

**MAI4CAREU**

Master programmes in Artificial  
Intelligence 4 Careers in Europe



Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

# MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

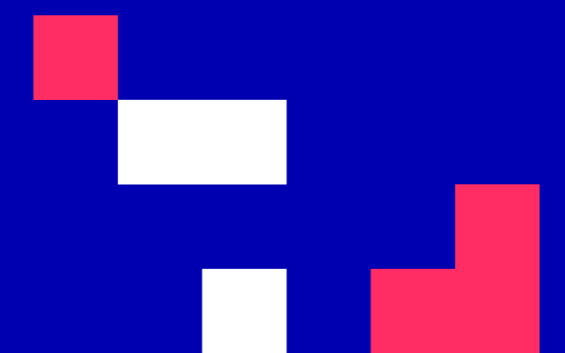
Διάλεξη 6: Δέντρα και δάση

Βασίλης Βασιλειάδης, PhD

Χειμερινό Εξάμηνο 2022/23



**CYENS**  
CENTRE OF EXCELLENCE





## Διάλεξη 6: Δέντρα και δάση

### Μαθησιακά αποτελέσματα

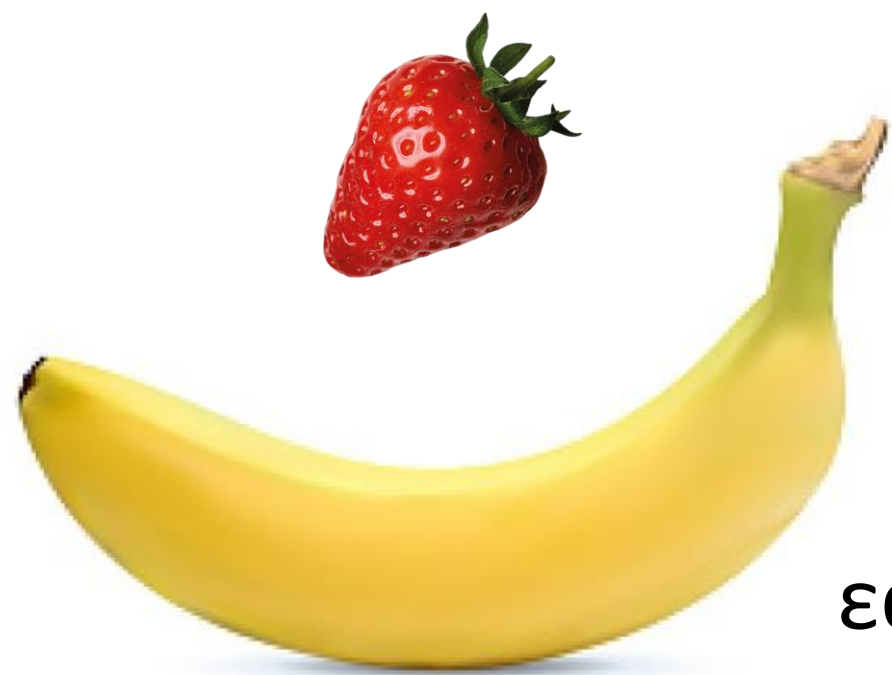
Θα καταλάβετε:

1. πώς λειτουργούν τα μοντέλα δέντρων αποφάσεων
2. πώς να εκπαιδεύσετε δέντρα αποφάσεων χρησιμοποιώντας τις έννοιες της εντροπίας και του κέρδους πληροφοριών
3. πώς να χρησιμοποιήσετε συνεχείς μεταβλητές στα δέντρα αποφάσεων
4. δέντρα ταξινόμησης και παλινδρόμησης
5. γιατί τα σύνολα μοντέλων μπορούν να επιτύχουν χαμηλότερα σφάλματα γενίκευσης
6. μέθοδοι ensemble όπως bagging, τυχαία δάση, boosting, XGBoost and stacking





