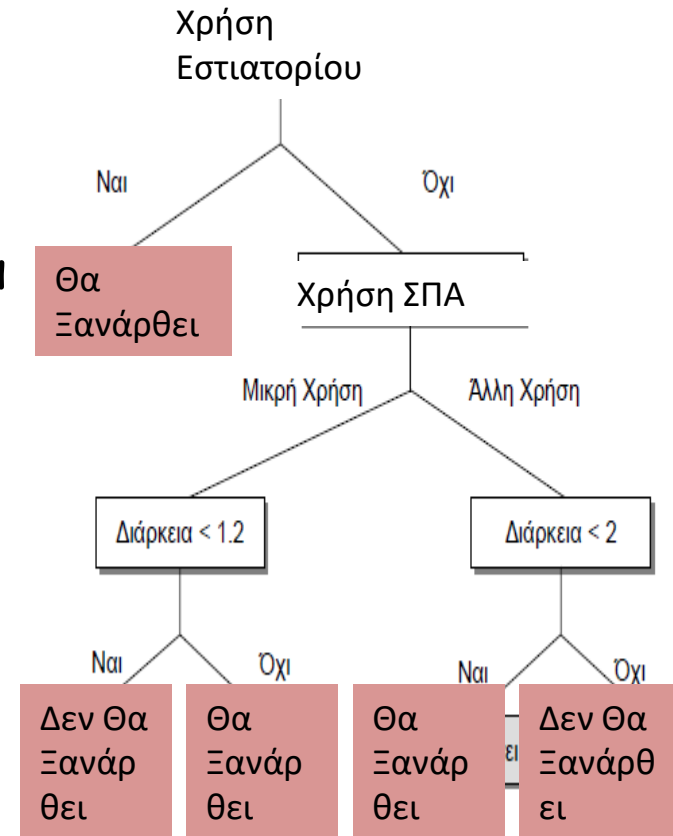


ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ ΣΤΟΝ ΤΟΥΡΙΣΜΟ

Δέντρα Απόφασης

Δέντρα απόφασης

- Το μοντέλο που επάγεται είναι μια αφαιρετική δενδροειδής δομή που αναπαριστά τα δεδομένα.
- Έστω τα δεδομένα ενός ξενοδοχείου που περιγράφουν περιπτώσεις πελατών που ξαναήρθαν να μείνουν στο ξενοδοχείο, βάσει της χρήσης του εστιατορίου και του ΣΠΑ που έκαναν.
- 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 την τιμή της εξόδου.



Ο αλγόριθμος ID3

- Είναι ο πιο γνωστός αλγόριθμος μάθησης δένδρων ταξινόμησης.
- Είναι αναδρομικός
 - Βρες το χαρακτηριστικό οι τιμές του οποίου διαχωρίζουν βέλτιστα τα παραδείγματα εκπαίδευσης σε σχέση με την τιμή τους στην κλάση ταξινόμησης και τοποθέτησέ το στη ρίζα του δέντρου
 - Κάνε τον διαχωρισμό των παραδειγμάτων εκπαίδευσης βάσει των τιμών του χαρ/κου
 - Για κάθε τιμή του χαρ/κού δημιούργησε έναν καινούριο κόμβο στο δέντρο. Για κάθε κόμβο επανέλαβε την διαδικασία. Επέλεξε δηλ. το χαρακτηριστικό εκείνο που διαχωρίζει τα παραδείγματα του συγκεκριμένου κόμβου βέλτιστα.
 - Η διαδικασία τερματίζει όταν
 - Όλα τα παραδείγματα που ανήκουν σε αυτόν έχουν ίδια τιμή στην κλάση ταξινόμησης
 - Έχω ξεμείνει από χαρακτηριστικά προς έλεγχο
- Η επιλογή του πιο κατάλληλου χαρ/κου σε κάθε κόμβο πραγματοποιείται με κάποιο στατιστικό μέτρο

