

Κεφάλαιο 18

Μηχανική Μάθηση (Machine Learning)

Τεχνητή Νοημοσύνη - Β' Έκδοση

Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου

Εισαγωγή

- ❖ Η μάθηση σε ένα γνωστικό σύστημα, όπως γίνεται αντιληπτή στην καθημερινή ζωή, μπορεί να συνδεθεί με δύο βασικές ιδιότητες:
 - ❑ την ικανότητά στην πρόσκτηση γνώσης κατά την αλληλεπίδρασή του με το περιβάλλον,
 - ❑ την ικανότητά να βελτιώνει με την επανάληψη τον τρόπο εκτέλεσης μία ενέργειας.
- ❖ Έχουν προταθεί διάφοροι ορισμοί για τη μάθηση:
 - ❑ Simon ('83), "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".
 - ❑ Minsky ('85), "... είναι να κάνουμε χρήσιμες αλλαγές στο μυαλό μας".
 - ❑ Michalski ('86), "... είναι η δημιουργία ή η αλλαγή της αναπαράστασης των εμπειριών".
- ❖ Για τα συστήματα που ανήκουν στην συμβολική TN, η μάθηση προσδιορίζεται ως πρόσκτηση επιπλέον γνώσης, που επιφέρει μεταβολές στην υπάρχουσα γνώση.
 - ❑ Τα τεχνητά νευρωνικά δίκτυα (που ανήκουν στην μη συμβολική TN) έχουν δυνατότητα μάθησης μετασχηματίζοντας την εσωτερική τους δομή, παρά καταχωρώντας κατάλληλα αναπαριστάμενη γνώση.

Ορισμός Μηχανικής Μάθησης

- ❖ Ο άνθρωπος προσπαθεί να κατανοήσει το περιβάλλον του παρατηρώντας το και δημιουργώντας μια απλοποιημένη (αφαιρετική) εκδοχή του που ονομάζεται **μοντέλο (model)**.
 - ❑ Η δημιουργία ενός τέτοιου μοντέλου, ονομάζεται **επαγωγική μάθηση (inductive learning)** ενώ η διαδικασία γενικότερα ονομάζεται **επαγωγή (induction)**.
- ❖ Επιπλέον ο άνθρωπος έχει τη δυνατότητα να οργανώνει και να συσχετίζει τις εμπειρίες και τις παραστάσεις του δημιουργώντας νέες δομές που ονομάζονται **πρότυπα (patterns)**.

Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται **μηχανική μάθηση (machine learning)**.

- ❖ Διάφοροι ορισμοί:
 - ❑ 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".
 - ❑ Witten & Frank (2000), "Κάτι μαθαίνει όταν αλλάζει τη συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον".

Είδη μηχανικής μάθησης

- ❖ Έχουν αναπτυχθεί πολλές τεχνικές μηχανικής μάθησης που χρησιμοποιούνται ανάλογα με τη φύση του προβλήματος και εμπίπτουν σε ένα από τα παρακάτω δυο είδη:
 - ❑ μάθηση με επίβλεψη (supervised learning) ή μάθηση με παραδείγματα (learning from examples),
 - ❑ μάθηση χωρίς επίβλεψη (unsupervised learning) ή μάθηση από παρατήρηση (learning from observation).
- ❖ Στη μάθηση με επίβλεψη το σύστημα καλείται να "μάθει" μια έννοια ή συνάρτηση από ένα σύνολο δεδομένων, η οποία αποτελεί περιγραφή ενός μοντέλου.
- ❖ Στη μάθηση χωρίς επίβλεψη το σύστημα πρέπει μόνο του να ανακαλύψει συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων, δημιουργώντας **πρότυπα**, χωρίς να είναι γνωστό αν υπάρχουν, πόσα και ποια είναι.

A) Μάθηση με Επίβλεψη

- ❖ Στη μάθηση με επίβλεψη το σύστημα πρέπει να "μάθει" επαγωγικά μια συνάρτηση που ονομάζεται **συνάρτηση στόχος (target function)** και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα.
 - ❑ Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής, που ονομάζεται **εξαρτημένη μεταβλητή** ή **μεταβλητή εξόδου**, βάσει των τιμών ενός συνόλου μεταβλητών, που ονομάζονται **ανεξάρτητες μεταβλητές** ή **μεταβλητές εισόδου** ή **χαρακτηριστικά**.
- ❖ Η επαγωγική μάθηση στηρίζεται στην "**υπόθεση επαγωγικής μάθησης**" (**inductive learning hypothesis**), σύμφωνα με την οποία:

Κάθε υπόθεση h που προσεγγίζει καλά τη συνάρτηση στόχο για ένα αρκετά μεγάλο σύνολο παραδειγμάτων, θα προσεγγίζει το ίδιο καλά τη συνάρτηση στόχο και για περιπτώσεις που δεν έχει εξετάσει.

- ❖ Στην μάθηση με επίβλεψη διακρίνονται δυο είδη προβλημάτων (learning tasks), τα προβλήματα ταξινόμησης και τα προβλήματα παρεμβολής.
 - ❑ Η **ταξινόμηση¹ (classification)** αφορά στη δημιουργία μοντέλων πρόβλεψης διακριτών τάξεων (κλάσεων/κατηγοριών) (π.χ. ομάδα αίματος).
 - ❑ Η **παρεμβολή (regression)** αφορά στη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών (π.χ. πρόβλεψη ισοτιμίας νομισμάτων ή τιμής μετοχής).

¹ Ο όρος *ταξινόμηση* συναντάται στην ελληνική βιβλιογραφία και ως *κατηγοριοποίηση*

1) Μάθηση Εννοιών (Concept Learning)

- ❖ Η **έννοια** (**concept**) είναι ένα υποσύνολο αντικειμένων που ορίζονται σε σχέση με ένα μεγαλύτερο σύνολο.
 - ❑ π.χ. η έννοια "πουλί" ορίζεται ως "το υποσύνολο των ζώων που έχουν φτερά"
- ❖ Εναλλακτικά, η έννοια είναι μια συνάρτηση που επιστρέφει λογική τιμή: *αληθής* για τα αντικείμενα ενός συνόλου που ανήκουν σε αυτήν και *ψευδής* για τα αντικείμενα που δεν ανήκουν.
- ❖ Η **μάθηση εννοιών** είναι τυπικό παράδειγμα επαγωγικής μάθησης κατά την οποία:
 - ❑ Το σύστημα τροφοδοτείται με παραδείγματα που ανήκουν (θετικά παραδείγματα) ή δεν ανήκουν (αρνητικά παραδείγματα) στη συγκεκριμένη έννοια.
 - ❑ Ακολούθως πρέπει να παραχθεί κάποια γενικευμένη περιγραφή της έννοιας,
 - δηλαδή να δημιουργηθεί ένα μοντέλο, ώστε να είναι δυνατό στη συνέχεια να αποφασιστεί αν μια άγνωστη περίπτωση ανήκει σε αυτήν την έννοια.
 - ❑ Ο πιο γνωστός αλγόριθμος μάθησης εννοιών είναι ο **αλγόριθμος απαλοιφής υποψηφίων**.



Ο Αλγόριθμος Απαλοιφής Υποψηφίων

Candidate Elimination Algorithm

- ❖ Περιορίζει το χώρο αναζήτησης επιτελώντας γενικεύσεις και εξειδικεύσεις σε κάποιες αρχικές υποθέσεις (έννοιες) με βάση τα δεδομένα εκπαίδευσης.
- ❖ Διατηρεί δύο σύνολα, G και S , που από κοινού περιγράφουν όλο το χώρο αναζήτησης και ορίζονται ως εξής:
 - ❑ G : το σύνολο των πιο γενικών (maximally general) υποψήφια υποθέσεων (δηλαδή εννοιών).
 - ❑ S : το σύνολο των πιο εξειδικευμένων (maximally specific) υποψήφια υποθέσεων.

Περιγραφή του Αλγορίθμου

Αρχικοποίηση:

Το G στο σύνολο όλων των υποθέσεων.

Το S στο κενό σύνολο.

Για κάθε δεδομένο εκπαίδευσης x :

Αν το x είναι θετικό:

i) Διέγραψε τα μέλη του G που δεν ικανοποιούν το x .

ii) Για κάθε υπόθεση $s \in S$ που δεν ικανοποιεί το x :

α) Διέγραψε την s από το S .

β) Πρόσθεσε στο S όλες τις ελάχιστες γενικεύσεις h της s , έτσι ώστε κάθε υπόθεση h να ικανοποιεί το x και να υπάρχει κάποια υπόθεση του G που να είναι πιο γενική.

γ) Διέγραψε από το S όποια υπόθεση είναι πιο γενική από κάποια άλλη υπόθεση του S .

Αν το x είναι αρνητικό:

i) Διέγραψε τα μέλη του S που δεν ικανοποιούν το x .

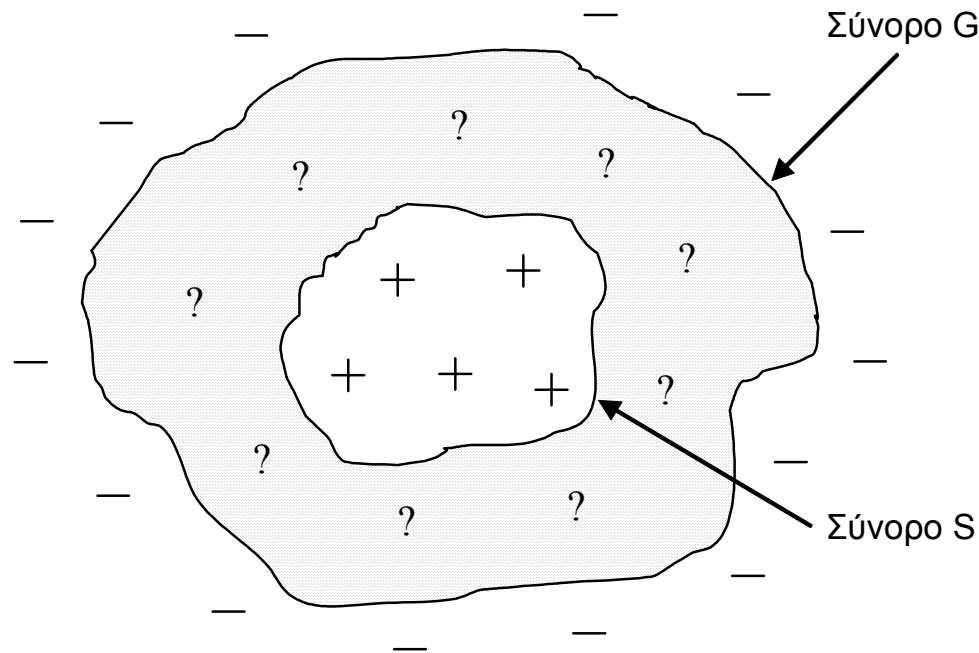
ii) Για κάθε υπόθεση $g \in G$ που δεν ικανοποιεί το x :

α) Διέγραψε την g από το G .

β) Πρόσθεσε στο G όλες τις ελάχιστες ειδικεύσεις h της g , έτσι ώστε κάθε υπόθεση h να ικανοποιεί το x και να υπάρχει κάποια υπόθεση του S που να είναι πιο ειδική.

γ) Διέγραψε από το G όποια υπόθεση είναι πιο ειδική από κάποια άλλη υπόθεση του G .

