Εξερεύνηση δεδομένων

Ιωάννης Μαδεμλής

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Ο υπολογισμός περιγραφικών στατιστικών συνόψισης ενός συνόλου δεδομένων μας επιτρέπει να «δούμε» περιληπτικά τα δεδομένα μας.

- Στην περίπτωση κατηγορικών γνωρισμάτων μπορούμε να υπολογίσουμε:
  - Τη **συχνότητα εμφάνισης** κάθε τιμής του γνωρίσματος.
    - Μία ξεχωριστή συχνότητα για κάθε δυνατή τιμή.
  - Τη **συχνότερα εμφανιζόμενη τιμή** του γνωρίσματος.
    - Μία μόνο τιμή για όλο το γνώρισμα.

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Στην περίπτωση διατακτικών ή αριθμητικών γνωρισμάτων μπορούμε να υπολογίσουμε τα **εκατοστημόρια**:
  - Έστω το i-οστό γνώρισμα των n προτύπων μας.
  - Το j-οστό εκατοστημόριο είναι μία τιμή από το πεδίο ορισμού του γνωρίσματος, τέτοια ώστε $j\%$ των παρατηρούμενων στο σύνολο δεδομένων τιμών του γνωρίσματος να είναι μικρότερες από αυτήν.
  - Άρα το j είναι πάντα ένας αριθμός στο διάστημα (0, 100).
  - Το $50^{ο}$ εκατοστημόριο είναι ο γνωστός **διάμεσος** (median).

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Για συνεχή γνωρίσματα, χρησιμοποιούνται τα συνήθη στατιστικά μέτρα θέσης και διασποράς.
- Τα κυριότερα μέτρα θέσης είναι ο **μέσος** και ο **διάμεσος**.
  - Αν διατάξουμε όλα τα n πρότυπα με κριτήριο την τιμή τους στο επίμαχο γνώρισμα, τότε ο διάμεσος είναι:
    - Η τιμή του γνωρίσματος στο μεσαίο πρότυπο, αν το n είναι περιττός αριθμός.
    - Η μέση τιμή των τιμών του γνωρίσματος στα δύο μεσαία πρότυπα, αν το n είναι άρτιος αριθμός.

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Ο διάμεσος είναι περισσότερο ευσταθής παρουσία ανωμαλιών.
- Εναλλακτικά, μπορούμε να υπολογίσουμε τον **περικομμένο μέσο** (trimmed mean) με βάση μία παράμετρο p.
  - Αγνοούμε το $(p/2)\%$ των προτύπων με τις μέγιστες τιμές στο γνώρισμα.
  - Αγνοούμε το $(p/2)\%$ των προτύπων με τις ελάχιστες τιμές στο γνώρισμα.
  - Υπολογίζουμε τον μέσο των τιμών του γνωρίσματος στα εναπομείναντα πρότυπα.
  - Είναι ο πλέον απλός τρόπος για να εξαλείψουμε τις ανωμαλίες πριν υπολογίσουμε τον μέσο.
    - Το p συνήθως είναι μικρό (π.χ., 1 – 5).

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Τα συνηθέστερα μέτρα διασποράς είναι το **εύρος**, η **διακύμανση** και η **τυπική απόκλιση**.
  - Το εύρος είναι το μήκος του τμήματος της πραγματικής ευθείας το οποίο καλύπτουν οι τιμές των προτύπων μας στο επίμαχο γνώρισμα.
    - Άρα, ισούται με τη διάφορα της μέγιστης μείον την ελάχιστη τιμή του γνωρίσματος στο σύνολο δεδομένων μας.
  - Το εύρος είναι ευαίσθητο σε ανωμαλίες, οπότε συνήθως προτιμάται η διακύμανση.
    - Η μέση τιμή της τετραγωνικής διαφοράς των τιμών του γνωρίσματος στα *n* πρότυπα από τον μέσο του γνωρίσματος.
    - Η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης.
  - Όμως ο μέσος επίσης είναι ευαίσθητος σε ανωμαλίες.
  - Έτσι υπάρχουν εναλλακτικά μέτρα διασποράς, περισσότερο ευσταθή σε ανωμαλίες.
    - Π.χ., η απόλυτη μέση απόκλιση (AAD) αντικαθιστά το τετράγωνο με απόλυτη τιμή στον τύπο της διακύμανσης.

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Τα ανωτέρω μέτρα διασποράς πρέπει να εφαρμοστούν ξεχωριστά σε κάθε γνώρισμα.
- Για m-διάστατα πρότυπα, μπορούμε να περιγράψουμε τη συνολική διασπορά των δεδομένων με τον **πίνακα συνδιακύμανσης C**, διάστασης m x m.
  - Το στοιχείο c_{ij} εκφράζει την τάση συνδιακύμανσης του i-οστού και του j-οστού γνωρίσματος στο σύνολο των n προτύπων.
  - Το στοιχείο c_{ii} ισούται με τη διακύμανση του i-οστού γνωρίσματος.
  - Θετική συνδιακύμανση σημαίνει ότι στα πρότυπα όπου το i-οστό γνώρισμα έχει υψηλή τιμή, συνήθως έχει και το j-οστό γνώρισμα υψηλή τιμή.
  - Αρνητική συνδιακύμανση σημαίνει ότι στα πρότυπα όπου το i-οστό γνώρισμα έχει υψηλή τιμή, το j-οστό γνώρισμα συνήθως έχει χαμηλή τιμή.