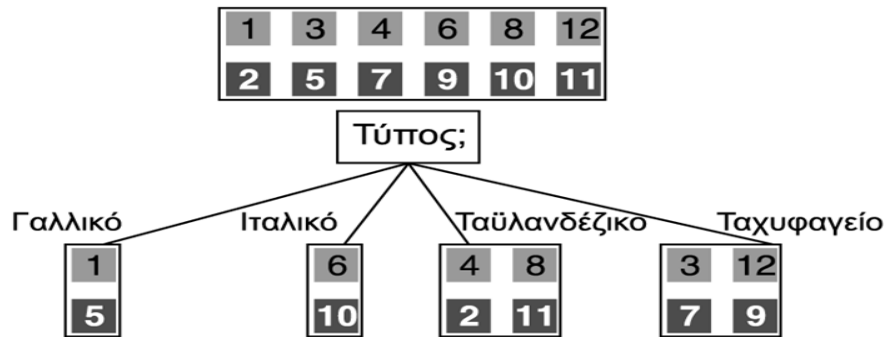
Το Παράδειγμα του Εστιατορίου

- 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, Ταϊλανδέζικο, ή ταχυφαγείο.
 - Εκτίμηση Αναμονής: 0'-10', 10'-30', 30'-60', >60'.

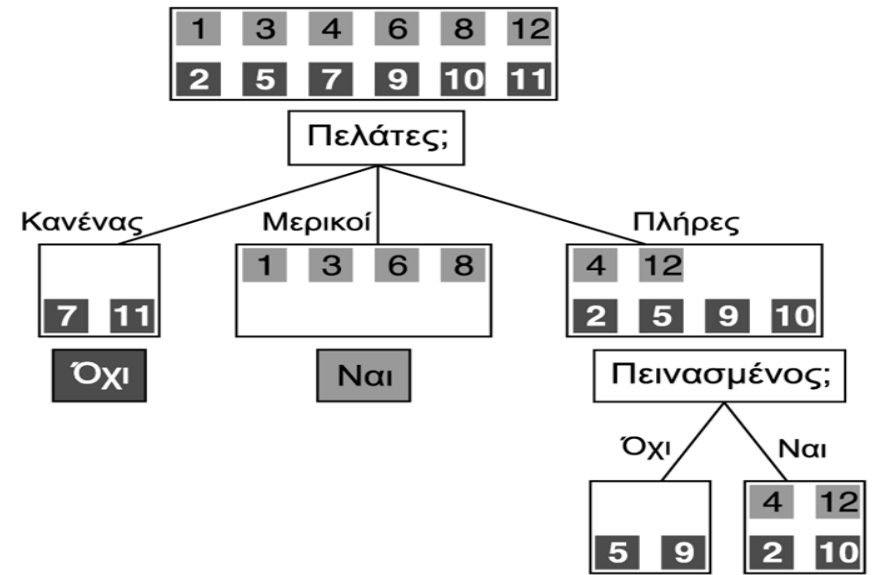
Παραδείγματα εκπαίδευσης

#	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 δένδρου αποφάσεων

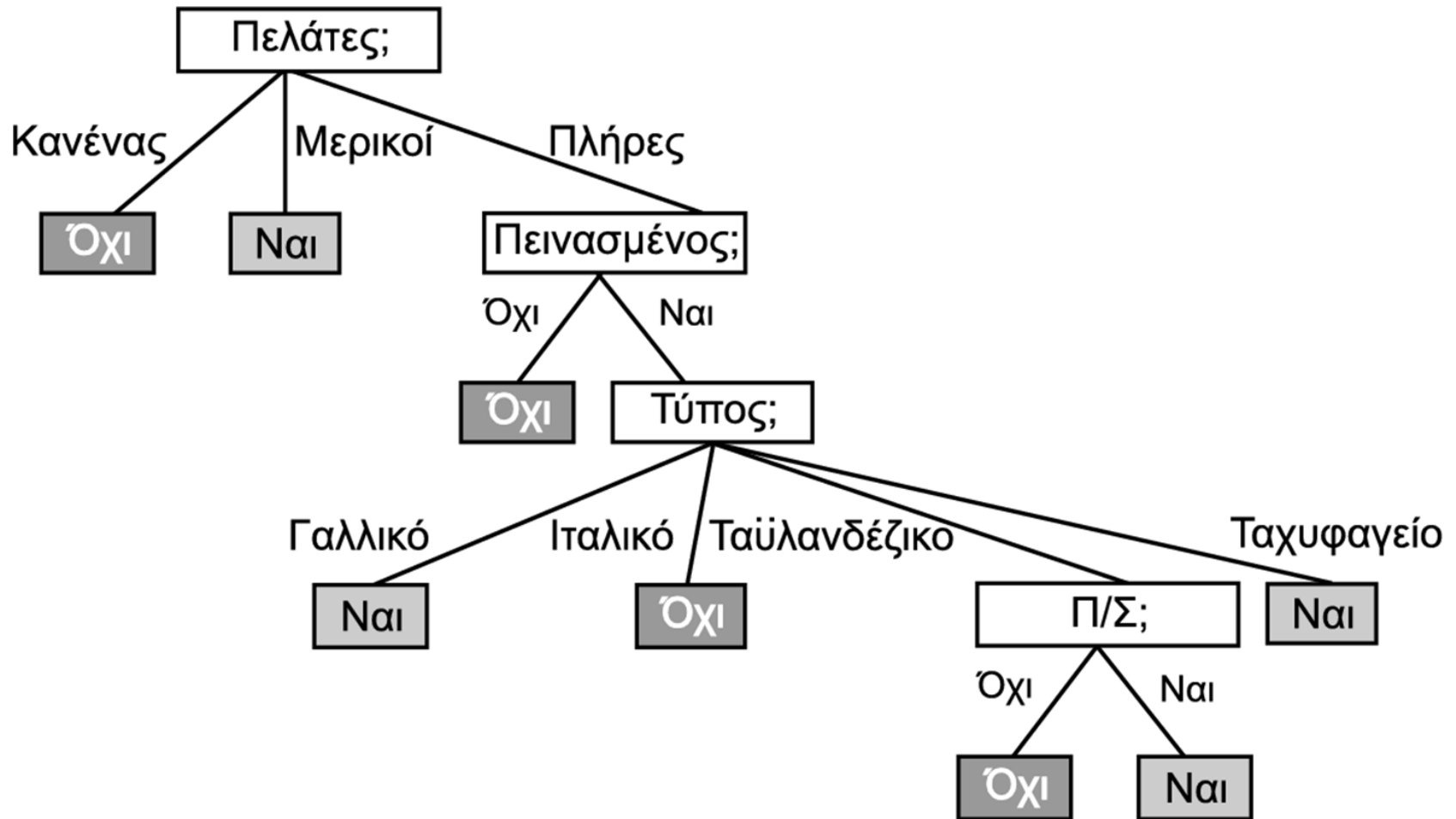


(α)