Σχηματική Περιγραφή του Τρόπου Λειτουργίας



- ❖ Τα σύνολα G και S ορίζουν κάποια σύνορα στο χώρο των υποθέσεων/εννοιών, τα οποία τον χωρίζουν σε περιοχές με θετικά, αρνητικά και απροσδιόριστης φύσης παραδείγματα.
- ❖ Κατά την εκπαίδευση το σύνορο G συρρικνώνεται ενώ το S επεκτείνεται μέχρις ότου εξαντληθούν τα παραδείγματα.
- ❖ Ο αλγόριθμος παρέχει ανά πάσα στιγμή μία αποδεκτή (αλλά όχι την καλύτερη) περιγραφή του σταδίου της εκπαίδευσης καθώς χρησιμοποιεί τα δεδομένα εκπαίδευσης σταδιακά.

Παράδειγμα Προβλήματος

- ❖ Χρησιμοποιώντας δύο θετικές και τρεις αρνητικές περιπτώσεις πελατών μιας τράπεζας που δανειοδοτήθηκαν, ζητείται μία περιγραφή της έννοιας "καλός υποψήφιος για δανειοδότηση".
 - ❑ Οι παράμετροι που έχουν καταγραφεί για τις παρελθούσες περιπτώσεις και οι δυνατές τους τιμές είναι:
 - Τρέχουσες Οφειλές: Υψηλές, Χαμηλές
 - Εισόδημα: Υψηλό, Χαμηλό
 - Παντρεμένος(η): Ναι, Όχι
 - Χαρακτηρισμός: Καλός (θετικό παράδειγμα),
Κακός (αρνητικό παράδειγμα)
- ❖ Ζητούμενο είναι η περιγραφή των πελατών με χαρακτηρισμό "Καλός" με βάση τις τρέχουσες οφειλές, το εισόδημα και την οικογενειακή τους κατάσταση, με αυτή τη σειρά.

Πελάτης	Τρέχουσες Οφειλές	Εισόδημα	Παντρεμένος(η)	Χαρακτηρισμός
1	Υψηλές	Υψηλό	Ναι	Καλός (p)
2	Χαμηλές	Υψηλό	Όχι	Κακός (n)
3	Χαμηλές	Υψηλό	Ναι	Καλός (p)
4	Υψηλές	Χαμηλό	Ναι	Κακός (n)
5	Χαμηλές	Χαμηλό	Ναι	Κακός (n)

Παράδειγμα: Εκτέλεση Αλγόριθμου Απαλοιφής Υποψηφίων

- ❖ Στο Σχήμα δίνονται οι τελικές καταστάσεις σε κάθε κύκλο εκπαίδευσης, μετά δηλαδή από τη χρήση ενός από τα δεδομένα του παρελθόντος.

G: {(X,Y,Z)}
S: {}

(Υψηλές, Υψηλό, Ναι) (#1 p)

G: {(X,Y,Z)}
S: {(Υψηλές, Υψηλό, Ναι)}

(Χαμηλές, Υψηλό, Όχι) (#2 n)

G: {(Υψηλές,Υ,Z), (X, Χαμηλό,Z), (X,Y, Ναι) }
S: {(Υψηλές, Υψηλό, Ναι)}

(Χαμηλές, Υψηλό, Ναι) (#3 p)

G: {(X,Y, Ναι) }
S: {(X, Υψηλό, Ναι)}

(Υψηλές, Χαμηλό, Ναι) (#4 n)

G: {(Χαμηλές,Υ, Ναι), (X, Υψηλό, Ναι)}
S: {(X, Υψηλό, Ναι)}

(Χαμηλές, Χαμηλό, Ναι) (#5 n)

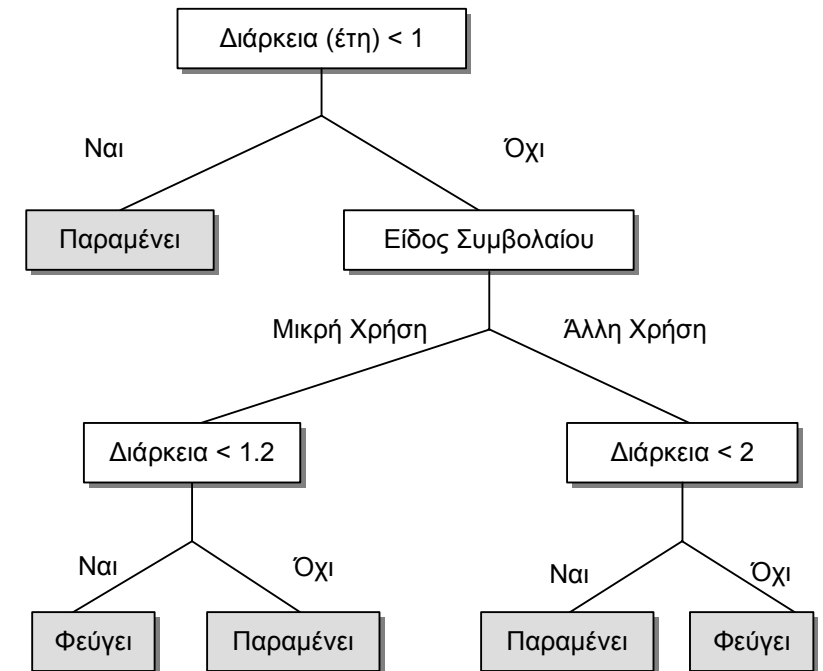
G: {(X, Υψηλό, Ναι)}
S: {(X, Υψηλό, Ναι)}

- ❖ Συμπέρασμα: ο "καλός υποψήφιος για δανειοδότηση" πρέπει να έχει σχετικά υψηλό εισόδημα και να είναι παντρεμένος.

❑ **Παρατήρηση:** οι τρέχουσες οφειλές του υποψήφιου δεν "φαίνεται" να αποτελούν αποτρεπτικό παράγοντα για δανειοδότηση (φυσικά με βάση τα δεδομένα εκπαίδευσης).

2) Δένδρα Ταξινόμησης/Απόφασης

- ❖ Δενδροειδής δομή που με γραφικό τρόπο περιγράφει τα δεδομένα.
- ❖ Έστω τα δεδομένα εταιρείας κινητής τηλεφωνίας που περιγράφουν περιπτώσεις συνδρομητών που παρέμειναν ή έφυγαν μετά τη λήξη του συμβολαίου τους, με βάση τη διάρκεια και το είδος αυτού.
- ❖ Μια αναπαράσταση σε δένδρο θα μπορούσε να έχει την μορφή του σχήματος δεξιά.
 - ❑ Κάθε κόμβος ορίζει μια συνθήκη ελέγχου της τιμής κάποιου χαρακτηριστικού των περιπτώσεων
 - ❑ Κάθε κλαδί που φεύγει από ένα κόμβο αντιστοιχεί σε μια διαφορετική διακριτή τιμή του χαρακτηριστικού που σχετίζεται με τον κόμβο.
 - ❑ Στα κλαδιά φύλλα έχουμε το τι συνέβη.
- ❖ Τα δένδρα ταξινόμησης χρησιμοποιούνται για να προβλέψουν, με κάποιο βαθμό ακρίβειας, την τιμή της μεταβλητής που μοντελοποιούν με βάση τις τιμές των θεωρούμενων ανεξάρτητων μεταβλητών (χαρακτηριστικών).



Αναπαράσταση με Κανόνες

❖ Εναλλακτικά, το δένδρο μπορεί να αναπαρασταθεί και ως σύνολο κανόνων if-then, που ονομάζονται **κανόνες ταξινόμησης (classification rules)**.

❖ Με χρήση μόνο AND προκύπτουν τόσοι κανόνες όσα και τα φύλλα του δένδρου.

□ π.χ. για τα δεδομένα της εταιρίας κινητής τηλεφωνίας προκύπτουν 5 κανόνες:

1) **if** Διάρκεια<1 **then** Παραμένει

2) **if** Διάρκεια>1 **and** Είδος_Συμβολαίου=Μικρή Χρήση **and** Διάρκεια<1.2 **then** Φεύγει

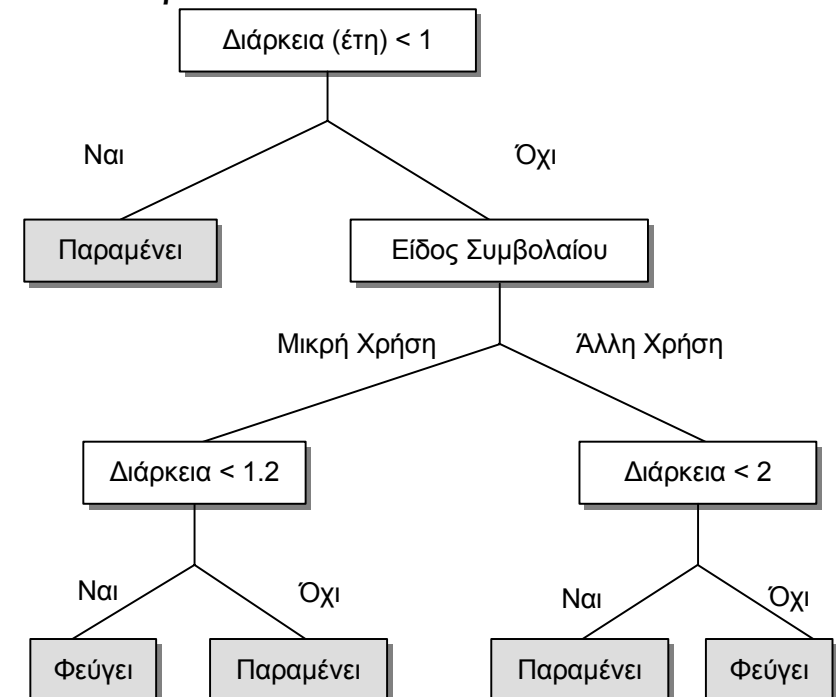
3) **if** Διάρκεια>1 **and** Είδος_Συμβολαίου=Μικρή Χρήση **and** Διάρκεια>1.2 **then** Παραμένει

4) **if** Διάρκεια>1 **and** Είδος_Συμβολαίου=Άλλη Χρήση **and** Διάρκεια<2 **then** Παραμένει

5) **if** Διάρκεια>1 **and** Είδος_Συμβολαίου=Άλλη Χρήση **and** Διάρκεια>2 **then** Φεύγει

❖ Αν γίνει χρήση και του τελεστή OR, κάποιοι από τους παραπάνω κανόνες ενοποιούνται.

□ τελικά μένουν τόσοι κανόνες όσες είναι και οι διαθέσιμες κατηγορίες.





Ο Αλγόριθμος ID3 (1/2)

- ❖ Είναι ο πιο γνωστός αλγόριθμος μάθησης δένδρων ταξινόμησης.
- ❖ Είναι αναδρομικός και στη γενική του μορφή περιγράφεται ως εξής:

1. Βρες την ανεξάρτητη μεταβλητή η οποία αν χρησιμοποιηθεί ως κριτήριο διαχωρισμού των δεδομένων εκπαίδευσης θα οδηγήσει σε κόμβους κατά το δυνατό διαφορετικούς σε σχέση με την εξαρτημένη μεταβλητή.
2. Κάνε το διαχωρισμό.
3. Επανάλαβε τη διαδικασία για κάθε έναν από τους κόμβους που προέκυψαν μέχρι να μην είναι δυνατός περαιτέρω διαχωρισμός.

