



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά
μαθήματα ΠΠ

Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων



Σημείωμα Αδειοδότησης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧ/ΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Εξόρυξη Δεδομένων και Ανακάλυψη Γνώσης

Β. Μεγαλοικονόμου

Κίνητρο – Γιατί χρειάζεται η εξόρυξη δεδομένων;

- Έχουμε πολλά δεδομένα αλλά μικρή πληροφορία!
- Πρόβλημα έκρηξης δεδομένων
 - Τα εργαλεία αυτοματοποιημένης συλλογής δεδομένων και η τελευταία τεχνολογία βάσεων δεδομένων οδηγεί σε ένα μεγάλο πλήθος δεδομένων το οποίο αποθηκεύεται σε βάσεις δεδομένων, data warehouses και άλλες αποθήκες δεδομένων
- Λύση: Εξόρυξη δεδομένων
 - Εξόρυξη ενδιαφέρουσας γνώσης (κανόνες, πρότυπα, περιορισμοί) από δεδομένα μεγάλων βάσεων δεδομένων

Εξέλιξη της τεχνολογίας βάσεων δεδομένων

- 1960:
 - Συλλογή δεδομένων, δημιουργία βάσης δεδομένων, συστήματα διαχείρισης πληροφορίας (IMS) και δικτυωτές βάσεις δεδομένων
- 1970:
 - Σχεσιακό μοντέλο δεδομένων, υλοποίηση σχεσιακού ΣΔΒΔ
- 1980:
 - RDBMS, προηγμένα μοντέλα δεδομένων (extended-relational, OO, deductive, κτλ.) και συστήματα προσανατολισμένα στην εφαρμογή (spatial, scientific, engineering, κτλ.)
- 1990—2000:
 - Εξόρυξη δεδομένων και αποθηκών δεδομένων (Data mining, data warehousing), βάσεις δεδομένων πολυμέσων, βάσεις δεδομένων στο Παγκόσμιο Ιστό (Web databases)

Ανακάλυψη Γνώσης- Εξόρυξη δεδομένων ή γνώσης

- ✓ Τι είναι **ανακάλυψη γνώσης από βάσεις δεδομένων** (knowledge discovery from databases-KDD):
 - Αποκάλυψη ή παραγωγή λειτουργικής γνώσης, μέσω της ανάλυσης δεδομένων από μεγάλες αποθήκες δεδομένων.
 - Εύρεση δομών γνώσης που αναδεικνύουν γνώση (π.χ. συσχετίσεις ή κανόνες) που είναι κρυμμένη μέσα στα δεδομένα και δεν μπορούν να εξαχθούν από τον άνθρωπο με ευκολία.
 - Αναφέρεται στην όλη διαδικασία.
- ✓ Τι είναι **εξόρυξη δεδομένων** (data mining) και **εξόρυξη γνώσης** (knowledge mining):
 - Τα ίδια με παραπάνω.
 - Αναφέρεται στις επί μέρους τεχνικές.

Ανακάλυψη γνώσης από δεδομένα

Ορισμός

- ✓ «KDD είναι η ντετερμινιστική διαδικασία αναγνώρισης έγκυρων, καινοτόμων, ενδεχομένως χρήσιμων και εν τέλει κατανοητών προτύπων στα δεδομένα.» (*Frawley, Piatesky-Shapiro and Matheus, 1991*).

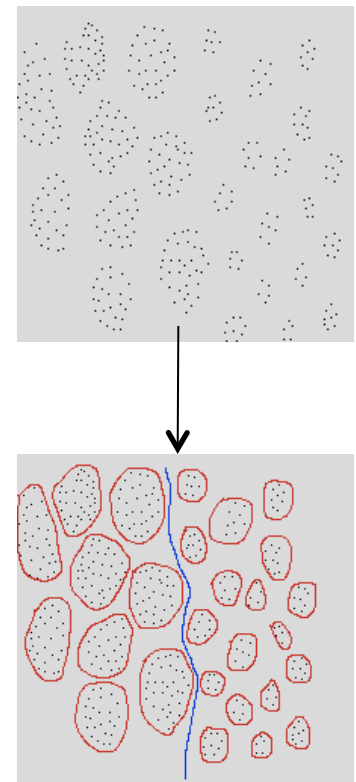
Δεδομένα: Οντότητες ή συσχετίσεις του πραγματικού κόσμου (π.χ., εγγραφές συναλλαγών τραπεζής, supermarket κλπ).

Πρότυπο: Μια έκφραση σε μια γλώσσα που χαρακτηρίζει ένα υποσύνολο των δεδομένων (π.χ., ένας κανόνας).

Εγκυρότητα: Το πρότυπο είναι συνεπές σε νέα δεδομένα.

Πιθανή χρησιμότητα: Να μπορεί να χρησιμοποιηθεί για κάποιο σκοπό (π.χ., λήψη αποφάσεων).

Τελικά κατανοητό: Κοινώς κατανοητό, ώστε να είναι κοινώς χρήσιμο.



Τι είναι η εξόρυξη δεδομένων?

- Τώρα που έχουμε συλλέξει τόσα δεδομένα, τι κάνουμε με αυτά;
- Εξαγωγή **χρήσιμων προτύπων** (αυτόματα)
 - Συσχετίσεις (π.χ., ψωμί + βούτυρο --> γάλα)
 - Ακολουθίες (π.χ., χρονικά δεδομένα που σχετίζονται με το χρηματιστήριο)
 - Κανόνες που διαμοιράζουν τα δεδομένα (π.χ. πρόβλημα τοποθεσίας αποθήκευσης)
- Ποια πρότυπα έχουν **“ενδιαφέρον”**;

Περιεχόμενο πληροφορίας, εμπιστοσύνη και υποστήριξη, απρόσμενο (χρησιμότητα στην λήψη απόφασης))

Γιατί εξόρυξη δεδομένων; —

Πιθανές εφαρμογές

- Ανάλυση βάσης δεδομένων και υποστήριξη απόφασης
 - Ανάλυση και διαχείριση αγοράς
 - Μάρκετινγκ στόχου, διαχείριση σε σχέση με τον πελάτη, ανάλυση καλαθιού αγορών, cross selling, τμηματοποίηση αγοράς
 - Ανάλυση και διαχείριση κινδύνου/ρίσκου
 - Πρόβλεψη, έλεγχος ποιότητας, ανάλυση ανταγωνισμού
 - Εντοπισμός οικονομικού εγκλήματος
- Άλλες εφαρμογές
 - Εξόρυξη κειμένου (newsgroup, email, documents), ανάλυση παγκόσμιου ιστού και κοινωνικών δικτύων.
 - Εξόρυξη χωρικών, χρονικών και χωροχρονικών δεδομένων
 - Έξυπνη απάντηση ερωτημάτων

Ανάλυση και διαχείριση αγοράς

- Ποιές είναι οι πηγές των δεδομένων για την ανάλυση; (Δοσοληψίες με κάρτες, εκπτωτικά κουπόνια, κλήσεις παραπόνων από πελάτες, κλπ.)
- Μάρκετινγκ στόχου (Εύρεση ομάδων “μοντέλων” πελατών που έχουν κοινά χαρακτηριστικά: εισόδημα, συνήθης αγορές, κλπ.)
- Καθορισμός προτύπων συναλλαγών των πελατών με το χρόνο (Μετατροπή ενός λογαριασμού ενός μόνο ατόμου σε κοινό: γάμος, κλπ.)
- «Cross-market» ανάλυση (Συσχετίσεις μεταξύ προϊόντων πωλήσεων και προβλέψεις με βάση τις πωλήσεις)
- Προφίλ πελάτη (Ποια προϊόντα αγοράζει ο πελάτης)
- Προσδιορισμός απαιτήσεων πελάτη (Τα καλύτερα προϊόντα για διαφορετικούς πελάτες)
- Παροχή συνοπτικής πληροφορίας (πολυδιάστατες αναφορές – multidimensional summary reports)

Ανάλυση και διαχείριση κινδύνου/ρίσκου

- Οικονομικός προγραμματισμός
 - Ανάλυση και πρόβλεψη κίνησης μετρητών
 - Ανάλυση χρονοσειρών και cross-sectional (ανάλυση τάσης (trend analysis), κλπ.)
- Προγραμματισμός πηγών:
 - Σύνοψη και σύγκριση των πηγών και των εξόδων
- Ανταγωνισμός:
 - Έλεγχος ανταγωνιστών και τάσεων της αγοράς
 - Ομαδοποίηση πελατών και απόδοση τιμών με βάση τις ομάδες αυτές
 - Ορισμός στρατηγικής απόδοσης τιμών σε μία ιδιαίτερα ανταγωνιστική αγορά

Εντοπισμός και διαχείριση οικονομικού εγκλήματος/απάτης

- Εφαρμογές
 - Υγεία, υπηρεσίες πιστωτικών καρτών, τηλεπικοινωνίες κτλ.
- Προσέγγιση
 - Χρήση ιστορικών στοιχείων για την δόμηση μοντέλων κανονικής και λανθασμένης συμπεριφοράς και χρήση τεχνικών εξόρυξης δεδομένων για τον προσδιορισμό λανθασμένων στιγμιοτύπων
- Παραδείγματα
 - Ασφάλεια αυτοκινήτων: εντοπισμός ομάδων που προκαλούν ατυχήματα για την είσπραξη της ασφάλειας
 - Ξέπλυμα χρήματος: εντοπισμός ύποπτων δοσοληψιών χρημάτων
 - Ιατρική ασφάλεια: εντοπισμός επαγγελματιών ασθενών, ακατάλληλων ιατρικών θεραπειών

Ανακάλυψη Ιατρικής/ Βιολογικής Γνώσης

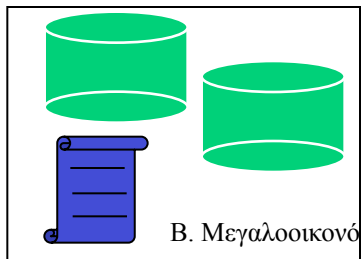
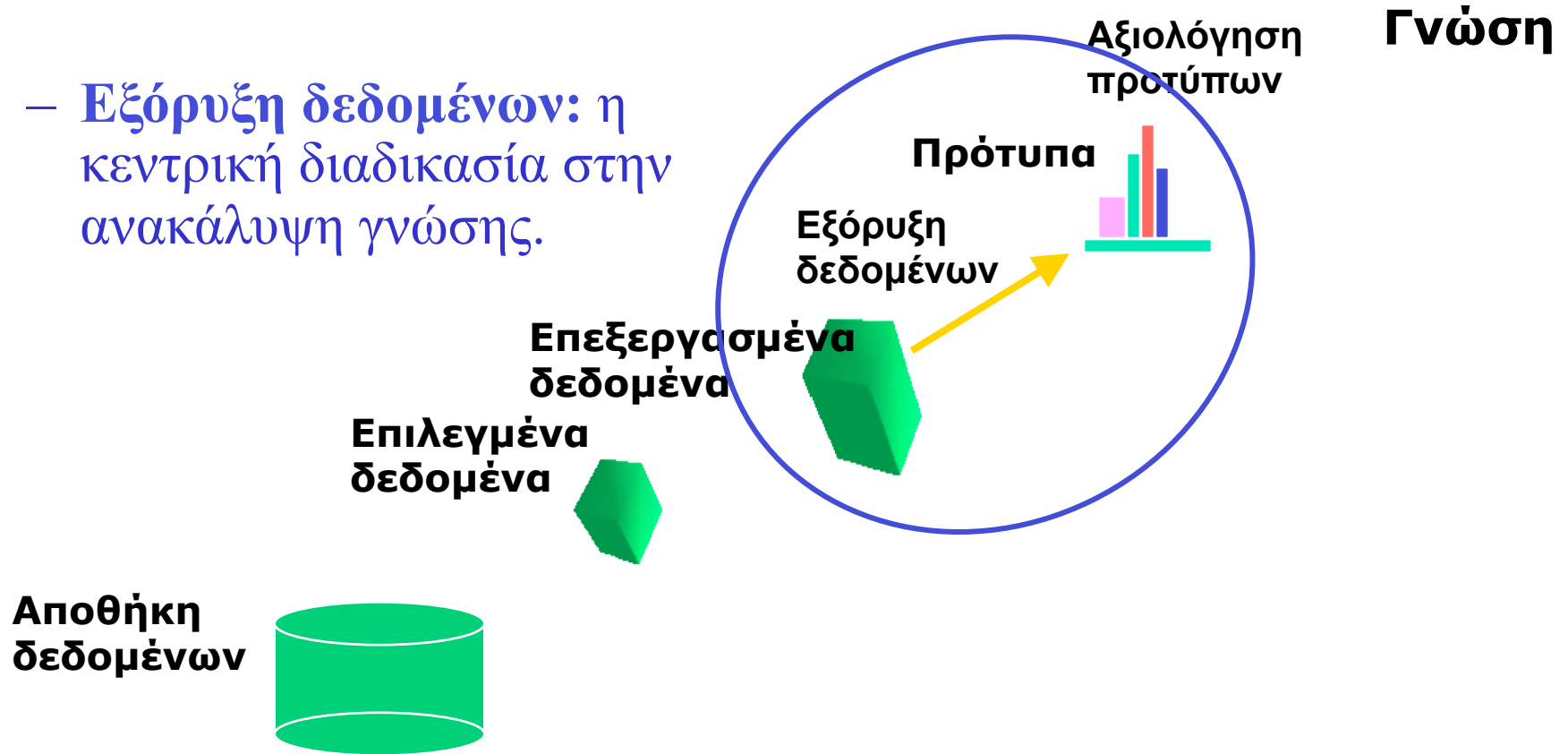
- Ανακάλυψη συσχετίσεων δομής-λειτουργίας
 - Χαρτογράφηση ανθρώπινου εγκεφάλου (lesion-deficit, task-activation associations)
 - Δομή κυττάρου (cytoskeleton) και λειτουργικότητα ή παθολογία
 - Δομή και λειτουργικότητα πρωτεϊνών
- Ανακάλυψη αιτιακών σχέσεων (causal relationships)
 - Συμπτώματα και κλινικές καταστάσεις ασθενών
- Ανάλυση ακολουθίας DNA
 - Βιοπληροφορική (microarrays, κτλ.)

Άλλες εφαρμογές

- Αθλήματα
 - IBM Advanced Scout ανέλυσαν στατιστικά παιχνιδιών του NBA (shots blocked, assists, and fouls) για να κερδίσουν ανταγωνιστικό πλεονέκτημα για New York Knicks και Miami Heat
- Αστρονομία
 - Τα JPL και Palomar Observatory ανακάλυψαν 22 quasars με την βοήθεια της εξόρυξης δεδομένων
- Internet Web Surf-Aid
 - IBM Surf-Aid χρησιμοποίησε data mining αλγορίθμους σε Web access logs για ιστοσελίδες αγοράς προϊόντων για να ανακαλύψουν προτιμήσεις και συμπεριφορές πελατών, αναλύοντας την αποτελεσματικότητα του Web marketing, βελτιώνοντας Web site organization, κλπ.

