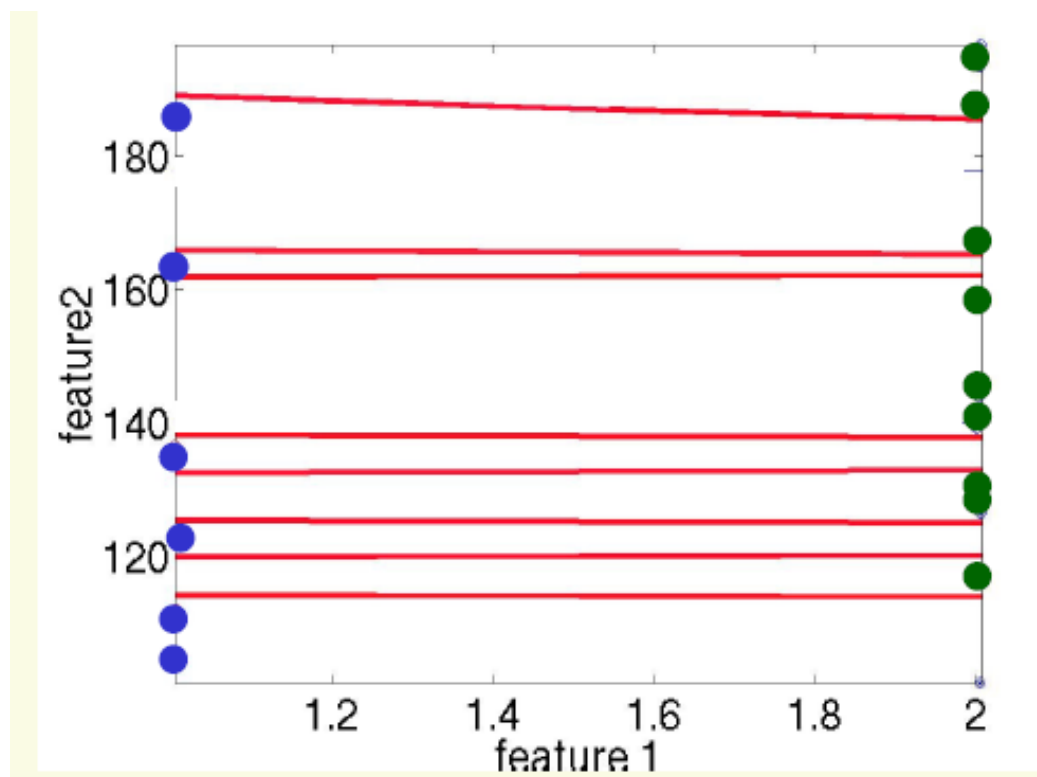
- Μέχρι τώρα υποθέσαμε ότι χρησιμοποιούμε την Ευκλείδεια Απόσταση

$$D(a, b) = \sqrt{\sum_k (a_k - b_k)^2}$$

Όμως κάποια χαρακτηριστικά μπορεί να παρέχουν περισσότερη πληροφορία για την διάκριση των κλάσεων