- ❖ Δηλαδή ο ID3 κατασκευάζει το δένδρο άπληστα (greedy) από πάνω προς τα κάτω επιλέγοντας αρχικά το πιο κατάλληλο χαρακτηριστικό για έλεγχο στη ρίζα.
 - ❑ Η επιλογή βασίζεται σε κάποιο στατιστικό μέτρο που υπολογίζεται από τα δεδομένα.
 - ❑ Στη συνέχεια, για κάθε δυνατή τιμή του χαρακτηριστικού δημιουργούνται οι αντίστοιχοι απόγονοι της ρίζας και τα δεδομένα μοιράζονται στους νέους κόμβους ανάλογα με την τιμή που έχουν για το χαρακτηριστικό που ελέγχεται στη ρίζα.

Ο Αλγόριθμος ID3 (2/2)

- ❖ Η όλη διαδικασία επαναλαμβάνεται για κάθε νέο κόμβο. Η επιλογή όμως του κατάλληλου χαρακτηριστικού σε κάθε νέο κόμβο αποφασίζεται χρησιμοποιώντας μόνο τα δεδομένα που ανήκουν σε αυτόν τον κόμβο.
- ❖ Η διαδικασία τερματίζει όταν οι κόμβοι γίνουν τερματικοί (ή φύλλα).
- ❖ Ένας κόμβος γίνεται τερματικός όταν:
 - ❑ Όλα τα δεδομένα που ανήκουν σε αυτόν ανήκουν στην ίδια κατηγορία.
 - Η κατηγορία αυτή γίνεται και η τιμή του κόμβου.
 - Ο κόμβος ονομάζεται **αμιγής κόμβος (pure node)**.
 - ❑ Σε κάποιο βάθος τελειώσουν τα χαρακτηριστικά προς έλεγχο.
 - Τιμή του κόμβου είναι η κατηγορία στην οποία ανήκει η πλειοψηφία των δεδομένων του κόμβου αυτού.
- ❖ Το βασικότερο στάδιο του αλγορίθμου είναι η επιλογή της ανεξάρτητης μεταβλητής πάνω στην οποία θα συνεχιστεί η ανάπτυξη του δένδρου.
- ❖ Απαιτείται ο ορισμός κάποιου μηχανισμού που θα καθοδηγήσει την αναζήτηση προς το καλύτερο δένδρο (περιγραφή) μέσα στο σύνολο των δυνατών δένδρων.
- ❖ Η συνάρτηση αξιολόγησης που καθοδηγεί την αναζήτηση είναι το **Κέρδος Πληροφορίας** που με τη σειρά του βασίζεται στο μέγεθος **Εντροπία Πληροφορίας** (αναπτύσσονται παρακάτω)

Εντροπία Πληροφορίας (Information Entropy)

❖ Μέγεθος στο οποίο βασίζεται το κλασικότερο κριτήριο διαχωρισμού.

□ Η τιμή της δίνεται από τη σχέση:

$$E(S) = -p_+ \cdot \log_2(p_+) - p_- \cdot \log_2(p_-)$$

□ όπου S είναι το σύνολο των δεδομένων εκπαίδευσης στο στάδιο (κόμβο) του διαχωρισμού, p_+ είναι το κλάσμα των θετικών παραδειγμάτων του S και p_- είναι το κλάσμα των αρνητικών παραδειγμάτων του S .

❖ Γενικότερα, για c διαφορετικές κατηγορίες, η εντροπία ορίζεται από τη σχέση:

$$E(S) = -\sum_{i=1}^c p_i \cdot \log_2(p_i)$$

□ όπου p_i το ποσοστό των παραδειγμάτων του S που ανήκουν στην κατηγορία i .

Κέρδος Πληροφορίας (Information Gain)

- ❖ $\text{Gain}(S,A)$ ή $G(S,A)$ – χρησιμοποιείται ως κριτήριο διαχωρισμού.
 - ❑ Αναπαριστά τη μείωση της εντροπίας του συνόλου εκπαίδευσης S αν επιλεγεί ως παράμετρος διαχωρισμού η μεταβλητή A .
- ❖ Όταν μειώνεται η πληροφοριακή εντροπία, αυξάνεται η πυκνότητα πληροφορίας και άρα η περιγραφή γίνεται περισσότερο συμπαγής.
- ❖ Δίνεται από τη σχέση:

$$G(S, A) = E(S) - \sum_{u \in \text{Values}(A)} \frac{|S_u|}{|S|} \cdot E(S_u)$$

- ❑ $E(S)$ είναι η εντροπία πληροφορίας του υπό εξέταση κόμβου, A είναι η ανεξάρτητη μεταβλητή, με τιμές $\text{Values}(A)$, βάσει της οποίας επιχειρείται ο επόμενος διαχωρισμός, u είναι μία από τις δυνατές τιμές του A , S_u είναι το πλήθος των εγγραφών με $A=u$ και $E(S_u)$ η εντροπία πληροφορίας του υπό εξέταση κόμβου ως προς την τιμή $A=u$.

Παράδειγμα (εκτέλεση) (1/2)

- ❖ ...για το σύνολο δεδομένων εκπαίδευσης του προβλήματος δανειοδότησης
- ❖ 1^{ος} κύκλος εκτέλεσης με κριτήριο διαχωρισμού το Gain

$$❖ S_0 = 5, p_+ = 2/5, p_- = 3/5$$

$$❖ E(S) = -p_+ \cdot \log_2 p_+ - p_- \cdot \log_2 p_- =$$

$$-2/5 \cdot \log_2(2/5) - 3/5 \cdot \log_2(3/5) =$$

$$\mathbf{0.97}$$

- ❖ Αν **A=Εισόδημα**, οι δυνατές τιμές του είναι:

V(A)={Υψηλό, Χαμηλό}. Οπότε:

- ❑ Για u=υψηλό έχουμε $p_+=2/3, p_-=1/3, S_{u=υψηλό}=3$
- ❑ $E(S_{u=υψηλό}) = -p_+ \cdot \log_2 p_+ - p_- \cdot \log_2 p_- = \mathbf{0.92}$
- ❑ Για u=χαμηλό έχουμε $p_+=0, p_-=2/2=1, S_{u=χαμηλό}=2$
- ❑ $E(S_{u=χαμηλό}) = -p_+ \cdot \log_2 p_+ - p_- \cdot \log_2 p_- = \mathbf{0}$

- ❖ Υπολογίζουμε το Gain(S_0, A).

$$❑ \text{Gain}(S_0, A=\text{εισόδημα}) = 0.97 - (3/5 \cdot 0.92 + 2/5 \cdot 0) = \mathbf{0.42}$$

- ❖ Όμοια: Gain($S_0, A=\text{Χρεός}$)=**0.02** και Gain($S_0, A=\text{Παντρεμένος}$)=**0.17**

- ❖ Άρα ο κόμβος-ρίζα θα διαχωριστεί με βάση τις τιμές του πεδίου **Εισοδημα** καθώς το Κέρδος Πληροφορίας (άρα και η μείωση της Εντροπίας Πληροφορίας) για αυτό το πεδίο ήταν το μεγαλύτερο.

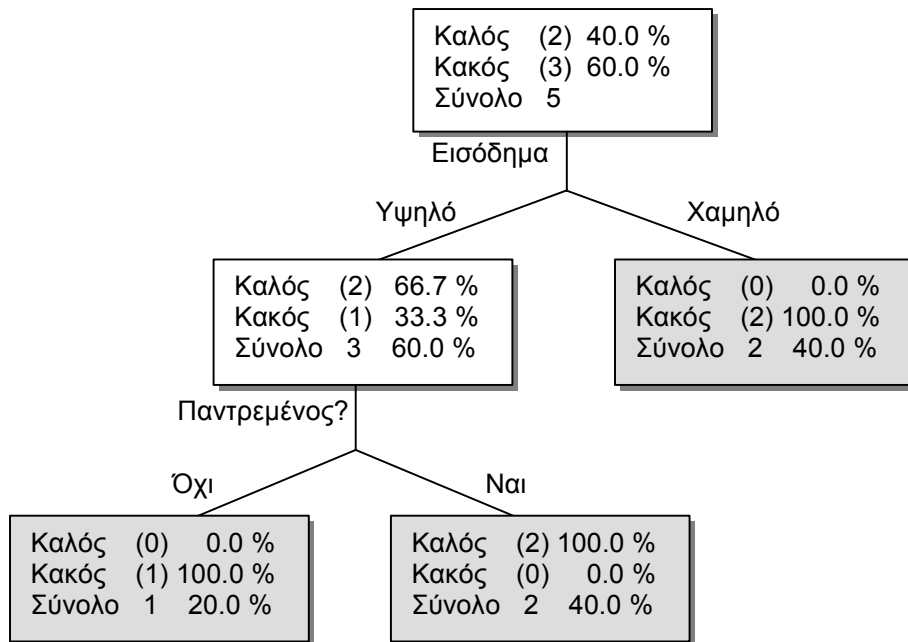
	Σύνολα	Εισόδημα		Χρεός		Παντρεμένος	
		Υψηλό	Χαμηλό	Υψηλό	Χαμηλό	Ναι	Όχι
S	5	3	2	2	3	4	1
+	2	2	0	1	1	2	0
-	3	1	2	1	2	2	1
p+	2/5	2/3	0	1/2	1/3	1/2	0
p-	3/5	1/3	1	1/2	2/3	1/2	1
E	0.97	0.92	0.00	1.00	0.92	1.00	0.00
Gain		0.42		0.02		0.17	

Παράδειγμα (εκτέλεση) (2/2)

❖ Παρατηρήσεις:

- ❑ Σε κάθε κύκλο τα παραπάνω μεγέθη επαναυπολογίζονται για τον πληθυσμό S των δεδομένων εκπαίδευσης που ανήκουν στον υπό εξέταση κόμβο.
- ❑ Σε δύο διαφορετικούς κόμβους μπορεί να επιλεγεί διαφορετική μεταβλητή διαχωρισμού.

❖ Το τελικό δένδρο απόφασης στο πρόβλημα δανειοδότησης



❖ Η ανάπτυξη του δένδρου τερματίστηκε διότι οι τερματικοί του κόμβοι έχουν την ίδια τιμή στην εξαρτημένη μεταβλητή για όλα τα παραδείγματα τους, δηλαδή είναι αμιγείς (φυσική συνθήκη τερματισμού ανάπτυξης).

❖ Άλλη φυσική συνθήκη τερματισμού:

- ❑ Όταν δεν υπάρχουν άλλες, προς εξέταση, μεταβλητές.