Ανακάλυψη γνώσης από δεδομένα

- **Εξόρυξη δεδομένων:** η κεντρική διαδικασία στην ανακάλυψη γνώσης.



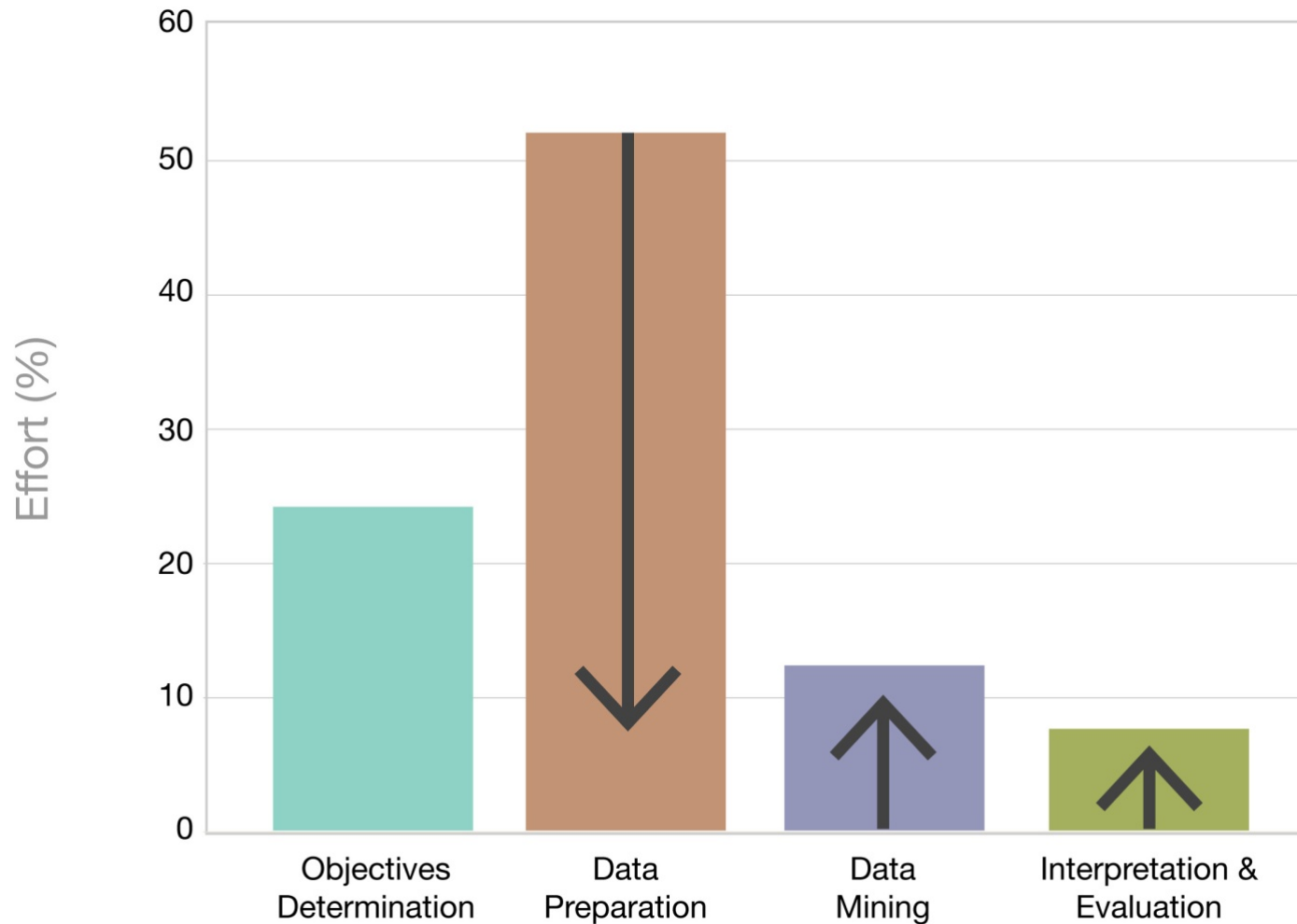
Βάσεις δεδομένων

Αρχεία

Βασικά Βήματα Διαδικασίας Ανακάλυψης Γνώσης

- *Επιλογή (Selection)*: Απόκτηση δεδομένων από διάφορες ετερογενείς πηγές.
- *Προεπεξεργασία (Preprocessing)*: Εσφαλμένα, προβληματικά ή ελλείποντα δεδομένα τακτοποιούνται (~50-60% της προσπάθειας!)
- *Μετασχηματισμός (Transformation)*: Μετατροπή ετερογενών δεδομένων σε κοινή τυποποίηση για επεξεργασία.
- *Εξόρυξη δεδομένων (Data mining)*: Εφαρμογή αλγορίθμων για παραγωγή μοντέλου.
- *Διερμηνεία/Αξιολόγηση (Interpretation/Evaluation)*: Παρουσίαση των αποτελεσμάτων της εξόρυξης δεδομένων στους χρήστες για αξιολόγηση (χρήση μεθόδων οπτικοποίησης και GUI).

Απαιτούμενη Προσπάθεια για κάθε Βήμα της Διαδικασίας Ανακάλυψης Γνώσης



Τα βέλη δείχνουν την κατεύθυνση στην οποία θέλουμε να κινηθούν οι προσπάθειες

Εξαγωγή Χαρακτηριστικών

Τα χαρακτηριστικά (features) διακρίνονται σε:

- Ποσοτικά χαρακτηριστικά :
 - continuous values (π.χ. βάρος),
 - discrete values (π.χ. ο αριθμός των υπολογιστών),
 - interval values (π.χ. η διάρκεια ενός γεγονότος).
- Ποιοτικά χαρακτηριστικά:
 - Ονομαστικά (nominal) or unordered (π.χ. χρώμα),
 - Ordinal (π.χ. η ποιοτική αξιολόγηση της θερμοκρασίας (“ζεστό” or “κρύο”) ή της έντασης του ήχου (“ήσυχα” or “δυνατά”)).
- Χαρακτηριστικά δομής: (trees, symbolic objects)

Αλγόριθμοι Εξόρυξης Δεδομένων

- ✓ Μας ενδιαφέρουν τα παρακάτω:
 - ✓ Συνθετικά μέρη
 - ✓ Περιγραφή μοντέλου
 - Τύπος ή στόχος μοντέλου (π.χ. ταξινόμηση, συσταδοποίηση)
 - Μορφή μοντέλου (π.χ. δένδρα, κανόνες, νευρωνικά δίκτυα)
 - ✓ Κριτήρια αξιολόγησης μοντέλου (εγκυρότητα, ακρίβεια, κλπ)
 - ✓ Τεχνική αναζήτησης:
 - Παραμέτρων ή
 - Μοντέλου

Τύποι Μοντέλων

- ✓ Προβλεπτικό (predictive)
 - Κάνει πρόβλεψη συμπεριφοράς κάποιων μεταβλητών που παρουσιάζουν ενδιαφέρον και οι οποίες βασίζονται στη συμπεριφορά άλλων μεταβλητών. Η πρόβλεψη μπορεί να στηρίζεται σε ιστορικά δεδομένα.
- ✓ Περιγραφικό (descriptive)
 - Βρίσκει πρότυπα (patterns) ή σχέσεις (relationships) που ενυπάρχουν στα δεδομένα. Ερευνά τις ιδιότητες των υπό εξέταση δεδομένων, εξηγεί τη συμπεριφορά τους.

Διεργασίες Εξόρυξης Δεδομένων

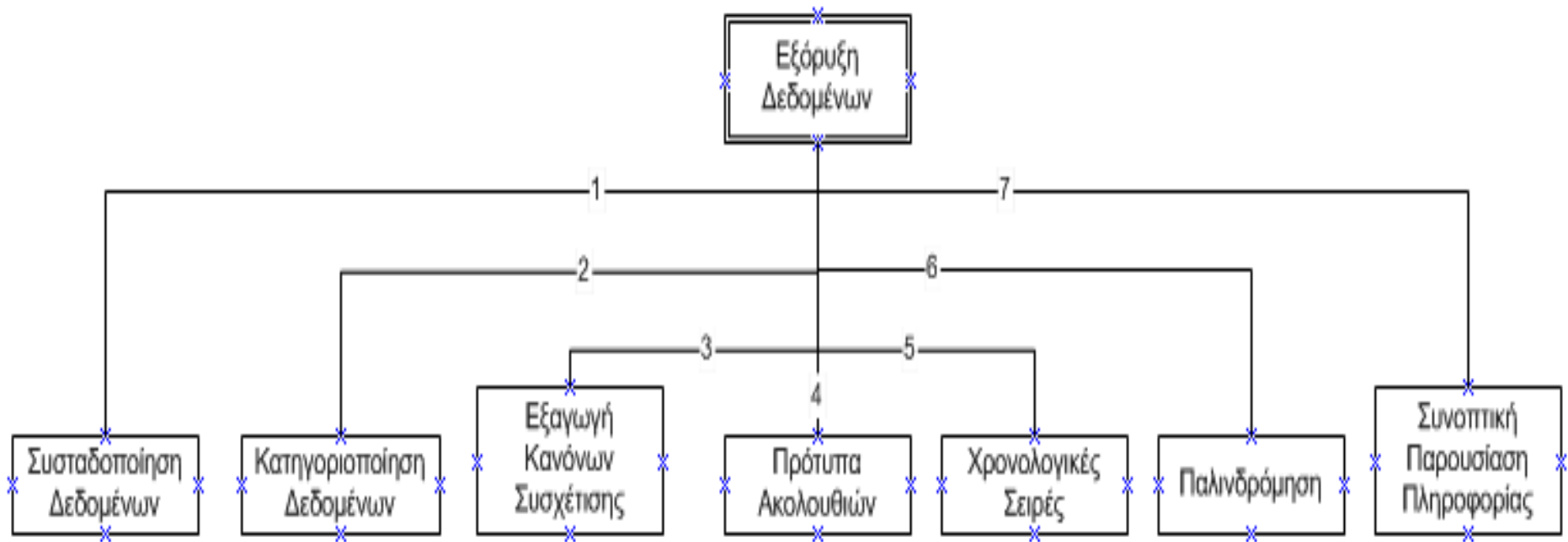
Προβλεπτικού τύπου ή μοντέλου

- ✓ Κατηγοριοποίηση (Classification)
- ✓ Παλινδρόμηση ή Παρεμβολή (Regression)
- ✓ Ανάλυση χρονοσειρών (Time series analysis)
- ✓ Πρόβλεψη (Prediction)

Περιγραφικού τύπου ή μοντέλου

- ✓ Συσταδοποίηση (Clustering)
- ✓ Σύνοψη (Summarization) ή Γενίκευση (Generalization)
- ✓ Εύρεση Κανόνων Συσχέτισης (Association Rules)
- ✓ Ανακάλυψη Συσχετίσεων σε Ακολουθίες (Pattern Discovery in Sequences)

Επισκόπηση Εργασιών Εξόρυξης



Κατηγοριοποίηση Δεδομένων - Πρόβλεψη

Κατηγοριοποίηση δεδομένων (data classification):

Βασίζεται στην εξέταση των χαρακτηριστικών ενός νέου αντικειμένου (μη κατηγοριοποιημένο) το οποίο με βάση τα χαρακτηριστικά αυτά αντιστοιχίζεται σε ένα προκαθορισμένο σύνολο κλάσεων (κατηγοριών).

Βασικές μέθοδοι: Μέθοδος Bayes, Δέντρα Αποφάσεων,
Νευρωνικά Δίκτυα

Πρόβλεψη:

Μοντελοποίηση συνεχών συναρτήσεων, π.χ., πρόβλεψη άγνωστων ή απώντων τιμών

Τυπικές Εφαρμογές

credit approval

target marketing

medical diagnosis

treatment effectiveness analysis

Κατηγοριοποίηση — Μια διαδικασία δύο βημάτων

- Κατασκευή μοντέλου: περιγράφοντας ένα σύνολο προκαθορισμένων κατηγοριών
 - Καθε μια από τις πλειάδες (tuples) υποτίθεται ότι ανήκει σε μια προκαθορισμένη κατηγορία, όπως καθορίζεται από το γνώρισμα κατηγορίας (κλάσης) (**supervised learning**)
 - Χρήση ενός training set για κατασκευή μοντέλου
 - Αναπαράσταση μοντέλου σαν classification rules, decision trees, ή μαθηματικές εξισώσεις
- Χρήση μοντέλου: για κατηγοριοποίηση άγνωστων αντικειμένων
 - Υπολογισμός ακρίβειας του μοντέλου κάνοντας χρήση ενός test set
 - Η γνωστή ετικέτα του test sample συγκρίνεται με το αποτέλεσμα κατηγοριοποίησης του μοντέλου
 - Η τιμή της ακρίβειας είναι το ποσοστό των test set samples που κατηγοριοποιήθηκαν σωστά απο το μοντέλο
 - Το test set είναι ανεξάρτητο από το training set, αλλιώς μπορεί να συμβεί over-fitting

Παράδειγμα... Δένδρα Αποφάσεων

- Πρόβλημα κατηγοριοποίησης: δοσμένου ενός συνόλου δεδομένων εκπαίδευσης (N εγγραφές, με M γνωρίσματα, και ένα γνώρισμα ετικέτας) βρες κανόνες για την πρόβλεψη της ετικέτας (κλάσης) μιας νέας εγγραφής
- Τεχνική: κατηγοριοποίηση με χρήση δένδρων
 - Φάση δημιουργίας - πολιτικές διαμερισμού (splitting policies)
 - Φάση περικοπής - για την αποφυγή υπερπροσαρμογής (over-fitting)

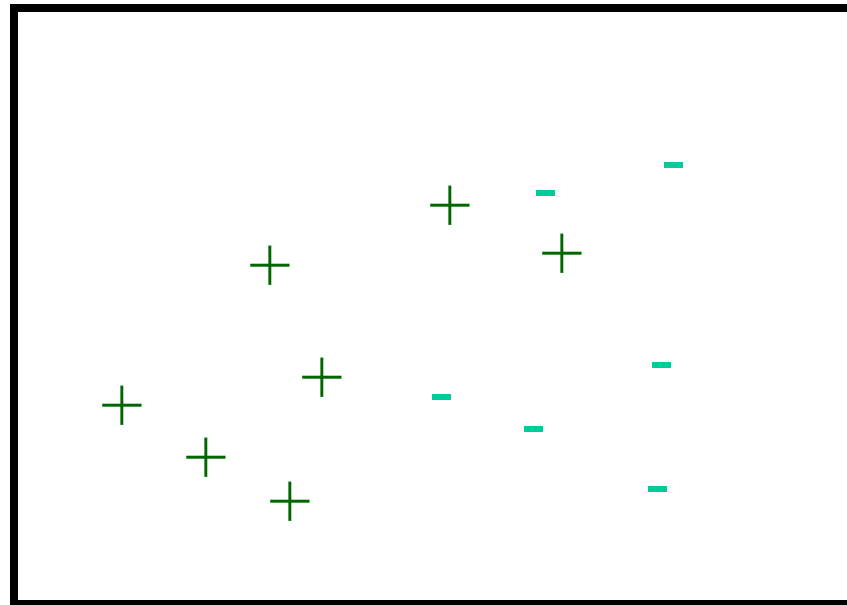
Δένδρα Απόφασης

Age	Chol-level	Gender	...	CLASS-ID
30	150	M		+
				...
				-
				??

