



Τεχνητή Νοημοσύνη

Μάθηση από Παρατηρήσεις

Διδάσκων: Τσίπουρας Μάρκος
Εκπαιδευτικό Υλικό: Τσίπουρας Μάρκος
<http://ai.uom.gr/aima/>

Μορφές μάθησης (1 / 3)

- Ευφυής πράκτορας:
 - Στοιχείο εκτέλεσης
 - Στοιχείο μάθησης
- Παράγοντες που επηρεάζουν τη σχεδίαση:
 - Συνιστώσες του στοιχείου εκτέλεσης που πρέπει να μαθευτούν.
 - Διαθέσιμη ανάδραση.
 - Αναπαράσταση.

Μορφές μάθησης (2/3)

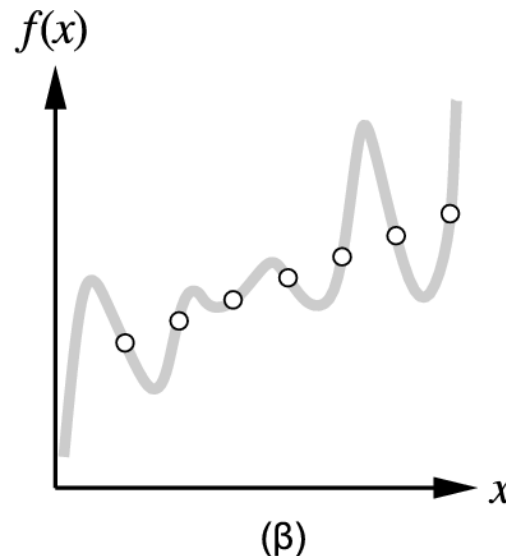
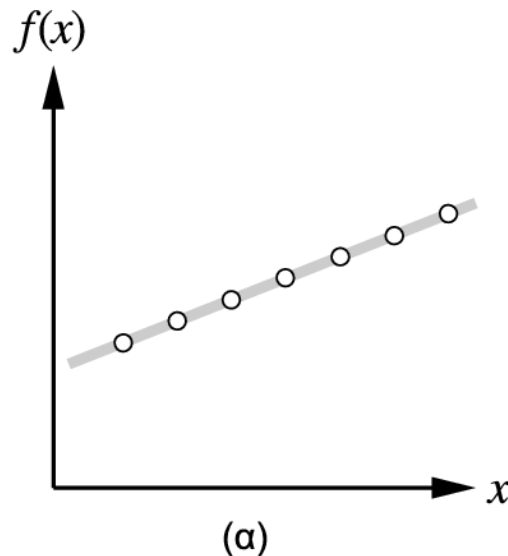
- Αντικείμενο μάθησης:
 - Άμεση αντιστοίχιση από τις συνθήκες της τρέχουσας κατάστασης στις ενέργειες.
 - Ιδιότητες του κόσμου από την αλληλουχία των αισθήσεων.
 - Εξέλιξη του κόσμου και αποτελέσματα των δυνατών ενεργειών.
 - Πληροφορίες χρησιμότητας.
 - Πληροφορίες αποτίμησης των ενεργειών.
 - Καταστάσεις στόχου.

Μορφές μάθησης (3/3)

- Ανάδραση:
 - Επιβλεπόμενη μάθηση
 - Μη-επιβλεπόμενη μάθηση
 - Ενισχυτική μάθηση
- Αναπαράσταση:
 - Λογική
 - Πιθανοτικά δίκτυα (Bayes, νευρωνικά)

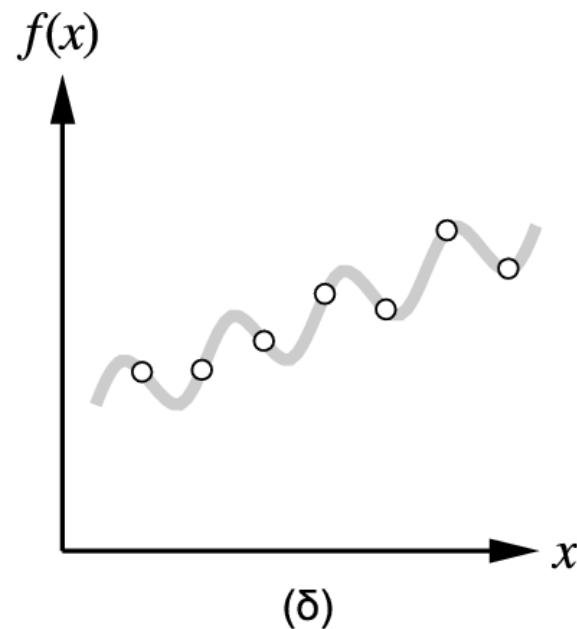
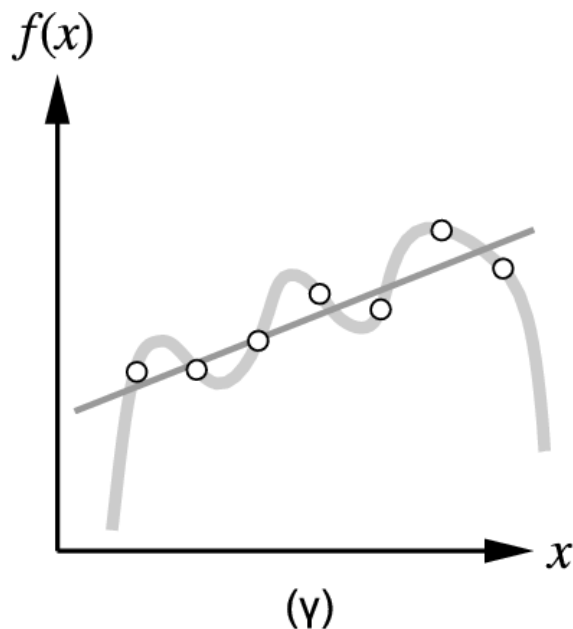
Επαγωγική μάθηση (1/2)

- Παραδείγματα: $(x, f(x))$
- Υπόθεση: h
 - Χώρος υποθέσεων
- Συνεπείς (με τα παραδείγματα) υποθέσεις
- Ευράφι του Ockham



