

Τεχνολογίες Ευφυούς Διαχείρισης Ανθρωπιστικών Δεδομένων

Εισαγωγή

Κάτια Κερμανίδου

kerman@ionio.gr

Ανθρωπιστικές & Κοινωνικές Επιστήμες

- Οι Ανθρωπιστικές (Humanistic) και Κοινωνικές (Social) Επιστήμες είναι

- Η Φιλολογία
- Η Γλωσσολογία
- Η Φιλοσοφία
- Η Ιστορία
- Η Αρχαιολογία
- Η Ανθρωπολογία
- Η Παιδαγωγική
- Η Ψυχολογία

- Η Νομική
- Οι Καλές τέχνες
- Η Μουσική
- Η Κοινωνιολογία

Ανθρωπιστικές - Κοινωνικές Επιστήμες & Πληροφορική

- Η Πληροφορική στις Επιστήμες αυτές
 - Εστιάζει στην εφαρμογή ΤΠΕ για την εξαγωγή, αναπαράσταση και επεξεργασία δεδομένων από τις Επιστήμες αυτές
 - Εστιάζει στην προσαρμογή μεθοδολογιών και τεχνικών αυτών των επιστημών ώστε να δημιουργηθούν προϊόντα και υπηρεσίες ΤΠΕ με βέλτιστη απόδοση και ανθρωποκεντρική σχεδίαση

Ενδεικτικές Περιοχές Ενδιαφέροντος

- Πληροφορική και Πολιτισμός (Cultural Informatics)
- Μουσική Πληροφορική (Music Informatics)
- Πληροφορική και Ιστορία (History Informatics)
- Υπολογιστική Γλωσσολογία (Computational Linguistics)
- Ψυχαγωγικό Λογισμικό (Game Software Design and Development)
- Σημασιολογικός Ιστός (Semantic Web)
- Εφαρμογές και Τεχνολογίες Κοινωνικού Ιστού (Social Web)
- Εικονικοί Κόσμοι
- ΤΠΕ στην Εκπαίδευση (e-Learning)
- Αναπαράσταση Γνώσης (Knowledge Representation and Management)
- Κοινωνικά-Συνεργατικά Συστήματα (Computer-supported Cooperative Systems)
- Ψηφιακά Μέσα (Digital Media)

Χαρακτηριστικά Παραδείγματα Ανθρωπιστικών-Κοινωνικών Δεδομένων

- Δεδομένα κίνησης/προτιμήσεων επισκέπτη σε ένα μουσείο
- Δεδομένα συμπεριφοράς ενός παίκτη βιντεοπαιχνιδιού
- Δεδομένα κινήσεων χρήστη κατά την πλοήγησή του σε μια ιστοσελίδα
- Δεδομένα διεκπεραίωσης ηλεκτρονικών αγορών
- Δεδομένα του κοινωνικού ιστού
- Δημογραφικά δεδομένα
- Γλωσσολογικά/κειμενικά δεδομένα
- Εκπαιδευτικά δεδομένα
- Μουσικά δεδομένα

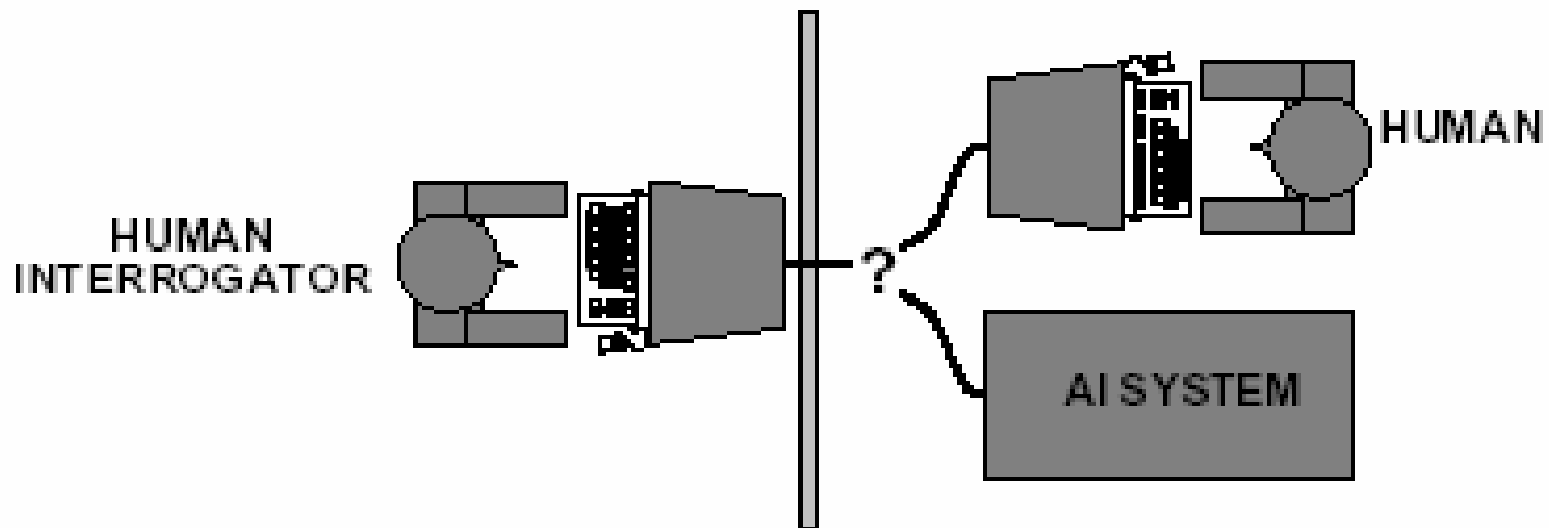
Τι θα πει «Ευφυής» Διαχείριση των Δεδομένων αυτών;

- Η επεξεργασία τους με τέτοιο τρόπο ώστε να εξαχθεί καινούρια γνώση από αυτά που δεν είναι ορατή με γυμνό μάτι (εξόρυξη δεδομένων - data mining)
- Υποπεριοχή της Τεχνητή Νοημοσύνης

Τεχνητή Νοημοσύνη

- Μετάφραση του όρου 'Artificial Intelligence'
- Η επιστημονική περιοχή που προσπαθεί να κατανοήσει και να κατασκευάσει οντότητες με νοημοσύνη
- Τι είναι νοημοσύνη;
 - Homo sapiens
 - Συλλογισμός (Reasoning)
 - Γλώσσα (language)
 - Ενδοσκόπηση (introspection)
 - Επίλυση προβλημάτων (problem solving)
- Δύο άξονες
 - Σκέψη - Συμπεριφορά
 - Άνθρωπος - Ορθολογικότητα (Rationality)
- Δοκιμασία Turing