Σύνοψη δεδομένων

Στατιστικά συνόψισης

- Η συσχέτιση του i-οστού και του j-οστού γνωρίσματος είναι απλώς η κανονικοποιημένη συνδιακύμανσή τους.
  - Προκύπτει από διαίρεση της συνδιακύμανσής τους με το γινόμενο των διακυμάνσεών τους.
  - Ως αποτέλεσμα, η συσχέτιση λαμβάνει πάντοτε τιμές στο διάστημα [-1, 1].
  - Συσχέτιση ίση με -1 είναι απολύτως αρνητική.
  - Συσχέτιση ίση με 1 είναι απολύτως θετική (αυτή είναι η τιμή των στοιχείων της διαγωνίου στον **πίνακα συσχέτισης**).
  - Συσχέτιση ίση με 0 δείχνει πως δεν υπάρχει κάποια *γραμμική* σχέση μεταξύ των δύο γνωρισμάτων.

Άλλες χρήσεις του πίνακα συνδιακύμανσης

Σύνοψη δεδομένων

- Ο πίνακας συνδιακύμανσης των δεδομένων μπορεί εναλλακτικά να χρησιμοποιηθεί όχι για συνόψιση των προτύπων, αλλά ως τμήμα ενός στατιστικού αλγορίθμου για την αυτόματη συναγωγή συμπερασμάτων περί των δεδομένων.
- Παράδειγμα: αλγόριθμοι οι οποίοι μοντελοποιούν την υποκείμενη γεννήτρια κατανομή των δεδομένων ως μία πολυδιάστατη γκαουσιανή κατανομή ορισμένη επί του m-διάστατου χώρου των προτύπων.
  - Μία συγκεκριμένη γκαουσιανή ορίζει ένα m-διάστατο ελλειψοειδές το οποίο περικλείει τα n δεδομένα.
  - Μία συγκεκριμένη γκαουσιανή ορίζεται πλήρως από έναν m-διάστατο μέσο και έναν πίνακα συνδιακύμανσης $m \times m$.
  - Ο πίνακας συνδιακύμανσης των δεδομένων είναι ο **δειγματικός πίνακας συνδιακύμανσης**.

Οπτικοποίηση

Οπτικοποίηση δεδομένων

- Στόχος όλων των τρόπων **οπτικοποίησης** δεδομένων είναι η οπτική αναπαράσταση προτύπων, γνωρισμάτων ή ομάδων, καθώς και των σχέσεών τους.
- Οι πραγματικές σχέσεις μεταξύ των ορατών αντικειμένων πρέπει να απεικονίζονται στις οπτικές σχέσεις μεταξύ των ορατών τους αναπαραστάσεων.
  - Συνήθως, στόχος μας είναι η άμεση/ρητή οπτικοποίηση των όχι άμεσα παρατηρήσιμων πραγματικών σχέσεων.
  - Για να επιτευχθεί αυτό, ίσως πρέπει να απορρίψουμε ορισμένα αντικείμενα (πρότυπα, γνωρίσματα ή ομάδες) και να δώσουμε έμφαση στα υπόλοιπα.
  - Εναλλακτικά, μπορούμε να εικονίσουμε διαφορετικές όψεις του συνόλου δεδομένων σε πολλαπλές, επιμέρους αναπαραστάσεις.

Οπτικοποίηση δεδομένων

- Παράδειγμα: έστω ένα σύνολο από n m-διάστατα πρότυπα με αριθμητικά γνωρίσματα. Αν το m είναι σχετικά μικρό, μπορούμε να δημιουργήσουμε ένα σύνολο από διδιάστατα διαγράμματα τα οποία να αναπαριστούν γραφικά τη σχέση κάθε γνωρίσματος με κάθε άλλο, σε ζεύγη ανά δύο.

- Αν το m είναι μεγάλο, συνήθως καταφεύγουμε σε μείωση διάστασης.
  - Π.χ., εφαρμόζουμε PCA και εικονίζουμε οπτικά μόνο τις σχέσεις μεταξύ των δύο μεγαλύτερων κυρίων συνιστωσών.

Οπτικοποίηση

Οπτικοποίηση δεδομένων

- Διακρίνουμε τρεις διαφορετικές κατηγορίες τρόπων οπτικοποίησης:
  - Οπτικοποίηση ενός **μικρού πλήθους γνωρισμάτων**.
  - Οπτικοποίηση δεδομένων **με χωρικά ή/και χρονικά γνωρίσματα**.
  - Οπτικοποίηση **πολυδιάστατων προτύπων**.

- Εναλλακτικές κατηγοριοποιήσεις επίσης είναι εφικτές, με κριτήριο, π.χ.:
  - Τύπος εφαρμογής (επιστημονική, στατιστική, κλπ.).
  - Δομή των δεδομένων (ιεραρχική, γράφοι, κλπ.).
  - Πλήθος γνωρισμάτων (1, 2, 3 ή περισσότερα).