Επαγωγική μάθηση (2/2)

- Επιλογή του κατάλληλου χώρου υποθέσεων
 - Εφικτά / ανέφικτα προβλήματα μάθησης



Δένδρα αποφάσεων

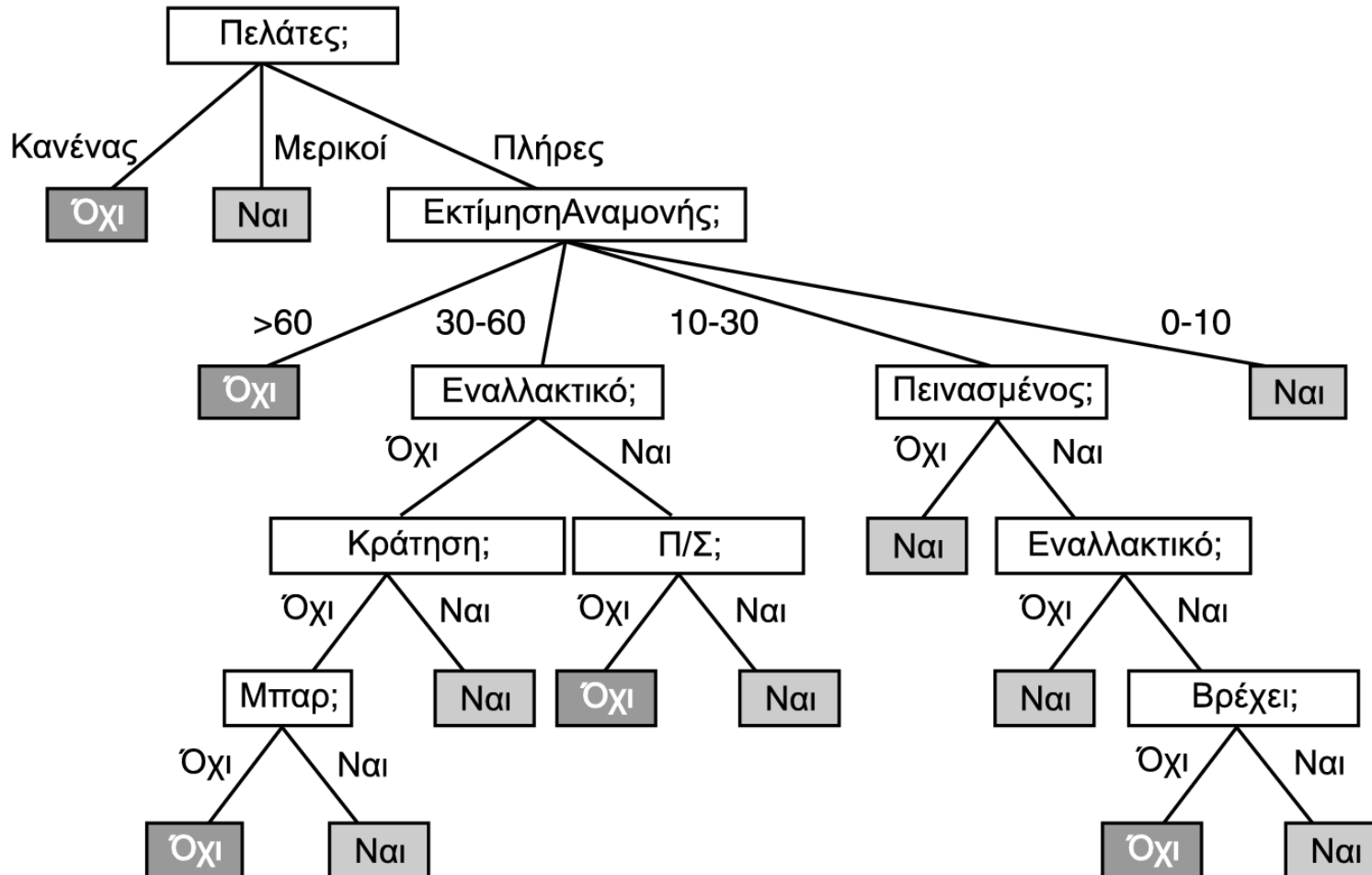
Decision Trees

A series of horizontal lines of varying lengths and colors (teal, light blue, white) extending from the left edge of the slide towards the right, positioned below the title and subtitle.

Το πρόβλημα του εστιατορίου

- Χαρακτηριστικά του προβλήματος:
 - *Εναλλακτικό*: Ναι, Όχι.
 - *Μπαρ*: Ναι, Όχι.
 - *Π/Σ*: Ναι, Όχι.
 - *Πεινασμένος*: Ναι, Όχι.
 - *Πελάτες*: Κανένας, Μερικοί, και Πλήρες.
 - *Τιμή*: \$, \$\$, \$\$\$.
 - *Βρέχει*: Ναι, Όχι.
 - *Κράτηση*: Ναι, Όχι.
 - *Τύπος*: Γαλλικό, Ιταλικό, Ταϋλανδέζικο, ή ταχυφαγείο.
 - *Εκτίμηση Αναμονής*: 0'–10', 10'–30', 30'–60', >60'.

Παράδειγμα



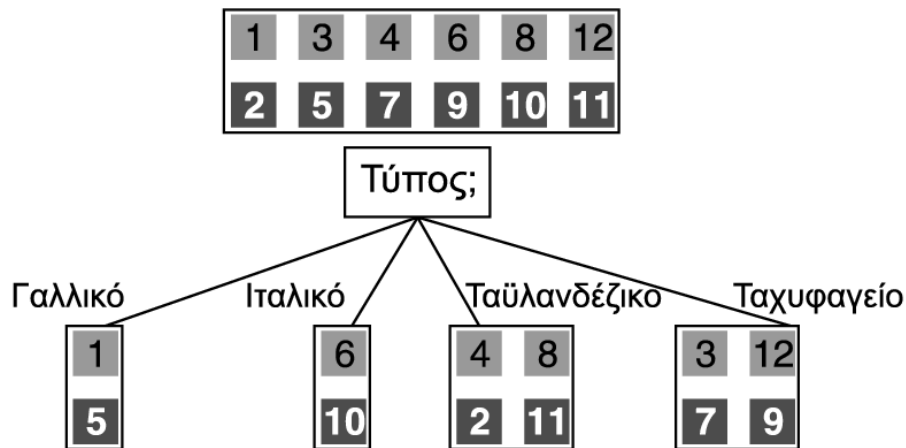
Εκφραστικότητα των δένδρων αποφάσεων

- Ισοδύναμα με την προτασιακή λογική.
 - Δεν μπορούν να αναφέρονται σε δύο διαφορετικά αντικείμενα ταυτόχρονα, άρα δεν επεκτείνονται στη λογική πρώτης τάξης.
- Υπερβολικά μεγάλα δένδρα σε κάποιες περιπτώσεις.
 - Συναρτήσεις ισοτιμίας (parity), πλειοψηφίας (majority)