Δένδρα Απόφασης

- Σχηματικά έχουμε:

num. attr#2
(e.g., chol-level)

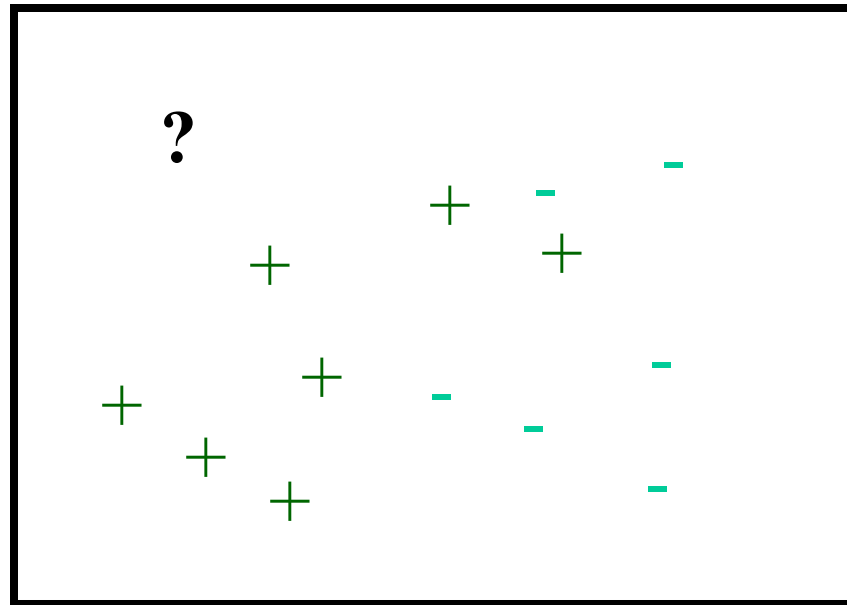


num. attr#1 (e.g., 'age')

Δένδρα Απόφασης

- Και θέλουμε να δώσουμε ετικέτα στο ‘?’

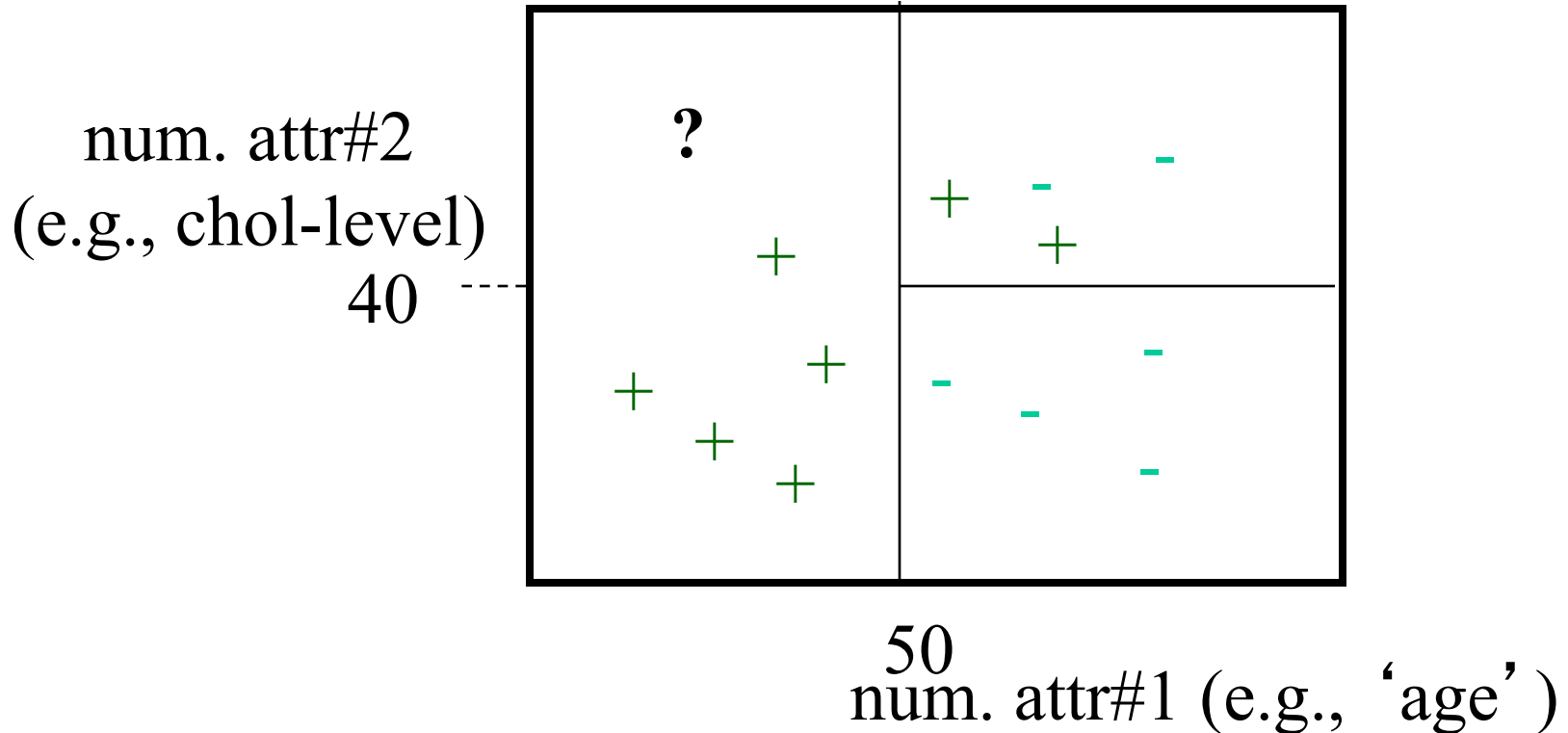
num. attr#2
(e.g., chol-level)



num. attr#1 (e.g., ‘age’)

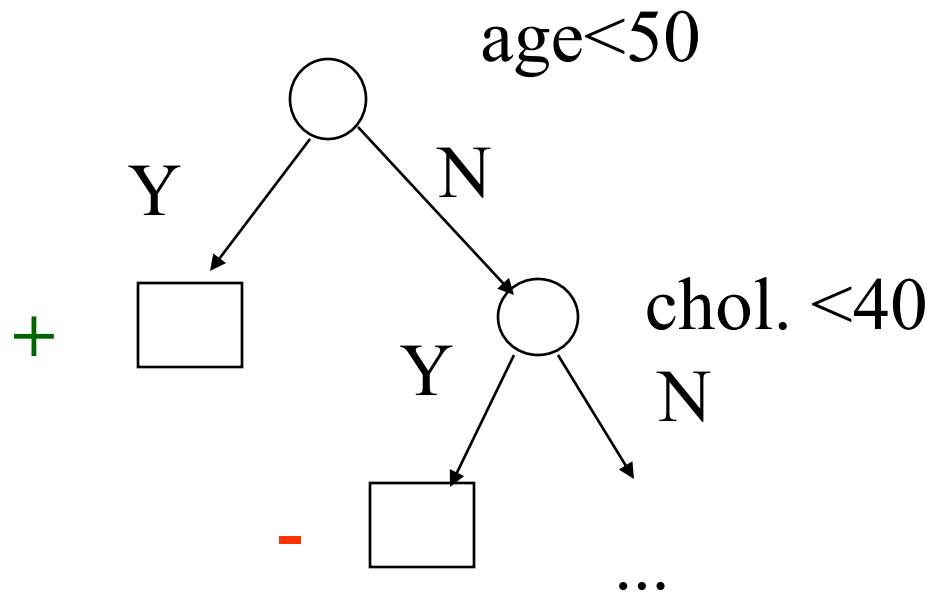
Δένδρα Απόφασης

- Οπότε δημιουργούμε το παρακάτω δένδρο:



Δένδρα Απόφασης

- Οπότε δημιουργούμε το παρακάτω δένδρο:

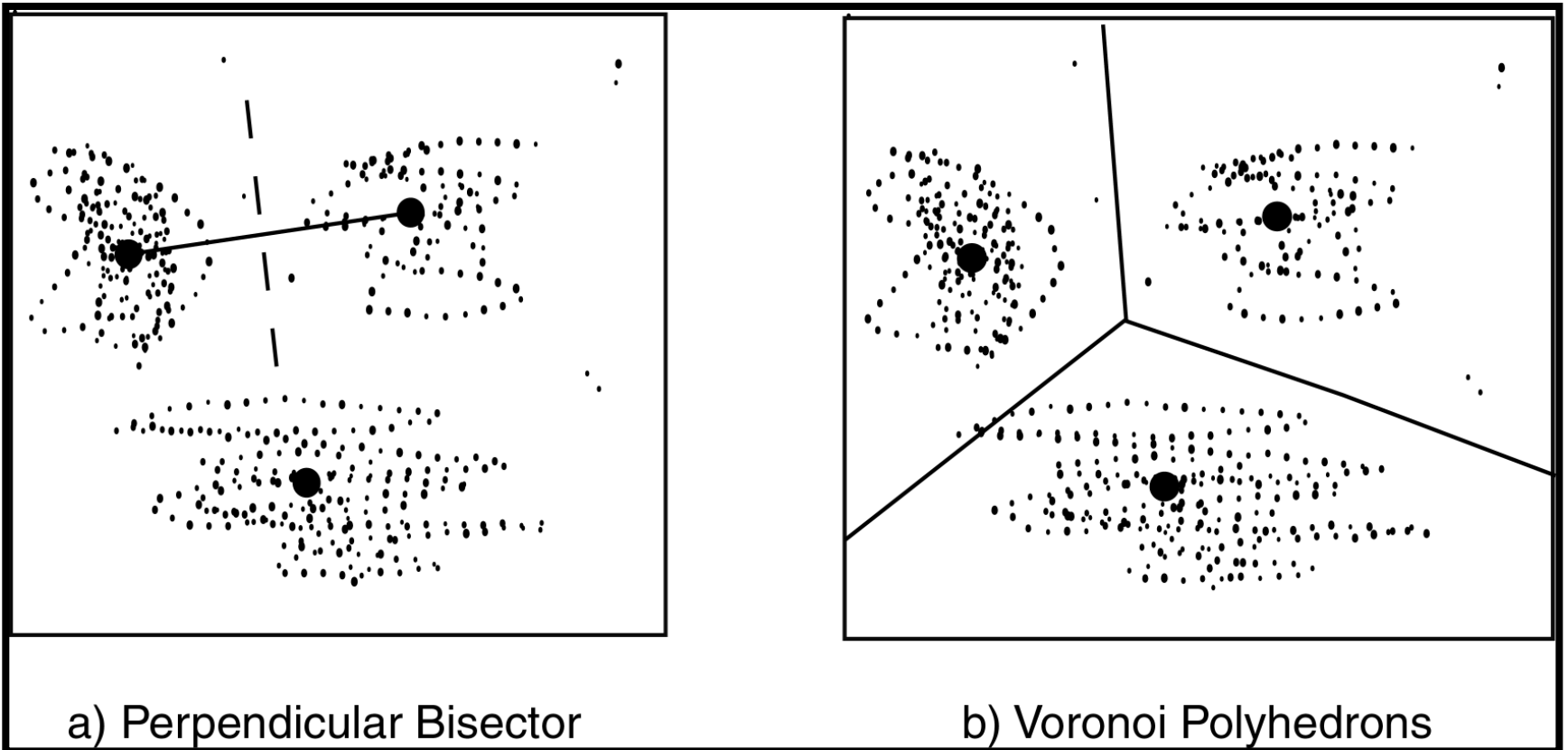


Συσταδοποίηση Δεδομένων

- ✓ Η εργασία του καταμερισμού ενός ετερογενούς πληθυσμού σε ένα σύνολο συστάδων (clusters).
- ✓ Στην συσταδοποίηση δεν υπάρχουν προκαθορισμένες κατηγορίες. Οι εγγραφές ομαδοποιούνται σε σύνολα με βάση την ομοιότητα που παρουσιάζουν μεταξύ τους.
- ✓ Αντιπροσωπευτικοί αλγόριθμοι: K-Means και παραλλαγές, PAM

Συσταδοποίηση Δεδομένων

- Διαγράμματα Voronoi



Κανόνες Συσχέτισης και Εξαγωγή τους

- ✓ Ένας συνοπτικός τρόπος για να εκφραστούν οι ενδεχομένως χρήσιμες πληροφορίες ώστε να γίνουν εύκολα κατανοητές.
- ✓ Ανακάλυψη κρυμμένων «συσχετίσεων» μεταξύ των γνωρισμάτων ενός συνόλου δεδομένων. Συσχετισμοί της μορφής: $A \rightarrow B$ όπου το A και το B αναφέρονται στα σύνολα γνωρισμάτων που υπάρχουν στα υπό ανάλυση δεδομένα
- ✓ Λαμβάνοντας υπόψη ένα σύνολο από n συναλλαγές S , κάθε υποσύνολο ενός συνόλου $A = \{A_1, \dots, A_k\}$, ένα κατώτατο όριο υποστήριξης T και ένα κατώτατο όριο εμπιστοσύνης s , παράγονται όλοι οι κανόνες $A \rightarrow B$.
- ✓ **Σύνολο αντικειμένων:** $I = \{I_1, I_2, \dots, I_m\}$
- ✓ **Συναλλαγές:** $D = \{t_1, t_2, \dots, t_n\}$, $t_j \subseteq I$
- ✓ **Συνολοστοιχεία:** $\{I_{i1}, I_{i2}, \dots, I_{ik}\} \subseteq I$
- ✓ **Υποστήριξη συνολοστοιχείου:** ποσοστό συναλλαγών που περιέχουν συνολοστοιχείο.
- ✓ **Μεγάλα (Συχνά) συνολοστοιχεία:** συνολοστοιχεία ο αριθμός εμφανίσεων των οποίων είναι πάνω από κάποιο όριο.