Οπτικοποίηση

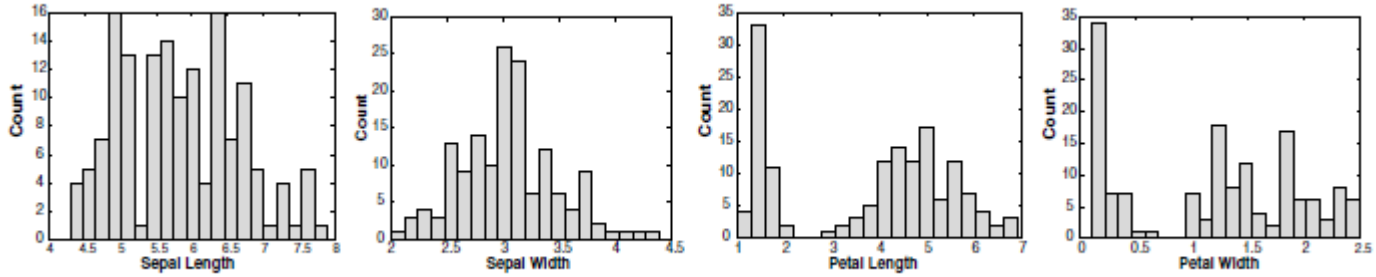
Μικρό πλήθος γνωρισμάτων

- Ο απλούστερος τρόπος οπτικοποίησης είναι το **ιστόγραμμα**.
  - Συνήθως ξεχωριστά ανά γνώρισμα.
  - Αναπαριστά την κατανομή των τιμών ενός γνωρίσματος σε όλα τα πρότυπα.
  - Διαιρούμε το πεδίο ορισμού του γνωρίσματος σε *κάδους* (bins, στον οριζόντιο άξονα) και δείχνουμε πόσα πρότυπα εμπίπτουν σε κάθε κάδο (κατακόρυφος άξονας).
    - Για κατηγορικά γνωρίσματα, κάθε δυνατή τιμή είναι ένας κάδος.
      - Αν προκύπτουν υπερβολικά πολλοί κάδοι, ομαδοποιούμε τις δυνατές τιμές σε λιγότερες.
    - Για αριθμητικά γνωρίσματα, διαιρούμε το εύρος τους (συνήθως ισόποσα) σε όσους κάδους θέλουμε.
      - Άρα το πλήθος των κάδων είναι υπερπαράμετρος επιλεγμένη από εμάς.
  - Η προκύπτουσα τιμή κάθε κάδου μπορεί να αναπαρασταθεί οπτικά ως μήκος της αντίστοιχης ράβδου σε **ραβδόγραμμα**.

Μικρό πλήθος γνωρισμάτων

Οπτικοποίηση

Ιστογράμματα των τεσσάρων γνωρισμάτων του συνόλου δεδομένων Iris, με 20 κάδους το καθένα.



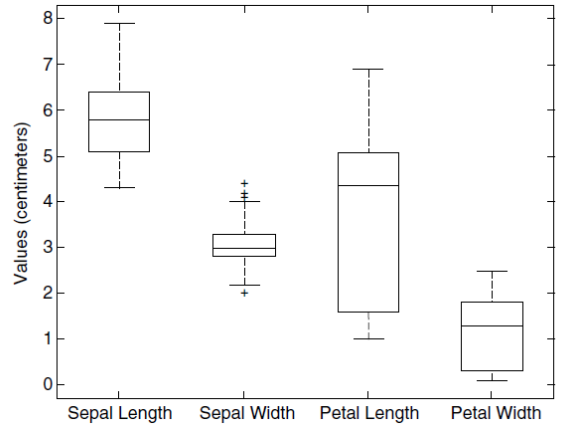
Πηγή: Tan, “Introduction to Data Mining”, 2006.

Οπτικοποίηση

Μικρό πλήθος γνωρισμάτων

- Με χρήση των εκατοστημορίων, μπορούμε να συνοψίσουμε την κατανομή πολλαπλών γνωρισμάτων στο ίδιο **θηκόγραμμα** (box plot).
  - Κάθε γνώρισμα αναπαρίσταται ξεχωριστά (οριζόντιος άξονας), με μία ορθογώνια *θήκη*.
  - Η επάνω και η κάτω οριζόντια γραμμή της θήκης τοποθετούνται στο 75ο και στο 25ο εκατοστημόριο, επί του κατακόρυφου άξονα, αντιστοίχως.
  - Η επάνω και η κάτω οριζόντια γραμμή της *ουράς* τοποθετούνται στο 90ο και στο 10ο εκατοστημόριο.
  - Η γραμμή μες στη θήκη τοποθετείται στον διάμεσο.
  - Οι ανωμαλίες σημειώνονται με +.

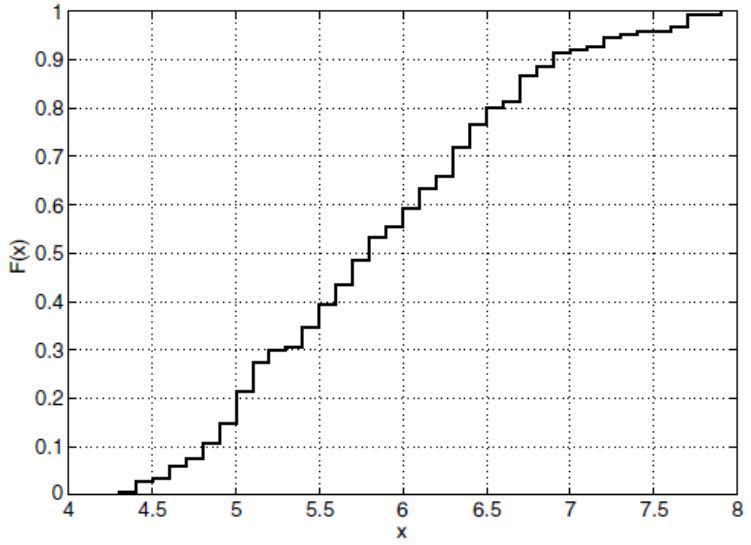
Θηκόγραμμα των τεσσάρων γνωρισμάτων του συνόλου δεδομένων Iris.



Πηγή: Tan, "Introduction to Data Mining", 2006.

Μικρό πλήθος γνωρισμάτων

- Εναλλακτικά, μπορεί να υπολογιστεί η **εμπειρική αθροιστική συνάρτηση κατανομής** (eCDF).
  - Για κάθε παρατηρούμενη τιμή του γνωρίσματος (οριζόντιος άξονας), δείχνει το ποσοστό των προτύπων (κατακόρυφος άξονας) τα οποία έχουν τιμή μικρότερη από αυτήν.
  - Ξεχωριστό διάγραμμα ανά γνώρισμα.



