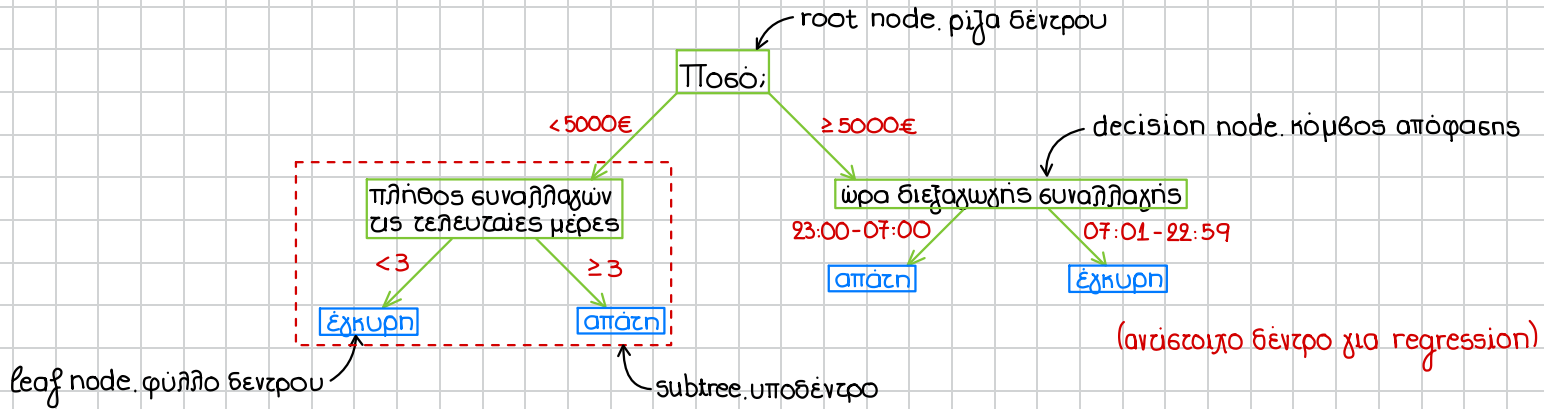


# DECISION TREES

Τα δέντρα αποφάσεων χρησιμοποιούνται σε προβλήματα τόσο regression όσο και classification.

**Πρόβλημα:** Μια τράπεζα βάσει δεδομένων όπως ο χρόνος και το πλήθος θέλει να κατηγοριοποιήσει ηλεκτρονικές συναλλαγές που πραγματοποιούνται μέσω αυτής σε απάτες και έγκυρες.

**Σκέψη:** Μπορούμε πραγματοποιώντας μία σειρά από ερωτήσεις να καταλήγουμε σε ένα συμπέρασμα σχετικά με τον χαρακτήρα της συναλλαγής. Η διαδικασία αυτή θα έχει τη δομή δέντρου. Δηλαδή:



Ουσιαστικά, ένα δέντρο απόφασης «επεί» ένα σύνολο δεδομένων σε μικρότερα υποσύνολα.

**Σκοπός:** Να καταλήγουμε σε όσο το δυνατόν πιο ομογενή σύνολα.

Αυτό επιτυγχάνεται μέσω μετρικών όπως το Gini impurity (αλγόριθμος CART) και η εντροπία entropy (αλγόριθμος ID3) (Ελαχιστοποίηση MSE σε περιπτώσεις regression) (+ search threshold)

**Ο σκοπός μας**, εάν δεν λάβουμε τα κατάλληλα μέτρα, ενδέχεται να οδηγήσει σε overfitting. Άρα θα χρησιμοποιήσουμε Regularization.

Τα δέντρα αποφάσεων επηρεάζονται από τη μορφή των δεδομένων άρα πραγματοποιούμε κλιμάκωση scaling. Επίσης, μικρές αλλαγές στα δεδομένα ή στις υπερπαραμέτρους ενδέχεται να οδηγήσουν σε πολύ διαφορετικά μοντέλα αποφάσεων.

↓ χρησιμοποιώ πολλά δέντρα αποφάσεων και βγάλω έναν «μέσον όρο»

## RANDOM FOREST

**Σκέψη:** Επιλέγουμε υποσύνολα των δεδομένων και κατασκευάζουμε με αυτά ως input δέντρα αποφάσεων. Υπολογίζουμε τον μέσον όρο των δέντρων (για regression) ή χρησιμοποιούμε ψηφοφορία voting (για classification).

## BOOSTING

Είναι μία μέθοδος που συνδυάζει «κακά» μοντέλα σε σειρά που μαθαίνουν το καθένα από τα λάθη του προηγούμενου. Τελικά προκύπτει ένα «καλό» μοντέλο.

Χρησιμοποιούνται AdaBoost και Gradient Boosting

# TUNING

Οι υπερπαραμέτροι *hyperparameters* είναι ένα σύνολο μεταβλητών που αφορούν τον αλγόριθμο μηχανικής μάθησης τον ίδιο και όχι το μοντέλο. Αυτές επηρεάζουν το training του μοντέλου.

*Tuning* είναι μία πειραματική διαδικασία μέσω της οποίας βρίσκουμε τις βέλτιστες τιμές των υπερπαραμέτρων για να εξασφαλίσουμε την καλύτερη δυνατή επίδοση του μοντέλου.