## Εισαγωγικό παράδειγμα: ταξινομητής εικόνας φρούτων



εάν τα περισσότερα pixels είναι κίτρινα:  
εάν το σχήμα είναι ωοειδές: λεμόνι  
διαφορετικά: μπανάνα



τα περισσότερα pixels είναι κόκκινα: φράουλα  
διαφορετικά : lime



**Θετικά:**

**+ Δυνατότητα ερμηνείας**

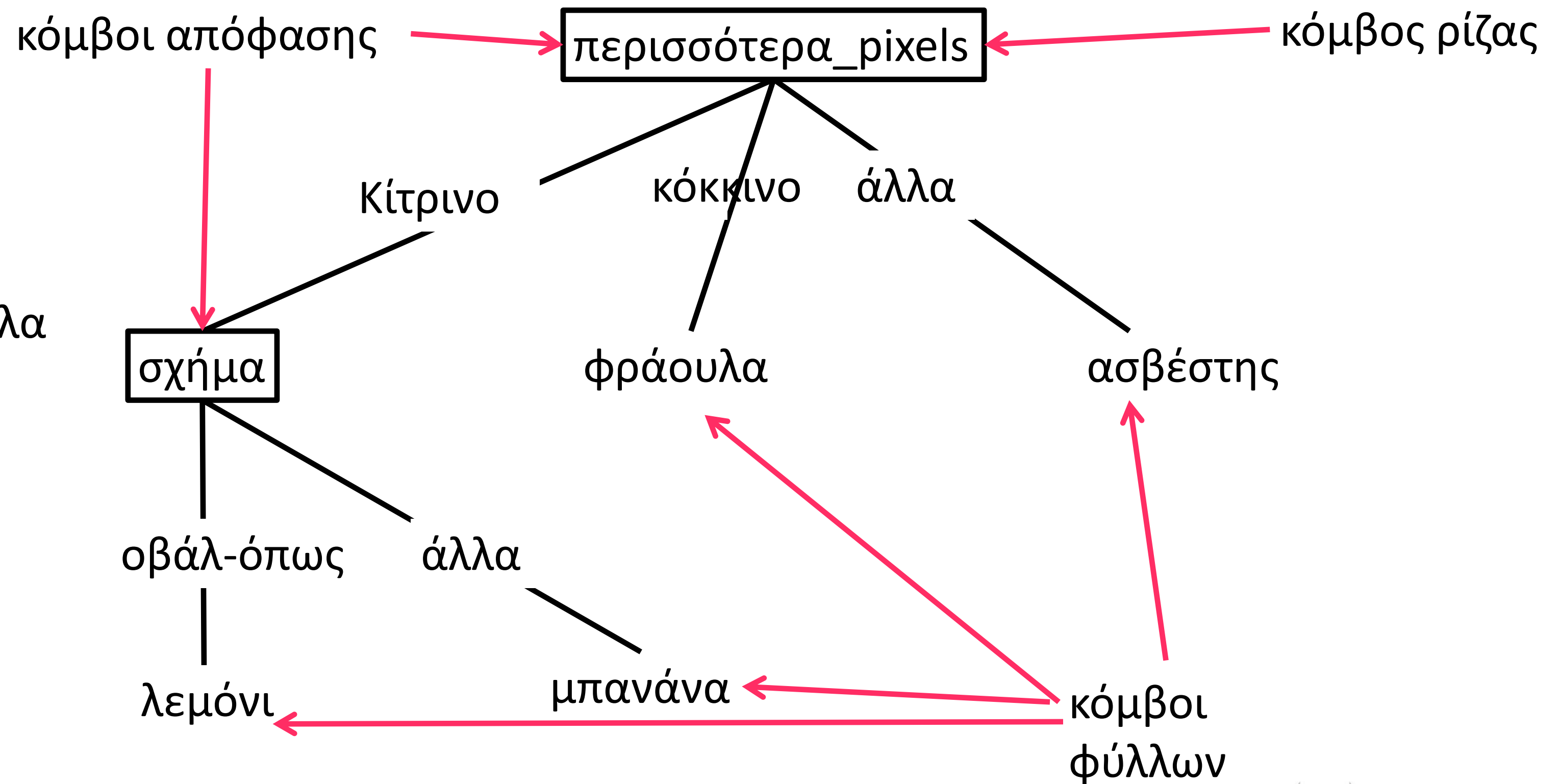
**+ Γρήγορη πρόβλεψη**

**Υπάρχει τρόπος να μάθουμε τέτοια  
προγράμματα από τα δεδομένα;**



# Αναπαράσταση δένδρων αποφάσεων για τον ταξινομητή εικόνας φρούτων

εάν τα περισσότερα pixels είναι κίτρινα:  
 εάν το σχήμα είναι ωειδές: λεμόνι  
 διαφορετικά : μπανάνα  
 τα περισσότερα pixels είναι κόκκινα: φράουλα  
