

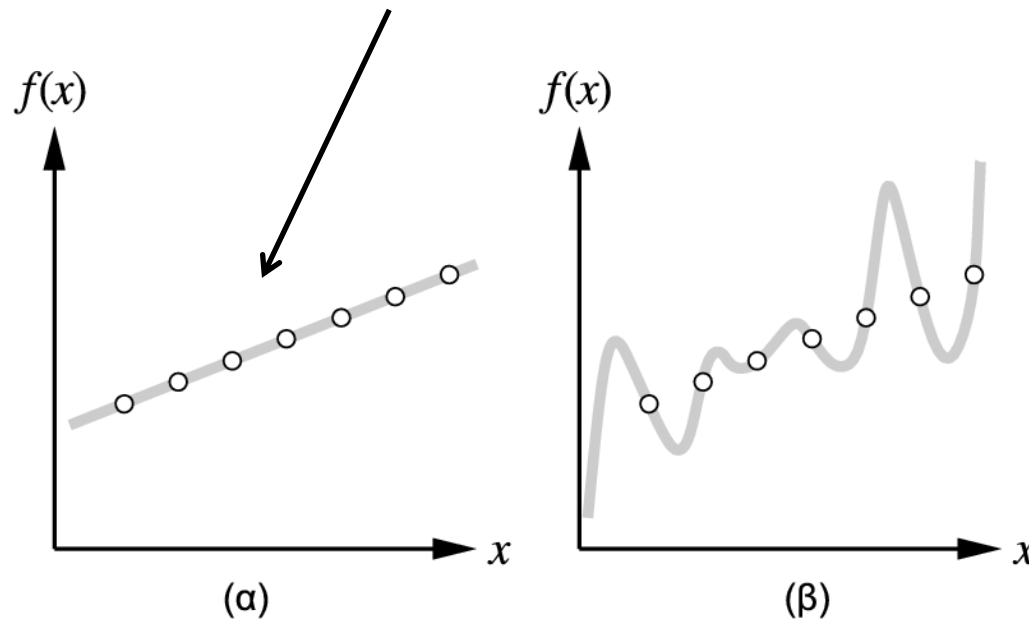
Μάθηση με παραδείγματα – Δέντρα Απόφασης

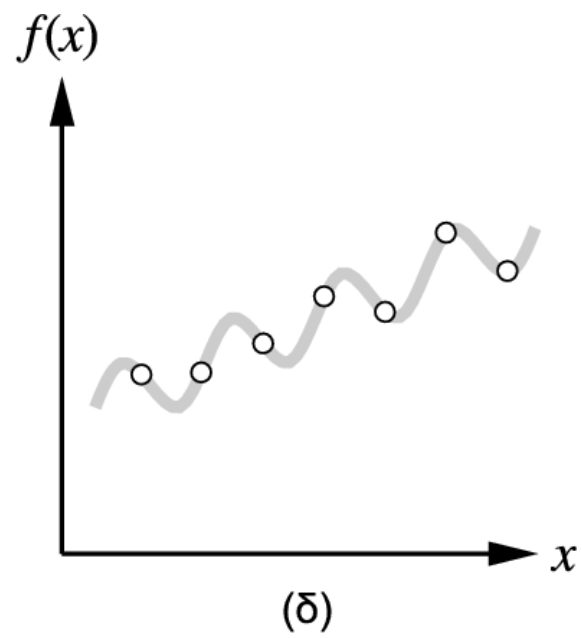
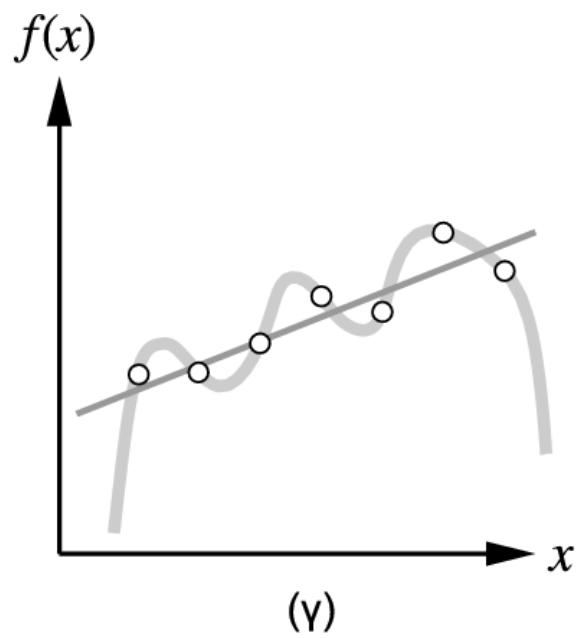
Μορφές μάθησης