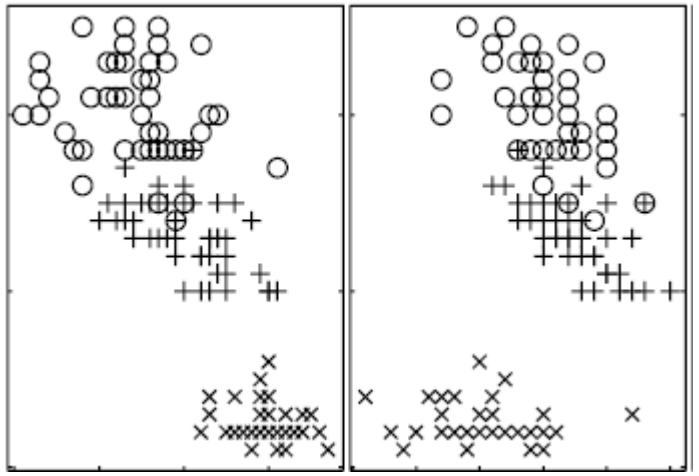
eCDF ενός γνωρίσματος του συνόλου δεδομένων Iris.

Πηγή: Tan, “Introduction to Data Mining”, 2006.

Οπτικοποίηση

Μικρό πλήθος γνωρισμάτων

- Τα πρότυπα μπορούν να εικονιστούν ως σημεία στο επίπεδο, λαμβάνοντας υπόψη μόνο δύο από τα γνωρίσματά τους.
  - **Διάγραμμα διασποράς** (scatter plot).
  - Για πληρότητα, πρέπει να κατασκευάσουμε ένα ξεχωριστό διάγραμμα για κάθε δυνατό ζεύγος διαφορετικών γνωρισμάτων.
  - Αν υπάρχουν ετικέτες κλάσης, μπορούμε να σημειώσουμε με διαφορετικό σύμβολο τα πρότυπα κάθε κλάσης.
  - <u>Δύο χρήσεις</u>: οπτικοποίηση σχέσεων μεταξύ γνωρισμάτων (π.χ., συσχέτιση), εύρεση του κατά πόσον τα δύο γνωρίσματα διαχωρίζουν τις κλάσεις.



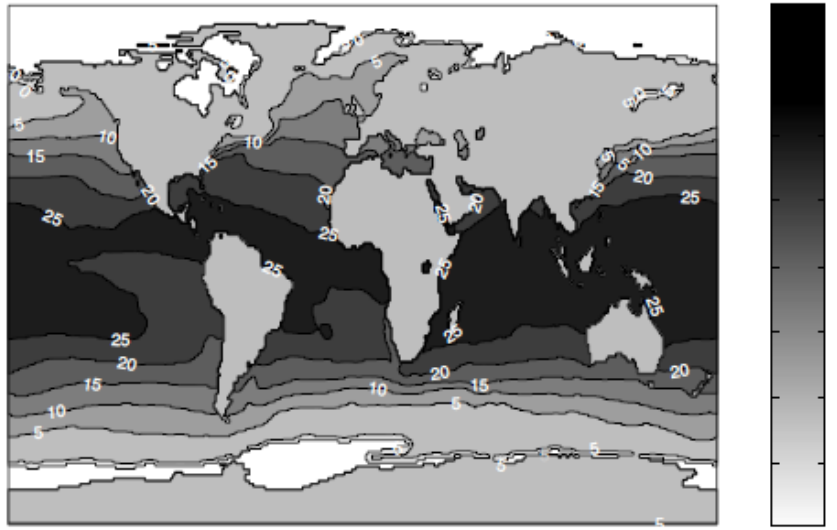
Δύο διαγράμματα διασποράς δύο διαφορετικών ζευγών γνωρισμάτων του συνόλου δεδομένων Iris (3 κλάσεις).

Πηγή: Tan, “Introduction to Data Mining”, 2006.

Χωροχρονικά δεδομένα

Οπτικοποίηση

- Τα χωροχρονικά δεδομένα μπορούν να οπτικοποιηθούν με ποικίλους τρόπους.
- Για γεωγραφικά δεδομένα, όπου μία μετρούμενη τιμή μεταβάλλεται από σημείο σε σημείο του επιπέδου, συνήθης πρακτική είναι τα **ισοϋψή διαγράμματα** (contour plots).
  - Η συνεχής περιοχή μεταξύ δύο ισοϋψών καμπυλών εικονίζεται με το ίδιο χρώμα, άρα έχει περίπου την ίδια τιμή.
  - Η μετρούμενη τιμή δεν είναι απαραιτήτως υψόμετρο.

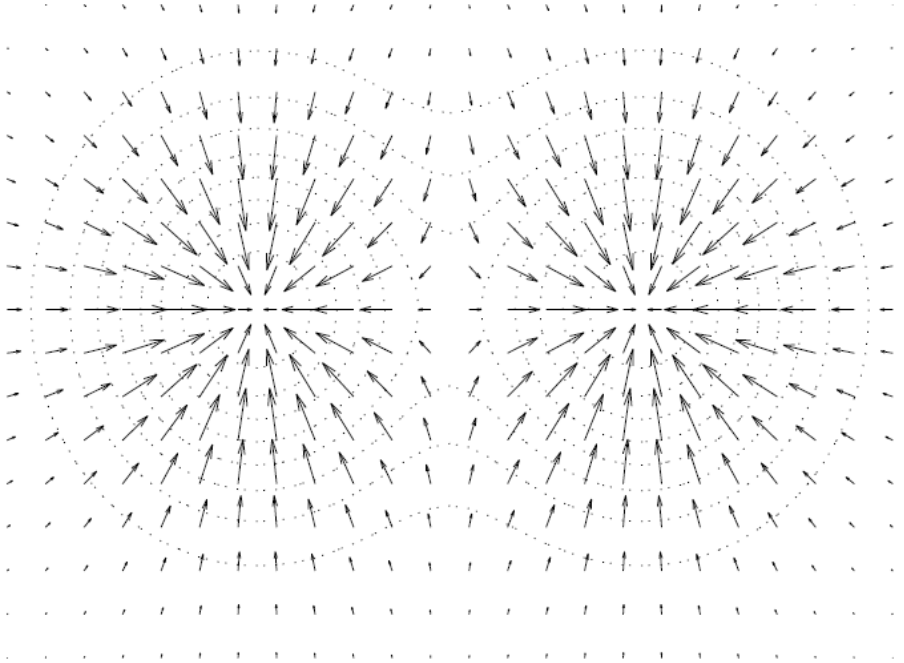


Ισοϋψές διάγραμμα της επιφανειακής θαλάσσιας θερμοκρατίας.