 διαφορετικά: lime







## Ένα άλλο παράδειγμα



[ΠΗΓΗ](#)

Nissi Beach

Αγία Νάπα

Κύπρος





## Συλλογή δεδομένων ηλιακού εγκαύματος

Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	Light	No	Sunburned
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Short	Average	Yes	None
Annie	Blonde	Short	Average	No	Sunburned
Emily	Red	Average	Heavy	No	Sunburned
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Kate	Blonde	Short	Light	Yes	None

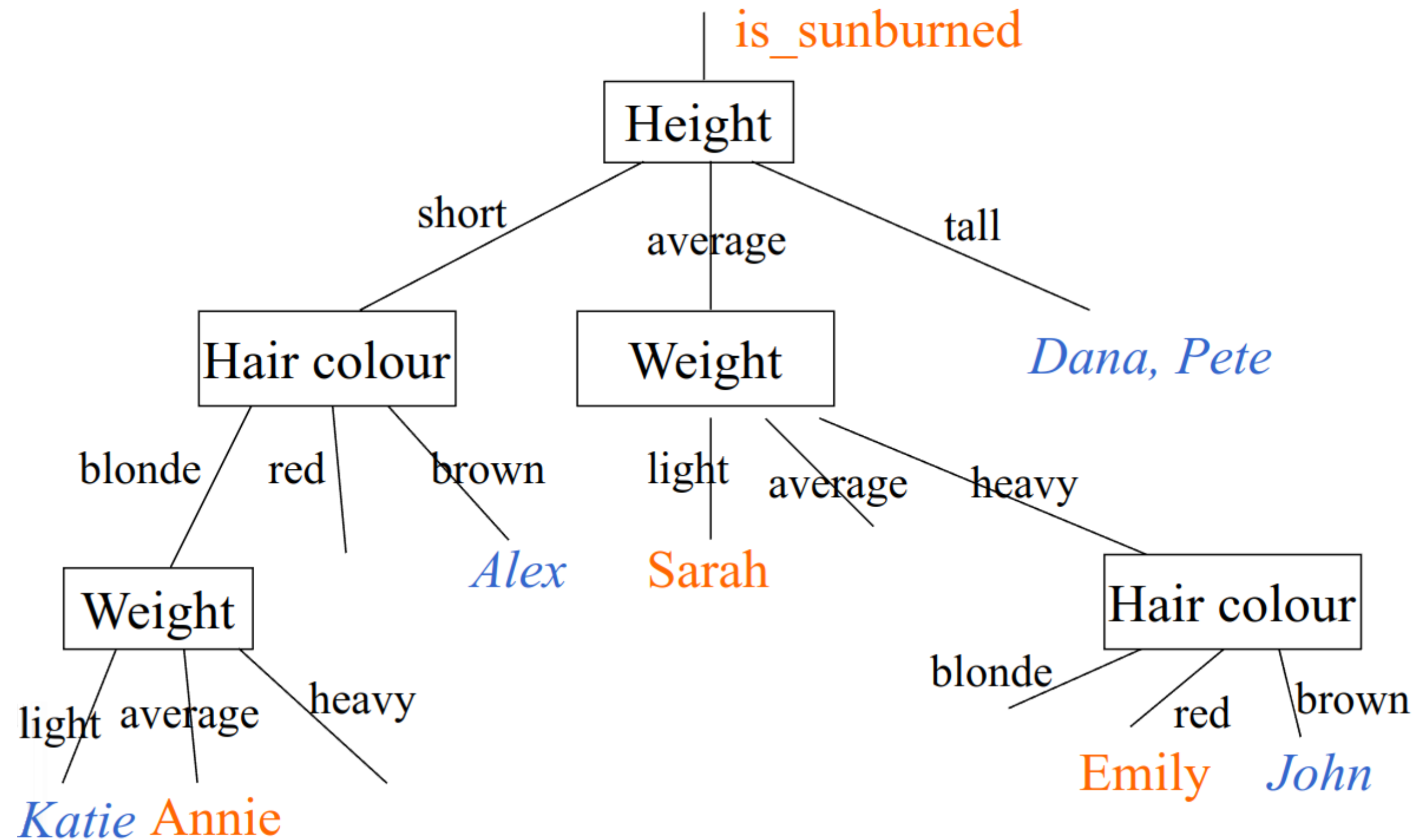
Θέλουμε να προβλέψουμε αν κάποιος άνθρωπος θα πάθει ένα ηλιακό έγκαυμα με βάση τα άλλα χαρακτηριστικά