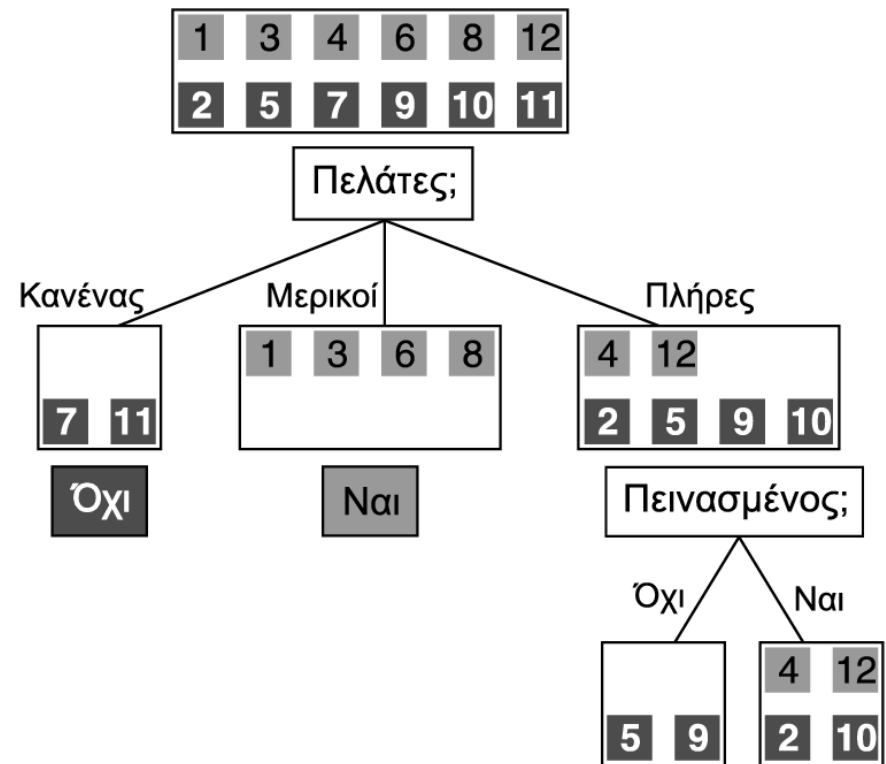
Σύνολο εκπαίδευσης

#	Εναλ	Μπαρ	Π/Σ	Πεινασμ	Πελατες	Τιμή	Βρέχει	Κράτηση	Τύπος	Εκτιμ	ΘαΠεριμένει
X_1	Ναι	Όχι	Όχι	Ναι	Μερικοί	\$\$\$	Όχι	Ναι	Γαλλικό	0-10	Ναι
X_2	Ναι	Όχι	Όχι	Ναι	Πλήρες	\$	Όχι	Όχι	Ταϋλ	30-60	Όχι
X_3	Όχι	Ναι	Όχι	Όχι	Μερικοί	\$	Όχι	Όχι	Ταχυφ.	0-10	Ναι
X_4	Ναι	Όχι	Ναι	Ναι	Πλήρες	\$	Ναι	Όχι	Ταϋλ	10-30	Ναι
X_5	Ναι	Όχι	Ναι	Όχι	Πλήρες	\$\$\$	Όχι	Ναι	Γαλλικό	>60	Όχι
X_6	Όχι	Ναι	Όχι	Ναι	Μερικοί	\$\$	Ναι	Ναι	Ιταλικό	0-10	Ναι
X_7	Όχι	Ναι	Όχι	Όχι	Κανένας	\$	Ναι	Όχι	Ταχυφ.	0-10	Όχι
X_8	Όχι	Όχι	Όχι	Ναι	Μερικοί	\$\$	Ναι	Ναι	Ταϋλ	0-10	Ναι
X_9	Όχι	Ναι	Ναι	Όχι	Πλήρες	\$	Ναι	Όχι	Ταχυφ.	>60	Όχι
X_{10}	Ναι	Ναι	Ναι	Ναι	Πλήρες	\$\$\$	Όχι	Ναι	Ιταλικό	10-30	Όχι
X_{11}	Όχι	Όχι	Όχι	Όχι	Κανένας	\$	Όχι	Όχι	Ταϋλ	0-10	Όχι
X_{12}	Ναι	Ναι	Ναι	Ναι	Πλήρες	\$	Όχι	Όχι	Ταχυφ.	30-60	Ναι

Επαγωγή δένδρων αποφάσεων (1/2)