Κανόνες Συσχέτισης

- ✓ *Κανόνας Συσχέτισης*: συνεπαγωγή $X \Rightarrow Y$
όπου $X, Y \subseteq I$ και $X \cap Y = \emptyset$
- ✓ *Υποστήριξη Κ.Σ.* $X \Rightarrow Y$: ποσοστό συναλλαγών που περιέχουν το σύνολο $X \cup Y$, δηλαδή ίσο με $P(X \cup Y)$
- ✓ *Εμπιστοσύνη Κ.Σ.* $X \Rightarrow Y$: πηλίκο του αριθμού των συναλλαγών που περιέχουν το $X \cup Y$ προς αυτά που περιέχουν το X , δηλαδή ίσο με $P(Y/X)$

Παράδειγμα Market Basket



- ? Σε ποιο μέρος στο Supermarket πρέπει να βάλουμε τα απορυπαντικά για να αυξήσουμε τις πωλήσεις τους;
- ? Αγοράζονται προϊόντα καθαρισμού παραθύρων όταν απορυπαντικά και πορτοκαλάδα αγοράζονται μαζί;
- ? Τα αναψυκτικά αγοράζονται συνήθως μαζί με τις μπανάνες; Κάνει διαφορά η μάρκα του αναψυκτικού;
- ? Πως επηρεάζουν τα demographics της περιοχής τα προϊόντα που αγοράζουν οι καταναλωτές;

Κανόνες Συσχέτισης

- Ανάλυση Market Basket: Τα αποτελέσματα εκφράζονται σαν Κανόνες Συσχέτισης
- Βρείτε όλους τους κανόνες $X \rightarrow Y$ που συσχετίζουν την εμφάνιση ενός συνόλου από αντικείμενα X με ένα άλλο σύνολο αντικειμένων Y
 - Παράδειγμα: Όταν ένας πελάτης αγοράζει ψωμί και βούτυρο, αγοράζει και γάλα σε 85% των περιπτώσεων



+



Αποτελέσματα: Χρήσιμα, Τετριμμένα ή Ανεξήγητα

- Ενώ οι κανόνες συσχέτισης είναι εύκολο να κατανοηθούν δεν είναι πάντα χρήσιμοι

Χρήσιμο: Τις Παρασκευές οι πελάτες ενός SuperMarket συχνά αγοράζουν diapers και μύρες μαζί.

Τετριμμένο: Πελάτες που αγοράζουν συμφωνίες συντήρησης είναι πολύ πιθανό να αγοράσουν μεγάλες οικιακές συσκευές.

Ανεξήγητο: Όταν ένα καινούργιο Super Store ανοίγει, ένα από τα πιο συχνά αγοραζόμενα προϊόντα είναι οι λάμπες.

Πως δουλεύει;

Transactions στο SuperMarket	
Πελάτης	Προϊόντα
1	Orange Juice, Soda
2	Milk, Orange Juice, Window Cleaner
3	Orange Juice, Detergent
4	Orange Juice, Detergent, Soda
5	Window Cleaner, Soda

Ομο-εμφάνιση προϊόντων					
	OJ	Window Cleaner	Milk	Soda	Detergent
OJ	4	1	1	2	1
Window Cleaner	1	2	1	1	0
Milk	1	1	1	0	0
Soda	2	1	0	3	1
Detergent	1	0	0	1	2

Πως δουλεύει; Πόσο καλοί είναι οι κανόνες;

- Ο πίνακας ομο-εμφάνισης περιέχει μερικά απλά πρότυπα:
 - Πορτοκαλάδα και αναψυκτικό είναι πολύ πιθανό να αγοραστούν μαζί από ότι άλλα προϊόντα
 - Απορυπαντικά δεν αγοράζονται μαζί με καθαριστικά τζαμιών ή γάλα
 - Γάλα δεν αγοράζεται μαζί με αναψυκτικό ή απορυπαντικό
- Αυτές οι παρατηρήσεις είναι παραδείγματα συχετίσεων που μπορεί να συνιστούν έναν τυπικό κανόνα:
 - Αν ένας πελάτης αγοράζει αναψυκτικό, ΤΟΤΕ αγοράζει και πορτοκαλάδα
- Στα δεδομένα, 2 από τις 5 αγορές (transactions) περιλαμβάνουν αναψυκτικό και πορτοκαλάδα. Αυτές οι δύο αγορές *υποστηρίζουν* τον κανόνα. Η **υποστήριξη** του είναι 2 στα 5 ή 40%.
- Και οι δύο αγορές που περιέχουν αναψυκτικό επίσης περιέχουν πορτοκαλάδα -> μεγάλος βαθμός *εμπιστοσύνης* στον κανόνα «αναψυκτικό -> πορτοκαλάδα». Κάθε αγορά που περιέχει αναψυκτικό επίσης περιέχει και πορτοκαλάδα. Ο κανόνας «αναψυκτικό -> πορτοκαλάδα» έχει **εμπιστοσύνη** 100%.



Λειτουργικότητες εξόρυξης δεδομένων— Πρότυπα που μπορούν να εξαχθούν

- Ανάλυση Outlier
 - Outlier: ένα αντικείμενο δεδομένων το οποίο δεν ακολουθεί την γενική συμπεριφορά των δεδομένων (μπορεί να εντοπισθεί με την χρήση στατιστικών ελέγχων που υιοθετούν ένα πιθανοτικό μοντέλο)
 - Συχνά θεωρείται θόρυβος αλλά είναι χρήσιμο στον εντοπισμό σφάλματος, την ανάλυση σπάνιων γεγονότων
- Ανάλυση τάσης και εξέλιξης (trend and evolution analysis)
 - Μελετά την κανονικότητα των αντικειμένων των οποίων η συμπεριφορά μεταβάλλεται με τον χρόνο
 - Τάση και απόκλιση: ανάλυση παλινδρόμησης (regression)
 - Εξόρυξη ακολουθιακού προτύπου, ανάλυση περιοδικότητας
 - Ανάλυση με βάση την ομοιότητα

Πότε είναι ενδιαφέρον ένα “Ανακαλυφθέν” πρότυπο;

- Ένα σύστημα εξόρυξης δεδομένων μπορεί να παράγει χιλιάδες πρότυπα, τα οποία δεν είναι όλα ενδιαφέροντα.
 - Προτεινόμενη προσέγγιση: Ανθρωποκεντρική, με βάση το ερώτημα, εστιασμένη εξόρυξη δεδομένων
- Μετρικές ενδιαφέροντος: Ένα πρότυπο είναι **ενδιαφέρον** εάν είναι εύκολα κατανοητό από τους ανθρώπους, έγκυρο σε νέα ή πειραματικά δεδομένα με κάποιο βαθμό βεβαιότητας, πιθανά χρήσιμο, καινούριο, ή επικυρώνει κάποια υπόθεση που ο χρήστης επιδιώκει να επαληθεύσει
- Αντικειμενικές vs. υποκειμενικές μετρικές ενδιαφέροντος:
 - Αντικειμενικές: βασίζονται σε στατιστικές και δομές προτύπων, π.χ., confidence and support
 - Υποκειμενικές: βασίζονται στην γνώμη του χρήστη για τα δεδομένα, π.χ., unexpectedness, actionability, novelty, κλπ

Μπορούμε να βρούμε όλα τα ενδιαφέροντα πρότυπα;

- Εύρεση όλων των ενδιαφερόντων προτύπων: Πληρότητα
 - Μπορεί ένα σύστημα εξόρυξης δεδομένων να βρει όλα τα ενδιαφέροντα πρότυπα;
 - Συσχέτιση (association) vs. Ταξινόμηση (classification) vs. Ομαδοποίηση (clustering)
- Αναζήτηση μόνο των ενδιαφερόντων προτύπων: Βελτιστοποίηση
 - Μπορεί ένα σύστημα εξόρυξης δεδομένων να βρει μόνο τα ενδιαφέροντα πρότυπα;
 - Προσεγγίσεις
 - Πρώτα παρήγαγε όλα τα πρότυπα και στην συνέχεια φιλτράρισε αυτά που δεν μας ενδιαφέρουν
 - Παρήγαγε μόνο τα ενδιαφέροντα πρότυπα - βελτιστοποίηση ερωτήματος

Χρονολογικές Σειρές

- Χρονολογική σειρά: μια ακολουθία αριθμών κάθε ένας από τους οποίους έχει ένα **timestamp** (ετικέτα χρόνου). Χαρακτηριστικά υποθέτουμε ότι οι διαδοχικοί αριθμοί χωρίζονται από ένα σταθερό χρονικό διάστημα, και το πραγματικό timestamp παραλείπεται.
- Διάφορες φυσικές διαδικασίες παράγουν δεδομένα υπό μορφή χρονολογικών σειρών (λ.χ. εφαρμογές στον οικονομικό και περιβαλλοντικό τομέα).

Χρονολογικές Σειρές

- ✓ Μοντελοποίηση χρονικών γεγονότων
- ✓ Ανάλυση χρονολογικών σειρών
- ✓ Ανάλυση τάσεων
- ✓ Μετασχηματισμός
- ✓ Ομοιότητα
- ✓ Πρόβλεψη
- ✓ Αλγόριθμοι παραγωγής κανόνων χρονικών συσχετίσεων

Ομοιότητα Χρονολογικών Σειρών

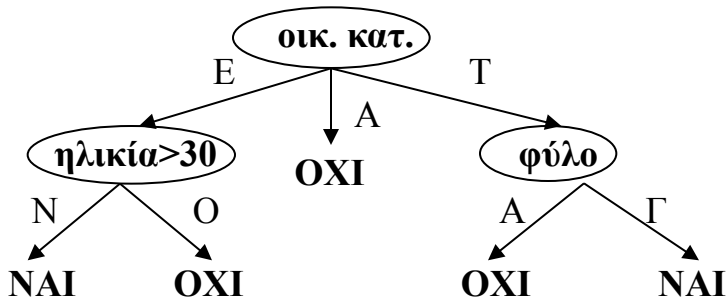
- Μια ρεαλιστική συνάρτηση απόστασης που θα ταιριάζει με την αντίληψη του χρήστη για το τι αυτός θεωρεί παρόμοιο.
- Ένα αποδοτικό σχέδιο indexing, το οποίο θα επιταχύνει τις ερωτήσεις των χρηστών.

Η **p-norm** απόσταση μεταξύ δύο n -διάστατων διανυσμάτων \mathbf{X} και \mathbf{Y} ορίζεται ως εξής:

$$L_p(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Παραδείγματα

Δέντρα Αποφάσεων

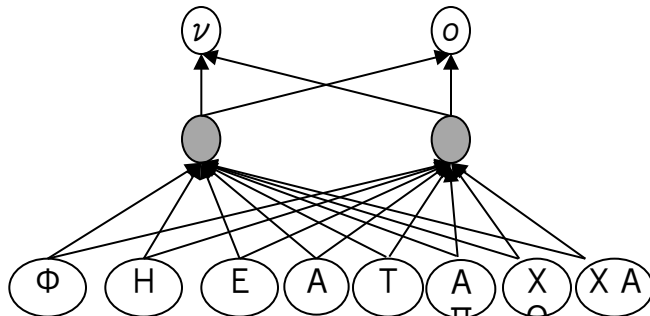


Κανόνες

$(\text{Ηλικία} \leq 30) \ \& \ (\text{Οικ:Τ}) \rightarrow (\text{Φύλο:Γ})$

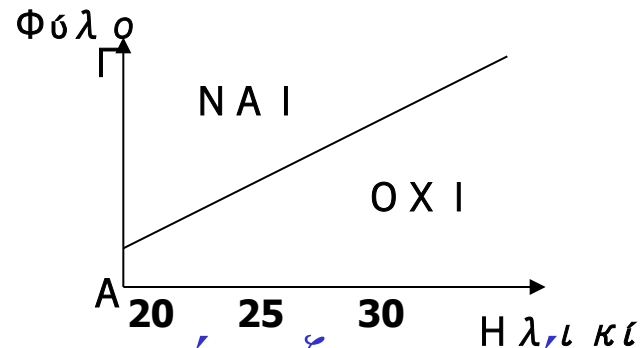
$(\text{Ηλικία} \leq 30) \ \& \ (\text{Οικ:Ε}) \rightarrow (\text{Φύλο:Γ})$

...



Νευρωνικά δίκτυα

Β. Μεγαλοικονόμου

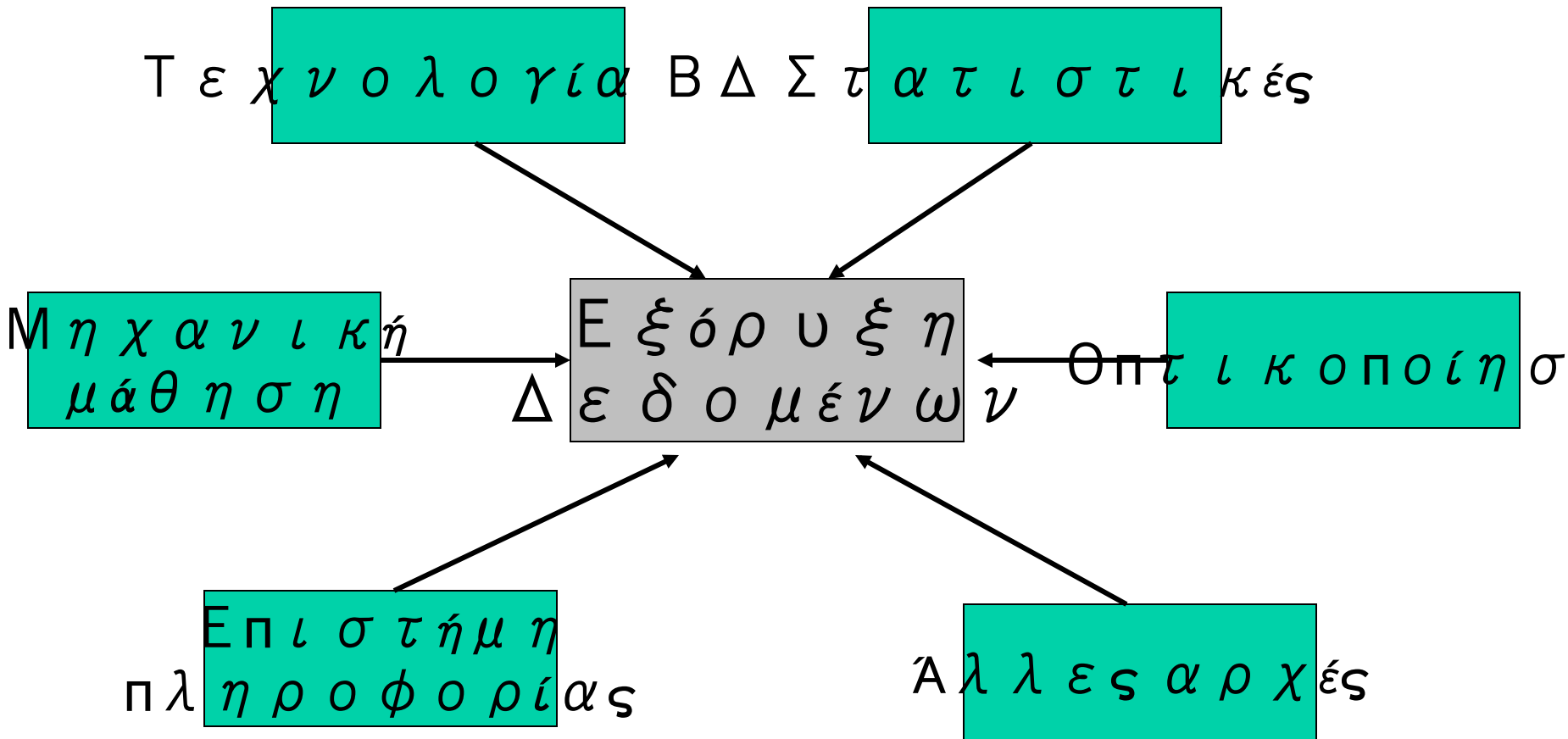


Στατιστικοί ταξινομητές

Γειτονικές Περιοχές

- ✓ Ανάκτηση Πληροφορίας
- ✓ Βάσεις Δεδομένων
- ✓ Τεχνητή Νοημοσύνη
- ✓ Ευφυής Υπολογισμός
- ✓ Στατιστική Ανάλυση Δεδομένων