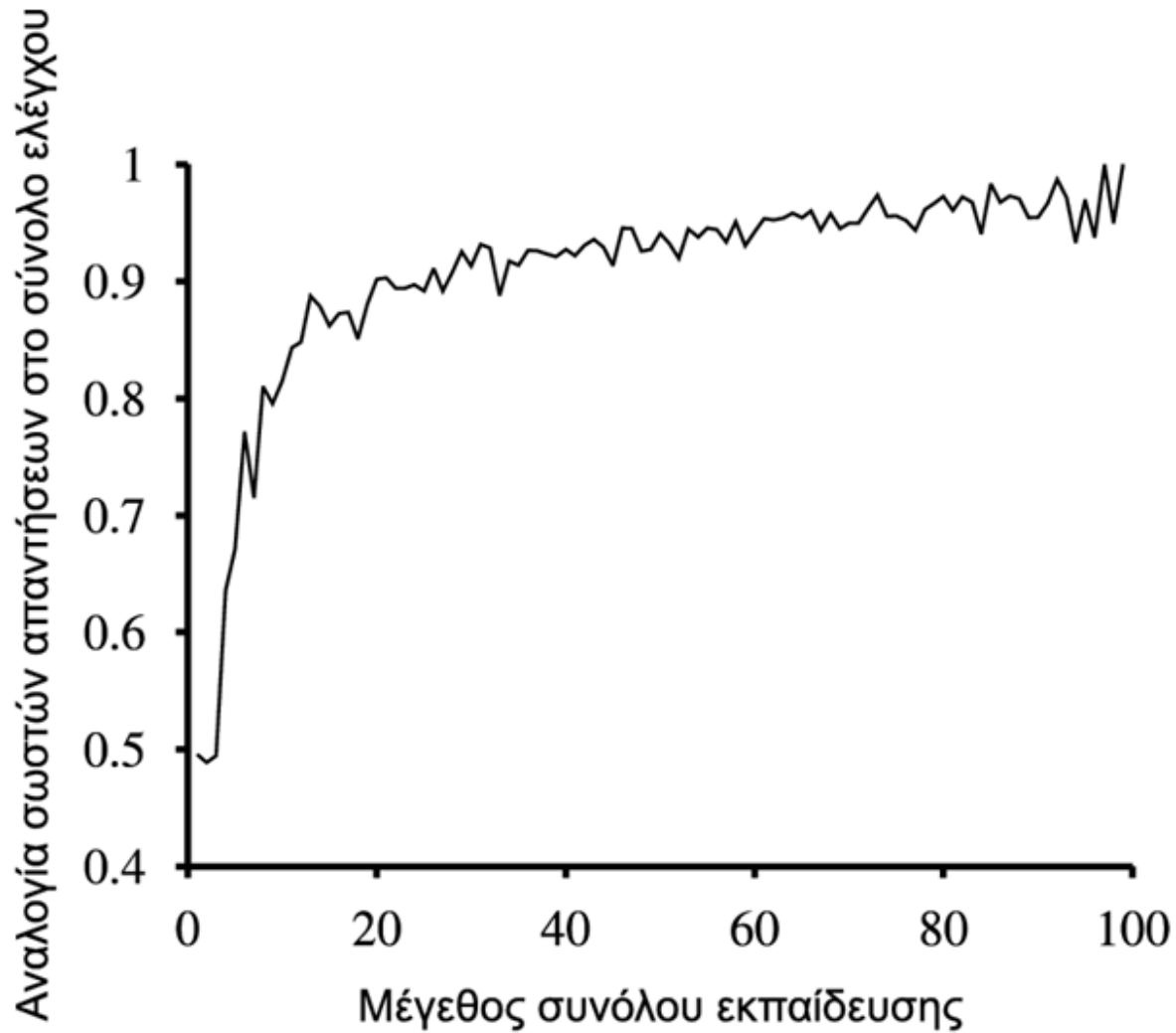


(β)