Test Turing

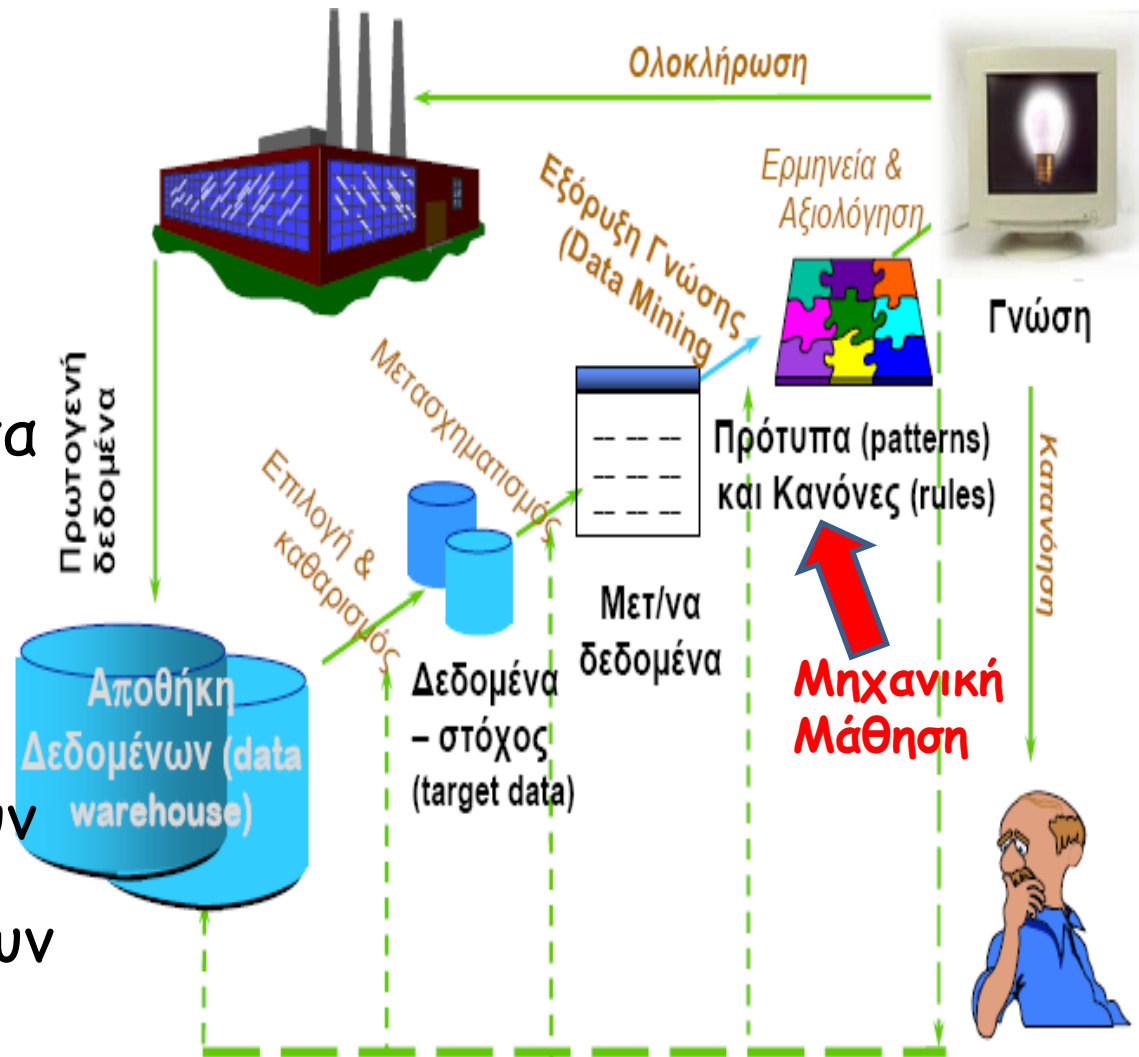


Περιοχές Τεχνητής Νοημοσύνης

- Αναζήτηση λύσης σε πρόβλημα (Search)
 - Παιχνίδια δύο αντιπάλων (Adversarial Search)
 - Γνώση, Εξαγωγή Συμπερασμού
 - Λογική (Logic)
 - Αναπαράσταση Γνώσης (Knowledge Representation)
 - Συμπερασμός (Reasoning)
 - Όλοι οι άνθρωποι είναι θνητοί
 - Ο Σωκράτης είναι άνθρωπος
- ➡ Ο Σωκράτης είναι θνητός
- Σχεδιασμός Ενεργειών (Planning)
 - Καταστάσεις (states)
 - Ενέργειες, Ακολουθίες ενεργειών (πλάνα)
 - Προϋποθέσεις (pre-conditions)
 - Αβεβαιότητα (Uncertainty)
 - Ασαφής Λογική (Fuzzy Logic)
 - υπεрсύνολο της Λογικής
 - Τιμές ανάμεσα στο «απολύτως αληθές» και το «απολύτως ψευδές»
 - **Εξόρυξη Δεδομένων - Μηχανική Μάθηση**
 - Επεξεργασία Φυσικής Γλώσσας
 - Ευφυείς πράκτορες (intelligent agents)
 - Αντιλαμβάνονται το περιβάλλον μέσω αισθητήρων και αντιδρούν με μηχανισμούς δράσεις

Εξόρυξη Δεδομένων (Data Mining)

- Πληθώρα ορισμών
 - Μη τετριμμένη εξαγωγή υποκρυπτόμενης, άγνωστης και εν δυνάμει χρήσιμης πληροφορίας από τα δεδομένα
 - Εξερεύνηση και ανάλυση, με αυτόματο ή ημι-αυτόματο τρόπο, μεγάλων ποσοτήτων δεδομένων για την ανακάλυψη χρήσιμων προτύπων