Εξόρυξη δεδομένων: Συμβολή πολλαπλών αρχών



Data Mining Development

- Relational Data Model
- SQL
- Association Rule Algorithms
- Data Warehousing
- Scalability Techniques

Databases

Data Mining

```
graph TD; Databases --> DM[Data Mining]; IR[Information Retrieval] --> DM; Statistics --> DM; ML[Machine Learning] --> DM; Algorithms --> DM;
```

Information Retrieval

- Similarity Measures
- Hierarchical Clustering
- IR Systems
- Imprecise Queries
- Textual Data
- Web Search Engines

Statistics

- Bayes Theorem
- Regression Analysis
- EM Algorithm
- K-Means Clustering
- Time Series Analysis

Machine Learning

- Neural Networks
- Decision Tree Algorithms

Algorithms

- Algorithm Design Techniques
- Algorithm Analysis
- Data Structures

Είδη Μηχανικής Μάθησης

- **Μάθηση υπό επίβλεψη ή από παραδείγματα:**
 - Μάθηση εννοιών (Concept Learning)
 - Δέντρα Ταξινόμησης ή Απόφασης (Classification or Decision Trees)
 - Μάθηση Κανόνων (Rule Learning)
 - Μάθηση κατά Bayes
 - Γραμμική Παλινδρόμηση/Παρεμβολή (Linear Regression)
 - Νευρωνικά Δίκτυα (Neural Networks)
 - Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)
- **Μάθηση χωρίς επίβλεψη ή από παρατήρηση:**
 - Εξαγωγή Ομάδων/Συστάδων (clustering)
 - Εξαγωγή συσχετίσεων (association rules)

Επιστημονικά Συνέδρια και Περιοδικά σε Εξόρυξη Δεδομένων

- KDD Conferences
 - ACM SIGKDD Int. Conf. on Knowledge Discovery in Databases and Data Mining (KDD)
 - SIAM Data Mining Conf. (SDM)
 - (IEEE) Int. Conf. on Data Mining (ICDM)
 - Conf. on Principles and practices of Knowledge Discovery and Data Mining (PKDD)
 - Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD)
- Other related conferences
 - ACM SIGMOD
 - VLDB
 - (IEEE) ICDE
 - WWW, SIGIR
 - ICML, CVPR, NIPS
- Journals
 - Data Mining and Knowledge Discovery (DAMI or DMKD)
 - IEEE Trans. On Knowledge and Data Eng. (TKDE)
 - KDD Explorations
 - ACM Trans. on KDD

Που μπορείτε να βρείτε Αναφορές; DBLP, CiteSeer, Google

- Data mining and KDD (SIGKDD: CDROM)
 - Conferences: ACM-SIGKDD, IEEE-ICDM, SIAM-DM, PKDD, PAKDD, etc.
 - Journal: Data Mining and Knowledge Discovery, KDD Explorations, ACM TKDD
- Database systems (SIGMOD: ACM SIGMOD Anthology—CD ROM)
 - Conferences: ACM-SIGMOD, ACM-PODS, VLDB, IEEE-ICDE, EDBT, ICDT, DASFAA
 - Journals: IEEE-TKDE, ACM-TODS/TOIS, JIIS, J. ACM, VLDB J., Info. Sys., etc.
- AI & Machine Learning
 - Conferences: Machine learning (ML), AAAI, IJCAI, COLT (Learning Theory), CVPR, NIPS, etc.
 - Journals: Machine Learning, Artificial Intelligence, Knowledge and Information Systems, IEEE-PAMIetc
- Web and IR
 - Conferences: SIGIR, WWW, CIKM, etc.
 - Journals: WWW: Internet and Web Information Systems,
- Statistics
 - Conferences: Joint Stat. Meeting, etc.
 - Journals: Annals of statistics, etc.
- Visualization
 - Conference proceedings: CHI, ACM-SIGGraph, etc.
 - Journals: IEEE Trans. visualization and computer graphics, etc.

Μέρος Β: Επιμέρους Εφαρμογές

K-means - Vector Quantization

- VQ: a generalization of scalar quantization (discretization) to quantization of a vector
- A vector quantizer Q of dimension k and size N is a mapping from a vector (or a “point” in \mathbb{R}^k), into a finite set $C = \{y_1, y_2, \dots, y_N\}$, $y_i \in \mathbb{R}^k$, the codebook of size N

$$Q: \mathbb{R}^k \rightarrow C$$

- It partitions \mathbb{R}^k into N regions or cells, R_i for $i \in J \equiv \{1, 2, \dots, N\}$

$$R_i = \{x \in \mathbb{R}^k: Q(x) = y_i\}$$

Vector Quantization - Σχεδίαση

- The goal is to find:
 - A codebook (decoder) – representation levels
 - A partition rule (encoder) – decision levels
- To maximize an overall measure of performance

VQ Design – Συνθήκες

Nearest Neighbor Condition

- For a given codebook, C , the optimal regions $\{R_i: i=1, \dots, N\}$ satisfy the condition:

$$R_i \subset \{x: d(x, y_i) \leq d(x, y_j); \forall j\}$$

That is

$$Q(x)=y_i \text{ only if } d(x, y_i) \leq d(x, y_j) \forall j$$

VQ Design – Συνθήκες

Centroid Condition

- For given partition regions $\{R_i: i=1, \dots, N\}$ the optimal codewords satisfy the condition:

$$y_i = \text{cent}(R_i)$$

For the SE measure, the centroid of a set R is the arithmetic average:

$$\text{cent}(R) = \frac{1}{|R|} \sum_{i=1}^{|R|} x_i$$

for $R = \{x_i : i = 1, \dots, |R|\}$

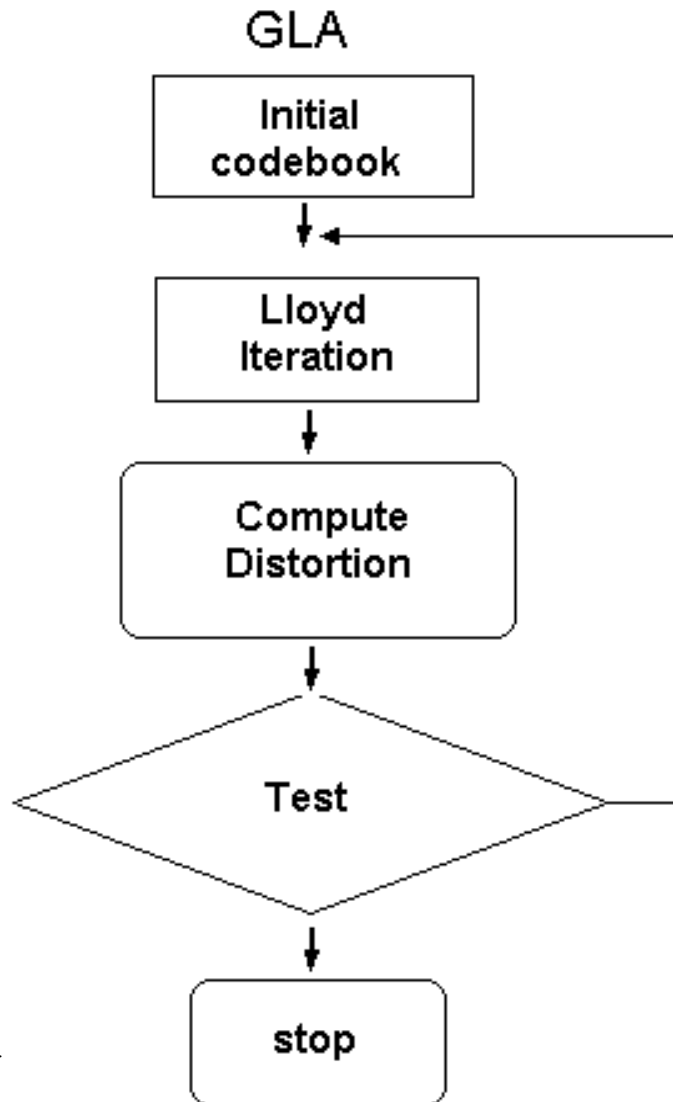
VQ Design – The Generalized Lloyd Algorithm (GLA)

- It produces a locally optimal codebook from a training sequence, T . It starts with an initial codebook and iteratively improves it.

Lloyd Iteration

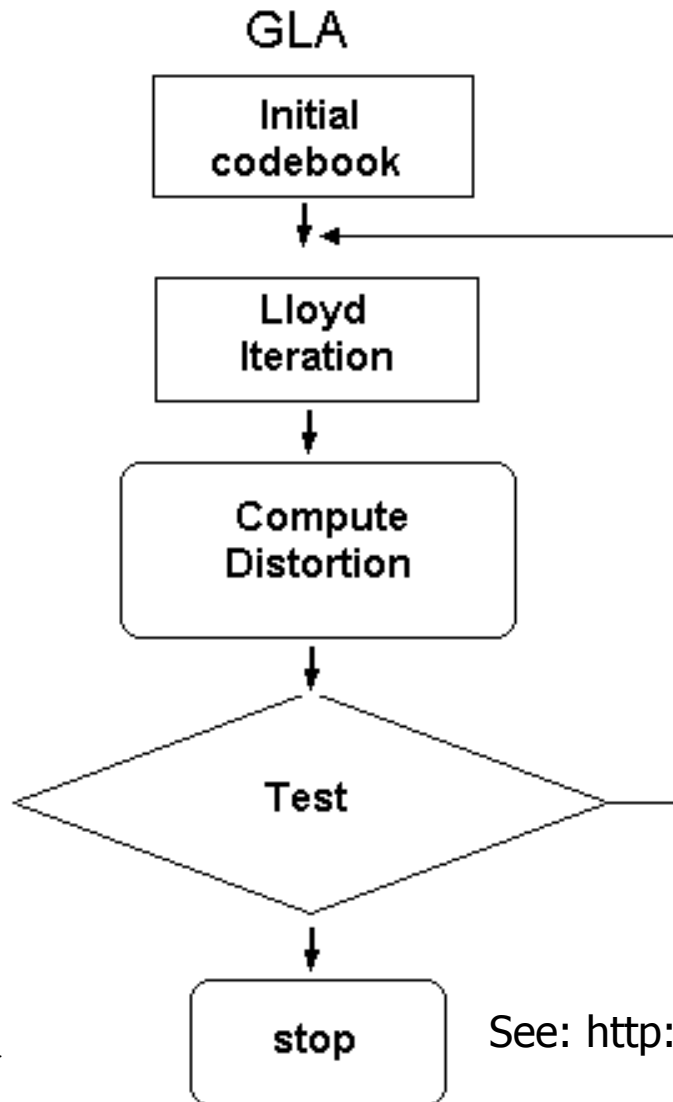
- Given a codebook $C_m = \{y_i\}$ generate an improved codebook C_{m+1} as follows:
 - Partition T into cells R_j using the Nearest Neighbor Condition:
$$R_i = \{x : d(x, y_i) \leq d(x, y_j); \forall j \neq i\}$$
 - Using the Centroid Condition compute the centroids of the cells just found to obtain the new codebook, $C_{m+1} = \{\text{cent}(R_i)\}$
- Compute the average distortion for C_{m+1} , D_{m+1} . If the fractional drop $(D_m - D_{m+1}) / D_m$ is below a certain threshold stop, else continue with $m \leftarrow m+1$

VQ Design – The Generalized Lloyd Algorithm (GLA)



To solve the initial codebook generation problem a partition split mechanism is used. How?

VQ Design – The Generalized Lloyd Algorithm (GLA)



To solve the initial codebook generation problem a partition split mechanism is used. How?

Starts with a codebook containing only one codeword (which one?) In each repetition and before the application of the Lloyd iteration, it doubles the number of codewords from the previous iteration

B.

See: <http://www.data-compression.com/vq.html#animation>

Παραδείγματα Εφαρμογών Εξόρυξης Βιοιατρικών Δεδομένων

Εξόρυξη Βιοιατρικών Δεδομένων

- Πειραματικές μέθοδοι στη βιοιατρική έχουν οδηγήσει σε μαζικά σύνολα δεδομένων - Επείγουσα η ανάγκη για αποδοτικές μεθόδους επεξεργασίας και ανάλυσης των δεδομένων για την κατανόηση τους
- Οι τομείς εφαρμογής περιλαμβάνουν:
 - υπολογιστική νευροεπιστήμη, μελέτες γονιδιώματος, καρδιολογία, διάγνωση καρκίνου, νευρολογία, κ.λπ.
- Παραδείγματα των εξεταζόμενων συνόλων δεδομένων περιλαμβάνουν:
 - Ακτινολογικά δεδομένα (MRI, fMRI)
 - Δεδομένα EEG, MEG
 - ακολουθίες γονιδιώματος,
 - στοιχεία έκφρασης γονιδίων,
 - πρωτεϊνικά δομικά στοιχεία,
 - στοιχεία αλληλεπίδρασης πρωτεϊνών,
 - στοιχεία αλληλεπίδρασης πρωτεϊνών-DNA, κλπ

Βιοιατρικά Σύνολα Δεδομένων

- Ελλιπή (ελλιπείς τιμές παραμέτρων)
- Λανθασμένα (συστηματικός ή τυχαίος θόρυβος στα στοιχεία)
- Αραιά (λίγες ή/και μη-αντιπροσωπευτικές εγγραφές ασθενών που είναι διαθέσιμες)
- Ανακριβή (ακατάλληλη επιλογή παραμέτρων για το δεδομένο στόχο)
- Περιλαμβάνουν μετρήσεις των παραμέτρων σε διαφορετικές χρονικές στιγμές

Εξόρυξη Βιοιατρικών Δεδομένων

- Στη βιοιατρική, η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί για την επίλυση διαφόρων στόχων:
 - Περιγραφικός στόχος: εύρεση ενδιαφέροντων προτύπων στα στοιχεία, ενδιαφεροντων συστάδων, κανόνων συσχέτισης, κ.λπ.
 - Προβλεπτικός στόχος: δημιουργώντας ένα προβλεπτικό μοντέλο που εξηγεί τα στοιχεία και μπορεί να χρησιμοποιηθεί για την πρόβλεψη ή την ταξινόμηση των απαρατήρητων περιπτώσεων
 - Χρησιμοποιούνται συχνά για την δημιουργία προτύπων ταξινόμησης που χρησιμοποιούνται για:
 - Διάγνωση
 - Πρόγνωση
 - Προγραμματισμός θεραπείας