Τελικό δένδρο

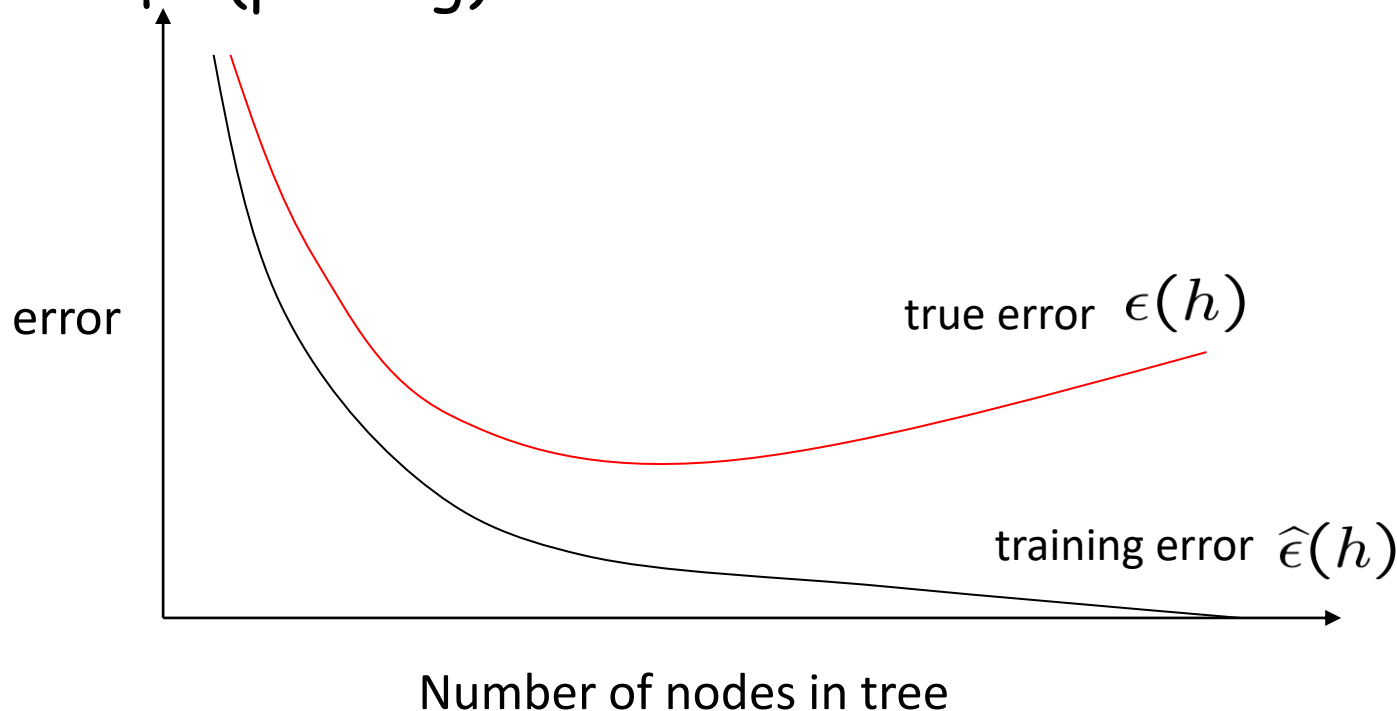


Αξιολόγηση ταξινόμησης



Το πρόβλημα της υπερπροσαρμογής (Overfitting)

- Το επαγόμενο δένδρο έχει την τάση να αποδίδει καλύτερα στην ταξινόμηση των παραδειγμάτων εκπαίδευσης, παρά των παραδειγμάτων ελέγχου.