Slide βασισμένο στο μάθημα μηχανικής μάθησης Ata Kaban — Πανεπιστήμιο του Μπέρμιγχαμ.





# Δέντρο απόφασης 1



## Οι πάσχοντες από ηλιακά εγκαύματα είναι...

If height="average" then

– if weight="light" then

- `return(true) ;;; Sarah`

– elseif weight="heavy" then

- if hair\_colour="red" then

- `return(true) ;;; Emily`

elseif height="short" then

– if hair\_colour="blonde" then

- if weight="average" then

- `return(true) ;;; Annie`

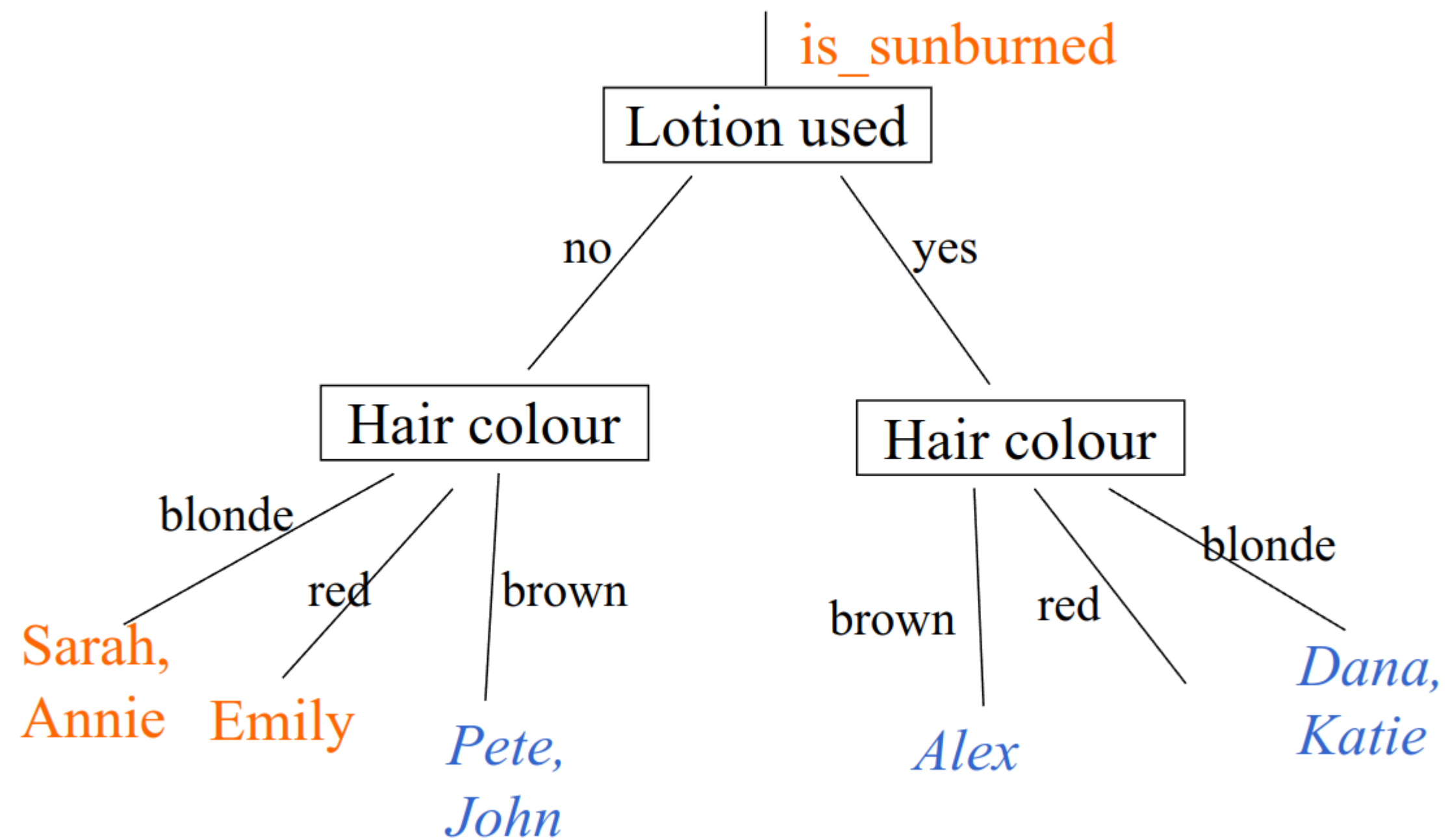
else `return(false) ;;;everyone else`

Slide βασισμένο στο μάθημα μηχανικής μάθησης Ata Kaban — Πανεπιστήμιο του Μπέρμιγχαμ.