Πηγή: Tan, “Introduction to Data Mining”, 2006.

Χωροχρονικά δεδομένα

- Αν σε κάθε σημείο του επιπέδου αντιστοιχεί ένα διδιάστατο διάνυσμα, με μέτρο **και** κατεύθυνση, χρησιμοποιούμε **διάγραμμα διανυσματικού πεδίου** (vector field).
  - Πρόκειται για την οπτική απεικόνιση μίας διδιάστατης διανυσματικής συνάρτησης δύο ανεξάρτητων μεταβλητών.



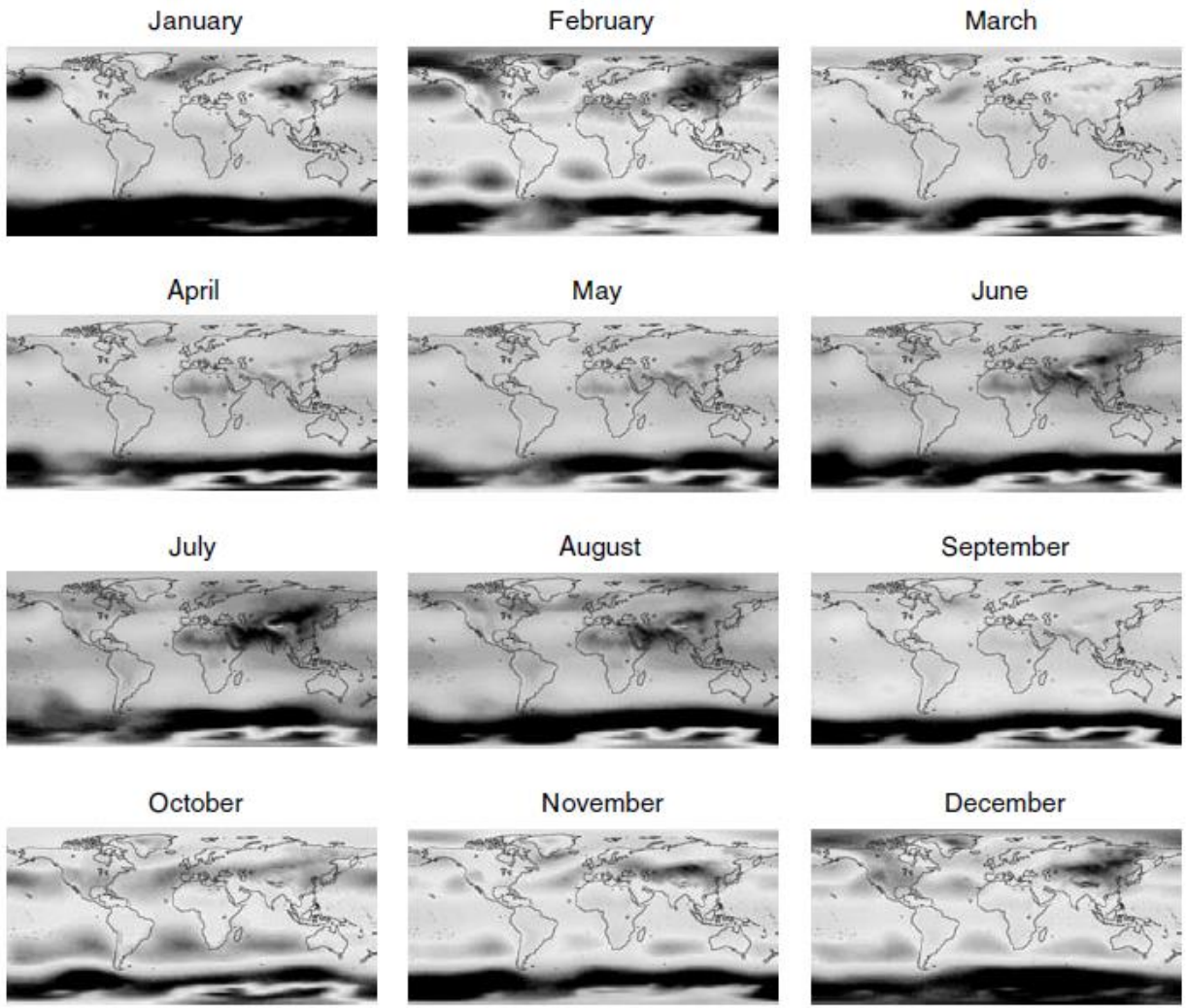
Πηγή: Tan, “Introduction to Data Mining”, 2006.

Χωροχρονικά δεδομένα

- Τα προηγούμενα δεδομένα δεν μεταβάλλονταν με τον χρόνο.
- Σε περιπτώσεις χωροχρονικών δεδομένων, μπορούμε να δείξουμε ξεχωριστά διαγράμματα για κάθε χρονική στιγμή.
- Παράδειγμα: πραγματική βαθμωτή συνάρτηση τριών ανεξάρτητων μεταβλητών, η οποία αποδίδει μία τιμή θαλάσσιας πίεσης σε ένα γεωγραφικό σημείο ανά μήνα.
  - Μπορούν να γίνουν ξεχωριστά ισοϋψή διαγράμματα για κάθε μήνα.

Χωροχρονικά δεδομένα

Οπτικοποίηση



Πηγή: Tan, “Introduction to Data Mining”, 2006.