Τι (ΔΕΝ) είναι εξόρυξη δεδομένων

ΔΕΝ είναι
εξόρυξη

Η εύρεση
ονομάτων σε ένα
τηλεφωνικό
κατάλογο

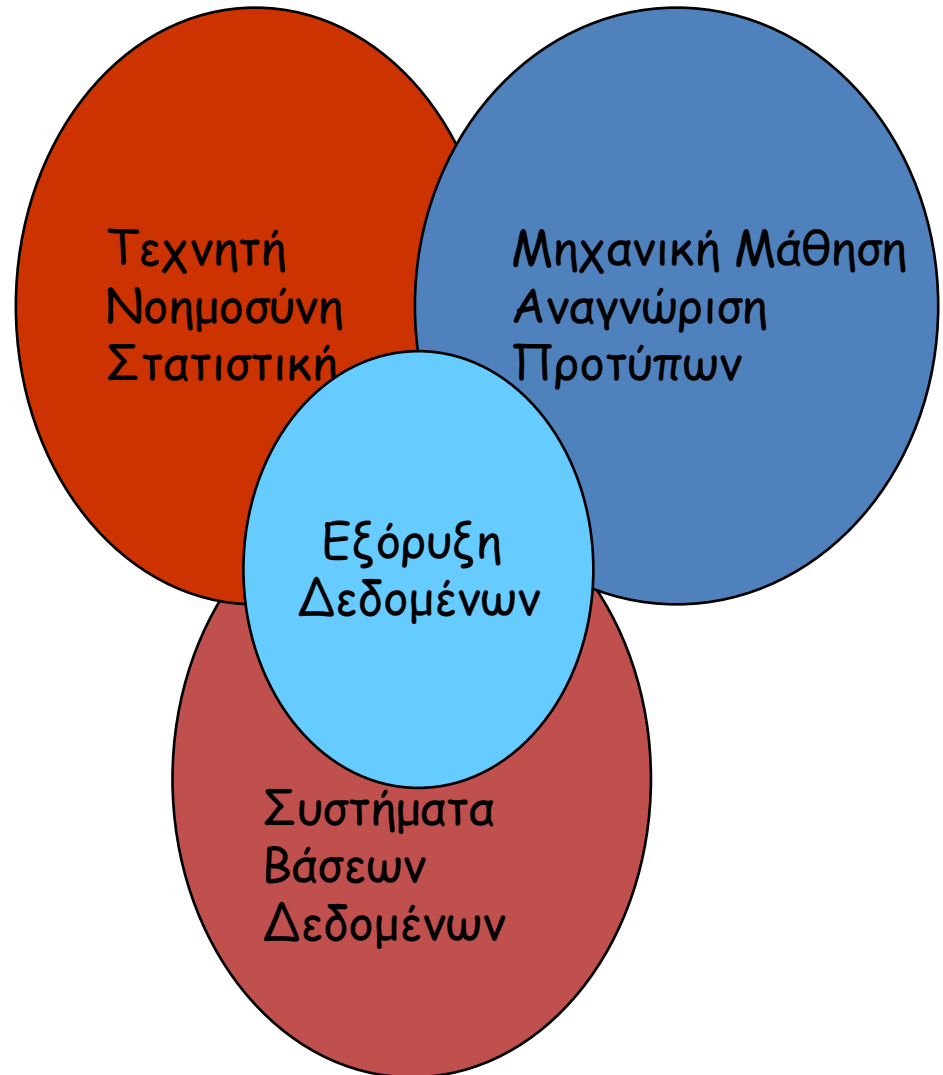
Η επερώτηση σε μια
μηχανή αναζήτησης
στο διαδίκτυο

ΕΙΝΑΙ
εξόρυξη

- Η εύρεση παρόμοιων βιβλίων σε ένα ηλ/κό κατάστημα με βάση το περιεχόμενό τους
- Ορισμένα ονόματα είναι πιο συχνά σε συγκεκριμένες περιοχές
- Το φιλτράρισμα ανεπιθύμητων μηνυμάτων

Προέλευση της εξόρυξης δεδομένων

- Σταχυολογεί ιδέες από
 - μηχανική μάθηση, τεχνητή νοημοσύνη, αναγνώριση προτύπων, στατιστική και βάσεις δεδομένων
- Οι παραδοσιακές τεχνικές δεν είναι κατάλληλες λόγω
 - Πληθώρας των δεδομένων
 - Μεγάλης διαστατικότητας των δεδομένων
 - Ετερογενείς, κατανεμημένες οργανώσεις δεδομένων



Εργασίες Εξόρυξης Δεδομένων

- Προβλεπτικές Μέθοδοι

- Χρησιμοποίηση ορισμένων μεταβλητών για να προβλεφθεί η τιμή (άγνωστη ή μελλοντική) τιμή άλλων μεταβλητών

- Π.χ. η μέτρηση του όγκου συναλλαγών μιας μετοχής για την πρόβλεψη της πορείας του ΓΔ την επόμενη ημέρα

- Περιγραφικές Μέθοδοι

- Εύρεση προτύπων που μπορεί να ερμηνεύσει ο άνθρωπος για την περιγραφή των δεδομένων

- Π.χ. η εύρεση ενός κανόνα με βάση τον οποίο ένα σύστημα θα κατατάσσει ένα μήνυμα ως ανεπιθύμητο

Εργασίες Εξόρυξης Δεδομένων...

- Κατηγοριοποίηση ή Ταξινόμηση-Classification [Προβλεπτική]
- Συσταδοποίηση ή Ομαδοποίηση-Clustering [Περιγραφική]
- Ανακάλυψη Κανόνων Συσχετίσεων-Association Rule Discovery [Περιγραφική]
- Ανακάλυψη Χρονολογικών Προτύπων-Sequential Pattern Discovery [Περιγραφική]
- Παλινδρόμηση-Regression [Προβλεπτική]
- Ανίχνευση Εξαιρέσεων-Deviation Detection [Προβλεπτική]

Μηχανική Μάθηση (Machine Learning)

- Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση.
- Simon (1983)
 - "η μάθηση σηματοδοτεί προσαρμοστικές αλλαγές σε ένα σύστημα με την έννοια ότι αυτές του επιτρέπουν να κάνει την ίδια εργασία, ή εργασίες της ίδιας κατηγορίας, πιο αποδοτικά και αποτελεσματικά την επόμενη φορά".
- Carbonell (1987)
 - "... η μελέτη υπολογιστικών μεθόδων για την απόκτηση νέας γνώσης, νέων δεξιοτήτων και νέων τρόπων οργάνωσης της υπάρχουσας γνώσης".
- Mitchell (1997)
 - "Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία E σε σχέση με μια κατηγορία εργασιών T και μια μετρική απόδοσης P , αν η απόδοση του σε εργασίες της T , όπως μετριοούνται από την P , βελτιώνονται με την εμπειρία E ".
 - Task T : playing chess
 - Performance measure P : percent of games won against opponents
 - Training Experience E : playing practice games against itself
- Witten & Frank (2000),
 - Κάτι μαθαίνει όταν αλλάζει τη συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον"

Μηχανική Μάθηση

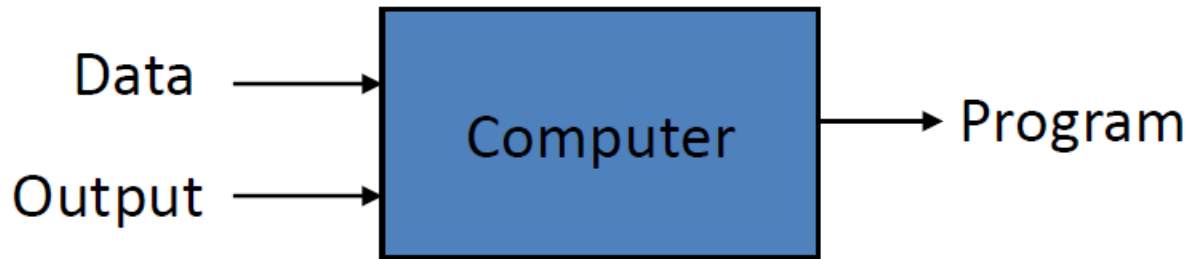
- Αλγόριθμοι για την εξαγωγή δομικών περιγραφών (προτύπων) από δεδομένα
- Οι περιγραφές αυτές
 - Μπορούν να χρησιμοποιηθούν για την πρόβλεψη (prediction) ενός αποτελέσματος σε καινούρια (άγνωστα) δεδομένα
 - μπορούν να χρησιμοποιηθούν για την ερμηνεία του τρόπου με τον οποίο επιλέγεται η πρόβλεψη
- Αντί να απαιτείται η συγγραφή κώδικα για να αναγνωριστούν αυτά τα πρότυπα **χειρωνακτικά**, οι αλγόριθμοι μάθησης τρέχουν πάνω στα δεδομένα και εξάγουν τα πρότυπα αυτά **αυτόματα**.