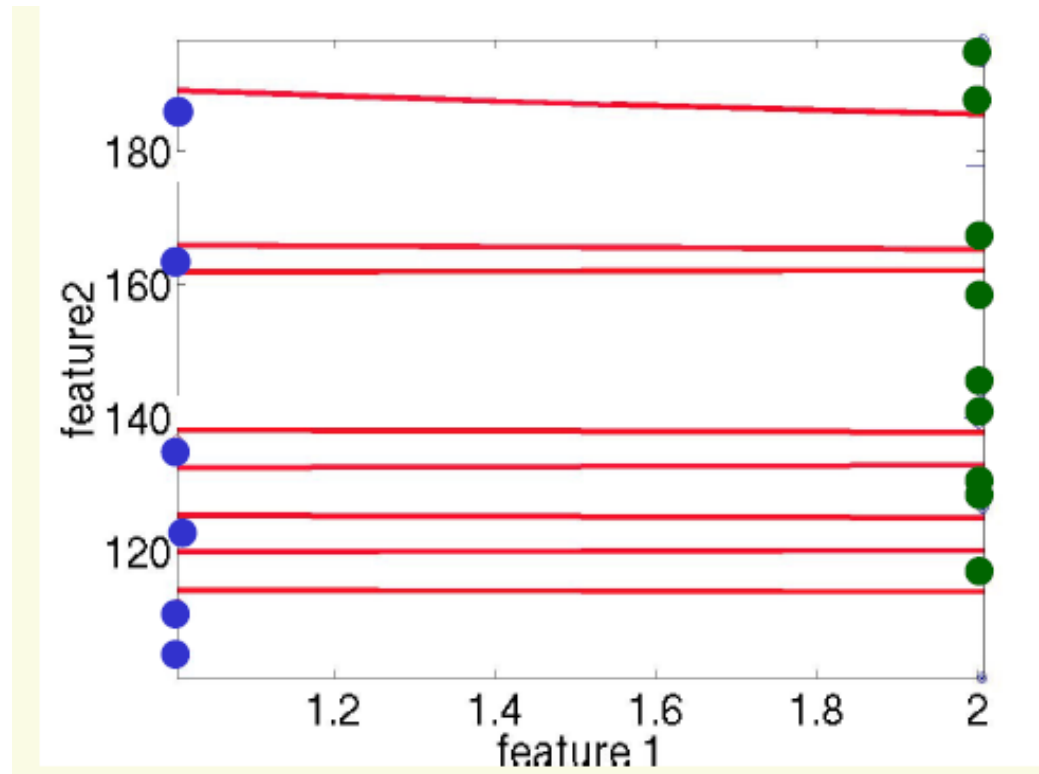
Η Ευκλείδεια απόσταση λαμβάνει με τον ίδιο τρόπο υπόψη το κάθε χαρακτηριστικό

Παράδειγμα



Τι παρατηρείτε?

Παράδειγμα



- Χαρακτηριστικό 1 πληροφορία, αλλά μικρή κλίμακα
- Χαρακτηριστικό 2 θόρυβος, αλλά μεγάλη κλίμακα

Παράδειγμα



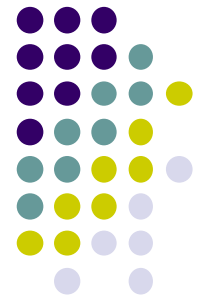
Το χαρακτηριστικό 1 παρέχει πληροφορία για τις κλάσεις 1 και 2
Το χαρακτηριστικό 2 έχει μια άσχετη τιμή μεταξύ 100 και 200

Έστω ότι θέλουμε να υπολογίσουμε την κλάση του $x=[1 \ 100]$ και έχουμε 2 δείγματα $[1 \ 150]$ και $[2 \ 110]$

$$D\left(\begin{bmatrix} 1 \\ 100 \end{bmatrix}, \begin{bmatrix} 1 \\ 150 \end{bmatrix}\right) = \sqrt{(1-1)^2 + (100-150)^2} = 50 \quad D\left(\begin{bmatrix} 1 \\ 100 \end{bmatrix}, \begin{bmatrix} 2 \\ 110 \end{bmatrix}\right) = \sqrt{(1-2)^2 + (100-110)^2} = 10.5$$

Το x θα ταξινομηθεί λάθος!