Πολυδιάστατα δεδομένα

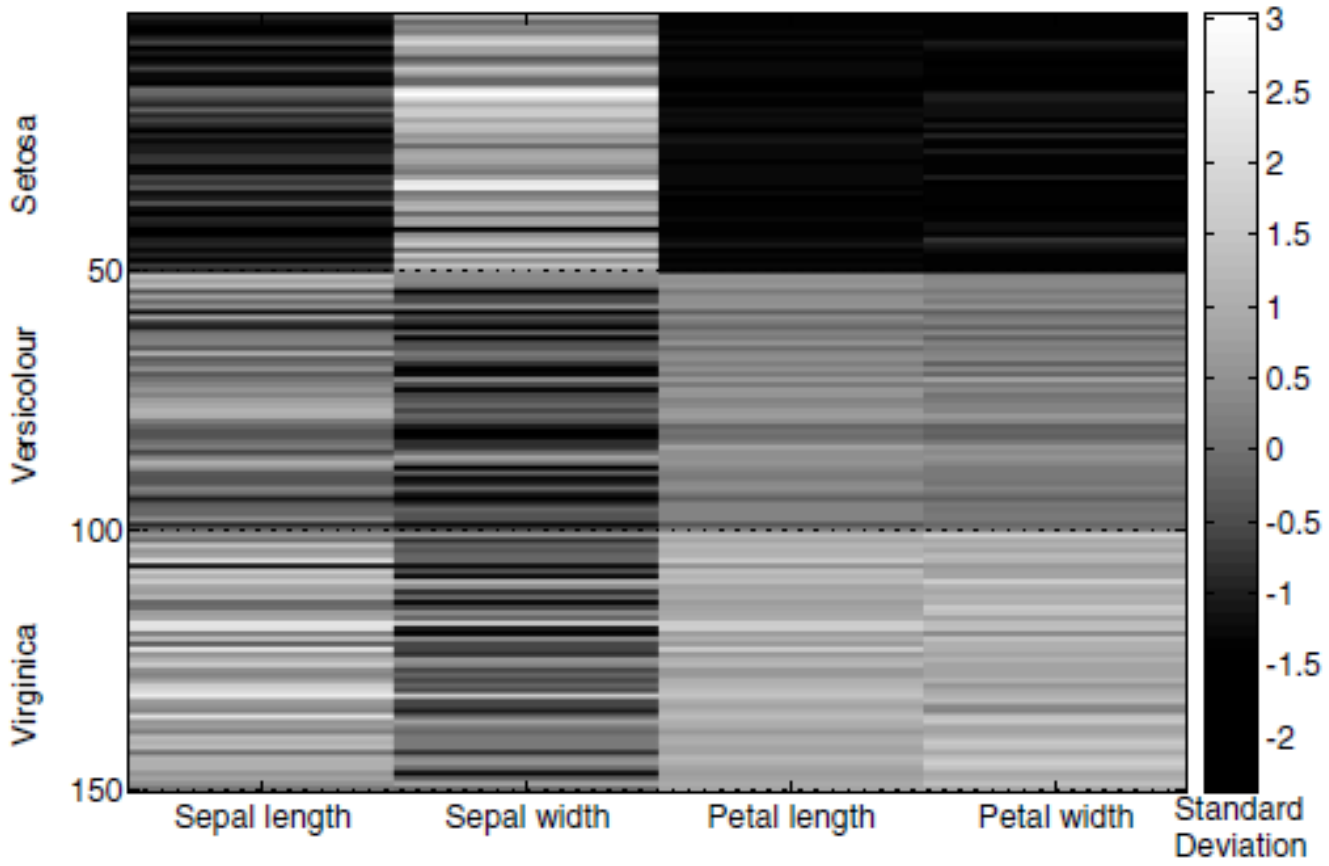
- Ο πιο απλός τρόπος να οπτικοποιηθούν πολυδιάστατα δεδομένα είναι να αναπαρασταθεί ως εικόνα ο ίδιος ο πίνακας δεδομένων.
  - Αναθέτουμε διαφορετικό χρώμα σε διαφορετικές τιμές των ποικίλων γνωρισμάτων.
  - Αν διαφορετικά γνωρίσματα έχουν διαφορετικά εύρη, μπορούν πρώτα να τυποποιηθούν όλα ώστε να έχουν μέση τιμή 0 και διακύμανση 1.
  - Αν το σύνολο δεδομένων διαθέτει ετικέτες κλάσεις, τότε ο πίνακας μπορεί να αναδιαταχθεί ώστε συνεχόμενες γραμμές να περιέχουν πρότυπα της ίδιας κλάσης.

Οπτικοποίηση

Πολυδιάστατα δεδομένα

Οπτικοποίηση

Οπτικοποίηση του πίνακα δεδομένων του συνόλου δεδομένων Iris (διάστασης n x m).



Πηγή: Tan, “Introduction to Data Mining”, 2006.