Παραδοσιακός Προγραμματισμός - Μηχανική Μάθηση

Traditional Programming



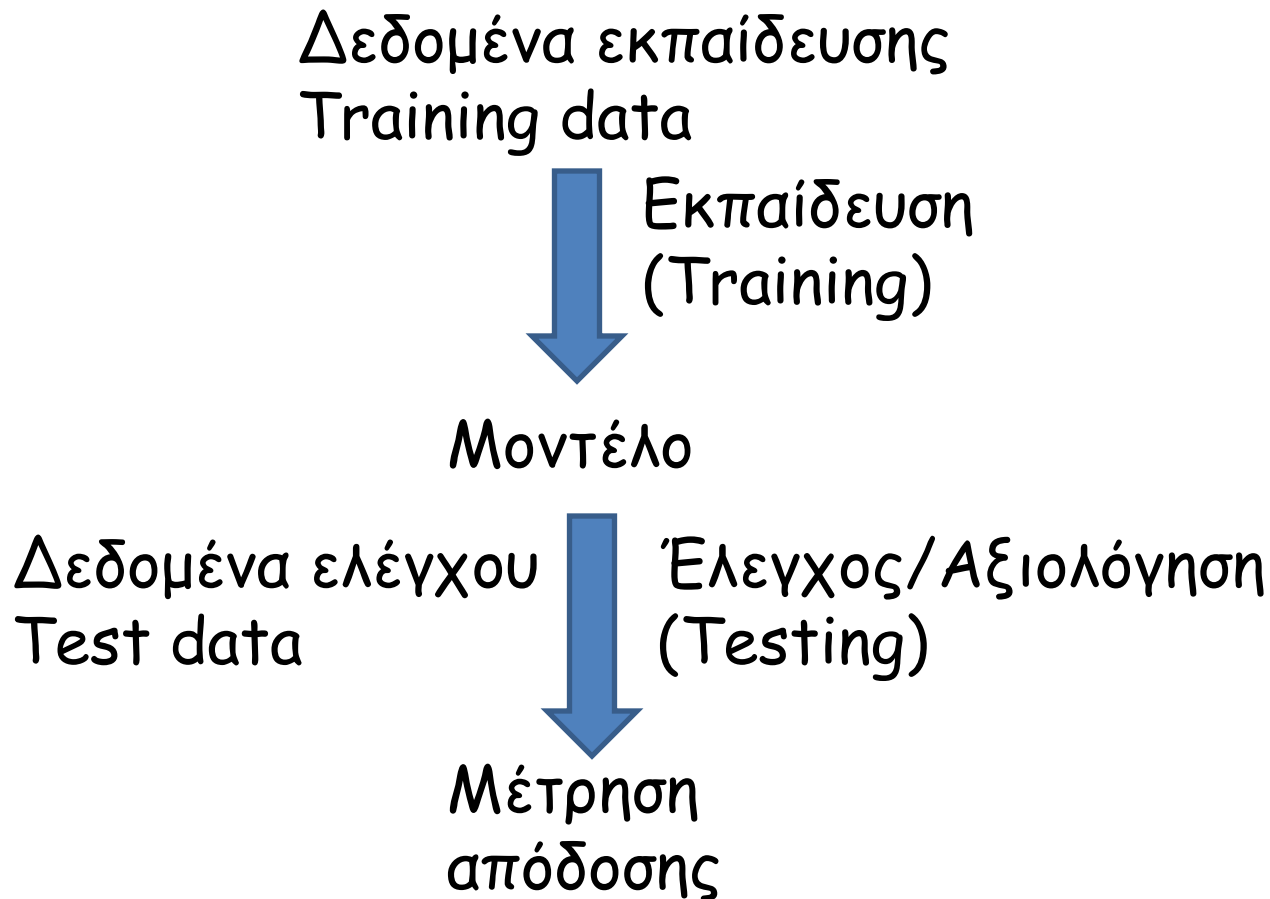
Machine Learning



Επαγωγή (Induction)

- Έννοια (concept) είναι αυτό που καλείται ο αλγόριθμος μάθησης να μάθει
- Επαγωγή είναι η διαδικασία δημιουργίας ενός **γενικευμένου** μοντέλου περιγραφής ή ορισμού μιας έννοιας από ένα σύνολο **ειδικών** παραδειγμάτων της έννοιας
- Προτάθηκε από τον Αριστοτέλη σαν την αντίστροφη διαδικασία από τον συλλογισμό
- Το παράδειγμα του Αρειανού
- Πιο μαθηματικοποιημένα
 - Ένα παράδειγμα-example (στιγμιότυπο-instance ή παρατήρηση-observation ή δεδομένο-data) είναι μια δυάδα $(x, f(x))$, όπου f είναι μια συνάρτηση, το x είναι η είσοδος, και το $f(x)$ η έξοδος της συνάρτησης όταν εφαρμόζεται στο x
 - Επαγωγή: Δοθέντος ενός συνόλου παραδειγμάτων μιας συνάρτησης f , βρες μια συνάρτηση h που προσεγγίζει την f
 - Η συνάρτηση h ονομάζεται υπόθεση και αποτελεί το γενικευμένο μοντέλο που περιγράφει τα παραδείγματα.