- ✓ Όσο μεγαλύτερη η πυκνότητα των δειγμάτων, τόσο καλύτερα τα αποτελέσματα
- ✓ Συνήθως όμως η πυκνότητα δεν είναι αρκετά μεγάλη



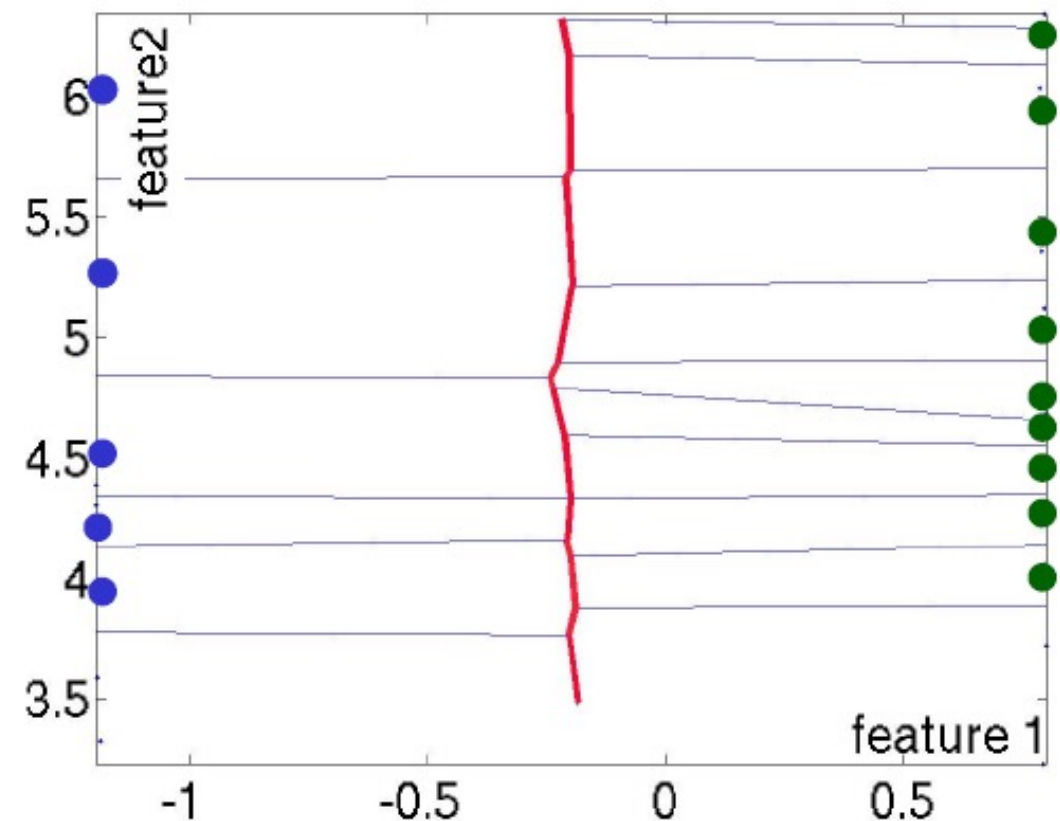
Κανονικοποίηση- Παράδειγμα

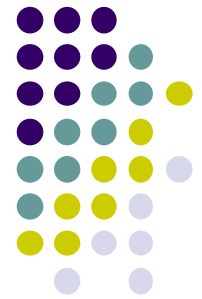
Αν η τυχαία μεταβλητή X έχει μέσο όρο μ και διασπορά σ^2

Τότε η τυχαία μεταβλητή $(X - \mu)/\sigma$ έχει μέσο όρο 0 και διασπορά 1

Αντικατάσταση του κάθε x_i με

$$[x_i - \text{mean}(x_i)]/\text{sqrt}[\text{var}(x_i)]$$





Feature Weighting

- Σε μεγάλες διαστάσεις όμως, όταν υπάρχουν πολλά «άσχετα» χαρακτηριστικά η κανονικοποίηση δεν βοηθάει.

$$D(a, b) = \sqrt{\sum_k (a_k - b_k)^2} = \sqrt{\underbrace{\sum_i (a_i - b_i)^2}_{\text{Σημαντικά}} + \underbrace{\sum_j (a_j - b_j)^2}_{\text{Άσχετα}}}$$