## Δέντρο Αποφάσεων 2



Απλούστερο από το προηγούμενο δέντρο αποφάσεων

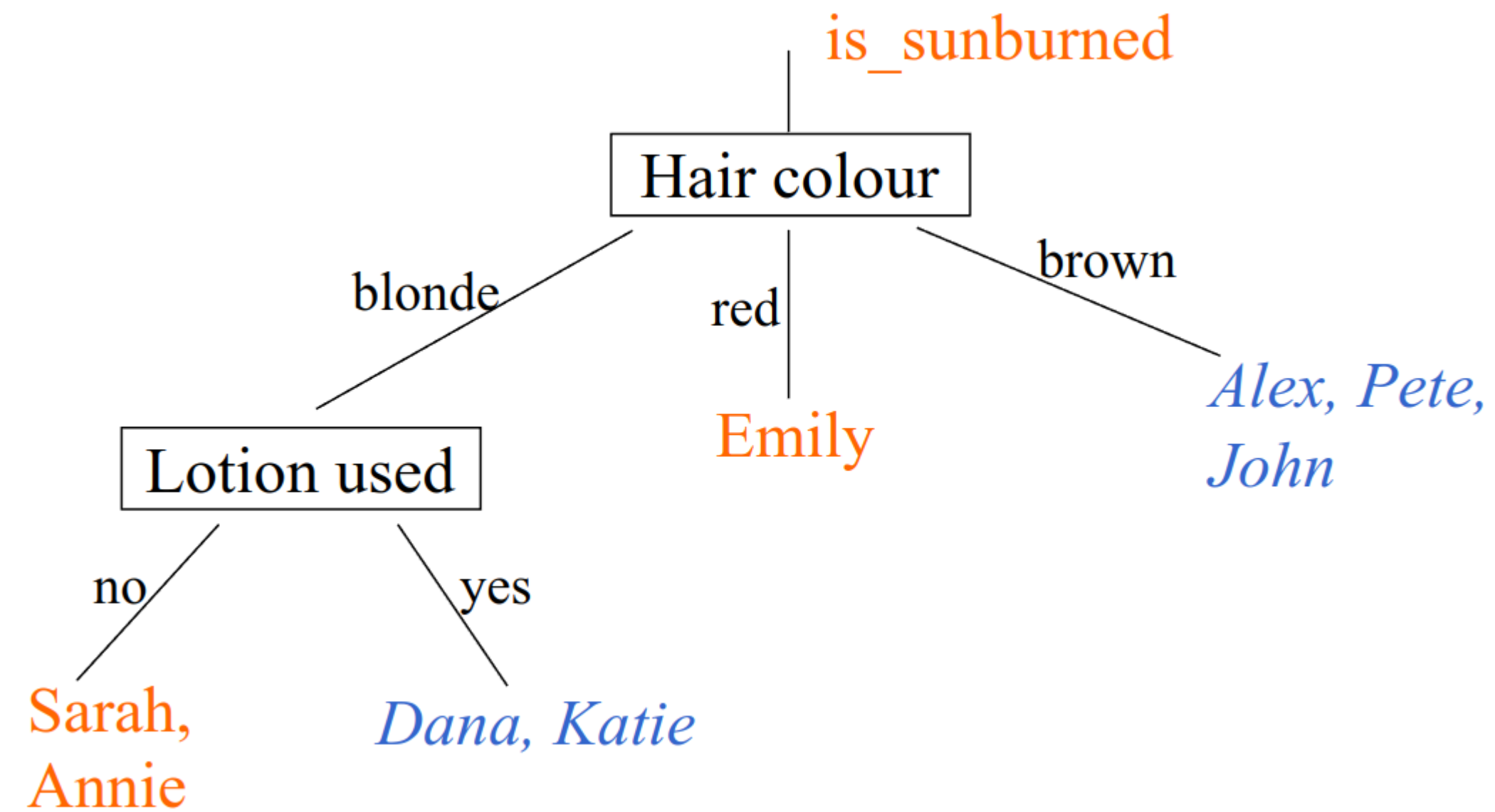
Slide βασισμένο στο μάθημα μηχανικής μάθησης Ata Kaban — Πανεπιστήμιο του Μπέρμιγχαμ.