Φάσεις Μηχανικής Μάθησης



Είδη Μηχανικής Μάθησης

- Επιβλεπόμενη Μάθηση
- Μη επιβλεπόμενη Μάθηση

Επιβλεπόμενη Μάθηση (Supervised Learning)

- Το σύστημα μάθησης καλείται να μάθει επαγωγικά την **συνάρτηση στόχο** (target function) που περιγράφει τα δεδομένα
- Η τιμή της συνάρτησης-στόχου ονομάζεται και **εξαρτημένη μεταβλητή** ενώ οι υπόλοιπες μεταβλητές που αναπαριστούν το παράδειγμα ονομάζονται **ανεξάρτητες μεταβλητές**.
- Κατά την εκπαίδευση η τιμή της συνάρτησης στόχου των παραδειγμάτων εκπαίδευσης είναι γνωστή, και καθοδηγεί τη διαδικασία μάθησης
- Κατά τον έλεγχο, η απόδοση μετράται σε καινούρια άγνωστα παραδείγματα (ελέγχου), για τα οποία ο αλγόριθμος δεν γνωρίζει την τιμή της συνάρτησης στόχου
- Δύο μορφές επιβλεπόμενης μάθησης
 - Ταξινόμηση (ή Κατηγοριοποίηση - Classification)
 - Η συνάρτηση-στόχος παίρνει διακριτές τιμές
 - Παλινδρόμηση (ή Παρεμβολή - Regression)
 - Η συνάρτηση-στόχος παίρνει αριθμητικές τιμές

Δεδομένα Μάθησης: Οι φακοί επαφής

Μεταβλητές εισόδου
Ανεξάρτητες μεταβλητές
Χαρακτηριστικά
Ιδιότητες
Features
Attributes

Μεταβλητή εξόδου
Εξαρτημένη μεταβλητή
Κλάση ταξινόμησης
Class attribute
Class variable

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
...	...	...	...	...

Παραδείγματα
μάθησης
Στιγμιότυπα
Learning
examples
Instances

Διανύσματα
χαρακτηριστικών-
τιμών
Feature value
vectors

Δεδομένα Μάθησης: The weather problem

Play tennis?

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	Normal	False	Yes
...	...	...	...	...

Outlook	Temperature	Humidity	Windy	Play
Sunny	85	85	False	No
Sunny	80	90	True	No
Overcast	83	86	False	Yes
Rainy	75	80	False	Yes
...	...	...	...	...

Δεδομένα Μάθησης: Iris flower



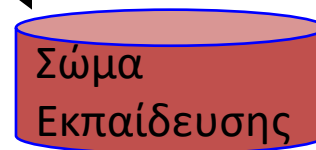
	Sepal length	Sepal width	Petal length	Petal width	Type
1	5.1	3.5	1.4	0.2	Iris setosa
2	4.9	3.0	1.4	0.2	Iris setosa
...					
51	7.0	3.2	4.7	1.4	Iris versicolor
52	6.4	3.2	4.5	1.5	Iris versicolor
...					
101	6.3	3.3	6.0	2.5	Iris virginica
102	5.8	2.7	5.1	1.9	Iris virginica
...					

Δεδομένα Μάθησης: Πελάτες ΔΟΥ

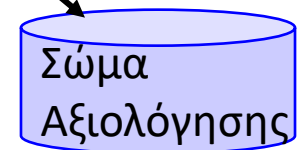
Διακριτή Διακριτή Συνεχής Κλάση

Tid	Επιστροφή	Οικογεν. Κατάσταση	Εισόδημα	Απάτη;
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Επιστροφή	Οικογεν. Κατάσταση	Εισόδημα	Απάτη;
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Εκμάθηση
Μοντέλου



Παλινδρόμηση

Outlook	Temperature	Humidity	Windy	Play-time
Sunny	Hot	High	False	5
Sunny	Hot	High	True	0
Overcast	Hot	High	False	55
Rainy	Mild	Normal	False	40
...	...	...	...	...

Μη επιβλεπόμενη μάθηση - Unsupervised Learning

- Ομαδοποίηση-Συσταδοποίηση (Clustering)
- Κατά την εκπαίδευση, η τιμή της συνάρτησης-στόχου των παραδειγμάτων δεν είναι γνωστή
- Δημιουργούνται ομάδες από παραδείγματα που είναι όμοια
- Η απόδοση μετράται υποκειμενικά

	Sepal length	Sepal width	Petal length	Petal width	Type
1	5.1	3.5	1.4	0.2	Iris setosa
2	4.9	3.0	1.4	0.2	Iris setosa
...					
51	7.0	3.2	4.7	1.4	Iris versicolor
52	6.4	3.2	4.5	1.5	Iris versicolor
...					
101	6.3	3.3	6.0	2.5	Iris virginica
102	5.8	2.7	5.1	1.9	Iris virginica
...					

Ανακάλυψη κανόνων συσχετίσεων: Ορισμός

- Δεδομένου ενός συνόλου από εγγραφών, κάθε μια εκ των οποίων περιέχει ένα αριθμό αντικειμένων από μια δεδομένη συλλογή
 - Παραγωγή κανόνων εξάρτησης οι οποίοι προβλέπουν την εμφάνιση ενός αντικειμένου με βάση την εμφάνιση άλλων αντικειμένων

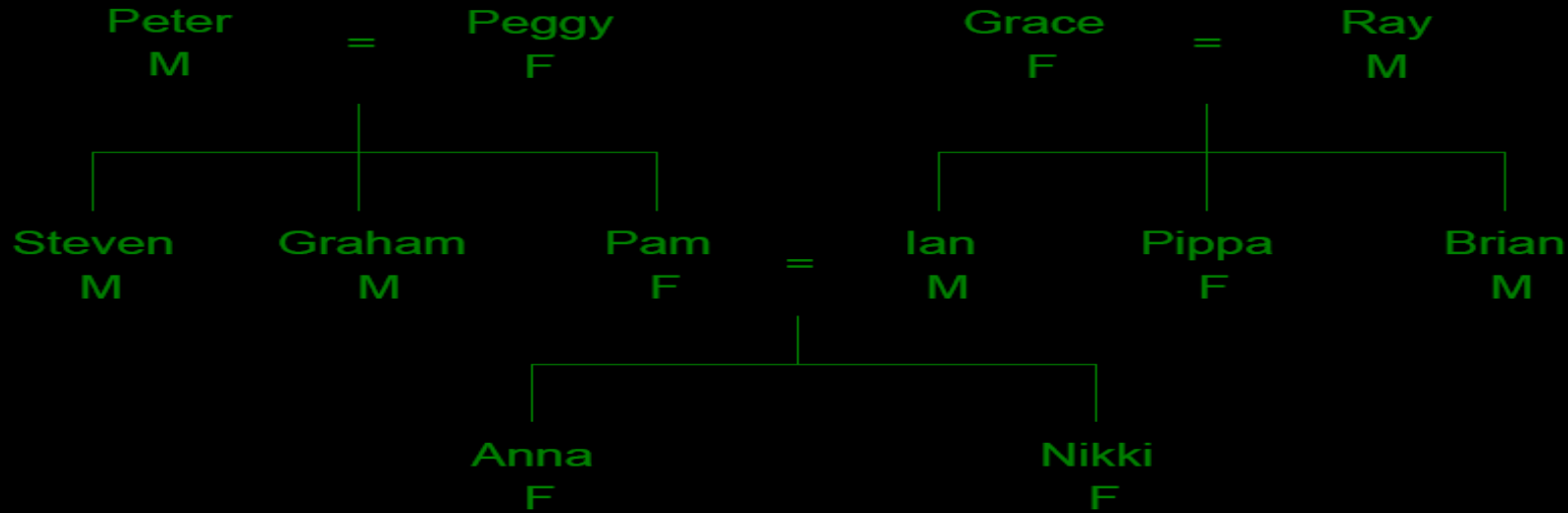
ID	Αντικείμενα
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Εξαγωγή Κανόνων:

{Milk} --> {Coke}

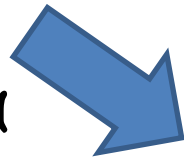
{Diaper, Milk} --> {Beer}

Αναπαράσταση δεδομένων



Missing values

Πώς
αναπαρίσταται μια
βάση γνώσης σαν
ένα σετ
διανυσμάτων
χαρακτηριστικών-
τιμών;



Name	Gender	Parent1	parent2
Peter	Male	?	?
Peggy	Female	?	?
Steven	Male	Peter	Peggy
Graham	Male	Peter	Peggy
Pam	Female	Peter	Peggy
Ian	Male	Grace	Ray
Pippa	Female	Grace	Ray
Brian	Male	Grace	Ray
Anna	Female	Pam	Ian
Nikki	Female	Pam	Ian

Παράδειγμα ταξινόμησης: Η σχέση ancestor-of

First person				Second person				Ancestor of?
Name	Gender	Parent1	Parent2	Name	Gender	Parent1	Parent2	
Peter	Male	?	?	Steven	Male	Peter	Peggy	Yes
Peter	Male	?	?	Pam	Female	Peter	Peggy	Yes
Peter	Male	?	?	Anna	Female	Pam	Ian	Yes
Peter	Male	?	?	Nikki	Female	Pam	Ian	Yes
Pam	Female	Peter	Peggy	Nikki	Female	Pam	Ian	Yes
Grace	Female	?	?	Ian	Male	Grace	Ray	Yes
Grace	Female	?	?	Nikki	Female	Pam	Ian	Yes
<i>Other positive examples here</i>								Yes
<i>All the rest</i>								No

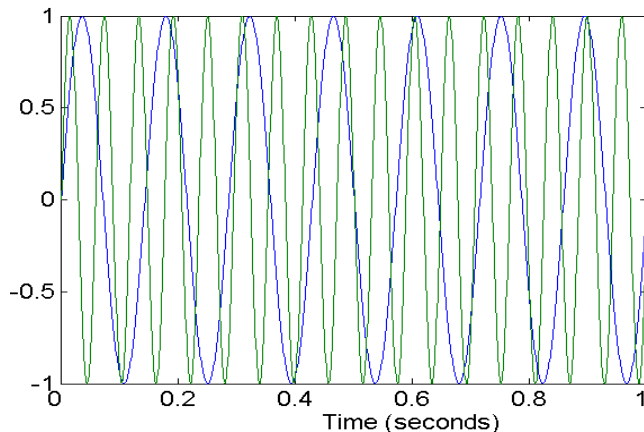
Είδη χαρακτηριστικών

- Ονοματικά (nominal)
 - Οι τιμές είναι διακριτές
 - Πχ «Δευτέρα», «Τρίτη», «Τετάρτη» κλπ
 - Δεν υπάρχει έννοια κατάταξης ή απόστασης
 - Μόνο ταύτιση ή μη-τάυτιση
- Τακτικά (ordinal)
 - Υπάρχει κατάταξη
 - "Hot"> "Mild"> "Cool"
 - Δεν υπάρχει δυνατότητα πρόσθεσης/αφαίρεσης
- Διαστήματα (Interval)
 - Υπάρχει κατάταξη
 - Οι τιμές είναι σε συγκεκριμένες και ίσες αποστάσεις μεταξύ τους
 - Πχ μεταβλητή: χρονιά, με τιμές «1992», «1993», «1994»...
 - Υπάρχει νόημα αφαίρεσης/διαφοράς δυο τιμών
 - Δεν υπάρχει νόημα πολλαπλασιασμού/διαίρεσης
- Αναλογικά (Ratio)
 - Οι τιμές είναι πραγματικοί αριθμοί που επιτρέπουν όλες τις μαθηματικές πράξεις

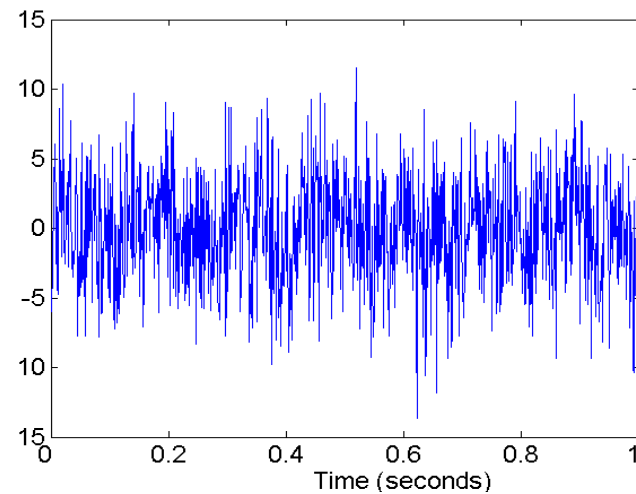
Ποιότητα στα δεδομένα

- Θόρυβος (Noise)

- αναφέρεται σε αλλοιώσεις των αρχικών τιμών μιας ιδιότητας
- Πχ παραμόρφωση στο σήμα ομιλίας λόγω κακής ποιότητας ενός μικροφώνου, χιόνια σε σήμα τηλεόρασης λόγω κακής λήψης, κτλ



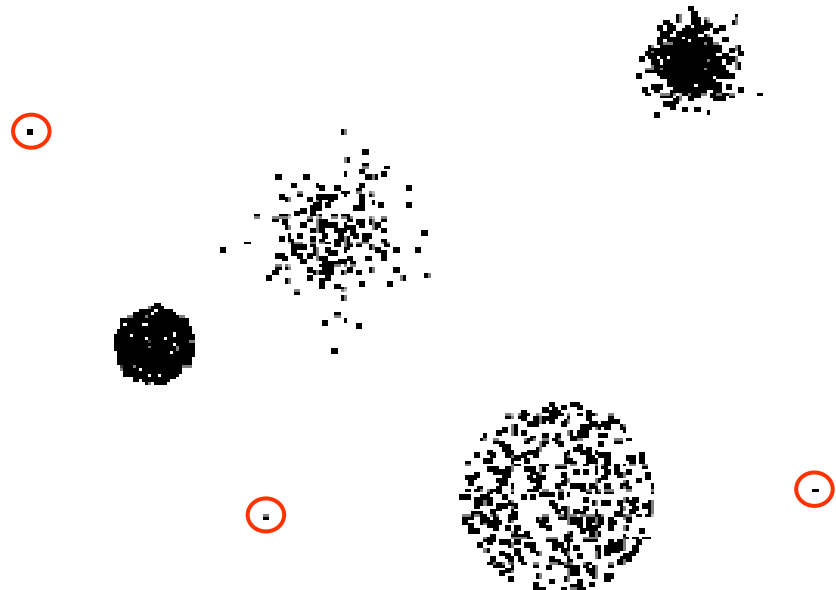
2 ημιτονοειδή
σήματα



2 ημιτονοειδή
σήματα με θόρυβο

Ποιότητα στα δεδομένα

- Εξαιρέσεις (Outliers)
 - δεδομένα με χαρακτηριστικά που διαφέρουν σημαντικά από τα περισσότερα δεδομένα του συνόλου