- Επιβλεπόμενη μάθηση (Ταξινόμηση – Πρόβλεψη)
 - Παραδείγματα: $\{(x_i, t_i)\}$
 - t κατηγορία \rightarrow ταξινόμηση
 - t αριθμός \rightarrow πρόβλεψη
- Μη-επιβλεπόμενη μάθηση (Ομαδοποίηση – Μείωση Διάστασης)
 - Παραδείγματα: $\{x_i\}$
- Ενισχυτική μάθηση
 - Παραδείγματα: $\{(x_i, r_i)\}$
- Επαγωγική Μάθηση (Αναπαράσταση):
 - Μαθηματικά Μοντέλα
 - Συστήματα Κανόνων (ερμηνευσιμότητα)

Επιβλεπόμενη μάθηση

- Υπόθεση (μοντέλο μάθησης): h (ορισμός χώρου υποθέσεων)
- ‘Συνεπείς’ (με τα παραδείγματα) υποθέσεις
- **Occam’s razor**: προτιμούμε την απλούστερη ‘συνεπή’ υπόθεση





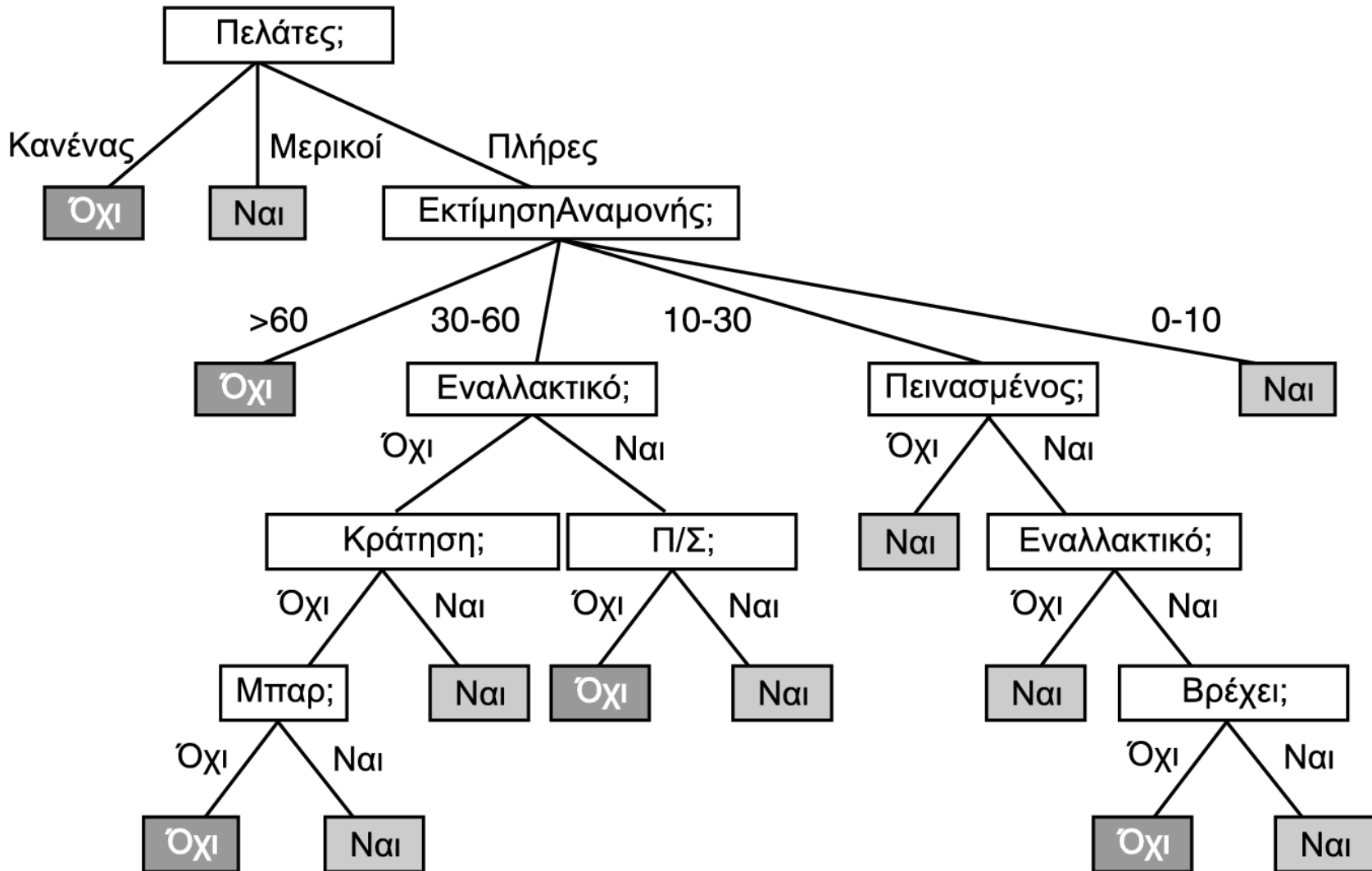
Δένδρα αποφάσεων για ταξινόμηση

Decision Trees for Classification

Το πρόβλημα του εστιατορίου

- Χαρακτηριστικά (attributes) του προβλήματος:
 - *Εναλλακτικό*: Ναι, Όχι.
 - *Μπαρ*: Ναι, Όχι.
 - *Π/Σ*: Ναι, Όχι.
 - *Πεινασμένος*: Ναι, Όχι.
 - *Πελάτες*: Κανέννας, 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'.
- Απόφαση για το αν ο πελάτης θα περιμένει: ΝΑΙ ή ΟΧΙ

Παράδειγμα



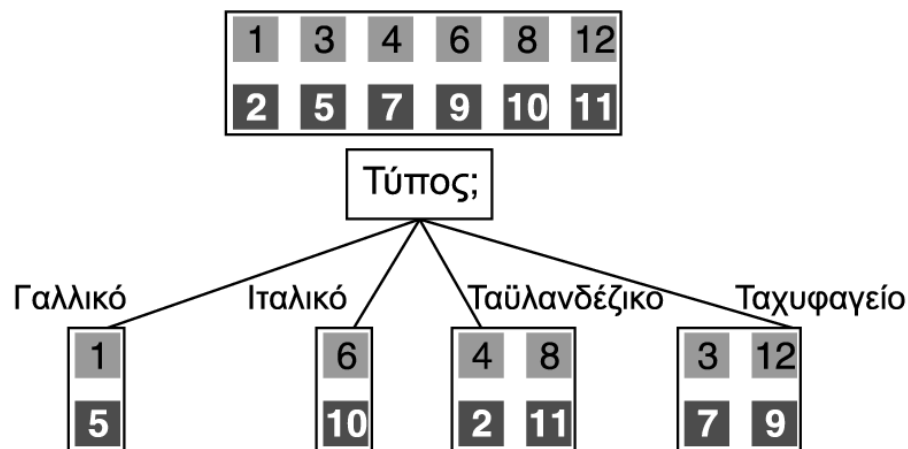
Δέντρα απόφασης

- Εσωτερικοί κόμβοι: έλεγχος και απόφαση με βάση την τιμή κάποιου χαρακτηριστικού (attribute test)
- Φύλλα: απόφαση ταξινόμησης σε κάποια κατηγορία