Κατηγοριοποίηση Όγκων στον Ανθρώπινο Εγκέφαλο

- Διάκριση μεταξύ καλοήθης (meningiomas) και κακοήθης όγκου (υποβοήθηση της ιατρικής διάγνωσης)
- Θεραπεία και πρόγνωση τελείως διαφορετική για τις δύο κατηγορίες
- Χρήση δεδομένων από πολλαπλά modalities (CT and conventional MR, MRS)
- Συνδυασμός χαρακτηριστικών γνωρισμάτων που προέρχονται από εικόνες και άλλα δεδομένα (ενσωμάτωση του μεταβολικού προφίλ των όγκων)

Κατηγοριοποίηση Όγκων στον Ανθρώπινο Εγκέφαλο

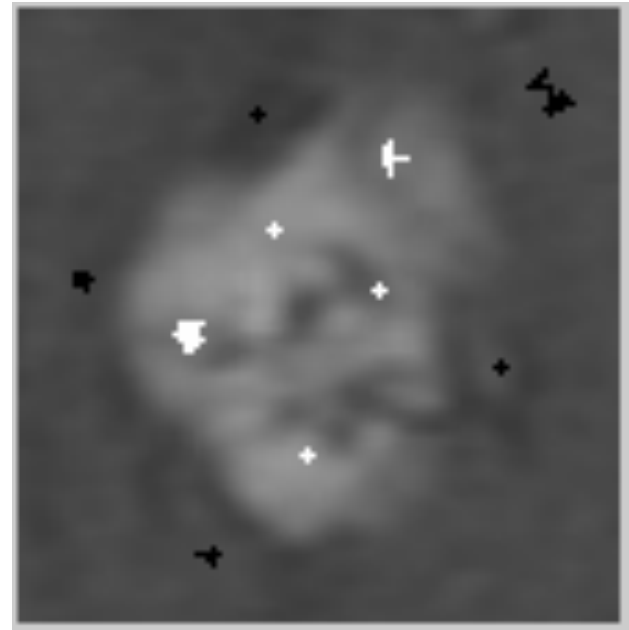
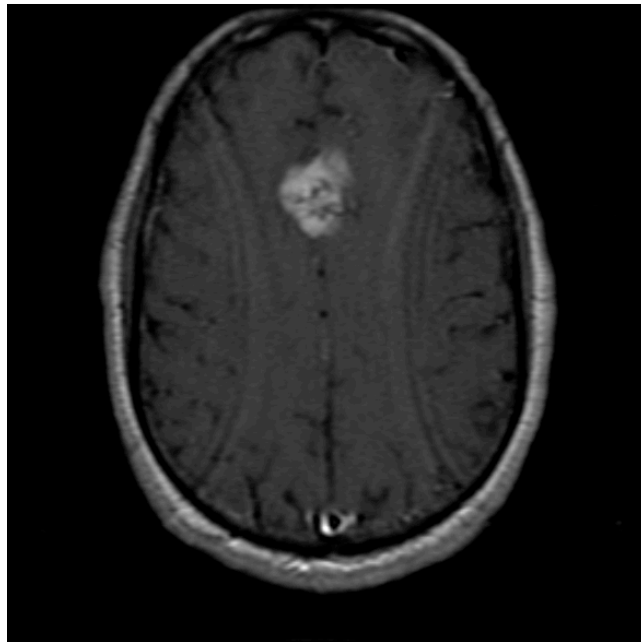
- Βήματα που συνήθως εμπλέκονται:
 - Segmentation of ROIs from MR images
 - ROI feature extraction
 - Fusion of MR image and MRS features
 - Informative feature selection
 - Classification model construction
- Προκλήσεις:
 - Uncertainty in identifying inter-tissue boundaries
 - Fusion of information from multiple modalities
 - Limited number of samples
 - Easy interpretability of classification model

Ημιαυτόματη κατάτμηση της περιοχής ενδιαφέροντος

- Χρήση της έννοιας του fuzzy-connectedness (Rosenfeld, 1979)
- Assign to every pair of voxels, x , y , in the image, a real number between 0 and 1 (the fuzzy-connectedness of x to y) denoting the strength of the link between x and y .
- The algorithm starts with a given set of seed points (more effective using some prior knowledge)
- Each point is assigned to the structure having a neighboring point with the highest fuzzy-connectedness value.
- The strongest connection is detected first, and the process repeats until the weakest connection is calculated. A sequence of points is called a **chain**, where its links are the ordered pairs of consecutive points in the sequence. The **strength of a chain** is the length of its weakest link.
- The segmented object is defined by the number of points that are connected through a chain to the selected seed point of the object.

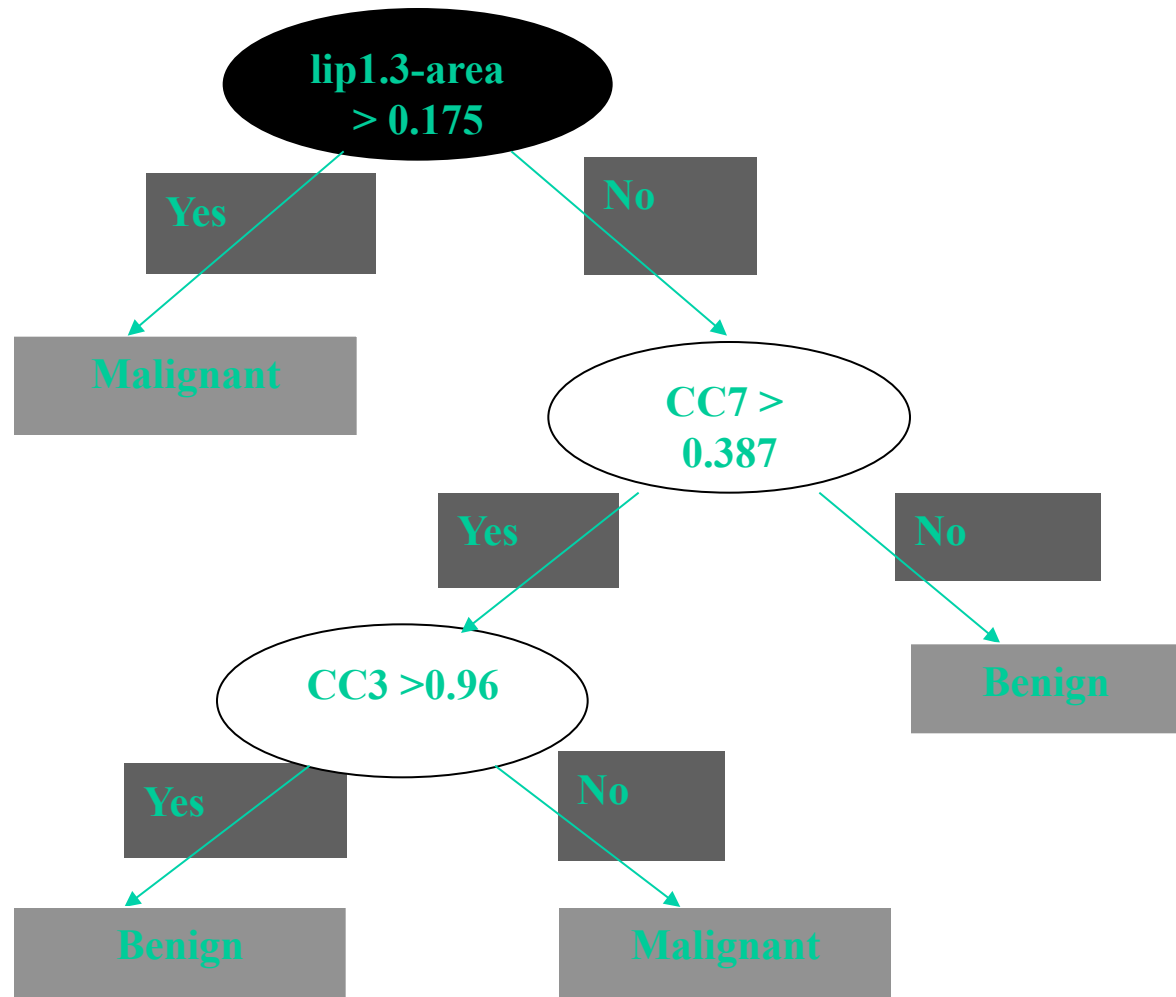
Ημιαυτόματη κατάτμηση της περιοχής ενδιαφέροντος

Ημιαυτόματη κατάτμηση σε λειτουργία



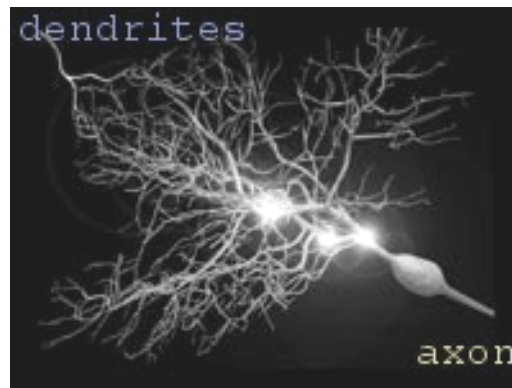
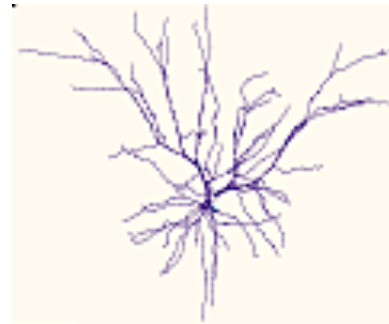
Μοντέλο κατηγοριοποίησης όγκων του ανθρώπινου εγκεφάλου

Δένδρο
Απόφασης

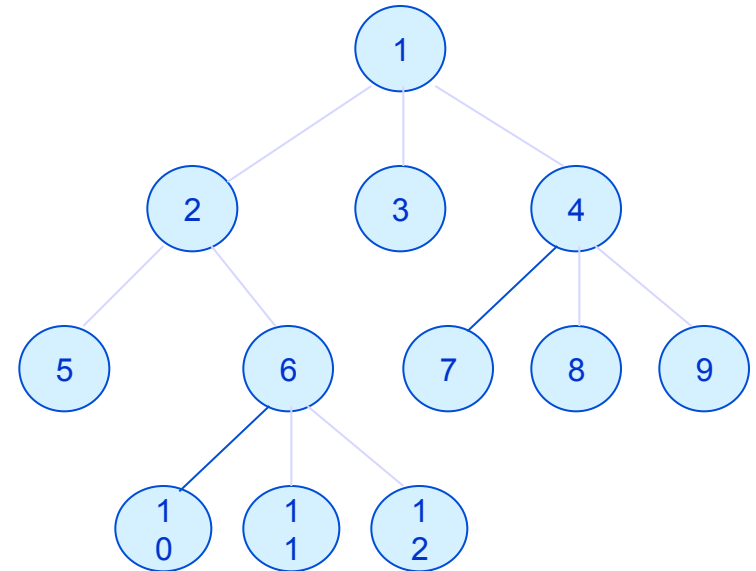
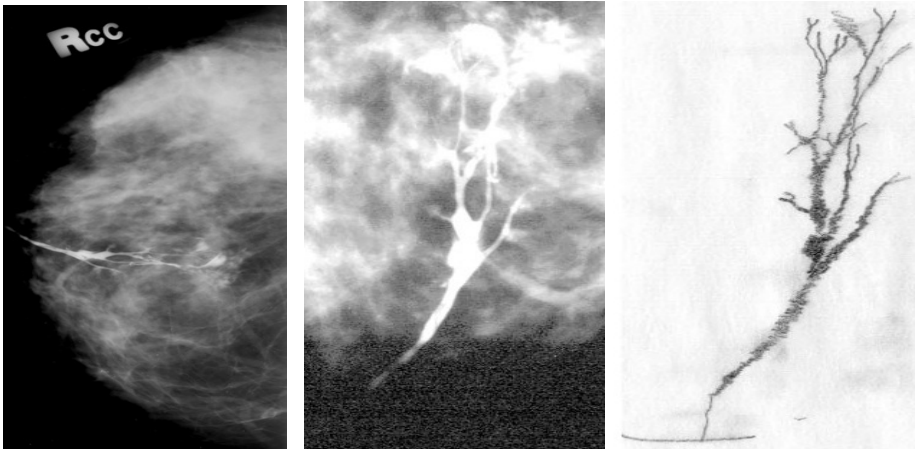


Ανάλυση Δενδρικών Δομών

- Τι είδους δενδρικές δομές;



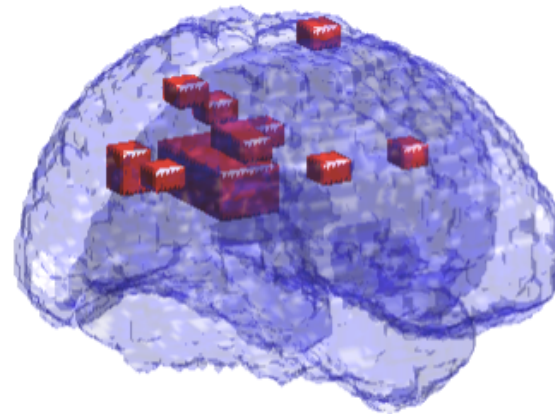
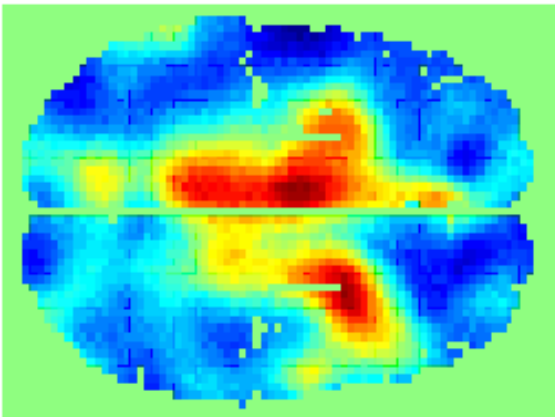
Ανάλυση μαστογραφίας, γαλακτογραφίας