Σημαντικά
χαρακτηριστικά

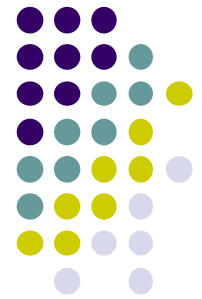
Άσχετα
χαρακτηριστικά

✓ Πρόβλημα όταν ο αριθμός των «άσχετων» χαρακτηριστικών, είναι μεγαλύτερος από των σημαντικών

$$D(a, b) = \sqrt{\sum_k w_k (a_k - b_k)^2}$$

Βάρη για κάθε χαρακτηριστικό τα οποία μαθαίνονται από τα δεδομένα

Μετρικές και KNN Ταξινόμηση



- Ιδιότητες:
- ↪ Nonnegativity: $D(\mathbf{a}, \mathbf{b}) \geq 0$
 - ↪ reflexivity: $D(\mathbf{a}, \mathbf{b}) = 0$ iff $\mathbf{a} = \mathbf{b}$
 - ↪ symmetry: $D(\mathbf{a}, \mathbf{b}) = D(\mathbf{b}, \mathbf{a})$
 - ↪ triangle inequality: $D(\mathbf{a}, \mathbf{b}) + D(\mathbf{b}, \mathbf{c}) \geq D(\mathbf{a}, \mathbf{c})$

- Ευκλείδεια Απόσταση:

$$L_2(\mathbf{a}, \mathbf{b}) = \left(\sum_{k=1}^d (a_k - b_k)^2 \right)^{1/2}$$

- Minkowski Μετρική:

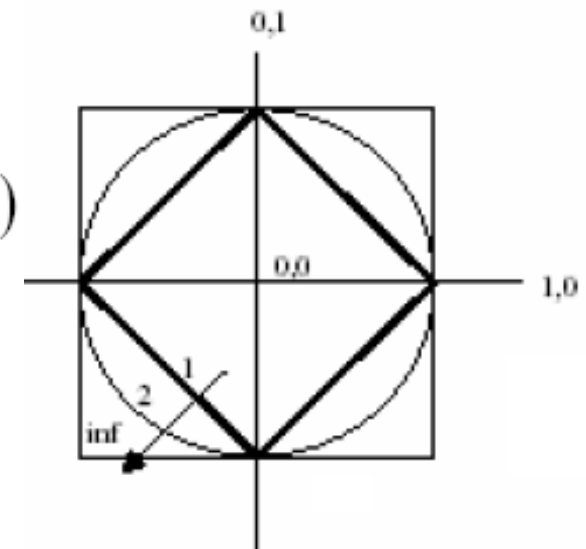
$$L_p(\mathbf{a}, \mathbf{b}) = \left(\sum_{k=1}^d |a_k - b_k|^p \right)^{1/p}$$

- Manhattan Απόσταση:

$$L_1(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^d |a_k - b_k|$$

- Chess-board Απόσταση:

$$L_\infty(\mathbf{a}, \mathbf{b}) = \max_{k=1, \dots, d} (|a_k - b_k|)$$



Απόσταση 1 από το κέντρο χρησιμοποιώντας κάθε μία από τις μετρικές L_p

Συμπεράσματα KNN



- Προτερήματα

- Μπορεί να χρησιμοποιηθεί για δεδομένα από οποιαδήποτε κατανομή
- Πολύ απλός
- Καλά αποτελέσματα ταξινόμησης, όταν ο αριθμός των δειγμάτων είναι αρκετά μεγάλος

- Μειονεκτήματα

- Δύσκολο να επιλεγθεί το κατάλληλο k
- Μεγάλος υπολογιστικός φόρτος, αν και υπάρχουν βελτιωμένες εκδοχές του αλγορίθμου
- Απαιτείται μεγάλος αριθμός δεδομένων

- **Βίντεο για Parzen και K-NN**

https://www.youtube.com/watch?v=UPXIdi_aTEg