Πολυδιάστατα δεδομένα

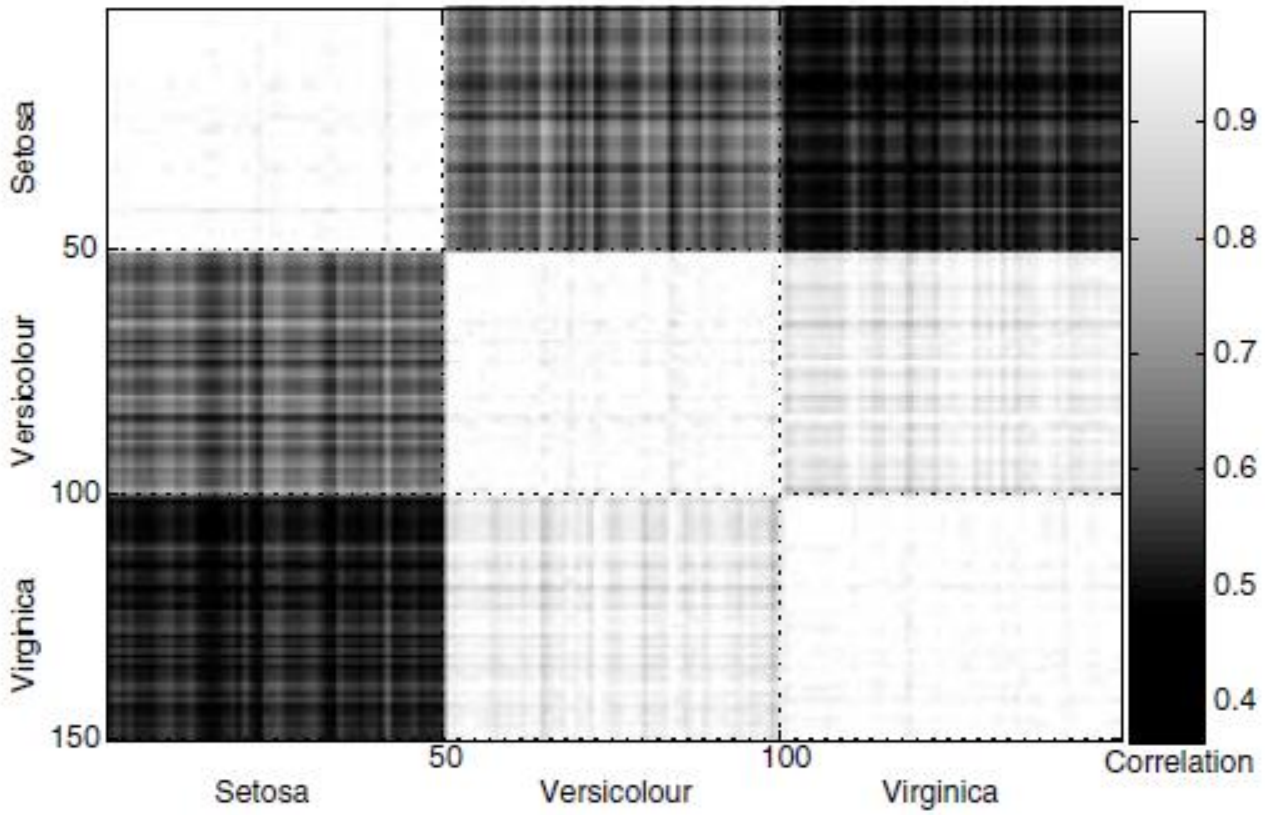
- Αντί για τον πίνακα δεδομένων μπορούμε να υπολογίσουμε τον πίνακα ομοιοτήτων (με βάση κάποιο επιλεγμένο μέτρο εγγύτητας), διάστασης $n \times n$.
  - Αν το σύνολο δεδομένων διαθέτει ετικέτες κλάσεις, τότε ο πίνακας μπορεί να αναδιαταχθεί ώστε συνεχόμενες γραμμές/στήλες να αφορούν πρότυπα της ίδιας κλάσης.
    - Αυτό επιτρέπει άμεση οπτική εκτίμηση του πόσο συμπαγής είναι μία κλάση, μέσω του χρώματος.
  - Οι πίνακες ομοιοτήτων είναι συμμετρικοί.

Οπτικοποίηση

Πολυδιάστατα δεδομένα

Οπτικοποίηση

Οπτικοποίηση του πίνακα ομοιοτήτων του συνόλου δεδομένων Iris (διάστασης n x n). Μέτρο εγγύτητας είναι η συσχέτιση μεταξύ δύο προτύπων.



Πηγή: Tan, "Introduction to Data Mining", 2006.

On-Line Analytical Processing

- Η προσέγγιση On-Line Analytical Processing (OLAP) είναι ένα σύνολο μηχανισμών αναπαράστασης και αλληλεπιδραστικού χειρισμού των δεδομένων, υπό μορφή **πολυδιάστατων πινάκων**.
  - Τα σχετικά εργαλεία είναι ενσωματωμένα συνήθως σε συστήματα διαχείρισης βάσεων δεδομένων.