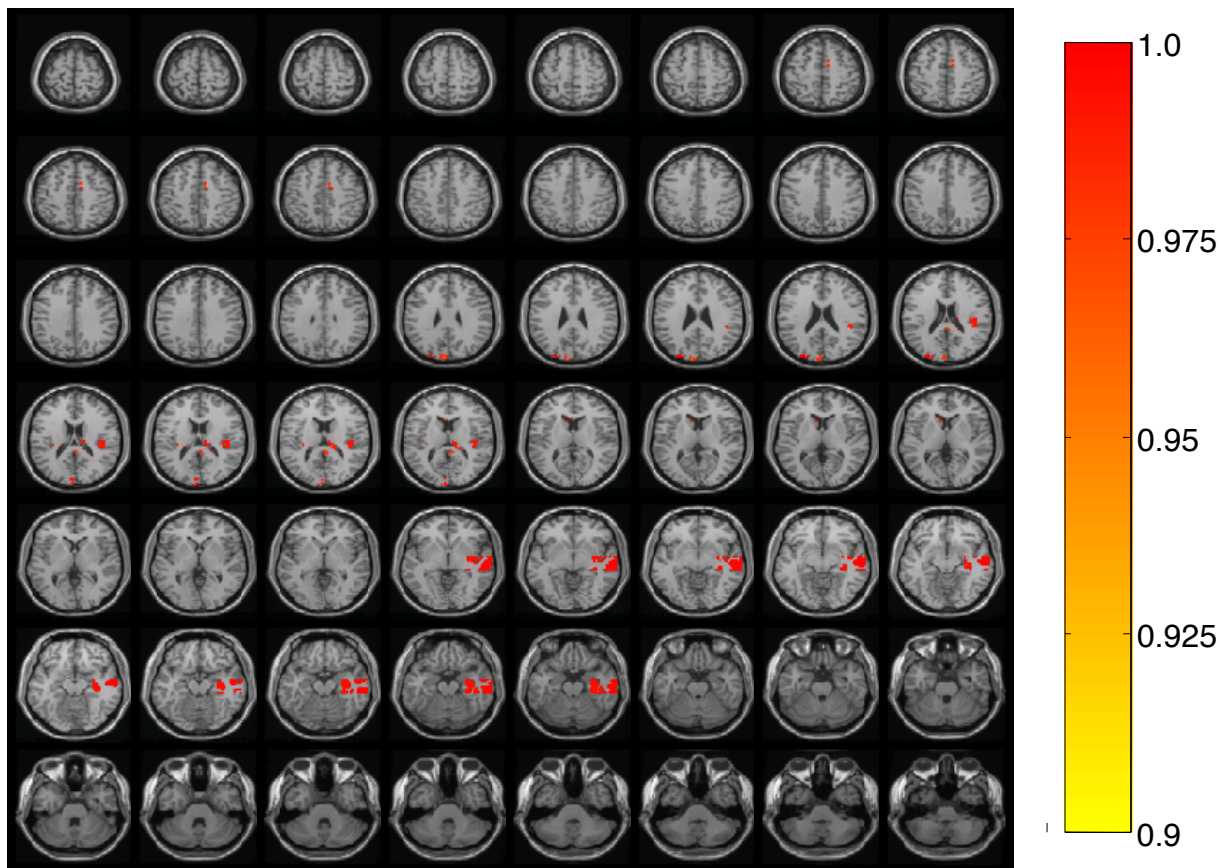
- Apply tree pattern analysis techniques to medical image data → galactograms
- Analysis of the ductal network

Εξόρυξη δεδομένων από fMRI

- Δημιουργία αποδοτικών και robust εργαλείων πληροφορικής για τις εικόνες του εγκεφάλου:
 - για να καταλάβουμε πώς λειτουργεί ο εγκέφαλος
 - για να βοηθήσουμε στην κλινική ερμηνεία των εικόνων
 - για να διευκολύνουμε την προόδο στην έγκαιρη διάγνωση και τη θεραπεία



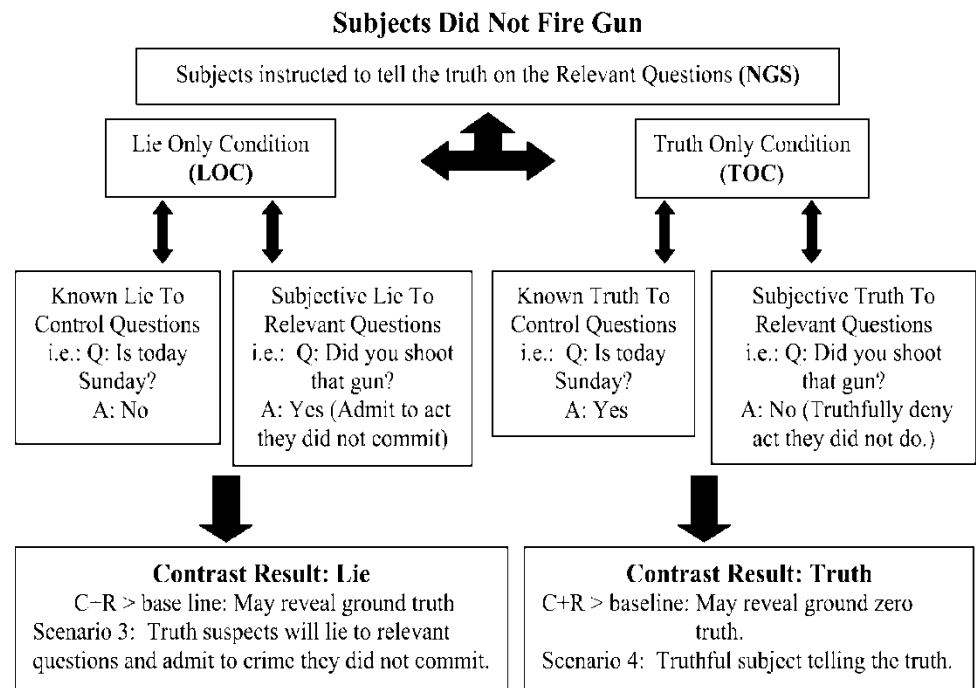
Πρώιμη Ανίχνευση της νόσου Alzheimer



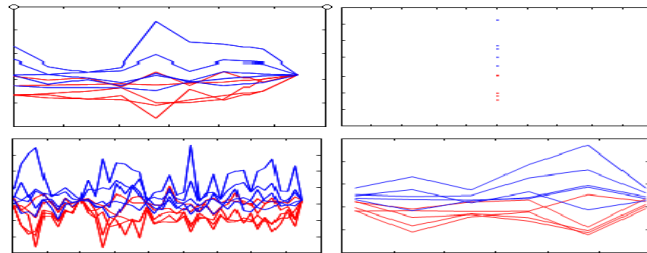
For a semantic memory test major differences were seen in the right posterior parietal and temporal lobe regions (colorbar shows significance)

Εφαρμογή τεχνικών εξόρυξης σε fMRI για ανίχνευση ψεύδους

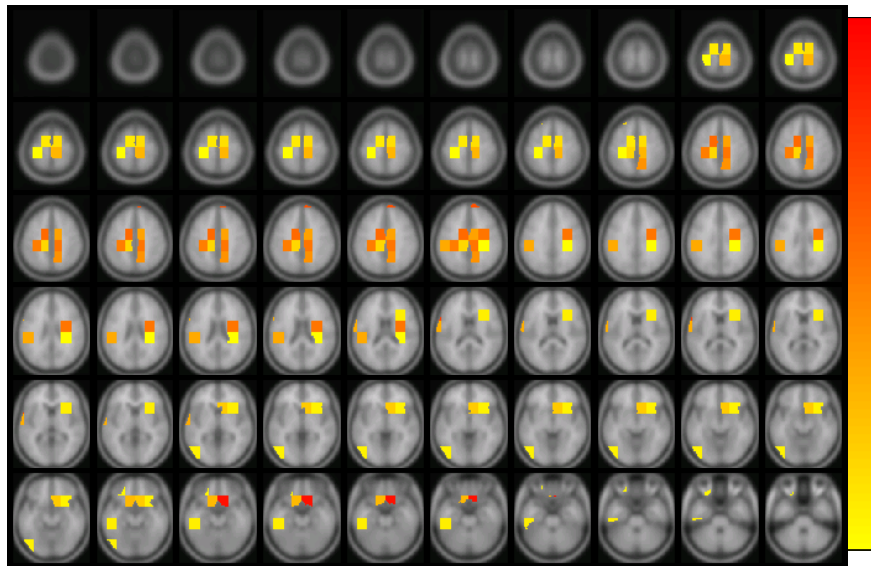
- 11 subjects μοιρασμένα σε δύο ομάδες.
- Πληροφορία fMRI συγκεντρώθηκε από δύο ομάδες σύμφωνα με τις απαντήσεις τους σε ερωτήσεις που τους τέθηκαν.
- Στόχος: ανάλυση των εικόνων για εύρεση περιοχών που μπορούν να διακρίνουν αν κάποιος λέει αλήθεια ή ψέματα



Εφαρμογή τεχνικών εξόρυξης σε fMRI για ανίχνευση ψεύδους



Activation values for the regions detected by DRP as discriminative (blue is for the guilty group, red the non-guilty group)



Colored areas are discriminative regions found by DRP with p -value = 0.05 and a maximum splitting level of 3

- Οι διακριτικές περιοχές που βρέθηκαν είναι στο frontal, parietal, and temporal lobes καθώς και στο deeper limbic system.
- Ο αριστερός medial inferior frontal lobe βρέθηκε σαν η περιοχή που ξεχωρίζει καλύτερα τις δύο ομάδες

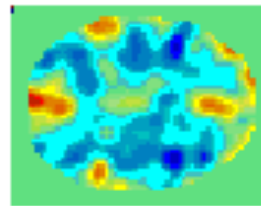
Parameters	Regions	Accuracy
p -value=.05, depth=4	56	97%
p -value=.01, depth=3	1	100%
p -value=.01, depth=4	6	97%

Classification accuracy for different parameters of DRP

Έυρεση ομοιότητας Ιατρικών Εικόνων

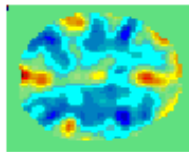
or ... Content-Based
Medical Image
Retrieval

Query Image:

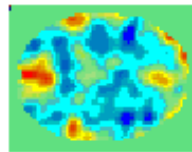


A-1-26

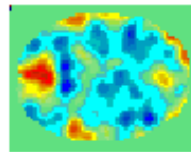
Top 6 matches:



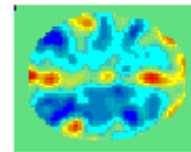
A-1-27



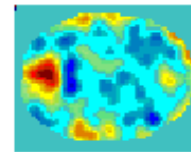
A-1-25



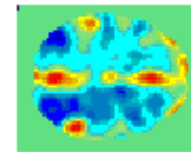
A-1-24



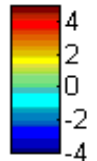
A-1-28



A-1-23



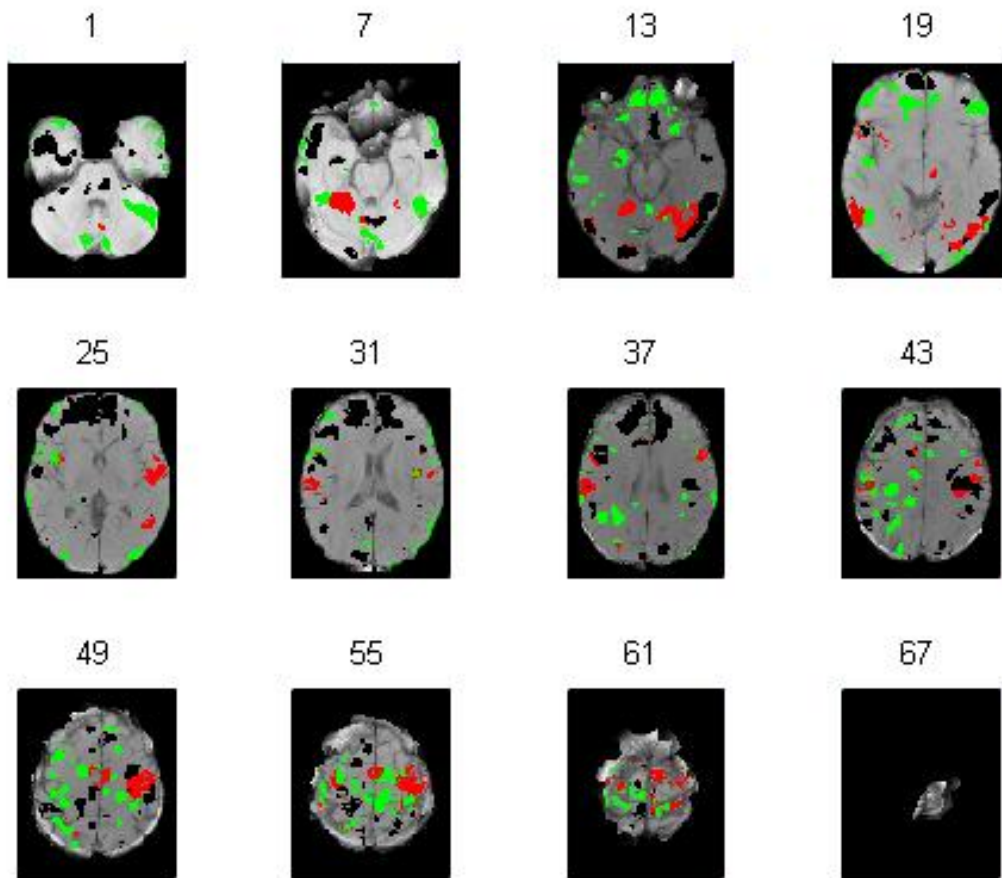
A-1-29



Ανάλυση και Συσταδοποίηση χρονοσειρών από fMRI

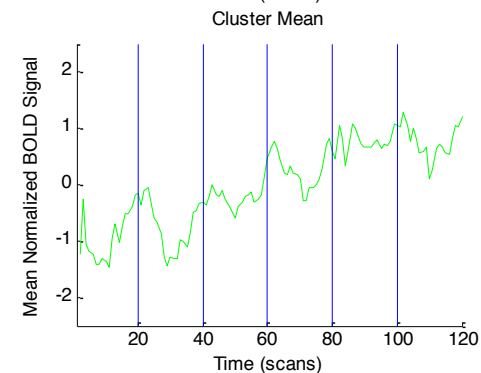
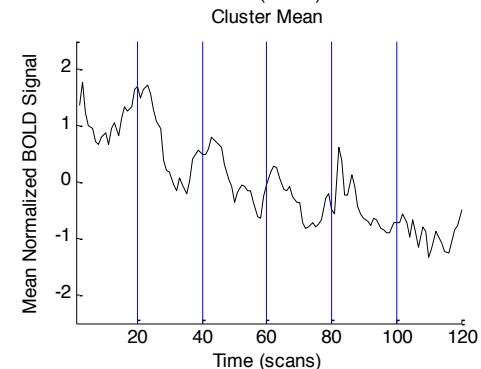
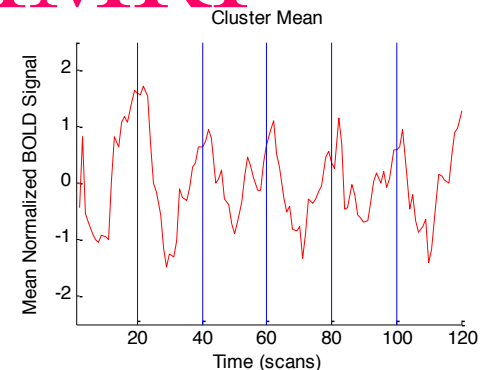
- Συσταδοποίηση, δηλ. Ομαδοποίηση εικονοστοιχείων (voxels) των οποίων οι χρονοσειρές έχουν παρόμοια συμπεριφορά στο χρόνο
- Μεθοδολογία:
 - Προεπεξεργασία, κανονικοποίηση, εξαγωγή θορύβου
 - Ανάλυση φασματικής ισχύς (power spectral analysis)
 - Συσταδοποίηση χρονοσειρών (k-means και άλλες τεχνικές)
 - Έλεγχος των συστάδων
 - Επαναληπτική αφαίρεση outliers