Ποιότητα στα δεδομένα

- Ανύπαρκτες τιμές (Missing values)
- Λόγοι ανύπαρκτων τιμών
 - Δεν συλλέγεται η πληροφορία
 - Π.χ. ορισμένοι αρνούνται να αποκαλύψουν την ηλικία ή το βάρος τους
 - Οι ιδιότητες δεν χαρακτηρίζουν όλα τα δεδομένα
 - Π.χ. το εισόδημα δεν εφαρμόζεται στα παιδιά
- Πως χειριζόμαστε τις ανύπαρκτες τιμές
 - Αφαίρεση δεδομένων που τις περιέχουν
 - Εκτίμηση των τιμών
 - Να τις αγνοήσουμε
 - Αντικατάσταση με όλες τις πιθανές τιμές (με στάθμιση των πιθανοτήτων τους)

Ποιότητα στα δεδομένα

- Διπλότυπα
- Το σύνολο δεδομένων μπορεί να περιλαμβάνει δεδομένα με διπλές εγγραφές ή σχεδόν διπλότυπα
 - Μεγάλο ζήτημα όταν συνενώνονται δεδομένα από ετερογενείς πηγές
- Παραδείγματα:
 - Ένα άτομο με πολλές διευθύνσεις ηλ. ταχυδρομείου
- Καθαρισμός δεδομένων
 - Η διαδικασία εξομάλυνσης του προβλήματος των διπλότυπων

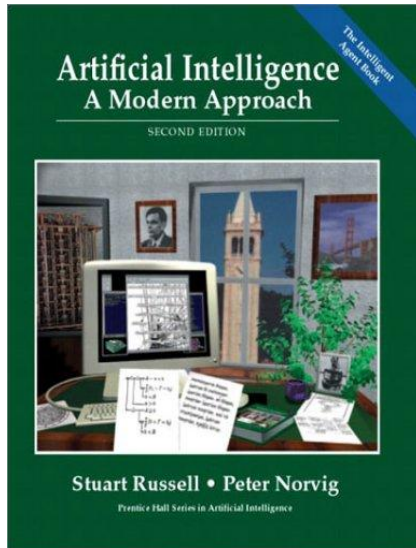
Το πρόβλημα της διαστατικότητας

- Ο χώρος του μικρόκοσμου που θέλω να κάνω εξόρυξη είναι τόσων διαστάσεων όσο είναι και το πλήθος των ανεξάρτητων μεταβλητών μου
- Όσο αυξάνουν οι διαστάσεις, τα δεδομένα γίνονται ολοένα και πιο αραιά στον πολυδιάστατο χώρο
- Η απόσταση και η πυκνότητα μεταξύ των σημείων που είναι κρίσιμες για την ομαδοποίηση ή την ανίχνευση εξαιρέσεων χάνουν τη σημασία τους

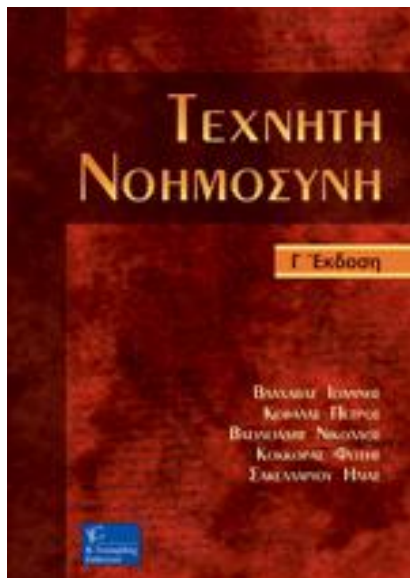
Μείωση της διαστατικότητας

- Μείωση χρόνου και χώρου για τους αλγόριθμους εξόρυξης δεδομένων
- Επιτρέπει την οπτικοποίηση των δεδομένων
- Βοηθάει στην αφαίρεση άσχετων χαρακτηριστικών ή στη μείωση του θορύβου
- Τεχνικές
 - Principle Component Analysis (PCA)
 - Singular Value Decomposition (SVD)
 - Άλλες

Βιβλία

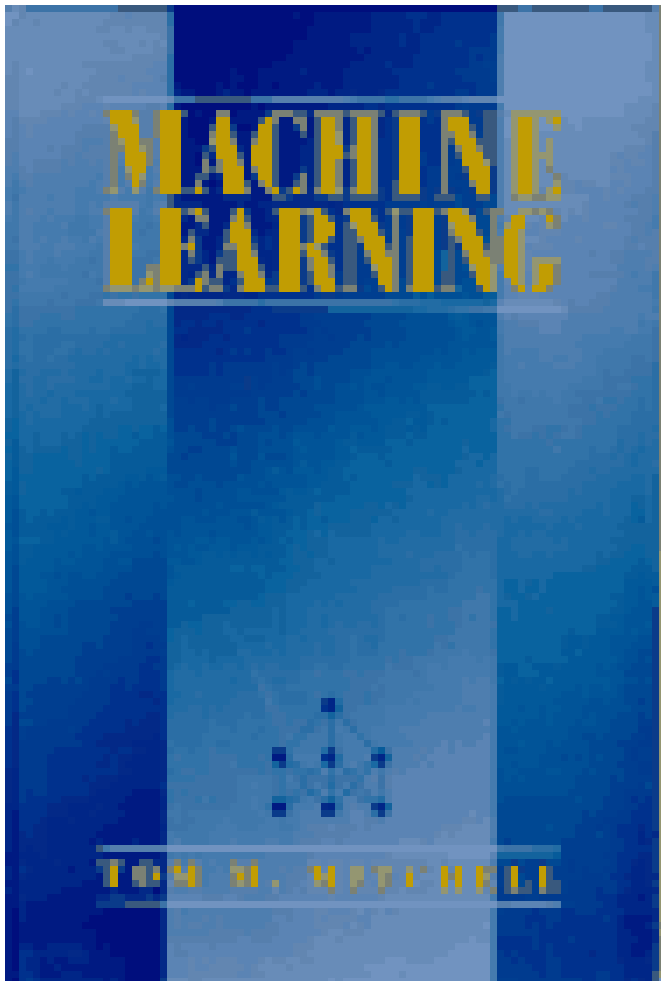


Τεχνητή Νοημοσύνη: Μια Σύγχρονη
Προσέγγιση
S. Russell, P. Norvig
Prentice Hall



Τεχνητή Νοημοσύνη
Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης,
Φ. Κόκκορας, Η. Σακελλαρίου
Γ' Έκδοση, 2006
Εκδόσεις Β.Γκιούρδας