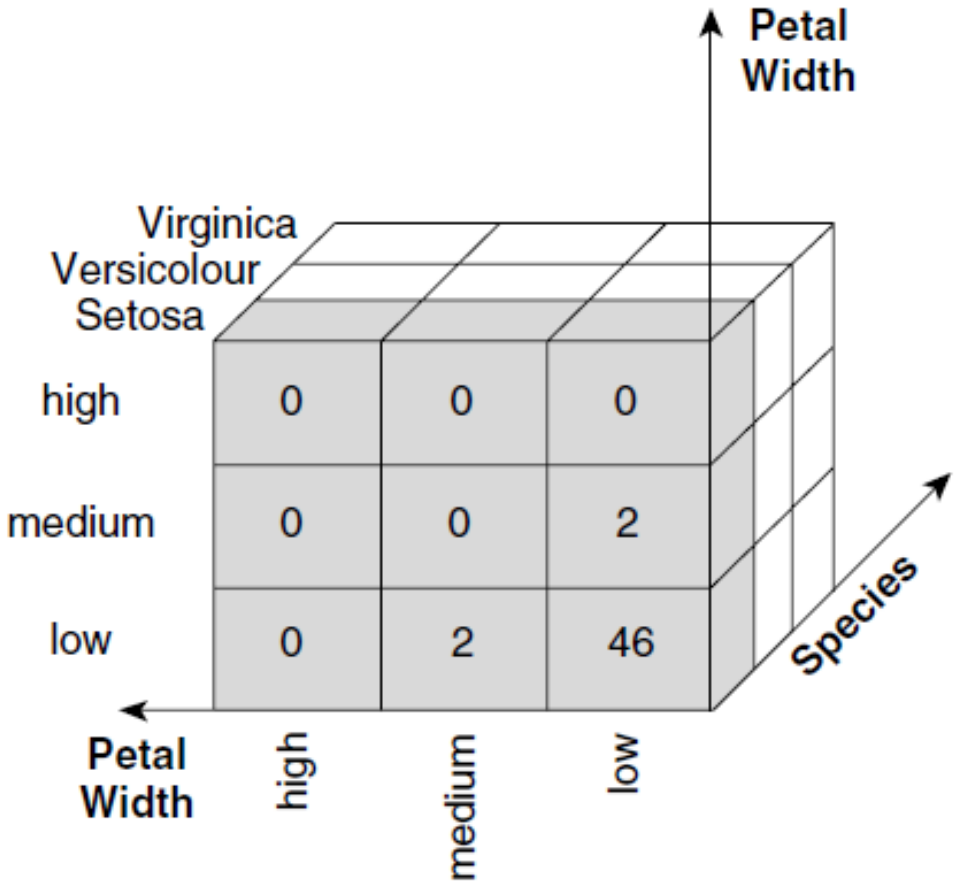
OLAP

- Κεντρική ιδέα είναι η αλλαγή της αναπαράστασης των δεδομένων, ενδεχομένως μετά από συνόψισή τους, ώστε να εικονίζονται ως **φέτες** ενός πολυδιάστατου πίνακα τον οποίον μπορούμε να χειριστούμε κατά βούληση.
  - Αντικαθιστά τον απλό πίνακα δεδομένων.

On-Line Analytical Processing

OLAP

Αναπαράσταση του συνόλου δεδομένων Iris (3 κλάσεων) με μόνο δύο από τα γνωρίσματά του, τα οποία έχουν διακριτοποιηθεί σε τρεις τιμές το καθένα (low, medium, high). Σε κάθε κελί του τριδιάστατου πίνακα αναγράφεται το πλήθος των προτύπων τα οποία εμπίπτουν σε αυτή τη θέση.



Πηγή: Tan, “Introduction to Data Mining”, 2006.

On-Line Analytical Processing

OLAP

- Ο πίνακας αυτός αποτελείται από τρεις διδιάστατες φέτες.
- Όμως μπορούμε να ορίσουμε τις φέτες κατά μήκος οποιασδήποτε από τις τρεις διαστάσεις.
- Έτσι, μπορούμε να αντλήσουμε εύκολα συμπεράσματα.
  - Π.χ., οι τρεις κλάσεις έχουν εμφανώς διαφορετικά χαρακτηριστικά όσον αφορά τα εν λόγω δύο γνωρίσματα.

Setosa

		Width		
		low	medium	high
Length	low	46	2	0
	medium	2	0	0
	high	0	0	0

Versicolour

		Width		
		low	medium	high
Length	low	0	0	0
	medium	0	43	3
	high	0	2	2

Virginica

		Width		
		low	medium	high
Length	low	0	0	0
	medium	0	0	3
	high	0	3	44

On-Line Analytical Processing

OLAP

- Στο ανωτέρω παράδειγμα η ποσότητα η οποία αποθηκεύεται σε κάθε κελί του πίνακα είναι το πλήθος των προτύπων τα οποία εμπίπτουν στην εν λόγω θέση.
  - Ωστόσο, αυτό δεν είναι απαραίτητο. Η μετρούμενη ποσότητα εξαρτάται από την εφαρμογή και το τι θέλουμε να πετύχουμε.

- Η λήψη μίας φέτας του πολυδιάστατου πίνακα καλείται *slicing*.
- Η λήψη ενός υποπίνακα κατά μήκος όλων των φετών, αλλά για ορισμένο μόνο εύρος εκ των δυνατών τιμών κάποιου ή κάποιων γνωρισμάτων, καλείται *dicing*.

On-Line Analytical Processing

OLAP

- Ένα βασικό πλεονέκτημα των μηχανισμών OLAP είναι ότι μας επιτρέπουν εύκολη **συνάθροιση** των μετρούμενων ποσοτήτων κατά βούληση.
  - Π.χ., άθροιση ή εύρεση μέσου όρου των φετών κατά μήκος όποιας διάστασης εμείς θέλουμε.
  - Πρόκειται για μία απλοϊκή εκδοχή *μείωσης διάστασης*.
    - Η συνάθροιση κατά μήκος μίας από τις 3 διαστάσεις εξαλείφει αυτή τη διάσταση.
- Ένας διαφορετικός τύπος συνάθροισης είναι όταν συναθροίζουμε κελιά στο εσωτερικό μίας διάστασης, αλλά όχι ενιαία καθ' όλη τη διάσταση, με κριτήριο μία ιεραρχία.
  - Π.χ., κάθε μήνας περιέχει 30 ή 31 ημέρες.
  - Ημερήσια δεδομένα μπορούν να μετατραπούν σε μηνιαία, με κατάλληλες επιμέρους συναθροίσεις κατά μήκος της διάστασης του χρόνου (*roll up*).
    - Συναθροίζουμε ξεχωριστά στο εσωτερικό κάθε μήνα.
  - Ή μπορεί να γίνει το αντίστροφο: ανάλυση των δεδομένων αντί για συνάθροιση (*drill down*).

On-Line Analytical Processing

OLAP

- Ένας πολυδιάστατος πίνακας μαζί με όλες τις πιθανές συναθροίσεις του καλείται **κύβος δεδομένων**.
  - Προσοχή: δεν χρειάζεται απαραιτήτως όλες οι διαστάσεις του να είναι ίσες, ούτε να είναι μόνο τρεις.

- Οι μηχανισμοί OLAP οι οποίοι είναι ενσωματωμένοι σε συστήματα βάσεων δεδομένων συνήθως επιτρέπουν κατασκευή, αλληλεπιδραστικό χειρισμό και οπτικοποίηση του κύβου δεδομένων, συμπληρωματικά με τον αρχικό πίνακα δεδομένων.

- Τα περισσότερα συστήματα επιτρέπουν προγραμματιστικά ερωτήματα OLAP μέσω της γλώσσας ερωτημάτων MDX.
  - Σε σχεσιακές βάσεις δεδομένων, η MDX λειτουργεί συμπληρωματικά ως προς την SQL.

Thank you for your attention!

Q & A

Contact: imademlis@aueb.gr