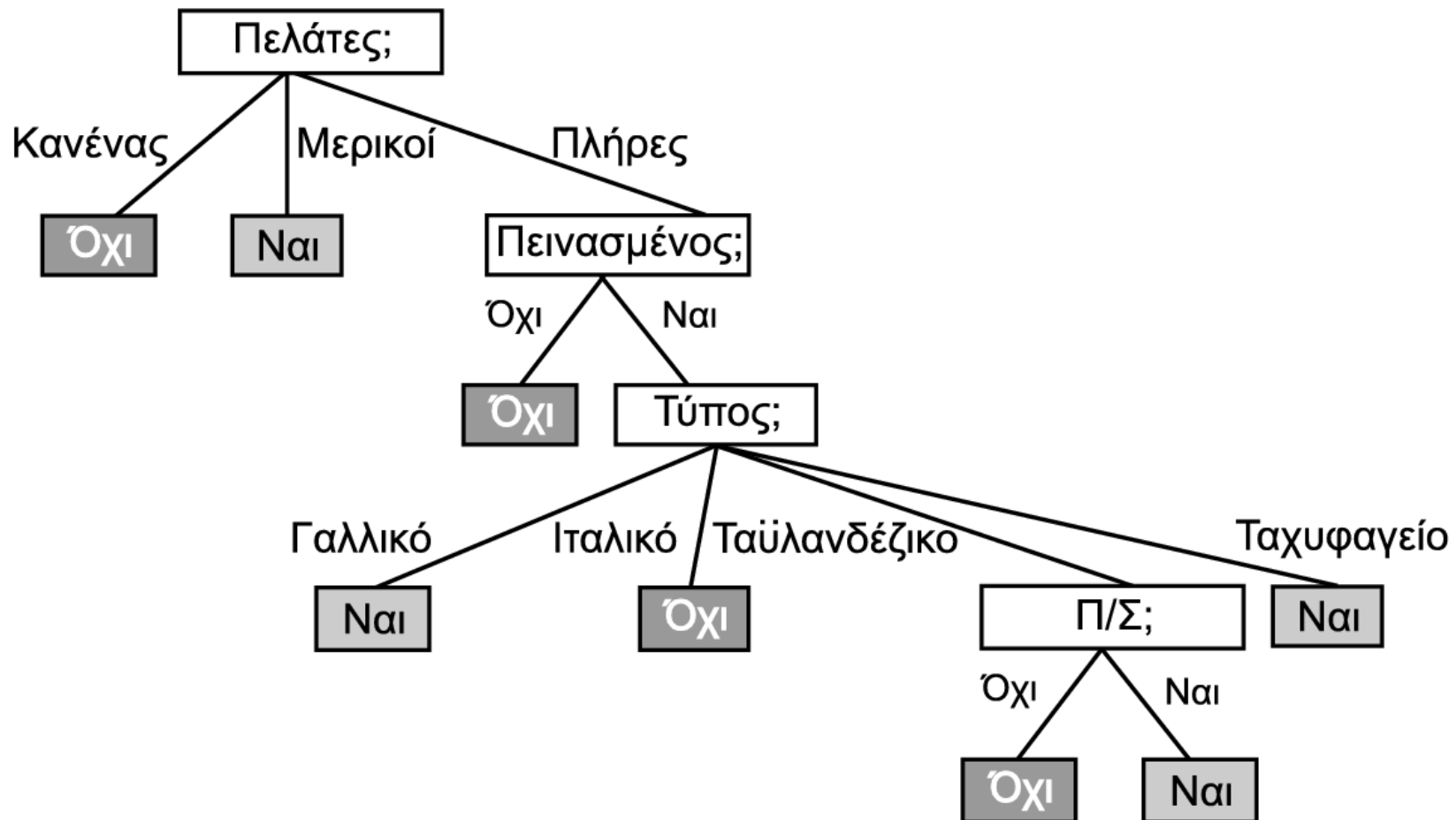


(α)



(β)

Επαγωγή δένδρων αποφάσεων (2/2)



Ο αλγόριθμος

- **function** Decision-Δένδρο-Learning
(*παραδείγματα, χαρακτηριστικά, ροεπιλογή*) **returns** δένδρο αποφάσεων
 - **inputs:** *παραδείγματα*, ένα σύνολο παραδειγμάτων
χαρακτηριστικά, ένα σύνολο χαρακτηριστικών
προεπιλογή, προεπιλεγμένη τιμή για το κατηγορημα στόχου
 - **if** *παραδείγματα* είναι κενό **then return** *προεπιλογή*
 - **else if** όλα στο *παραδείγματα* έχουν την ίδια ταξινόμηση **then return** την ταξινόμηση
 - **else if** *χαρακτηριστικά* είναι κενό **then return** Majority-Value (*παραδείγματα*)
 - **else**
 - *καλύτερο* $\leftarrow$ Choose-Attribute(*χαρακτηριστικά, παραδείγματα*)
 - *δένδρο* $\leftarrow$ νέο δένδρο αποφάσεων με έλεγχο ρίζας *καλύτερο*
 - *m* $\leftarrow$ Majority-Value(*παραδείγματα*)
 - **for each** τιμή v_i του βέλτιστο **do**
 - *παραδείγματα_i* $\leftarrow$ {στοιχεία από τα *παραδείγματα* με *καλύτερο* = v_i }
 - *υποδένδρο* $\leftarrow$ Decision-Δένδρο-Learning(*παραδείγματα_i, χαρακτηριστικά-καλύτερο, m*)
 - προσθήκη διακλάδωσης στο δένδρο με ετικέτα v_i και υποδένδρο *υποδένδρο*
 - **return** *δένδρο*

Πληροφορία

- Πληροφορία (Shannon & Weaver, 1949)

$$I(P(v_1), \dots, P(v_n)) = \sum_{i=1}^n -P(v_i) \log_2 P(v_i)$$

- π.χ. για δύο ισοπίθανα ενδεχόμενα:

$$I\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ δυαδικό ψηφίο}$$

- Για μη ισοπίθανα ενδεχόμενα, π.χ.