3) Μάθηση Κανόνων Ταξινόμησης

- ❖ Οι κανόνες if-then είναι από τις πιο εκφραστικές και κατανοητές για τον άνθρωπο αναπαραστάσεις.
- ❖ Κυριότερες κατηγορίες κανόνων
 - ❑ **προτασιακοί** (propositional rules)
 - ❑ κατηγορηματικοί κανόνες πρώτης τάξης (first order predicate rules)
- ❖ Οι **προτασιακοί κανόνες**:
 - ❑ Μπορεί να προκύψουν από άλλες μορφές αναπαράστασης (π.χ. δένδρα, γενετικούς αλγόριθμους) αλλά και από απ' ευθείας μάθηση με αλγόριθμους σειριακής κάλυψης.
 - ❑ Δεν περιλαμβάνουν μεταβλητές και έτσι δεν μπορεί να αναπαραστήσουν γενικές σχέσεις ανάμεσα στις τιμές των χαρακτηριστικών.
 - ❑ π.χ. ο παρακάτω κανόνας ισχύει για μία συγκεκριμένη οικογένεια:
 - if Father1=Bob and Name2=Bob and Female1=true then Daughter1_2=true
- ❖ Οι **κατηγορηματικοί κανόνες πρώτης τάξης**:
 - ❑ Περιέχουν μεταβλητές (μεγάλη εκφραστική ικανότητα).
 - ❑ Προκύπτουν με απ'ευθείας μάθηση μέσω αλγορίθμων μάθησης κανόνων 1ης τάξης.
 - ❑ π.χ. ο επόμενος κανόνας ισχύει για όλες τις οικογένειες:
 - if father(Y,X) and female(Y) then daughter(X,Y)

Μάθηση Προτασιακών Κανόνων Ταξινόμησης (1/2)

- ❖ Αφορά σε προβλήματα όπου δεν απαιτείται η αναπαράσταση σχέσεων ανάμεσα στις τιμές των διαφόρων χαρακτηριστικών (όπως και στα δένδρα ταξινόμησης/απόφασης).
- ❖ Ένας γενικός αλγόριθμος μάθησης τέτοιων κανόνων είναι ο **αλγόριθμος σειριακής κάλυψης** (Sequential Covering Algorithm).
 - ❑ Δημιουργεί ένα σύνολο προτασιακών κανόνων σταδιακά (incrementally) μαθαίνοντας έναν κανόνα κάθε φορά.
 - ❑ Κάθε κανόνας καλύπτει ένα σύνολο θετικών παραδειγμάτων, που είναι ξένο προς τα σύνολα των υπολοίπων κανόνων.
 - Ένας κανόνας καλύπτει ένα παράδειγμα όταν οι τιμές των χαρακτηριστικών της συνθήκης του κανόνα συμφωνούν με τις αντίστοιχες τιμές του παραδείγματος.
 - ❑ Το σύνολο των κανόνων καλύπτει το σύνολο των θετικών παραδειγμάτων.
- ❖ Τα βήματα του αλγορίθμου είναι:

Αλγόριθμος Σειριακής Κάλυψης

1. Αρχικοποίησε το Σύνολο_Κανόνων με το κενό σύνολο.
2. *Μάθε_έναν_Κανόνα(Εξαρτημένη_Μεταβλητή, Μεταβλητές, Παραδείγματα)* .
3. Αν ο Κανόνας ικανοποιεί το Κριτήριο Απόδοσης:
 - 3α. Αφαίρεσε τα θετικά παραδείγματα που κάλυψε ο Κανόνας αυτός.
 - 3β. Πρόσθεσε τον Κανόνα στο Σύνολο_Κανόνων.
4. Επανάλαβε από το 2, όσο ικανοποιείται το Κριτήριο Απόδοσης.

Μάθηση Προτασιακών Κανόνων Ταξινόμησης (2/2)

- ❖ Η συνάρτηση **Μάθε_έναν_Κανόνα** είναι κρίσιμη για την απόδοση του αλγορίθμου και έχει την παρακάτω μορφή:

Συνάρτηση Μάθε_έναν_Κανόνα (Εξαρτημένη_Μεταβλητή, Χαρακτηριστικά, Παραδείγματα)
//Αναζήτηση "Γενικό προς Ειδικό" (*General-to-Specific Search*)

1. Έστω η βέλτιστη υπόθεση (αρχικά, ο πιο γενικός κανόνας) που ταιριάζει με όλα τα παραδείγματα του συνόλου εκπαίδευσης.
2. Επανάλαβε, όσο υπάρχουν υποψήφια υποθέσεις:
Εξειδίκευσε τη βέλτιστη υπόθεση, προσθέτοντας το ζεύγος χαρακτηριστικού-τιμής που βελτιστοποιεί το κριτήριο απόδοσης.

- ❖ Το Κριτήριο_Απόδοσης είναι ένα μέτρο της απόδοσης του κανόνα που καθορίζει ο χρήστης, όπως η Εντροπία, η Σχετική Συχνότητα ή ο m-Εκτιμητής Ακρίβειας.

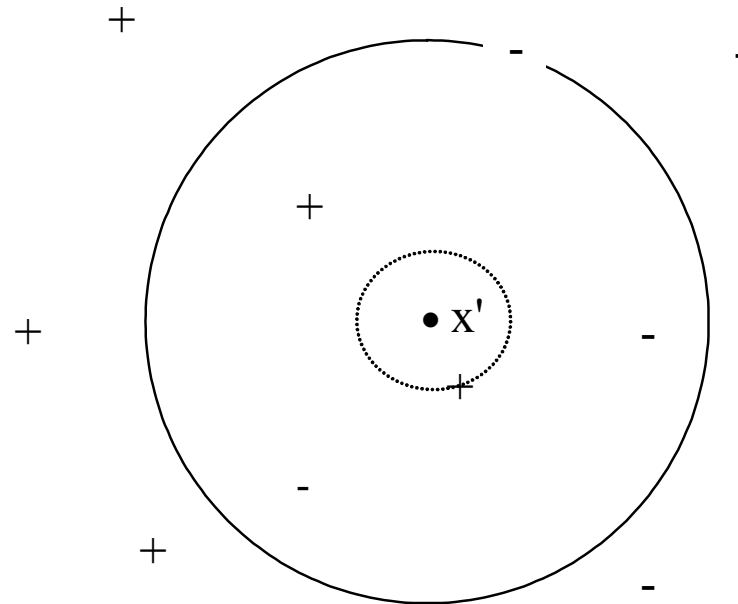
4) Μάθηση κατά Περίπτωση (Instance-based Learning)

- ❖ Τα δεδομένα εκπαίδευσης διατηρούνται αυτούσια...
 - ❑ ...σε αντίθεση με τις άλλες μεθόδους μηχανικής μάθησης οι οποίες κωδικοποιούν τα παραδείγματα εκπαίδευσης σε μια συμπαγή περιγραφή.
- ❖ Όταν ένα τέτοιο σύστημα κληθεί να αποφασίσει για την κατηγορία μιας νέας περίπτωσης, εξετάζει εκείνη τη στιγμή τη σχέση της με τα ήδη αποθηκευμένα παραδείγματα.
- ❖ Χαρακτηριστικότερος είναι ο αλγόριθμος των k-κοντινότερων γειτόνων (k-Nearest Neighbors)
 - ❑ Κάνει την παραδοχή ότι τα διάφορα παραδείγματα μπορεί να αναπαρασταθούν ως σημεία σε κάποιον n-διάστατο Ευκλείδειο χώρο R^n όπου n ο αριθμός των χαρακτηριστικών (ανεξάρτητων μεταβλητών).
 - ❑ Κάθε νέα περίπτωση τοποθετείται στο χώρο αυτό ως νέο σημείο και η τιμή του προσδιορίζεται με βάση το χαρακτηρισμό των k γειτονικών σημείων.
 - ❑ Οι κοντινότεροι γείτονες μιας περίπτωσης υπολογίζονται με βάση την Ευκλείδεια απόστασή τους.

$$d(x, x') = \sqrt{\sum_{r=1}^n (a_r(x) - a_r(x'))^2}$$

Παράδειγμα

Προσδιορισμός κατηγορίας με βάση τον 1 και τους 5 κοντινότερους γείτονες



- ❖ Στο Σχήμα όπου υπάρχουν παραδείγματα δυο κατηγοριών, η νέα περίπτωση x' χαρακτηρίζεται ως θετική, αν ληφθεί υπ' όψη μόνο ο πλησιέστερος γείτονας (1-Nearest Neighbor) και ως αρνητική αν ληφθούν υπ' όψη οι πέντε πλησιέστεροι γείτονες (5-Nearest Neighbors) καθώς η πλειοψηφία αυτών έχει αρνητικό χαρακτηρισμό (εξωτερικός κύκλος στο σχήμα).

5) Μάθηση κατά Bayes

- ❖ Στη μάθηση κατά Bayes (Bayesian learning) κάθε παράδειγμα εκπαίδευσης μπορεί σταδιακά να μειώσει ή να αυξήσει την πιθανότητα να είναι σωστή μια υπόθεση.
- ❖ Μια πρακτική δυσκολία στην εφαρμογή της μάθησης κατά Bayes είναι η απαίτηση για τη γνώση πολλών τιμών πιθανοτήτων.
- ❖ Όταν αυτές οι τιμές δεν είναι δυνατό να υπολογιστούν επακριβώς, υπολογίζονται κατ' εκτίμηση από παλαιότερες υποθέσεις, εμπειρική γνώση, κτλ.
- ❖ Η παραπάνω δυσκολία εφαρμογής έχει δώσει μεγάλη πρακτική αξία σε μια απλουστευμένη εκδοχή της μάθησης κατά Bayes, τον **απλό ταξινομητή Bayes**, στον οποίο γίνεται η παραδοχή ότι τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους.

Απλός ταξινομητής Bayes

- ❖ Ο απλός ταξινομητής Bayes (simple/naive Bayes classifier) είναι μια πρακτική μέθοδος μάθησης που στηρίζεται σε στατιστικά στοιχεία (κατανομές πιθανότητας).
- ❖ Η ποσότητα P που περιγράφει έναν απλό ταξινομητή Bayes για ένα σύνολο παραδειγμάτων εκφράζει την πιθανότητα να είναι c η τιμή της εξαρτημένης μεταβλητής C με βάση τις τιμές $x=(x_1, x_2, \dots, x_n)$ των χαρακτηριστικών $X=(X_1, X_2, \dots, X_n)$ και δίνεται από τη σχέση:

$$P(c | x) = P(c) \cdot \prod_i P(x_i | c)$$

- όπου τα χαρακτηριστικά X_i θεωρούνται ανεξάρτητα μεταξύ τους.
- ❖ Ο υπολογισμός της παραπάνω ποσότητας για ένα σύνολο N παραδειγμάτων γίνεται με βάση τις σχέσεις:
 - $P(c) = N(c) / N$,
 - $P(x_i|c) = N(x_i,c) / N(c)$, για χαρακτηριστικό X_i με διακριτές τιμές,
 - $P(x_i|c) = g(x_i, \mu_c, \sigma_c^2)$, για χαρακτηριστικό X_i με αριθμητικές τιμές,
 - $N(c)$ είναι ο αριθμός των παραδειγμάτων που έχουν στην εξαρτημένη μεταβλητή την τιμή c , $N(x_i,c)$ είναι ο αριθμός των παραδειγμάτων που έχουν για το χαρακτηριστικό X_i και την εξαρτημένη μεταβλητή, τιμές x_i και c αντίστοιχα, και $g(x_i, \mu_c, \sigma_c^2)$ είναι η συνάρτηση πυκνότητας πιθανότητας Gauss με μέσο όρο μ_c και διασπορά σ_c για το χαρακτηριστικό X_i .

6) Παρεμβολή ή Παλινδρόμηση (regression)

- ❖ Η διαδικασία προσδιορισμού της σχέσης μιας μεταβλητής y (εξαρτημένη μεταβλητή ή έξοδος) με μια ή περισσότερες άλλες μεταβλητές x_1, x_2, \dots, x_n (ανεξάρτητες μεταβλητές ή είσοδοι).
 - ❑ Σκοπός της είναι η πρόβλεψη της τιμής της εξόδου όταν είναι γνωστές οι είσοδοι.
- ❖ Το πιο διαδεδομένο μοντέλο είναι το **γραμμικό** (linear) όπου η αναμενόμενη τιμή της εξόδου μοντελοποιείται με μία γραμμική συνάρτηση ή σταθμισμένο άθροισμα (weighted sum) των παραμέτρων εισόδου.
 - $y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_n x_{nj} \quad j = 1, 2, \dots, m$
 - ❑ όπου m είναι ο αριθμός των δεδομένων (παραδειγμάτων) εκπαίδευσης, ενώ το ζητούμενο είναι να υπολογιστούν οι συντελεστές β_i .
 - ❑ Διαδεδομένη μέθοδος επίλυσης: η μέθοδος των ελαχίστων τετραγώνων (least squares).
 - Ελαχιστοποιεί το σφάλμα μεταξύ της εκτιμώμενης συνάρτησης και των πραγματικών δεδομένων.