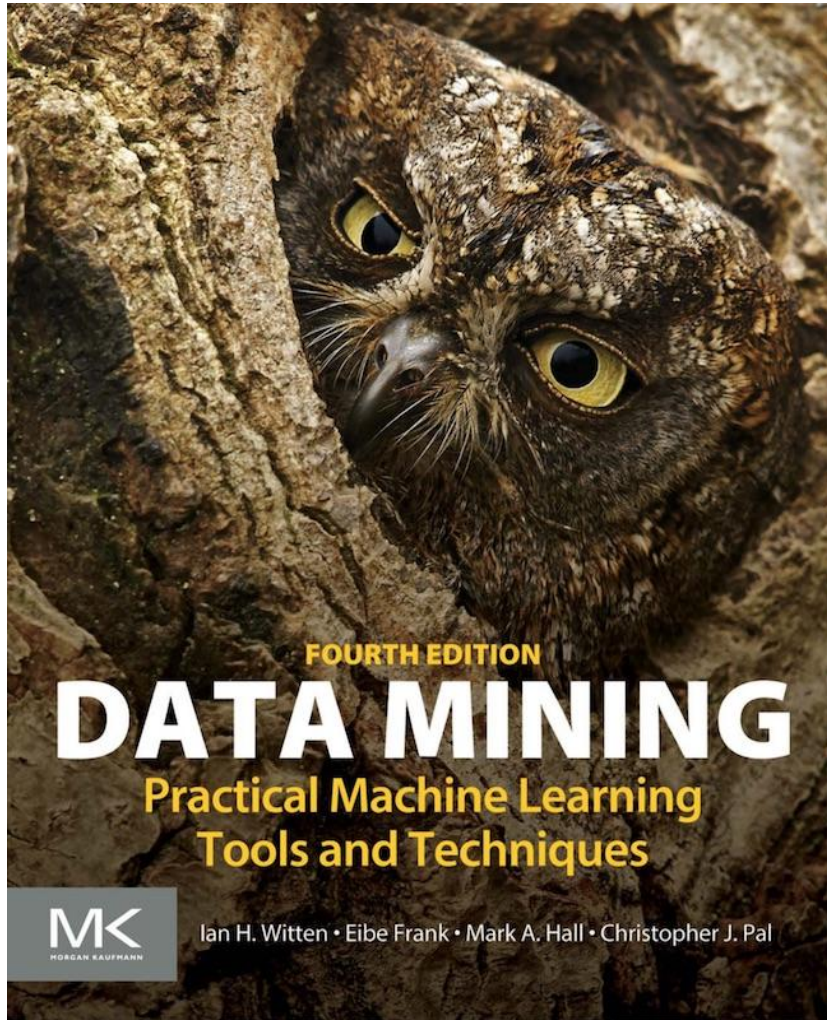
Βιβλία



Machine Learning,
Tom Mitchell,
McGraw Hill,
1997.

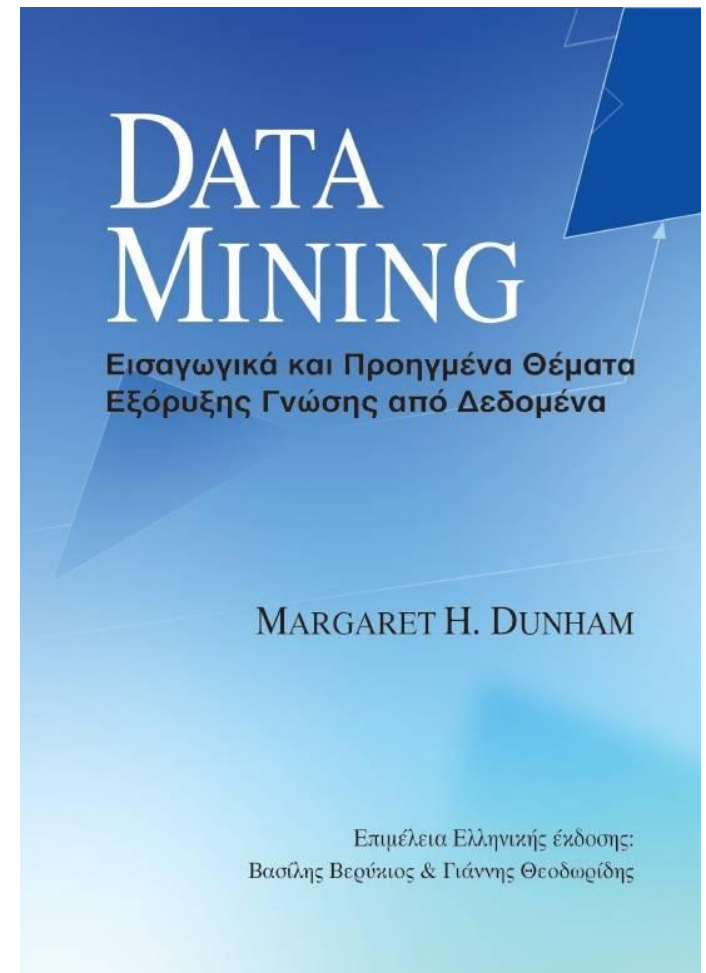
Βιβλία

Το βιβλίο του Weka



Βιβλία

- Data Mining (στα Ελληνικά)
 - Margaret H. Dunham
 - Έτος Έκδοσης: 2004
 - Εκδότης: ΕΚΔΟΣΕΙΣ ΝΕΩΝ ΤΕΧΝΟΛΟΓΙΩΝ
 - Αριθμός σελίδων: 315
 - Κωδικός ISBN: 960-8105-72-2



Βιβλία

- Εισαγωγή στην Εξόρυξη Δεδομένων και τις Αποθήκες Δεδομένων
- Αλ. Νανόπουλος - Γ. Μανωλόπουλος
 - Έτος Έκδοσης: 2008
 - Εκδότης: ΕΚΔΟΣΕΙΣ ΝΕΩΝ ΤΕΧΝΟΛΟΓΙΩΝ
 - Αριθμός σελίδων: 384
 - Κωδικός ISBN: 978-960-6759-17-8



Πηγές

- Διαδικτυακά μαθήματα και βιβλία
 - <http://www.cs.cmu.edu/~tom/mlbook.html>
 - <http://aima.cs.berkeley.edu/index.html>
 - <http://cs229.stanford.edu/materials.html>
 - <http://aibook.csd.auth.gr/index.html>
 - <http://www.cs.waikato.ac.nz/ml/weka/book.html>
 - www.hlt.utdallas.edu/~vgogate/ml/2012s/notes/