- Χαμηλή ικανότητα γενίκευσης
- Κλάδεμα (pruning)



Reduced error pruning

Table 1.2 **The weather data.**

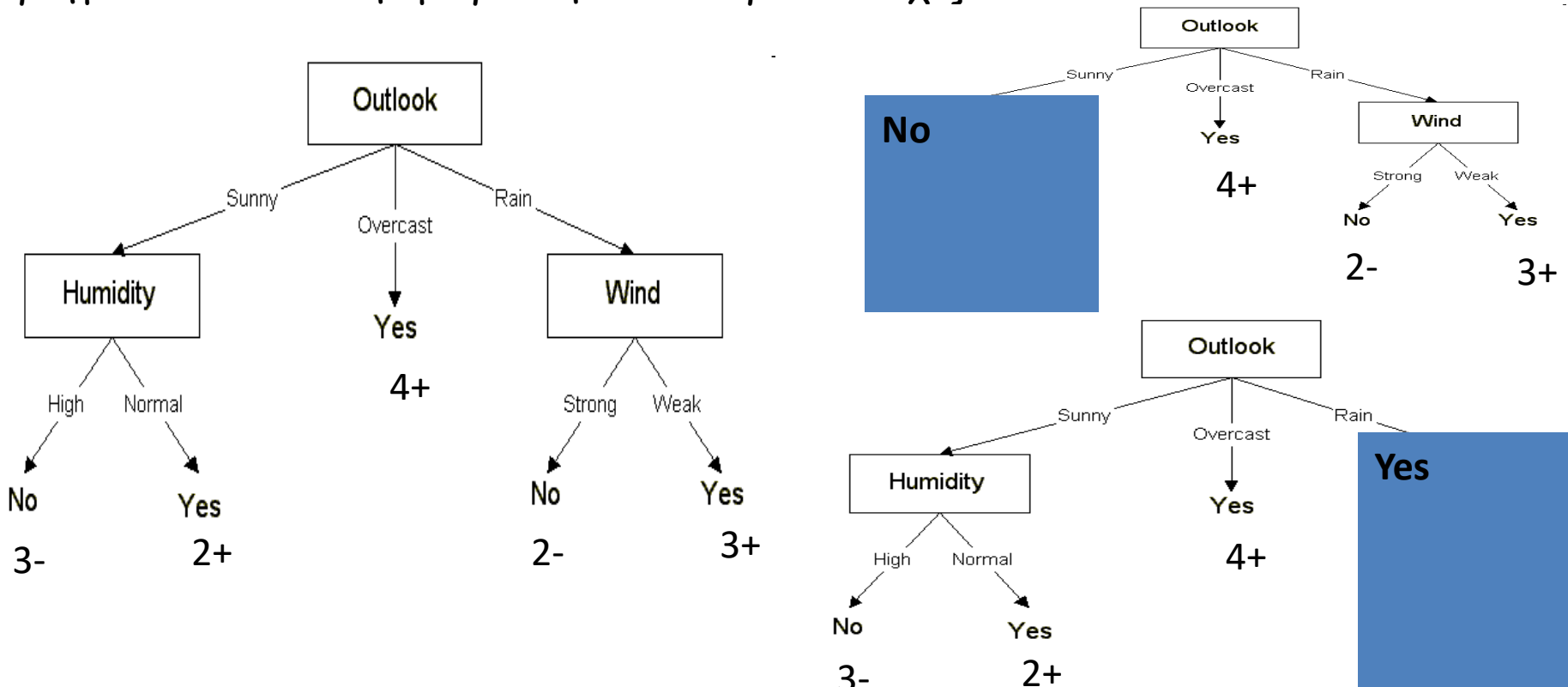
Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Reduced error pruning