7) Νευρωνικά Δίκτυα (Neural Networks)

- ❖ Παρέχουν ένα πρακτικό (εύκολο) τρόπο για την εκμάθηση αριθμητικών και διανυσματικών συναρτήσεων ορισμένων σε συνεχή ή διακριτά μεγέθη.
 - ❑ Χρησιμοποιούνται τόσο για παρεμβολή (γραμμική και μη γραμμική) όσο και για ταξινόμηση.
 - ❑ Έχουν το μεγάλο πλεονέκτημα της ανοχής που παρουσιάζουν σε δεδομένα εκπαίδευσης με θόρυβο, δηλαδή δεδομένα που περιστασιακά έχουν λανθασμένες τιμές (π.χ. λάθη καταχώρησης).
 - ❑ Αδυνατούν όμως να εξηγήσουν ποιοτικά τη γνώση που μοντελοποιούν.
- ❖ Λόγω της σημαντικότητάς τους, τα νευρωνικά δίκτυα παρουσιάζονται λεπτομερώς σε άλλο κεφάλαιο.

8) Μηχανές Διανυσμάτων Υποστήριξης (ΜΔΥ) Support Vector Machines (SVMs)

- ❖ Στηρίζονται στη Θεωρία Στατιστικής Μάθησης (Statistical Learning Theory) και στα νευρωνικά δίκτυα τύπου Perceptron
 - ❑ Προτάθηκαν από τον Vladimir Vapnik και τους συνεργάτες του το 1992.
 - ❑ Η γενικότερη ιδέα στην οποία στηρίζονται είχε προταθεί αρκετά νωρίτερα (δεκαετία '60)
- ❖ Έχουν εδραιωθεί ως μια από τις πιο διαδεδομένες μεθόδους (γραμμικής και μη) παρεμβολής και ταξινόμησης, με πλήθος εφαρμογών.
 - ❑ αναγνώριση γραφής (handwriting recognition),
 - ❑ ταξινόμηση κειμένων (text categorization),
 - ❑ ταξινόμηση δεδομένων έκφρασης γονιδίων (gene expression data),
 - ❑ κτλ
- ❖ Στην περίπτωση της ταξινόμησης, οι ΜΔΥ προσπαθούν να βρουν μια υπερ-επιφάνεια (hypersurface) που να διαχωρίζει στο χώρο των παραδειγμάτων τα αρνητικά από τα θετικά παραδείγματα.
 - ❑ Η υπερεπιφάνεια αυτή επιλέγεται έτσι, ώστε να απέχει όσο το δυνατόν περισσότερο από τα κοντινότερα θετικά και αρνητικά παραδείγματα (maximum margin hypersurface).

B) Μάθηση Χωρίς Επίβλεψη

- ❖ Στη μάθηση χωρίς επίβλεψη το σύστημα έχει στόχο να ανακαλύψει συσχετίσεις και ομάδες από τα δεδομένα, βασιζόμενο μόνο στις ιδιότητές τους.
- ❖ Σαν αποτέλεσμα προκύπτουν πρότυπα (περιγραφές), κάθε ένα από τα οποία περιγράφει ένα μέρος από τα δεδομένα.
- ❖ Παραδείγματα προτύπων πληροφόρησης είναι
 - ❑ οι κανόνες συσχέτισης (association rules) και
 - ❑ οι ομάδες (clusters), οι οποίες προκύπτουν από τη διαδικασία της ομαδοποίησης (clustering).

1) Κανόνες Συσχέτισης (1/2)

- ❖ Η ανακάλυψη ή εξόρυξη κανόνων συσχέτισης (association rule mining) εμφανίστηκε αρκετά αργότερα από τη μηχανική μάθηση και έχει περισσότερες επιρροές από την ερευνητική περιοχή των βάσεων δεδομένων.
 - ❑ Προτάθηκε στις αρχές της δεκαετίας του '90 από τον Rakesh Agrawal ως τεχνική ανάλυσης καλαθιού αγορών (market basket analysis) όπου το ζητούμενο είναι η ανακάλυψη συσχετίσεων ανάμεσα στα αντικείμενα μιας βάσης δεδομένων.
- ❖ Στο συγκεκριμένο πρόβλημα υπάρχει ένας μεγάλος αριθμός αντικειμένων (items), για παράδειγμα ψωμί, γάλα, κτλ. Οι πελάτες γεμίζουν τα καλάθια τους με κάποιο υποσύνολο αυτών των αντικειμένων και το ζητούμενο είναι να βρεθεί ποια από αυτά τα αντικείμενα αγοράζονται μαζί, χωρίς να ενδιαφέρει ποιος είναι ο αγοραστής.
- ❖ Οι κανόνες συσχέτισης είναι προτάσεις της μορφής $\{X_1, \dots, X_n\} \rightarrow Y$, που σημαίνει ότι αν βρεθούν όλα τα X_1, \dots, X_n στο καλάθι (στην ανάλυση καλαθιού αγορών) τότε είναι πιθανό να βρεθεί και το Y .
 - ❑ Για παράδειγμα, ένας τέτοιος κανόνας θα μπορούσε να λέει:
 - "όποιος αγοράζει καφέ (X_1) και ζάχαρη (X_2) αγοράζει και αναψυκτικά (Y)"

Κανόνες Συσχέτισης (2/2)

- ❖ Απλή αναφορά ενός τέτοιου κανόνα δεν έχει μεγάλη αξία αν δε συνοδεύεται από **κάποια ποσοτικά μεγέθη** που μετρούν την ποιότητα των ευρεθέντων κανόνων συσχέτισης.
- ❖ Τέτοια μεγέθη είναι τα:
 - **Υποστήριξη** (support) ή κάλυψη (coverage): εκφράζει την πιθανότητα να βρεθεί το καλάθι $\{X_1, \dots, X_n, Y\}$ στη βάση δεδομένων και ισούται με το λόγο των εγγραφών που περιλαμβάνουν το $\{X_1, \dots, X_n, Y\}$ προς το σύνολο των εγγραφών.
 - **Εμπιστοσύνη** (confidence) ή ακρίβεια (accuracy): εκφράζει την πιθανότητα να βρεθεί το Y σε ένα καλάθι που περιέχει τα $\{X_1, \dots, X_n\}$ και ισούται με το λόγο των εγγραφών που περιλαμβάνουν το $\{X_1, \dots, X_n, Y\}$ προς το σύνολο των εγγραφών που περιλαμβάνουν τα X_i .

Αλγόριθμοι Εύρεσης Κανόνων Συσχέτισης

- ❖ Για την ανακάλυψη κανόνων συσχέτισης χρησιμοποιείται η ιδιότητα της μονοτονίας (monotonicity property) ή αλλιώς ιδιότητα **a priori** σύμφωνα με την οποία:

"Αν ένα σύνολο αντικειμένων S είναι συχνό, τότε όλα τα υποσύνολα του S είναι επίσης συχνά".

- π.χ. αν είναι συχνό το {γάλα, ψωμί, λάδι} τότε είναι τουλάχιστον εξίσου συχνό και το {γάλα, ψωμί} (ή το {γάλα, λάδι}, ή το {ψωμί, λάδι})
- ❖ Συχνό είναι ένα σύνολο αντικειμένων όταν εμφανίζεται σε ποσοστό των καλαθιών ίσο ή μεγαλύτερο από ένα όριο που συνήθως ορίζει ο χρήστης.
- ❖ Σε έναν αλγόριθμο εύρεσης κανόνων συσχέτισης μας ενδιαφέρει κυρίως ο αριθμός των περασμάτων στα δεδομένα που απαιτείται κατά την εκτέλεσή του.

Ο αλγόριθμος Apriori

- ❖ Προτάθηκε από τον Rakesh Agrawal το 1994 και είναι ίσως ο κλασικότερος αλγόριθμος ανακάλυψης κανόνων συσχέτισης.
 - ❑ Περιλαμβάνει δυο βασικά βήματα, τη δημιουργία των συχνών συνόλων αντικειμένων και τη δημιουργία των κανόνων συσχέτισης.
- ❖ Η διαδικασία της δημιουργίας συχνών συνόλων αντικειμένων περιλαμβάνει δύο στάδια:
 - ❑ Αρχικά δημιουργείται ένα σύνολο υποψήφιων συχνών αντικειμένων C_i και στη συνέχεια, χρησιμοποιώντας το όριο υποστήριξης (support), δημιουργείται το σύνολο των συχνών συνόλων αντικειμένων L_i .
 - ❑ Η διαδικασία επαναλαμβάνεται πραγματοποιώντας διαδοχικά περάσματα στα δεδομένα μέχρι να βρεθούν είτε τα συχνά σύνολα αντικειμένων ενός προκαθορισμένου επιπέδου ή τα μέγιστα συχνά σύνολα αντικειμένων.
 - ❑ Το πρώτο στάδιο επιπλέον αποτελείται από ένα βήμα συνένωσης (join step) και ένα βήμα κλαδέματος (prune step) τα οποία συνήθως εκτελούνται στη μνήμη και έτσι δεν είναι ιδιαίτερα χρονοβόρα.
- ❖ Για τη δημιουργία των κανόνων συσχέτισης ελέγχεται η εμπιστοσύνη (confidence) όλων των πιθανών κανόνων που προκύπτουν από τα μέγιστα συχνά σύνολα αντικειμένων και στο τέλος μένουν εκείνοι των οποίων η εμπιστοσύνη ξεπερνά το όριο που τέθηκε από το χρήστη.

Δημιουργία Συχνών Συνόλων Αντικειμένων

- ❖ Έστω ότι το όριο υποστήριξης είναι sup . Στο πρώτο πέρασμα βρίσκονται τα αντικείμενα εκείνα που εμφανίζονται στη βάση δεδομένων σε ένα ποσοστό sup των καλαθιών ή μεγαλύτερο.
 - ❑ Αυτό το σύνολο ονομάζεται **σύνολο συχνών αντικειμένων** (frequent 1- itemset) και συμβολίζεται με $L1$.
 - ❑ Από το $L1$ προκύπτουν, κάνοντας όλους τους δυνατούς συνδυασμούς, τα υποψήφια (συχνά) ζεύγη αντικειμένων $C2$.
- ❖ Στο δεύτερο πέρασμα, τα ζεύγη του $C2$ των οποίων το πλήθος ικανοποιεί το κριτήριο sup , δημιουργούν το $L2$, δηλαδή τα συχνά ζεύγη (*frequent 2-itemsets*).
 - ❑ Από το $L2$ προκύπτουν οι υποψήφιες (συχνές) τριάδες αντικειμένων $C3$ που θα χρησιμοποιηθούν στο τρίτο πέρασμα.
 - ❑ Οι υποψήφιες τριάδες $C3$ είναι σύνολα του τύπου $\{A,B,C\}$ τέτοια, ώστε όλα τα υποσύνολά του μεγέθους δυο, δηλαδή τα $\{A,B\}$, $\{A,C\}$, $\{B,C\}$, να ανήκουν στο $L2$.
- ❖ Στο τρίτο πέρασμα, υπολογίζεται ο αριθμός των εμφανίσεων των τριάδων του $C3$ και με βάση το κριτήριο sup δημιουργείται το $L3$.
- ❖ Η διαδικασία συνεχίζεται για προκαθορισμένο αριθμό επιπέδων ή μέχρι να αδειάσουν τα υποψήφια συχνά σύνολα αντικειμένων και να δημιουργηθούν τα μέγιστα συχνά σύνολα αντικειμένων.