- $I(1/100, 99/100) = 0,08$ δυαδικά ψηφία

Επιλογή χαρακτηριστικών

- Αρχικά υπολοιπόμενη ποσότητα πληροφορίας:

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

- Αναμενόμενο υπόλοιπο πληροφορίας μετά από ερώτηση:

$$\text{Υπόλοιπο}(A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

- Αναμενόμενο κέρδος πληροφορίας:

$$\text{Κέρδος}(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \text{Υπόλοιπο}(A)$$

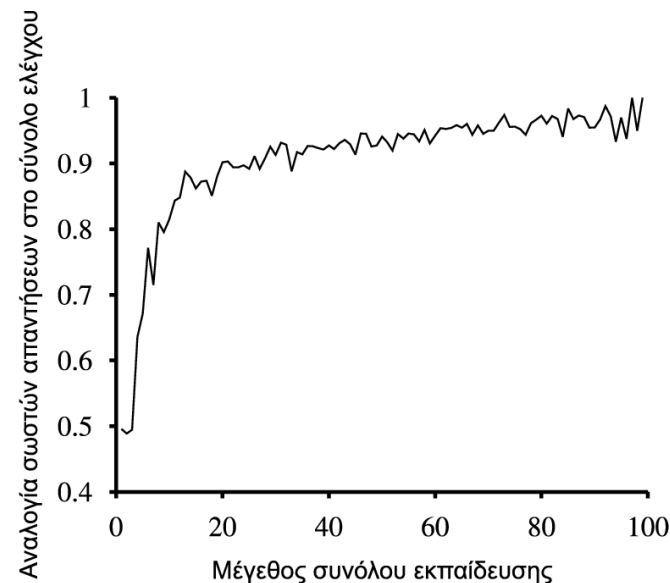
Παράδειγμα

$$Κέρδος(Πελάτες) = 1 - \left[\frac{2}{12} I(0,1) + \frac{4}{12} I(1,0) + \frac{6}{12} I\left(\frac{2}{6}, \frac{4}{6}\right) \right] \approx 0,541 \text{ δυαδικά ψηφία}$$

$$Κέρδος(Τύπος) = 1 - \left[\frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} I\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{4}{12} I\left(\frac{2}{4}, \frac{2}{4}\right) \right] = 0$$

Αποτίμηση αλγορίθμου μάθησης

- Σύνολο ελέγχου
 - Πολλαπλές δοκιμές για διαφορετικά μεγέθη και διαφορετικά στοιχεία του συνόλου ελέγχου.
- Καμπύλη μάθησης



Υπερπροσαρμογή

- Χρήση άσχετων χαρακτηριστικών για διαχωρισμό των παραδειγμάτων.
 - Ημερομηνίες, IDs κλπ
- Στατιστικός έλεγχος σημαντικότητας
 - Μηδενική υπόθεση
- Κλάδεμα χ^2
- Διασταυρωμένη επικύρωση

Ελλιπή δεδομένα

- Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί ως εξής:
 - Κάθε παράδειγμα συνοδεύεται με έναν συντελεστή βαρύτητας.
 - Όλα τα αρχικά παραδείγματα έχουν συντελεστή βαρύτητας 1.
 - Κάθε φορά που ένα παράδειγμα δεν έχει τιμή στο χαρακτηριστικό που ελέγχεται, αυτό αντιγράφεται σε όλους τους κόμβους παιδιά με την τιμή που αντιστοιχεί σε κάθε παιδί και συντελεστή βαρύτητας ίσο με το λόγο των παραδειγμάτων κάθε κόμβου παιδιού προς το σύνολο των παραδειγμάτων του κόμβου-γονέα.
 - Σταθμισμένος υπολογισμός πιθανοτήτων.

Χαρακτηριστικά με πολλές τιμές

- Ένας τρόπος να απομονώνονται τέτοια πεδία είναι να υπολογίσουμε το λόγο των κερδών (**gain ratio**) ως εξής:
 - Έστω I_1 το αναμενόμενο κέρδος πληροφορίας εάν επιλέξουμε κάποιο χαρακτηριστικό A .
 - Έστω I_2 το κέρδος πληροφορίας για να μάθουμε την τιμή του A .
 - Το A βαθμολογείται με I_1/I_2 .

Αριθμητικά χαρακτηριστικά εισόδου

- Δύο προσεγγίσεις
 - Διακριτοποίηση.
 - Υποκειμενικότητα
 - Συνεχείς τιμές και χρήση ανισοτικών ελέγχων στους κόμβους του δένδρου.
 - Πιθανές θέσεις για τους ελέγχους είναι εκείνες που οδηγούν σε διαφορετική κατηγοριοποίηση τουλάχιστον ενός παραδείγματος.
 - Αυξημένο υπολογιστικό κόστος κατασκευής του δένδρου.

Έξοδος με συνεχείς τιμές

- Η διαδικασία προσαρμόζεται ως εξής:
 - Κάποια από τα χαρακτηριστικά χρησιμοποιούνται για ελέγχους και κάποια όχι.
 - Τα χαρακτηριστικά που δεν χρησιμοποιήθηκαν σε ελέγχους χρησιμοποιούνται στα φύλλα για να οριστεί μια συνάρτηση ελαχίστων τετραγώνων.
- Το παραγόμενο δένδρο λέγεται δένδρο παλινδρόμησης (regression tree).
 - Το πρόβλημα είναι στο διαχωρισμό των χαρακτηριστικών.