## Δέντρο Αποφάσεων 3



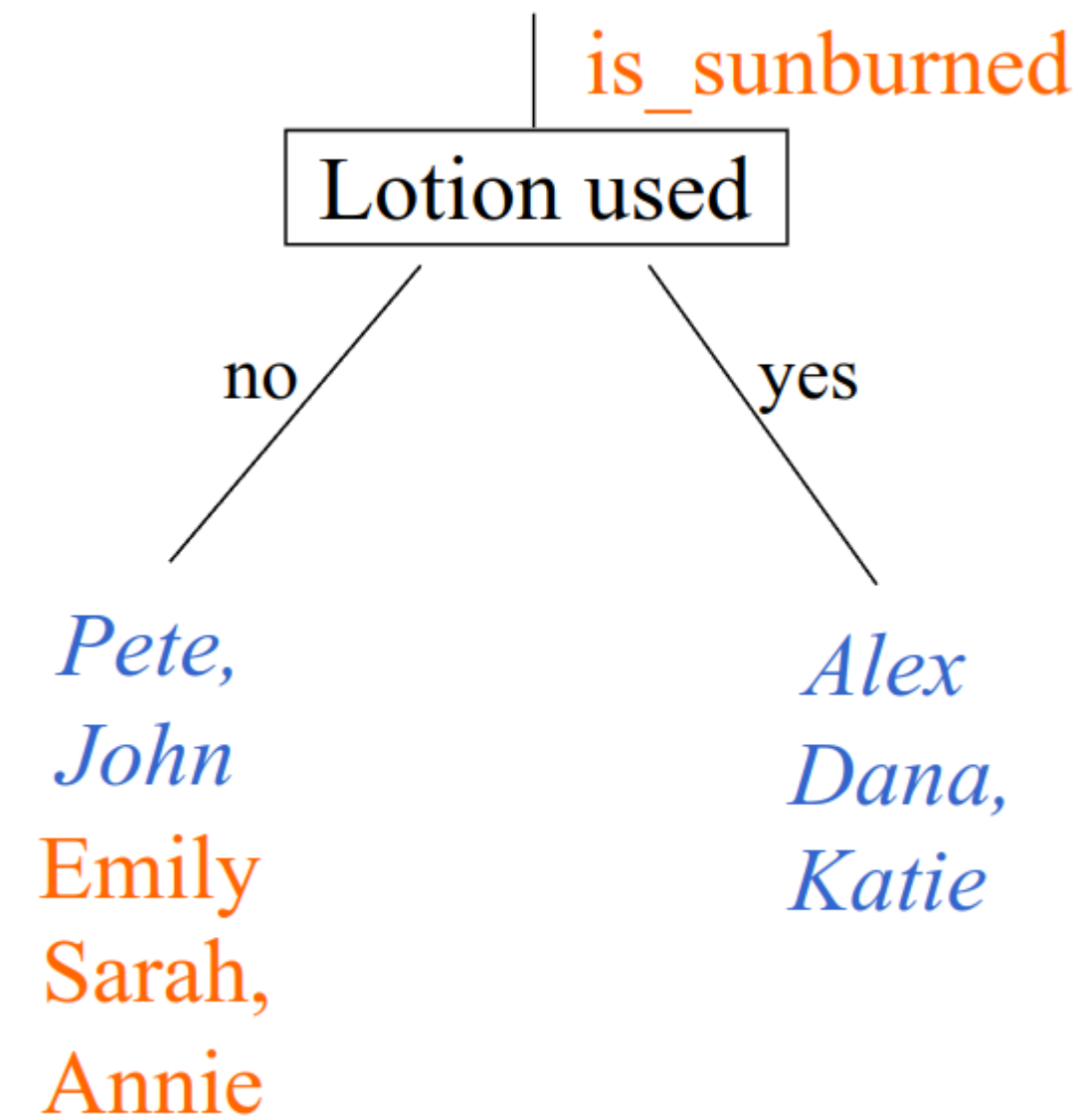
Ακόμα πιο απλό από το προηγούμενο δέντρο αποφάσεων