Παράδειγμα Εφαρμογής του Apriori (1/4)

- ❖ Έστω ένα σύνολο δεδομένων που αντιστοιχούν σε 10 διαφορετικά καλάθια αγορών από ένα super market.
 - ❑ Κάθε καλάθι περιλαμβάνει ένα υποσύνολο των προϊόντων του super market.
 - ❑ Για παράδειγμα, στο πρώτο καλάθι ο πελάτης αγόρασε μόνο ψωμί και γάλα.

Καλάθι	Ψωμί	Καφές	Γάλα	Ζάχαρη
#1	1	0	1	0
#2	0	1	0	0
#3	1	0	1	1
#4	0	1	0	1
#5	1	0	1	1
#6	1	1	1	0
#7	1	0	0	1
#8	1	1	1	1
#9	0	0	1	1
#10	1	1	0	1

- ❖ Έστω επίσης ότι η ζητούμενη υποστήριξη είναι $sup=40\%$ και η ζητούμενη εμπιστοσύνη $conf=80\%$.

Παράδειγμα Εφαρμογής του Apriori (2/4)

- ❖ Στο πρώτο βήμα, ο Apriori υπολογίζει την υποστήριξη όλων των αντικειμένων, δηλαδή δημιουργεί το σύνολο $L1$.
 - $S\{\Psi\omega\mu\acute{\iota}\} = 7/10 = 70\% \geq sup$
 - $S\{Καφές\} = 5/10 = 50\% \geq sup$
 - $S\{Γάλα\} = 6/10 = 60\% \geq sup$
 - $S\{Ζάχαρη\} = 7/10 = 70\% \geq sup$
- Συνεπώς $L1 = \{ \Psi\omega\mu\acute{\iota}, Καφές, Γάλα, Ζάχαρη \}$
- ❖ Στο δεύτερο βήμα, παράγονται όλοι οι δυνατοί συνδυασμοί των αντικειμένων, για να δημιουργηθεί το σύνολο υποψήφιων ζευγών αντικειμένων, δηλαδή το σύνολο $C2$.
 - $C2 = \{ \{ \Psi\omega\mu\acute{\iota}, Καφές \}, \{ \Psi\omega\mu\acute{\iota}, Γάλα \}, \{ \Psi\omega\mu\acute{\iota}, Ζάχαρη \}, \{ Καφές, Γάλα \},$
 - $\{ Καφές, Ζάχαρη \}, \{ Γάλα, Ζάχαρη \} \}$
- Κατόπιν, υπολογίζεται η υποστήριξη των μελών του $C2$ και απορρίπτονται εκείνα που δεν ξεπερνούν το όριο ελάχιστης υποστήριξης, ώστε να δημιουργηθεί το σύνολο συχνών ζευγών $L2$.
 - $S(\{ \Psi\omega\mu\acute{\iota}, Καφές \}) = 3/10 = 30\% < sup$ (απορρίπτεται)
 - $S(\{ \Psi\omega\mu\acute{\iota}, Γάλα \}) = 5/10 = 50\% \geq sup$
 - κτλ.
- Τελικά: $L2 = \{ \{ \Psi\omega\mu\acute{\iota}, Γάλα \}, \{ \Psi\omega\mu\acute{\iota}, Ζάχαρη \}, \{ Γάλα, Ζάχαρη \} \}$

Παράδειγμα Εφαρμογής του Apriori (3/4)

- ❖ Από το $L2$, με τον ίδιο τρόπο, θα δημιουργηθούν τα $C3$ και $L3$ (αν τελικά υπάρχουν "συχνές" τριάδες). Στο συγκεκριμένο παράδειγμά, το βήμα δημιουργίας υποψήφιων τριάδων έχει ως εξής:
 - ❑ Βήμα συνένωσης: $\{\Psiωμί, Γάλα\} \cup \{\Psiωμί, Ζάχαρη\} = \{\Psiωμί, Γάλα, Ζάχαρη\}$
 - ❑ Βήμα κλαδέματος: Οι επιμέρους δυάδες (ζεύγη) του $\{\Psiωμί, Γάλα, Ζάχαρη\}$ ανήκουν όλες στο $L2$, άρα:
 - $C3 = \{\{\Psiωμί, Γάλα, Ζάχαρη\}\}$
 - $S(\{\Psiωμί, Γάλα, Ζάχαρη\}) = 3/10 = 30\%$ (απορρίπτεται), άρα $L3 = \{\}$.
- ❖ Ο αλγόριθμος εύρεσης συχνών συνόλων αντικειμένων σταματά εδώ και συνεπώς το μέγιστο συχνό σύνολο αντικειμένων είναι το $L2$.
- ❖ Το επόμενο βήμα είναι η εξαγωγή των κανόνων από τα συχνά σύνολα (στο συγκεκριμένο παράδειγμα μόνο το $L2$), βάσει της εμπιστοσύνης τους.
 - $L2 = \{\{\Psiωμί, Γάλα\}, \{\Psiωμί, Ζάχαρη\}, \{\Gammaάλα, Ζάχαρη\}\}$
- ❖ Ελέγχεται η εμπιστοσύνη όλων των πιθανών κανόνων που μπορεί να προκύψουν από το $L2$.
 - ❑ $\{\Psiωμί, Γάλα\}$
 - $\Psiωμί \rightarrow \Gammaάλα$: εμπιστοσύνη = $5/7 = 71\% < conf$ (απορρίπτεται)
 - $\Gammaάλα \rightarrow \Psiωμί$: εμπιστοσύνη = $5/6 = 83\% > conf$ (εγκρίνεται)

Παράδειγμα Εφαρμογής του Apriori (4/4)

□ {Ψωμί, Ζάχαρη}

- $\Psi\omega\acute{\mu}\acute{\iota} \rightarrow \text{Ζάχαρη}$: εμπιστοσύνη = $5/7 = 71\% < conf$ (απορρίπτεται)
- $\text{Ζάχαρη} \rightarrow \Psi\omega\acute{\mu}\acute{\iota}$: εμπιστοσύνη = $5/7 = 71\% < conf$ (απορρίπτεται)

□ {Γάλα, Ζάχαρη}

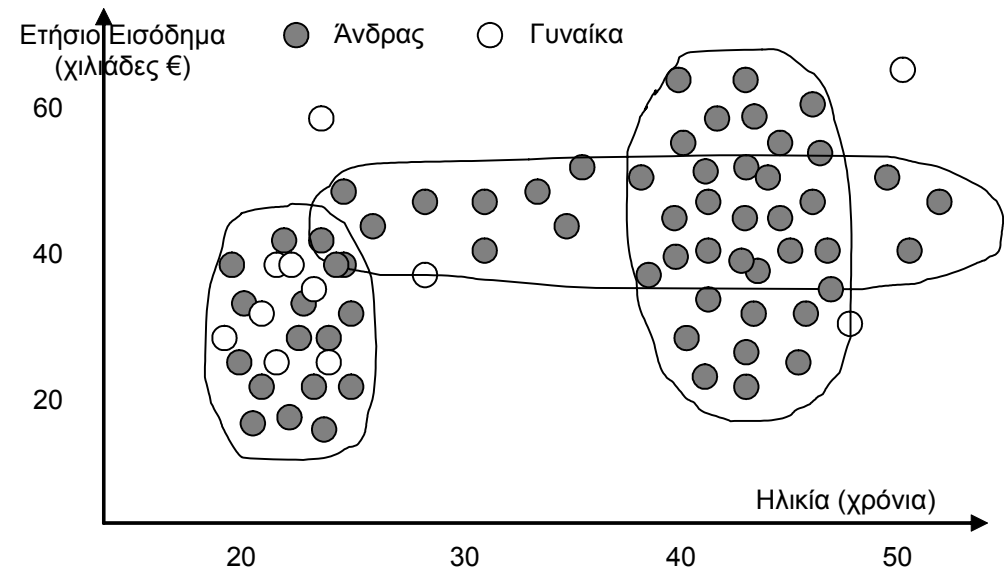
- $\Gamma\acute{\alpha}\lambda\alpha \rightarrow \text{Ζάχαρη}$: εμπιστοσύνη = $4/6 = 66\% < conf$ (απορρίπτεται)
- $\text{Ζάχαρη} \rightarrow \Gamma\acute{\alpha}\lambda\alpha$: εμπιστοσύνη = $4/7 = 57\% < conf$ (απορρίπτεται)

❖ Τελικά παράγεται μόνο ο κανόνας: $\Gamma\acute{\alpha}\lambda\alpha \rightarrow \Psi\omega\acute{\mu}\acute{\iota}$, δηλαδή όποιος αγοράζει $\Gamma\acute{\alpha}\lambda\alpha$ αγοράζει και $\Psi\omega\acute{\mu}\acute{\iota}$.

- Αν ελαττώσουμε τη ζητούμενη εμπιστοσύνη στο 70% τότε θα παραχθούν τέσσερις κανόνες.

2) Ομάδες (clusters)

- ❖ Είναι πρότυπα πληροφόρησης που προκύπτουν με **ομαδοποίηση** (clustering) δηλαδή διαχωρισμό ενός συνόλου (συνήθως πολυδιάστατων) δεδομένων σε ομάδες, ώστε:
 - ❑ σημεία που ανήκουν στην ίδια ομάδα να μοιάζουν όσο το δυνατόν περισσότερο και
 - ❑ σημεία που ανήκουν σε διαφορετικές ομάδες να διαφέρουν όσο το δυνατόν περισσότερο.
- ❖ Στο Σχήμα απεικονίζεται γραφικά μία υποθετική ομαδοποίηση σε δεδομένα αγοραστών σπορ αυτοκινήτων, με βάση την ηλικία (άξονας x), το ετήσιο εισόδημα (άξονας y) και το φύλο. Διακρίνονται τρεις ομάδες:
 - "αγοραστές νεαρής ηλικίας ανεξαρτήτως φύλλου",
 - "άνδρες αγοραστές με υψηλό εισόδημα, όλων των ηλικιών μέχρι τα 53 χρόνια" και
 - "άνδρες αγοραστές ηλικίας περίπου 44 ανεξαρτήτως εισοδήματος".



Αλγόριθμοι Ομαδοποίησης

- ❖ Υπάρχουν τρεις γενικές κατηγορίες αλγορίθμων ομαδοποίησης:
 - ❑ Οι αλγόριθμοι βασισμένοι σε διαχωρισμούς (partition based), που προσπαθούν να βρουν τον καλύτερο διαχωρισμό ενός συνόλου δεδομένων σε ένα συγκεκριμένο αριθμό ομάδων.
 - ❑ Οι ιεραρχικοί (hierarchical) αλγόριθμοι, που προσπαθούν με ιεραρχικό τρόπο να ανακαλύψουν τον αριθμό και τη δομή των ομάδων.
 - ❑ Οι πιθανοκρατικοί (probabilistic) αλγόριθμοι, που βασίζονται σε μοντέλα πιθανοτήτων.
- ❖ Η ομαδοποίηση απαιτεί κάποιο μέτρο της ομοιότητας ή διαφοράς μεταξύ των δεδομένων. Συνήθως υπολογίζεται η "απόσταση" μεταξύ των δεδομένων.
- ❖ Έστω ένα σύνολο δεδομένων D , και δύο δεδομένα του, x, y που περιγράφονται από m χαρακτηριστικά $(x_1, x_2, \dots, x_m), (y_1, y_2, \dots, y_m)$.
 - ❑ Τυπικά μέτρα απόστασης αυτών των δύο δεδομένων είναι η απόσταση *Μανχάταν* και η *Ευκλείδεια* απόσταση:

$$d(x, y) = \sum_i |x_i - y_i|$$

Απόσταση *Μανχάταν*

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Ευκλείδεια απόσταση

- ❑ Αν κάποια χαρακτηριστικά είναι διακριτά, τότε η απόσταση των τιμών τους θεωρείται 0 αν πρόκειται για την ίδια τιμή και 1 αν πρόκειται για διαφορετικές τιμές.
- ❑ Τα αριθμητικά χαρακτηριστικά θα πρέπει να ομογενοποιούνται ώστε η απόστασή τους να πέφτει μέσα στο διάστημα $[0, 1]$.