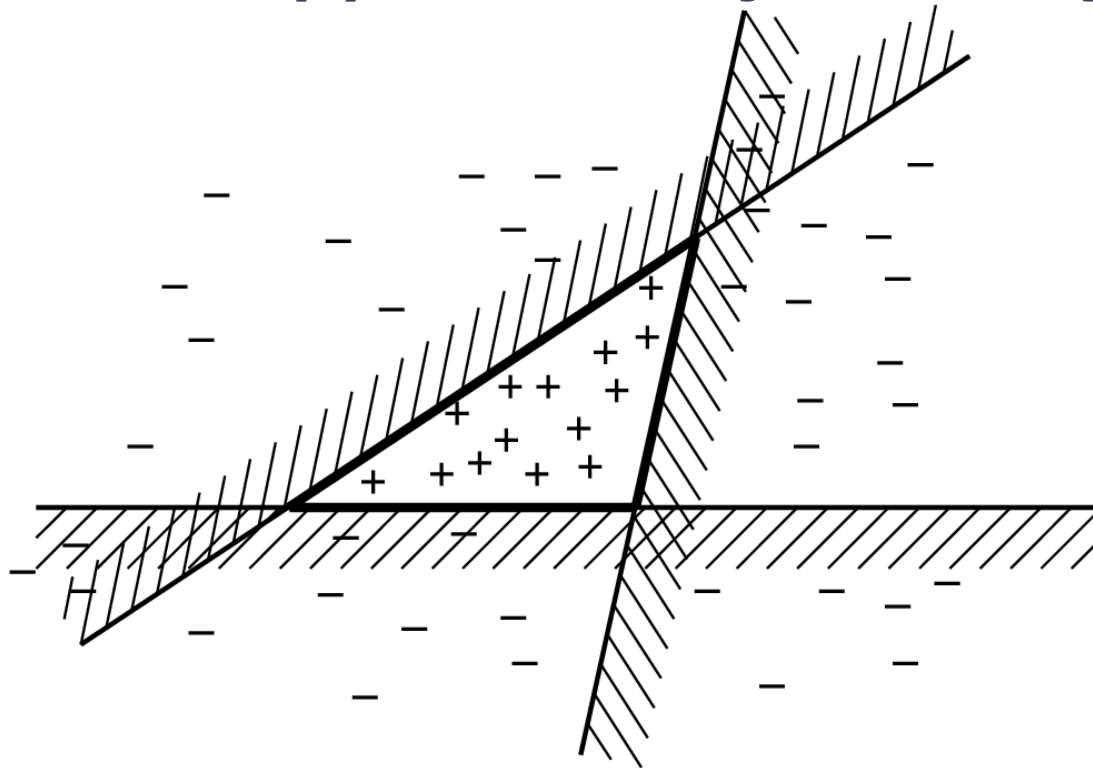
Συλλογική μάθηση

Ensemble Learning

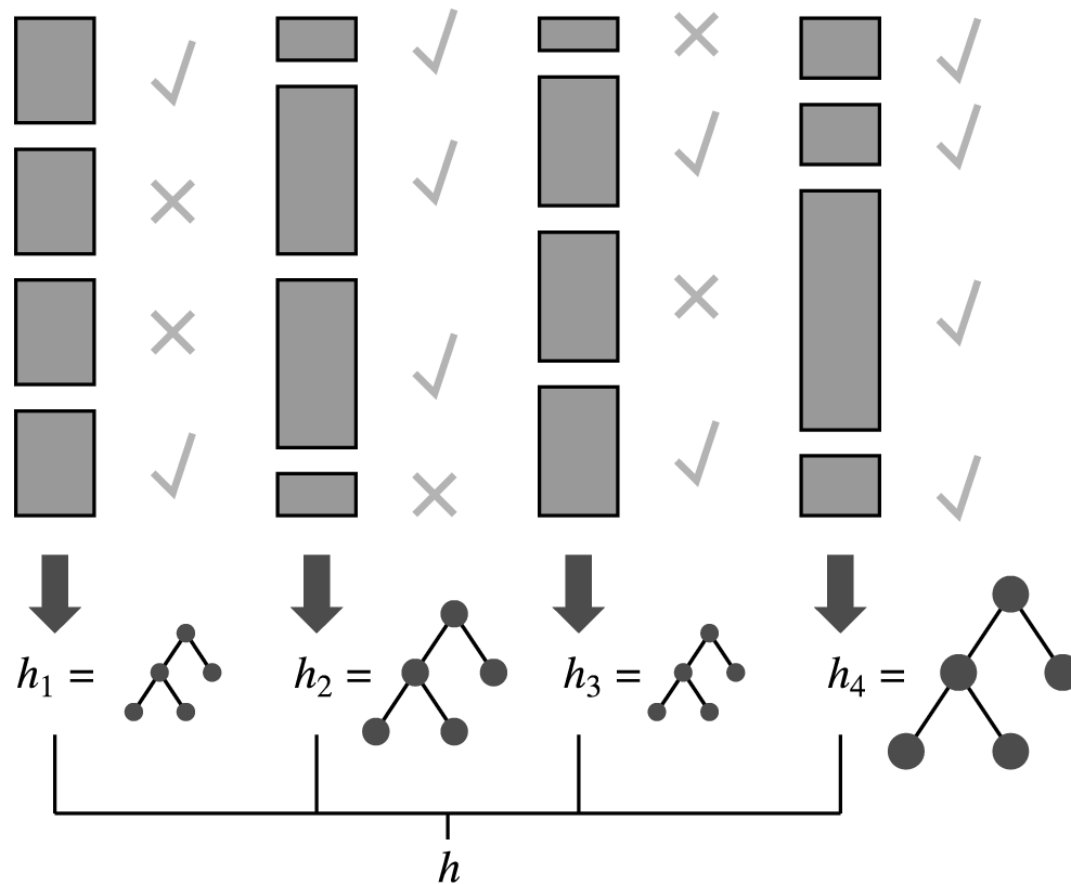
Κεντρική ιδέα

- Επιλογή μίας ομάδας ή **συλλογής** (ensemble) υποθέσεων από το χώρο υποθέσεων και ο συνδυασμός των προβλέψεών τους.
 - Χρήση πλειοψηφικής ψηφοφορίας
- Προϋπόθεση: Οι υποθέσεις της συλλογής να είναι όσο το δυνατόν πιο ανεξάρτητες.
- Παράδειγμα: Η χρήση πέντε υποθέσεων μειώνει ένα ρυθμό σφαλμάτων της τάξης του 1 στα 10 σε ρυθμό σφαλμάτων 1 στα 100.

Αύξηση εκφραστικής ισχύος



Ενδυνάμωση (boosting)