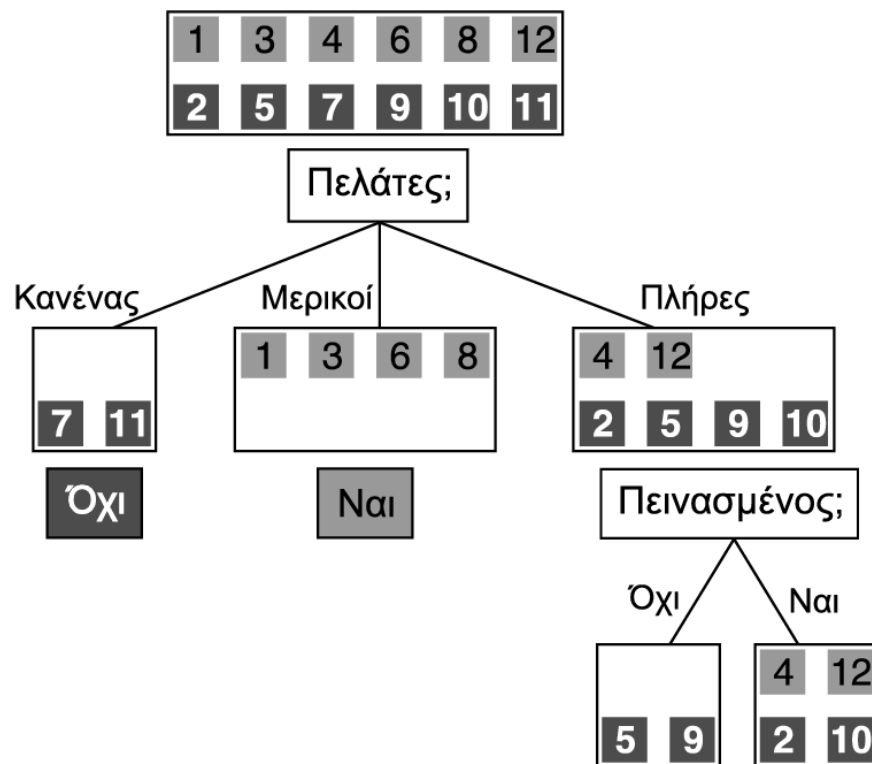
Σύνολο εκπαίδευσης

#	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

Κατασκευή δένδρων αποφάσεων (1/2)