Αλγόριθμοι Βασισμένοι σε Διαχωρισμούς

Ο αλγόριθμος των K-μέσων (1/2)

- ❖ Ένας από τους πιο γνωστούς αλγόριθμους ομαδοποίησης αυτής της κατηγορίας είναι ο αλγόριθμος των K-μέσων (K-means).
 - ❑ Ο αριθμός K των ομάδων καθορίζεται πριν την εκτέλεση του αλγορίθμου.
 - ❑ Ο αλγόριθμος ξεκινά διαλέγοντας K τυχαία σημεία από τα δεδομένα ως τα κέντρα των ομάδων.
 - ❑ Έπειτα αναθέτει κάθε σημείο στην ομάδα της οποίας το κέντρο είναι πιο κοντά (μικρότερη απόσταση) σε αυτό το σημείο.
 - ❑ Στη συνέχεια, υπολογίζει για κάθε ομάδα το μέσο όρο όλων των σημείων της (μέσο διάνυσμα) και ορίζει αυτό ως νέο κέντρο της.
 - ❑ Τα δύο τελευταία βήματα επαναλαμβάνονται για ένα προκαθορισμένο αριθμό βημάτων ή μέχρι να μην υπάρχει αλλαγή στο διαχωρισμό των σημείων σε ομάδες.

Ο αλγόριθμος των K-μέσων (2/2)

❖ Ο αλγόριθμος των των K-μέσων δίνεται σε ψευδογλώσσα στη συνέχεια:

Αλγόριθμος K-μέσων

είσοδος:

Σύνολο δεδομένων $D = \{x_1, \dots, x_n\}$

Αριθμός Ομάδων k

έξοδος:

Ομάδες C_i

1.//ανάθεση τυχαίων κέντρων

για $i = 1, \dots, k$ κάνε:

θεώρησε m_i ως ένα τυχαίο στοιχείο από το D ;

2.//ομαδοποίηση

όσο υπάρχουν αλλαγές στις ομάδες C_i κάνε:

2α.//δημιουργία ομάδων

για $i = 1, \dots, k$ κάνε

$C_i = \{x \in D \mid d(m_i, x) \leq d(m_j, x) \text{ για όλα τα } j = 1, \dots, k, j \neq i\}$;

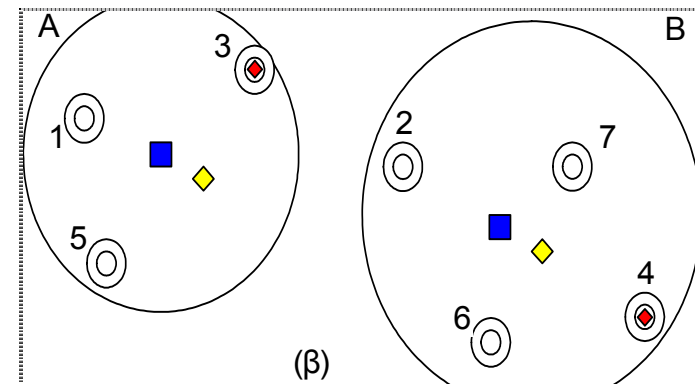
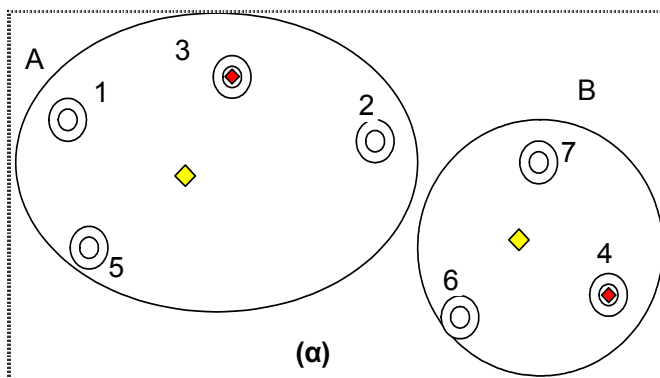
2β.//υπολογισμός νέων κέντρων

για $i = 1, \dots, k$ κάνε

$m_i =$ το μέσο διάνυσμα των σημείων που ανήκουν στην ομάδα C_i ;

Παράδειγμα Εκτέλεσης του Αλγορίθμου των K-μέσων

- ❖ Έστω ότι ο αλγόριθμος εκτελείται με $k=2$ για τα 7 σημεία του Σχήματος.
 - ❑ Αρχικά επιλέγονται τυχαία, έστω τα σημεία 3 και 4, ως κέντρα για τις δύο ομάδες A και B
 - ❑ Κάθε σημείο από τα υπόλοιπα ανατίθεται στην ομάδα της οποίας το κέντρο είναι πιο κοντά, άρα τα σημεία 1, 2, 3 και 5 θα ανήκουν στην ομάδα A ενώ τα 4, 6 και 7 στην B.
 - ❑ Ξαναυπολογίζονται τα κέντρα κάθε ομάδας και με αυτόν τον τρόπο ολοκληρώνεται ένας κύκλος υπολογισμών.
 - ❑ Τα νέα κέντρα απεικονίζονται στο Σχήμα (α) με ρόμβο.



- ❑ Ο παραπάνω κύκλος υπολογισμών επαναλαμβάνεται χρησιμοποιώντας αυτήν τη φορά τις αποστάσεις των σημείων από τα νέα κέντρα.
 - Μια μεταβολή που συντελείται είναι ότι το σημείο 2 αλλάζει ομάδα και ανήκει πλέον στη B.
 - Επιπλέον, τα κέντρα των ομάδων μετατοπίζονται σε νέες θέσεις που στο Σχήμα (β) σημειώνονται με τετράγωνο.
- ❑ Κατά την τρίτη επανάληψη των παραπάνω υπολογισμών, δε συντελείται καμία μεταβολή στον πληθυσμό κάθε ομάδας, οπότε η διαδικασία τερματίζει.

Αλγόριθμοι Ιεραρχικής Ομαδοποίησης (1/3)

- ❖ Οι αλγόριθμοι ιεραρχικής ομαδοποίησης συνδυάζουν ομάδες σε μεγαλύτερες ομάδες ή διαιρούν μεγάλες ομάδες σε μικρότερες.
- ❖ Το αποτέλεσμα των αλγορίθμων αυτών είναι μια ιεραρχία από διαφορετικές ομαδοποιήσεις των δεδομένων στο ένα άκρο της οποίας βρίσκεται μια μόνο ομάδα με όλα τα δεδομένα, και στο άλλο τόσες ομάδες όσες και ο αριθμός των δεδομένων.
- ❖ Με βάση την κατεύθυνση ανάπτυξης της ιεραρχίας που ακολουθούν, οι ιεραρχικοί αλγόριθμοι ομαδοποίησης χωρίζονται στους αλγορίθμους συγχώνευσης (agglomerative) και στους αλγορίθμους διαίρεσης (divisive).
 - ❑ Οι αλγόριθμοι συγχώνευσης είναι οι πιο σημαντικοί και διαδεδομένοι από τους δύο. Βασίζονται σε μετρικές απόστασης ανάμεσα σε ομάδες.
 - ❑ Δεδομένης μιας αρχικής ομαδοποίησης (για παράδειγμα, κάθε σημείο αποτελεί μια ομάδα), οι αλγόριθμοι αυτοί βρίσκουν τις δύο πιο κοντινές ομάδες και τις συγχωνεύουν με μία.
 - ❑ Η διαδικασία συνεχίζεται μέχρις ότου προκύψει μία μόνο ομάδα.

Αλγόριθμοι Ιεραρχικής Ομαδοποίησης (2/3)

- ❖ Δίνεται ο γενικός αλγόριθμος ιεραρχικής ομαδοποίησης, σε ψευδογλώσσα:

Ιεραρχικός Αλγόριθμος Ομαδοποίησης

είσοδος:

Σύνολο δεδομένων $D = \{x_1, \dots, x_n\}$

Συνάρτηση $d(C_i, C_j)$ απόστασης δύο ομάδων C_i, C_j

έξοδος:

Ομάδες C_i

1. για $i = 1, \dots, n$ κάνε:

 θεώρησε $C_i = \{x_i\}$;

2. όσο ο αριθμός των ομάδων είναι μεγαλύτερος από 1 κάνε:

 2α. Υπολόγισε την απόσταση μεταξύ όλων των ομάδων ανά δύο

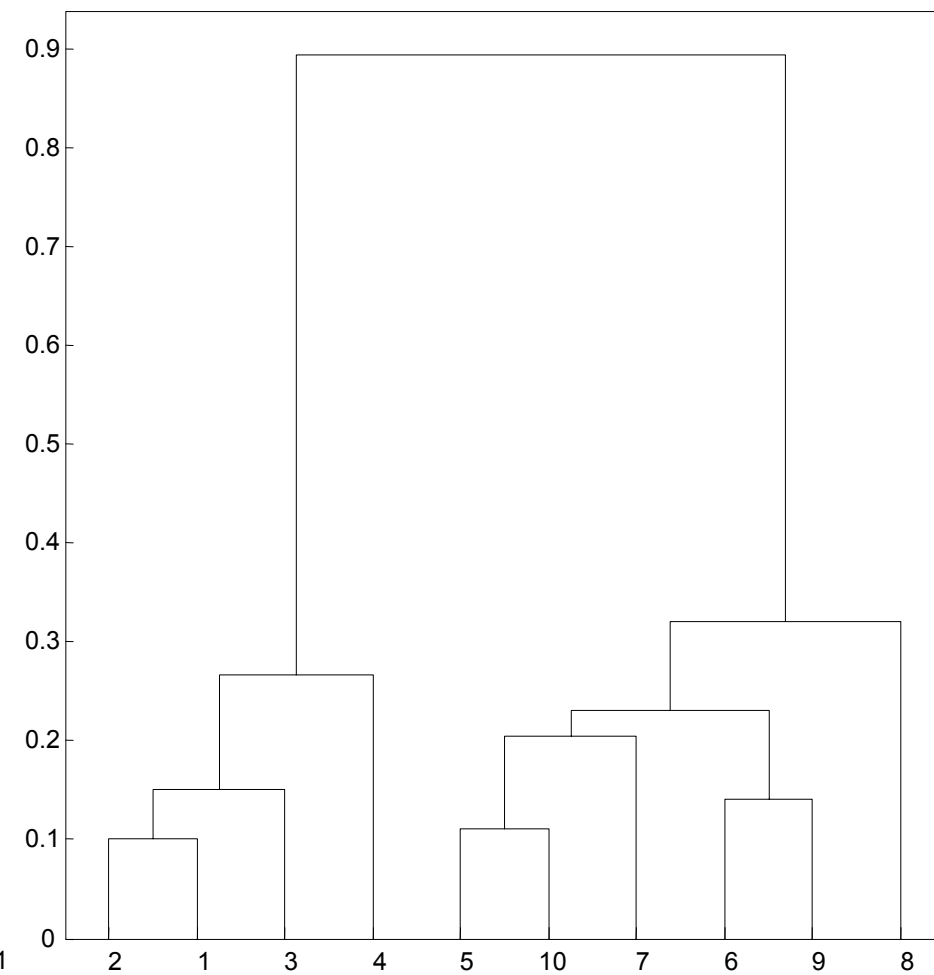
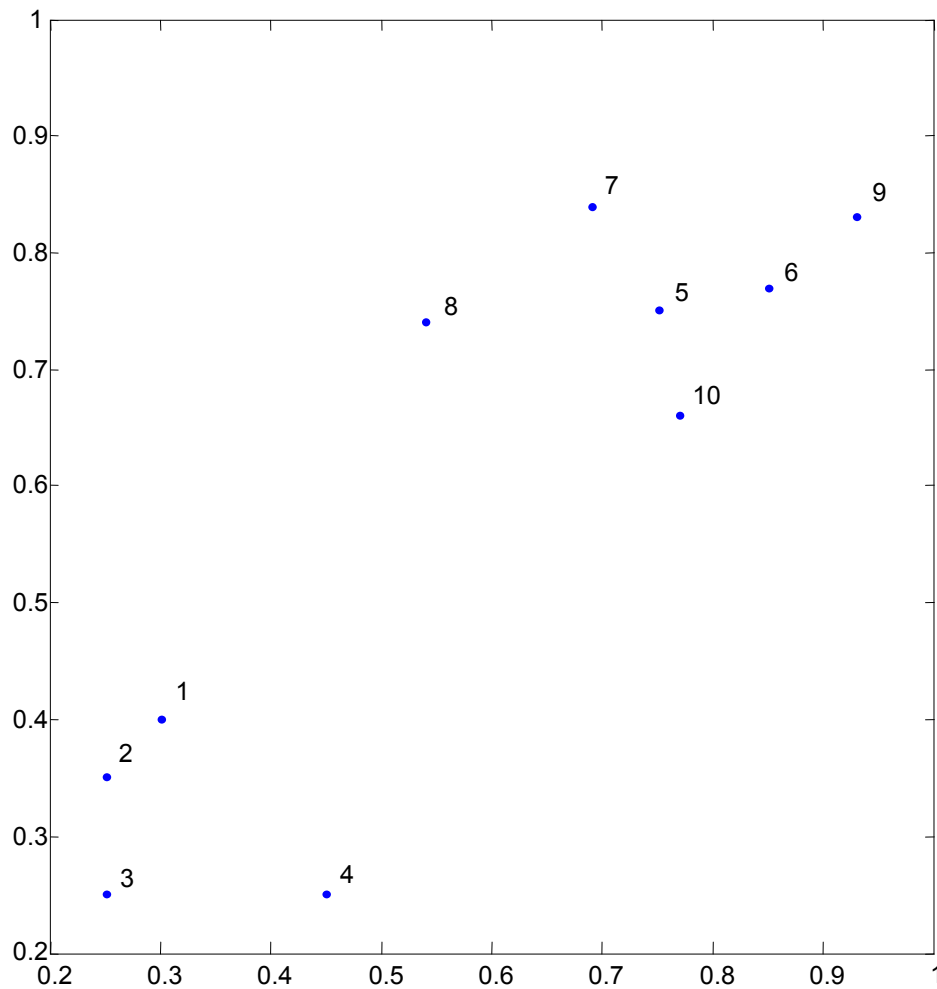
 2β. $C_i = C_i \cup C_j$, όπου C_i και C_j οι δυο πιο κοντινές ομάδες

 2γ. Αφαίρεσε την ομάδα C_j από το σύνολο των ομάδων

- ❖ Οι ιεραρχίες που προκύπτουν από τους αλγορίθμους ιεραρχικής ομαδοποίησης μπορεί να απεικονιστούν με έναν πρακτικό και εύκολο τρόπο μέσω ενός γραφήματος δενδρικής μορφής, το οποίο ονομάζεται δενδρόγραμμα.

Αλγόριθμοι Ιεραρχικής Ομαδοποίησης (3/3)

- ❖ Στο Σχήμα αριστερά, φαίνονται κάποια σημεία στο διδιάστατο χώρο, ενώ δεξιά παρουσιάζεται το δενδρόγραμμα που προκύπτει μέσω ιεραρχικής ομαδοποίησης.



Δενδρόγραμμα Ιεραρχικής Ομαδοποίησης

Άλλα Είδη Μάθησης

- ❖ Εκτός από τις μεθόδους που παρουσιάστηκαν, υπάρχουν και άλλες προσεγγίσεις στο πρόβλημα της μηχανικής μάθησης. Δύο από αυτές, είναι οι γενετικοί αλγόριθμοι και η ενισχυτική μάθηση.

Γενετικοί Αλγόριθμοι (Genetic Algorithms)

- ❖ Μέθοδος μάθησης που βασίζεται στην προσομοίωση του φυσικού φαινομένου της εξέλιξης (evolution) - (παρουσιάζονται αναλυτικά σε σχετικό κεφάλαιο).
 - ❑ Οι υποθέσεις συνήθως αναπαριστώνται από ακολουθίες bit (bit-strings).
 - ❑ Η αναζήτηση της κατάλληλης υπόθεσης ξεκινάει τυχαία με έναν πληθυσμό (μια συλλογή) αρχικών υποθέσεων, τα μέλη του οποίου παράγουν τη νέα "γενιά" μέσω διαδικασιών αναπαραγωγής αντίστοιχων των βιολογικών, όπως:
 - διασταύρωση (crossover)
 - τυχαία μετάλλαξη (random mutation)
 - ❑ Σε κάθε βήμα, οι υποθέσεις του τρέχοντος πληθυσμού αξιολογούνται βάσει μιας προκαθορισμένης συνάρτησης καταλληλότητας (fitness function).
 - Βάσει αυτής επιλέγονται για το αν θα υφίστανται ή όχι στην επόμενη γενιά.
 - ❑ **Συμπέρασμα:** Η μάθηση αντιμετωπίζεται σαν μία ειδική περίπτωση βελτιστοποίησης.



Ενισχυτική Μάθηση

Reinforcement Learning

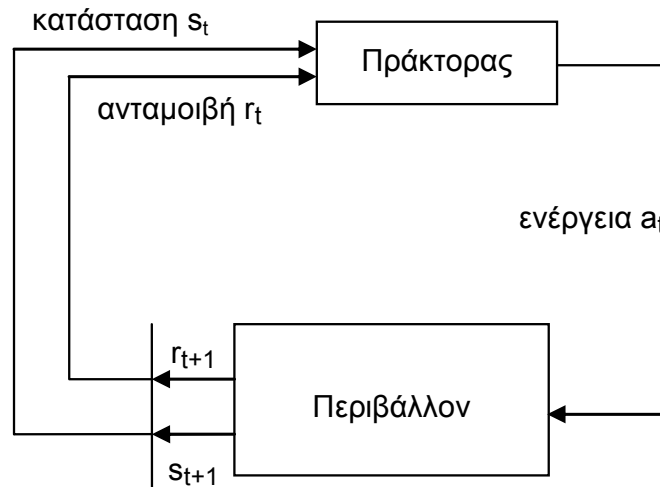
- ❖ Γενική περιγραφή οικογένειας τεχνικών στις οποίες το σύστημα μάθησης προσπαθεί να μάθει μέσω άμεσης αλληλεπίδρασης με το περιβάλλον.
- ❖ **Εφαρμογές:** έλεγχο κίνησης ρομπότ, βελτιστοποίηση εργασιών σε εργοστάσια, μάθηση επιτραπέζιων παιχνιδιών, κτλ.
- ❖ Είναι εμπνευσμένη από τα αντίστοιχα ανάλογα της μάθησης με **επιβράβευση** και **τιμωρία** που συναντώνται στα έμβια όντα.
- ❖ Σκοπός του συστήματος μάθησης: να μεγιστοποιήσει μια συνάρτηση του αριθμητικού σήματος ενίσχυσης (ανταμοιβή), για παράδειγμα την αναμενόμενη τιμή του σήματος ενίσχυσης στο επόμενο βήμα.
- ❖ Το σύστημα δεν καθοδηγείται από κάποιον εξωτερικό επιβλέποντα για το ποια ενέργεια θα πρέπει να ακολουθήσει αλλά πρέπει να ανακαλύψει μόνο του ποιες ενέργειες είναι αυτές που θα του αποφέρουν το μεγαλύτερο κέρδος.

Το Πλαίσιο της Ενισχυτικής Μάθησης (1/2)

- ❖ Στο βασικό πλαίσιο της ενισχυτικής μάθησης, η οντότητα που μαθαίνει και παίρνει αποφάσεις ονομάζεται πράκτορας (agent), ενώ οτιδήποτε άλλο εκτός του πράκτορα ονομάζεται περιβάλλον.
 - ❑ Ο πράκτορας και το περιβάλλον αλληλεπιδρούν συνεχώς, με τον πρώτο να επιλέγει ενέργειες και το δεύτερο να αποκρίνεται σε αυτές και να του παρουσιάζει καινούριες καταστάσεις.
 - ❑ Το περιβάλλον δίνει στον πράκτορα ανταμοιβές (rewards), ειδικές αριθμητικές τιμές τις οποίες αυτός προσπαθεί μακροπρόθεσμα να μεγιστοποιήσει.
- ❖ Ο πράκτορας και το περιβάλλον αλληλεπιδρούν σε μια ακολουθία διακριτών χρονικών στιγμών, $t=0, 1, 2, \dots$.
 - ❑ Σε μια χρονική στιγμή t , ο πράκτορας λαμβάνει μια αναπαράσταση της κατάστασης του περιβάλλοντος, $s_t \in \mathcal{S}$,
 - \mathcal{S} είναι το σύνολο των πιθανών καταστάσεων στις οποίες μπορεί να βρεθεί ο πράκτορας.
 - ❑ Ο πράκτορας διαλέγει μια ενέργεια, $a_t \in A(s_t)$,
 - $A(s_t)$ είναι το σύνολο των ενεργειών που είναι διαθέσιμες στην δεδομένη κατάσταση s_t
 - ❑ Την επόμενη χρονική στιγμή, σαν αποτέλεσμα της ενέργειάς του, ο πράκτορας λαμβάνει μια αριθμητική ανταμοιβή, $r_{t+1} \in \mathcal{R}$, και μεταβαίνει σε μια καινούρια κατάσταση, s_{t+1} .

Το Πλαίσιο της Ενισχυτικής Μάθησης (2/2)

- ❖ Το Σχήμα δείχνει την αλληλεπίδραση του πράκτορα με το περιβάλλον.



- ❖ Σε κάθε χρονική στιγμή ο πράκτορας πραγματοποιεί μια απεικόνιση από τις καταστάσεις σε πιθανότητες επιλογής κάθε δυνατής ενέργειας.
 - ❑ Η απεικόνιση αυτή ονομάζεται *πολιτική* του πράκτορα και υποδηλώνεται ως π ,
 - $\pi(s_t, a_t)$ είναι η πιθανότητα να επιλεγεί η ενέργεια a_t στη κατάσταση s_t .
- ❖ Άρα η ενισχυτική μάθηση πραγματοποιείται σε πραγματικό χρόνο μέσω της αλληλεπίδρασης του πράκτορα με το περιβάλλον.