Ανάλυση και Συσταδοποίηση χρονοσειρών από fMRI



Subject's clustering result overlaid on the T1 map

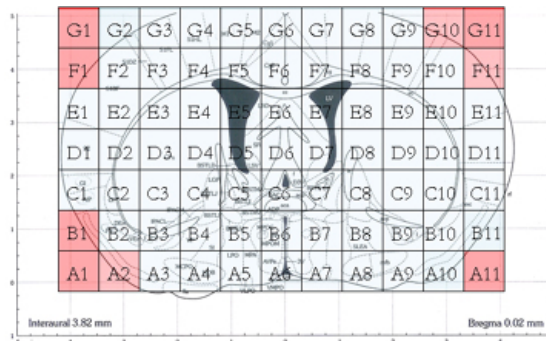
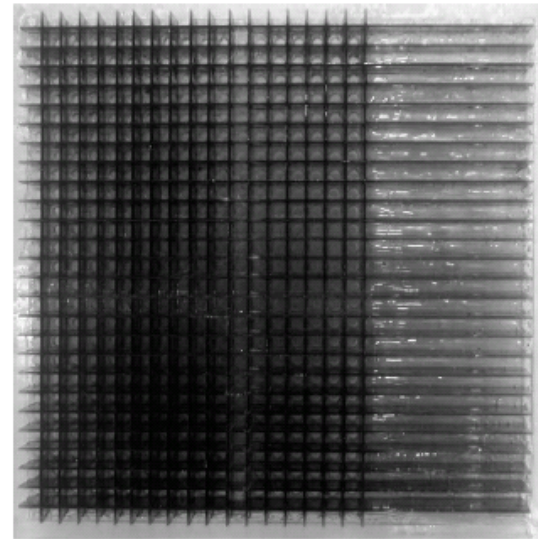
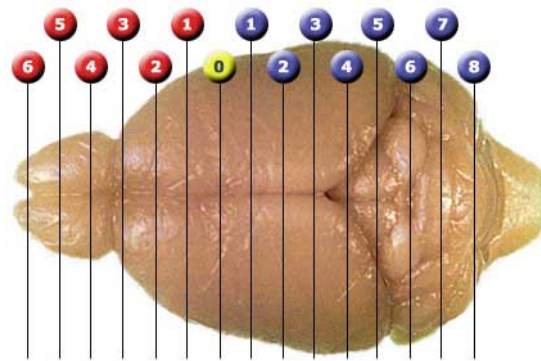
B. Μεγαλοικονόμου



Mean time series for each cluster

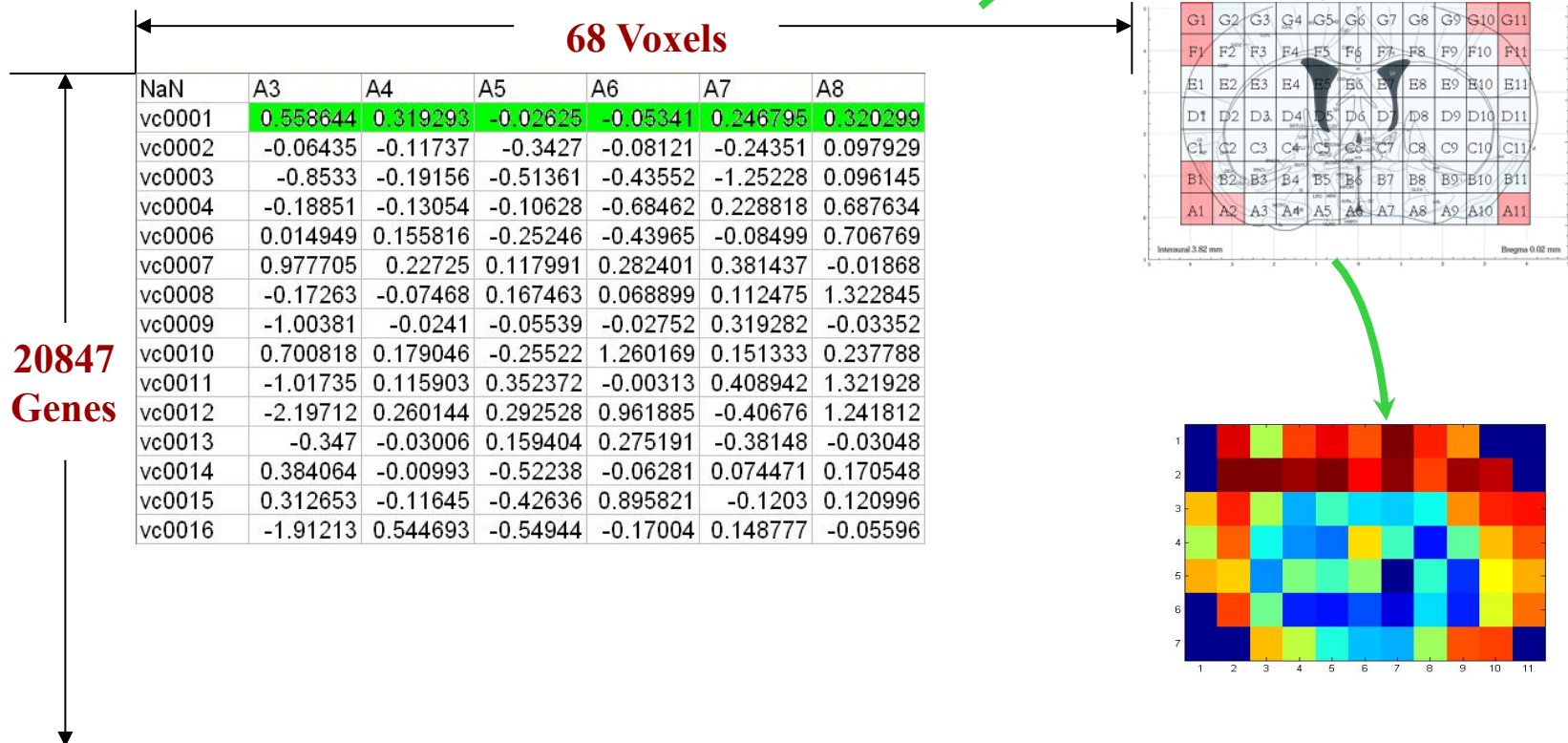
Ανάλυση Χαρτών Έκφρασης Γονιδίων (Gene Expression Maps)

Data from microarrays in a coronal slice of normal mice brain



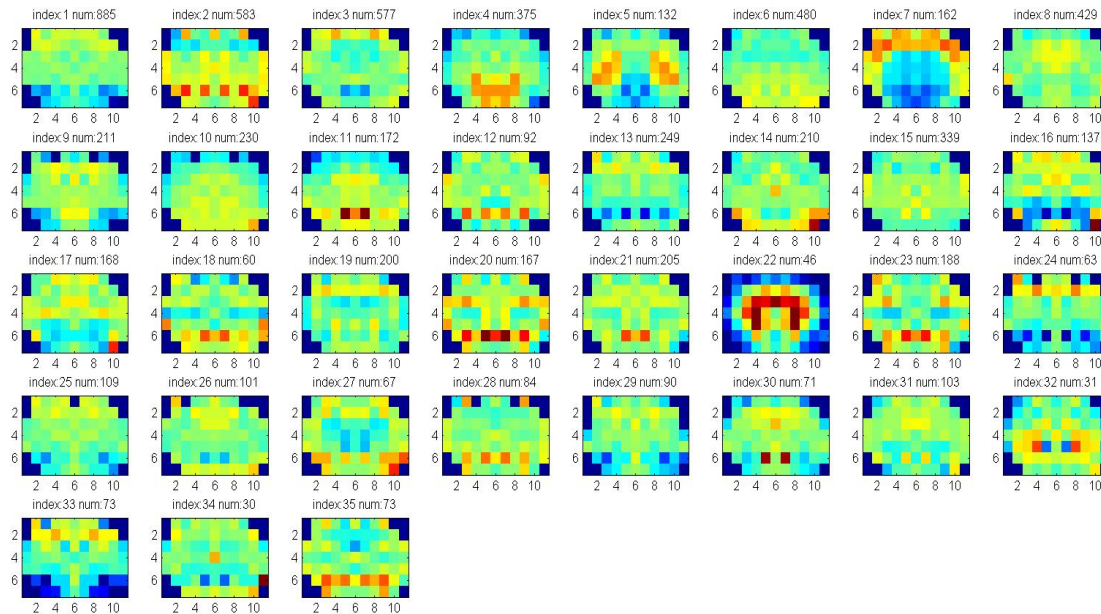
Ανάλυση Χαρτών Έκφρασης Γονιδίων

- 2-dimensional images of gene expression for 20,847 genes
- Each element of the matrix is the gene expression value for a particular gene in a given voxel



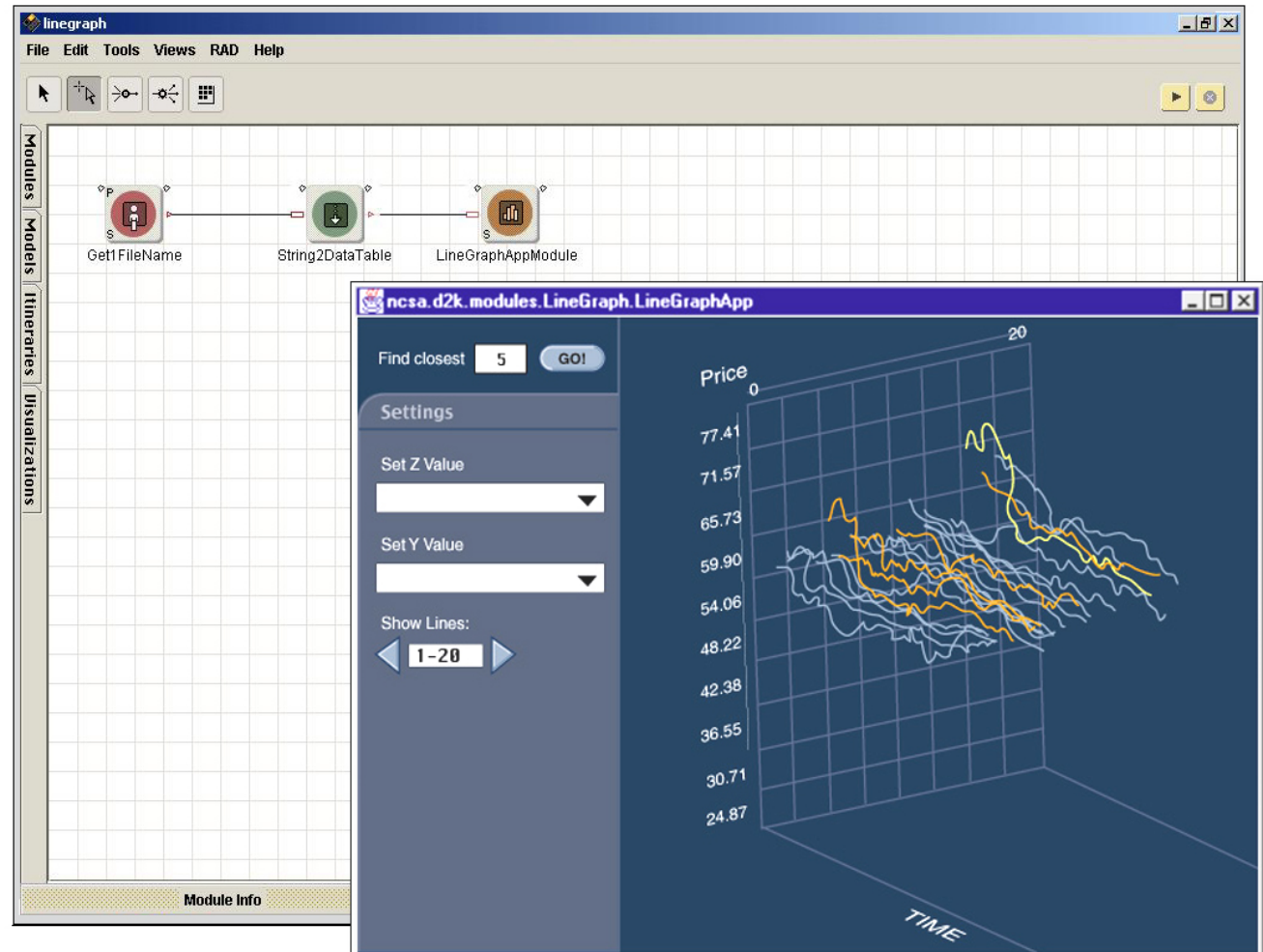
Συσταδοποίηση

- Cellular Component, 35 significant clusters



Σύγκριση Δεδομένων, Σύνθεση και Σύνοψη - Οπτικοποίηση

- Στόχος
 - Ανάπτυξη S/W εργαλείου για 3D οπτικοποίηση δεδομένων αισθητήρων πολλαπλών καναλιών. Αυτό το εργαλείο επιτρέπει την σύγκριση, σύνθεση και σύνοψη πολλαπλών χρονοσειρών.
- Συναφή Projects
 - Παρακολούθηση παραγώγων
 - Παρακολούθηση Συστήματος σε Πραγματικό Χρόνο



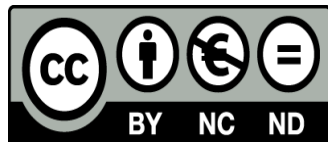
Επίδραση της Εξόρυξης Βιοιατρικών Δεδομένων

- Assist in clinical interpretation of biomedical images
- Facilitate advances in diagnosis and treatment
- Aid in early detection of diseases
- Enable researchers to integrate, manipulate and analyze large volumes of biomedical data
- Provide new insight into the relation of anatomy and function

Ευχαριστώ!

ΑΠΟΡΙΕΣ;

Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρών, Β. Μεγαλοικονόμου «Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων: Εξόρυξη Δεδομένων και Ανακάλυψη Γνώσης». Έκδοση: 1.0. Πάτρα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:

<https://eclass.upatras.gr/courses/MATH959/>

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση, Όχι Παράγωγα Έργα 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει) μαζί με τους συνοδευόμενους υπερσυνδέσμους.