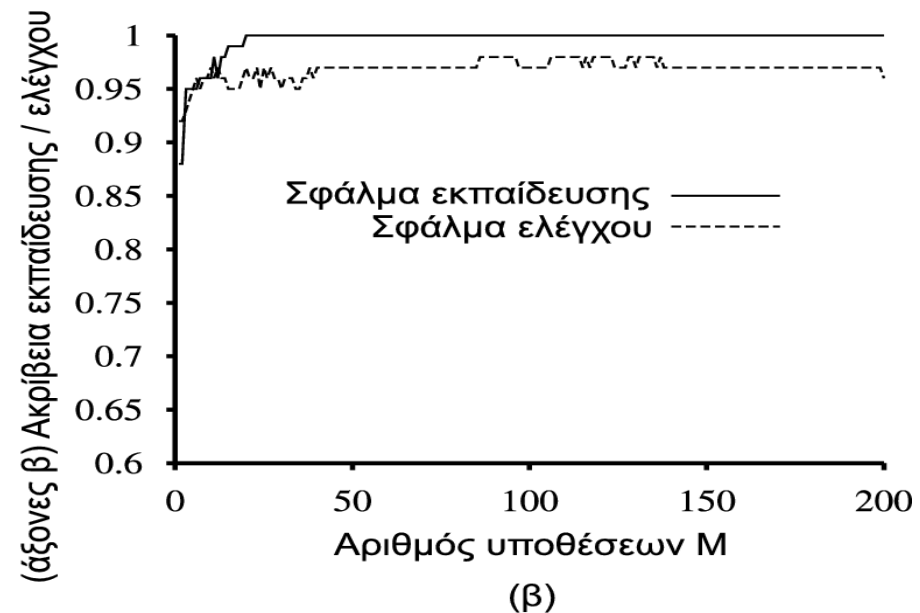
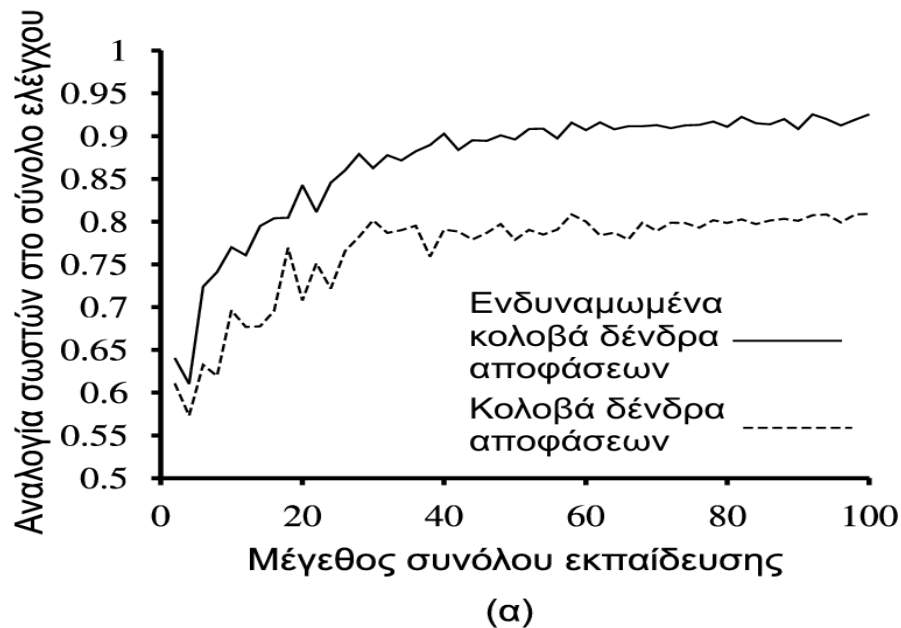


Αλγόριθμος AdaBoost

- **function** AdaBoost (παραδείγματα, L , M) **returns** μια υπόθεση σταθμισμένης πλειοψηφίας
 - **inputs:** παραδείγματα, σύνολο N παραδειγμάτων με ετικέτες $(x_1, y_1) \dots (x_N, y_N)$
 L , αλγόριθμος μάθησης
 M , ο αριθμός υποθέσεων στη συλλογή
 - **local variables:** $\mathbf{w}$, διάνυσμα N βαρυτήτων για παραδείγματα, αρχικά $1/N$
 $\mathbf{h}$, διάνυσμα M υποθέσεων
 $\mathbf{z}$, διάνυσμα βαρυτήτων των M υποθέσεων
 - **for** $m = 1$ **to** M **do**
 - $\mathbf{h}[m] \leftarrow L(\text{παραδείγματα}, \mathbf{w})$
 - $\text{σφάλμα} \leftarrow 0$
 - **for** $j = 1$ **to** N **do**
 - **if** $\mathbf{h}[m](x_j) \neq y_j$ **then** $\text{σφάλμα} \leftarrow \text{σφάλμα} + \mathbf{w}[j]$
 - **for** $j = 1$ **to** N **do**
 - **if** $\mathbf{h}[m](x_j) = y_j$ **then** $\mathbf{w}[j] \leftarrow \mathbf{w}[j] \cdot \text{σφάλμα} / (1 - \text{σφάλμα})$
 - $\mathbf{w} \leftarrow \text{Normalize}(\mathbf{w})$
 - $\mathbf{z}[m] \leftarrow \log(1 - \text{σφάλμα}) / \text{σφάλμα}$
 - **return** Weighted-Majority($\mathbf{h}$, $\mathbf{z}$)

Παράδειγμα

- Ενδυνάμωση με κολοβά δένδρα αποφάσεων (decision stumps) στα δεδομένα του εστιατορίου



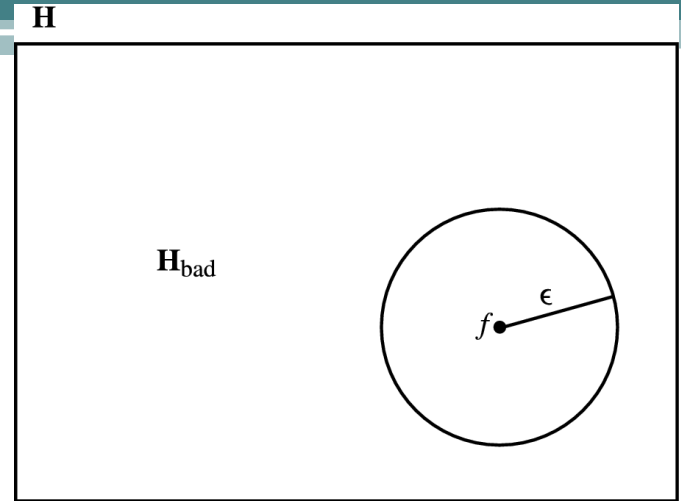
Θεωρία υπολογιστικής μάθησης

Computational Learning Theory

Πιθανώς προσεγγιστικά σωστή (probably approximately correct, PAC)