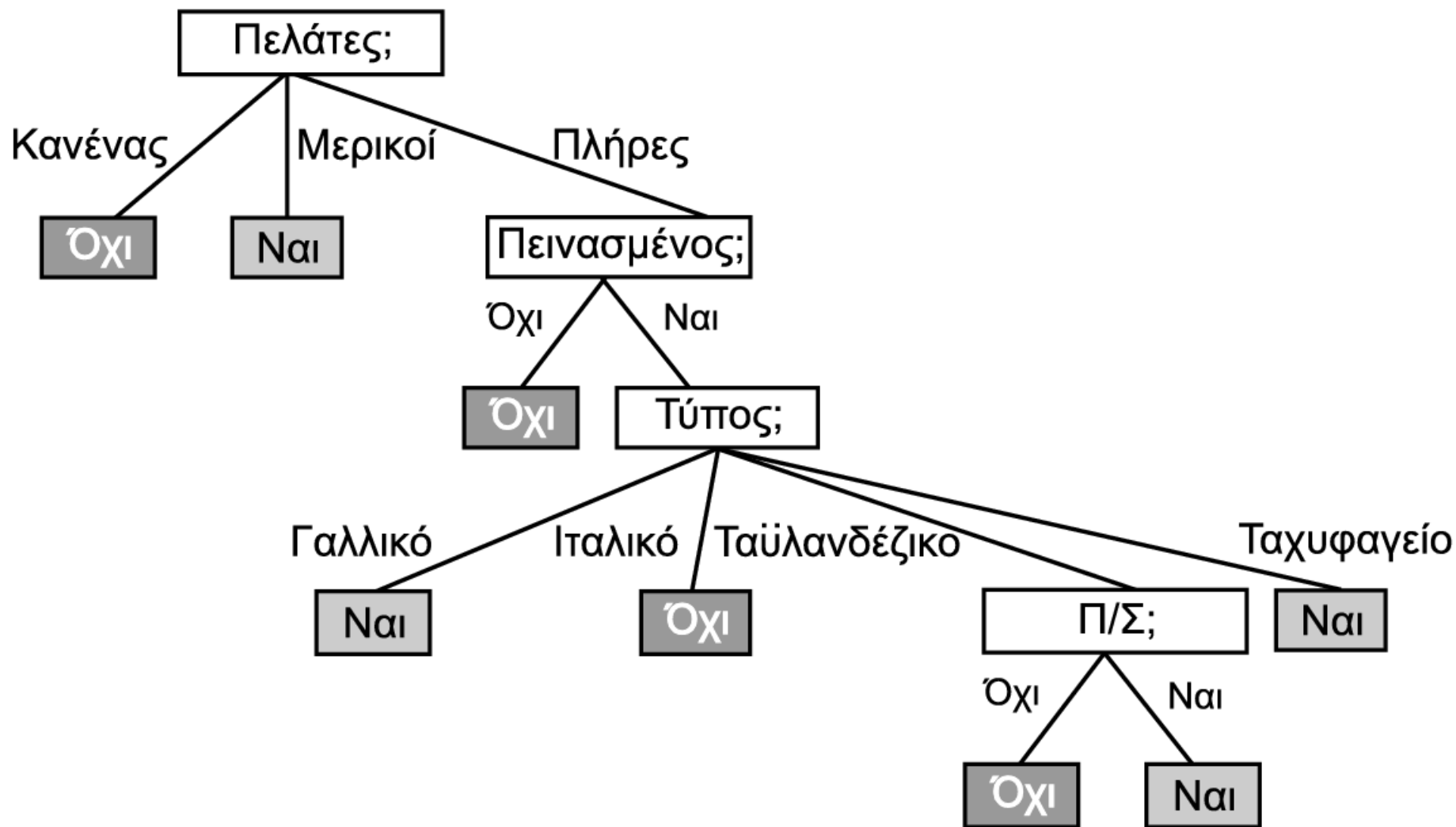


(α)



(β)

Κατασκευή δένδρων αποφάσεων (2/2)



Ο αλγόριθμος

- **function** Decision-Tree-Learning(*παραδείγματα, χαρακτηριστικά, προεπιλογή*)
returns δέντρο αποφάσεων
 - **inputs:** *παραδείγματα*, ένα σύνολο παραδειγμάτων
χαρακτηριστικά, ένα σύνολο χαρακτηριστικών
προεπιλογή, προεπιλεγμένη κατηγορία
 - **if** *παραδείγματα* είναι κενό **then return** *προεπιλογή*
 - **else if** όλα στο *παραδείγματα* έχουν την ίδια κατηγορία **then return** την κατηγορία
 - **else if** *χαρακτηριστικά* είναι κενό **then return** Majority-Class (*παραδείγματα*)
 - **else**
 - $best \leftarrow \text{Choose-Attribute}(\text{χαρακτηριστικά}, \text{παραδείγματα})$
 - $tree \leftarrow$ νέο δέντρο αποφάσεων με έλεγχο ρίζας το χαρακτηριστικό $best$
 - $m \leftarrow \text{Majority-Value}(\text{παραδείγματα})$
 - **for each** τιμή u_i του $best$ **do**
 - $\text{παραδείγματα}_i \leftarrow \{\text{στοιχεία από τα παραδείγματα με } best = u_i\}$
 - $subtree \leftarrow \text{Decision-Tree-Learning}(\text{παραδείγματα}_i, \text{χαρακτηριστικά}-best, m)$
 - προσθήκη διακλάδωσης στο $tree$ με ετικέτα u_i και υποδέντρο= $subtree$
 - **return** $tree$

Εντροπία

- **Εντροπία Πληροφορίας** (Shannon & Weaver, 1949)
(μέτρο της **αβεβαιότητας** ή **ανομοιογένειας** των δεδομένων)

- Μεταβλητή με n σύμβολα:

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

- 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, $I(1,0)=I(0,1)=0$ δυαδικά ψηφία

- **Θέλουμε η διάσπαση του συνόλου παραδειγμάτων με βάση κάποιο χαρακτηριστικό να οδηγεί σε όσο το δυνατό μεγαλύτερη μείωση της εντροπίας.**

Επιλογή χαρακτηριστικών

- Εντροπία συνόλου παραδειγμάτων πριν τη διάσπαση ($p: C_1, n: C_2$):

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

- Μέση εντροπία μετά τη διάσπαση με βάση το χαρακτηριστικό A με v δυνατές τιμές (διακλαδώσεις):

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

- Κέρδος πληροφορίας (Information Gain):

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

Παράδειγμα

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

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

Αποτίμηση συστήματος μάθησης

- **Σύνολο ελέγχου**
 - Αξιολόγηση σε παραδείγματα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση → ικανότητα γενίκευσης
- **Υπερεκπαίδευση:** το σύστημα είναι πιο ευέλικτο απότι χρειάζεται (μαθαίνει και το θόρυβο που συνήθως υπάρχει στα παραδείγματα)
- **Υποεκπαίδευση:** το σύστημα δεν είναι επαρκώς ευέλικτο
- Υπάρχει ένα βέλτιστο μοντέλο: το μικρότερο 'συνεπές' σύστημα (occam's razor).
- Για τα δέντρα απόφασης μπορούμε να κάνουμε κλάδεμα ενός μεγάλου δέντρου που κατασκευάζουμε αρχικά.