Ξεχωρίζεται ένα σετ επικύρωσης (validation set) πριν την εκπαίδευση
Πραγματοποιείται εκπαίδευση επομένως με λιγότερα δεδομένα
Υπολογίζεται το σφάλμα ταξινόμησης του δέντρου που προκύπτει στα δεδομένα επικύρωσης

Από κάτω προς τα πάνω, ένας-ένας κόμβος αφαιρείται και αντικαθίσταται με φύλλο
Υπολογίζεται ξανά το σφάλμα ταξινόμησης του μικρότερου δέντρου στα δεδομένα επικύρωσης

Αν είναι μικρότερο από το σφάλμα στο μεγαλύτερο δέντρο, η αντικατάσταση πραγματοποιείται & η σμίκρυνση του δέντρου συνεχίζεται.



Παράδειγμα

- Έστω ένα σύστημα που προβλέπει αν ένας πελάτης ξενοδοχείου θα ξανάρθει να μείνει στο ξενοδοχείο ή όχι. Τα χαρακτηριστικά που παίζουν ρόλο είναι η χρήση του ΣΠΑ, η χρήση του εστιατορίου και το μέγεθος του δωματίου. Έστω τα παρακάτω παραδείγματα εκπαίδευσης

	ΧΡΗΣΗ ΣΠΑ	ΧΡΗΣΗ ΕΣΤΙΑΤΟΡΙΟΥ	ΜΕΓΕΘΟΣ ΔΩΜΑΤΙΟΥ	ΘΑ ΞΑΝΑΡΘΕΙ:
1	ΠΟΛΥ	ΠΟΛΥ	2ΚΛΙΝΟ	ΝΑΙ
2	ΛΙΓΟ	ΠΟΛΥ	1ΚΛΙΝΟ	ΝΑΙ
3	ΛΙΓΟ	ΛΙΓΟ	1ΚΛΙΝΟ	ΟΧΙ
4	ΚΑΘΟΛΟΥ	ΛΙΓΟ	3ΚΛΙΝΟ	ΟΧΙ

- Σχεδιάστε το δέντρο απόφασης που ο id3 θα επάγει από τα παραπάνω παραδείγματα εκπαίδευσης.

ΣΠΑ ΠΟΛΥ: 1ΝΑΙ 0ΟΧΙ

ΛΙΓΟ: 1ΝΑΙ 1ΟΧΙ

ΚΑΘΟΛΟΥ: 0ΝΑΙ 1ΟΧΙ

ΕΣΤΙΑΤΟΡΙΟ ΠΟΛΥ: 2ΝΑΙ 0ΟΧΙ

ΛΙΓΟ: 0ΝΑΙ 2ΟΧΙ

ΜΕΓΕΘΟΣ 2ΚΛΙΝΟ: 1ΝΑΙ 0ΟΧΙ

1ΚΛΙΝΟ: 1ΝΑΙ 1ΟΧΙ

3ΚΛΙΝΟ: 0ΝΑΙ 1ΟΧΙ

- Έστω πως ένας πελάτης κάνει λίγη χρήση του εστιατορίου, πολλή χρήση του ΣΠΑ και κλείνει 2κλινο Δωμάτιο. Τι θα προβλέψει για τον πελάτη ο id3;