- Υπόθεση PAC
 - *Κάθε υπόθεση που είναι σημαντικά εσφαλμένη είναι σχεδόν σίγουρο ότι θα "ανακαλυφθεί" με μεγάλη πιθανότητα μετά από ένα μικρό αριθμό παραδειγμάτων, επειδή θα κάνει εσφαλμένη πρόβλεψη. Κατά συνέπεια, κάθε υπόθεση που είναι συνεπής με ένα επαρκώς μεγάλο σύνολο παραδειγμάτων εκπαίδευσης δεν είναι πιθανό να είναι σημαντικά εσφαλμένη: δηλαδή, πρέπει να είναι **πιθανώς προσεγγιστικά σωστή** (probably approximately correct).*
- Υπόθεση στασιμότητας

Σφάλμα υπόθεσης



- Σημειογραφία:
 - $\mathbf{X}$: το σύνολο όλων των δυνατών παραδειγμάτων.
 - D : η κατανομή από την οποία λαμβάνονται τα παραδείγματα.
 - $\mathbf{H}$: το σύνολο των δυνατών υποθέσεων
 - N : ο αριθμός των παραδειγμάτων στο σύνολο εκπαίδευσης.
- $\text{σφάλμα}(h) = P(h(x) \neq f(x) \mid x \text{ από το } D)$
- Προσεγγιστικά σωστή υπόθεση: $\text{σφάλμα}(h) \leq \epsilon$

Πιθανότητα παραπλάνησης

- Έστω $h_b \in \mathbf{H}_{\text{bad}}$
- $P(h_b \text{ συμφωνεί με } N \text{ παραδείγματα}) \leq (1 - \epsilon)^N$
- $P(\mathbf{H}_{\text{bad}} \text{ περιέχει μία συνεπή υπόθεση}) \leq |\mathbf{H}_{\text{bad}}|(1 - \epsilon)^N$
 $\leq |\mathbf{H}|(1 - \epsilon)^N$
- $|\mathbf{H}|(1 - \epsilon)^N \leq \delta \Rightarrow$

$$\begin{array}{c} \dots \\ (1 - \epsilon \leq e^{-\epsilon}) \\ \dots \Rightarrow \end{array} N \geq \frac{1}{\epsilon} \left(\ln \frac{1}{\delta} + \ln |\mathbf{H}| \right) \quad \begin{array}{l} \text{Πολυπλοκότητα} \\ \text{δείγματος} \end{array}$$

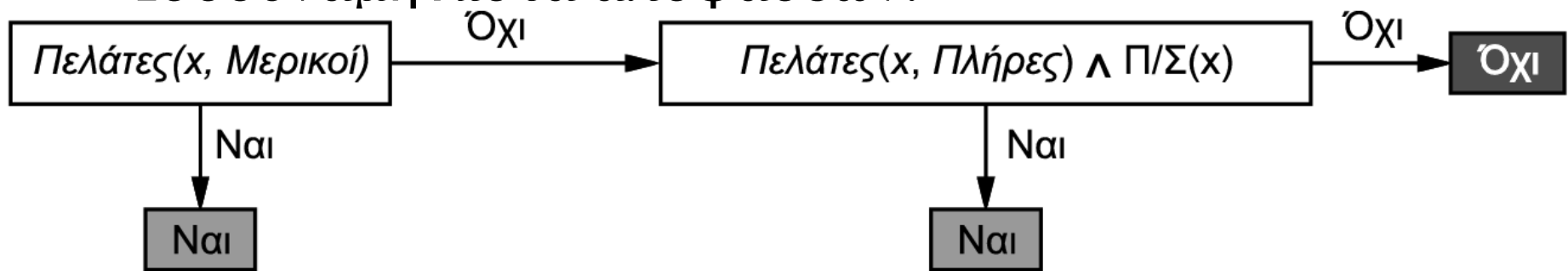
- Αν ένας αλγόριθμος μάθησης επιστρέψει μία υπόθεση που είναι συνεπής με τόσα πολλά παραδείγματα, τότε με πιθανότητα τουλάχιστον $1 - \delta$ έχει σφάλμα το πολύ ϵ .

Περιορισμός χώρου υποθέσεων

- Η πολυπλοκότητα δείγματος εξαρτάται από το μέγεθος του χώρου υποθέσεων.
 - $|\mathbf{H}| = 2^{2^n}$ για boolean συναρτήσεις
- Πρέπει να περιορίσουμε το μέγεθος του χώρου υποθέσεων.
 - Χρήση περιορισμένων γλωσσών

Λίστες αποφάσεων (Decision Lists)

- Λογική έκφραση:
 - $\forall x \Theta\alpha\text{Περιμένει}(x) \Leftrightarrow \text{Πελάτες}(x, \text{Μερικοί}) \vee (\text{Πελάτες}(x, \text{Πλήρες}) \wedge \Pi/\Sigma(x))$
- Ισοδύναμη λίστα αποφάσεων:



- k-DL: Λίστες αποφάσεων με k λεκτικά ανά σύζευξη.
 - Περιλαμβάνει τα k-DT.
- k-DL(n): n Boolean χαρακτηριστικά

Μάθηση με λίστες αποφάσεων (1/2)

- $\text{Συζεύξεις}(n, k)$: Η γλώσσα των συζεύξεων k λεκτικών από n χαρακτηριστικά.

$$|\text{Συζεύξεις}(n, k)| = \sum_{i=0}^k \binom{2n}{i} = O(n^k)$$

- Πλήθος συνδυασμών συζεύξεων:
 - $3^{|\text{Συζεύξεις}(n, k)|}$

- Πλήθος λιστών αποφάσεων (διαφορετικές σειρές ελέγχων):
 - $|k\text{-dl}(n)| \leq 3^{|\text{Συζεύξεις}(n, k)|} |\text{Συζεύξεις}(n, k)|!$

Μάθηση με λίστες αποφάσεων (2/2)

- Τελικά: $|k_{-DL}(n)| = 2^{O(n^k \log_2(n^k))}$
- και ο απαιτούμενος αριθμός παραδειγμάτων:

$$N \geq \frac{1}{\epsilon} \left(\ln \frac{1}{\delta} + O(n^k \log_2(n^k)) \right)$$

- Κατασκευή λιστών αποφάσεων με άπληστο τρόπο.