Slide βασισμένο στο μάθημα μηχανικής μάθησης Ata Kaban — Πανεπιστήμιο του Μπέρμιγχαμ.





## Δέντρο Αποφάσεων 4



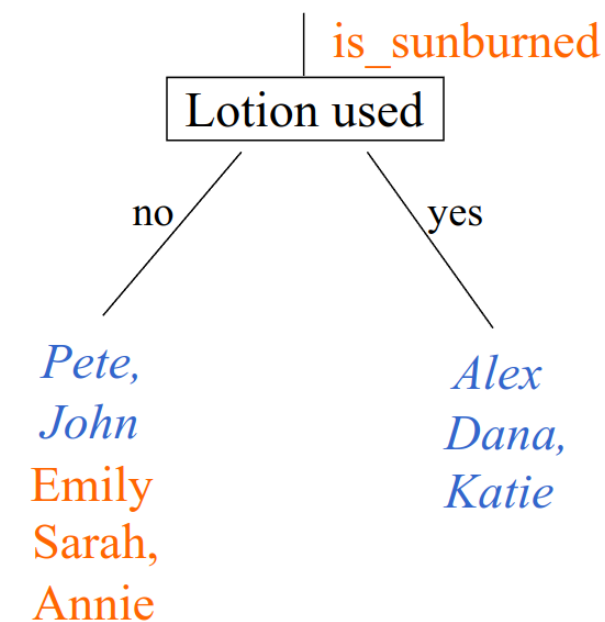
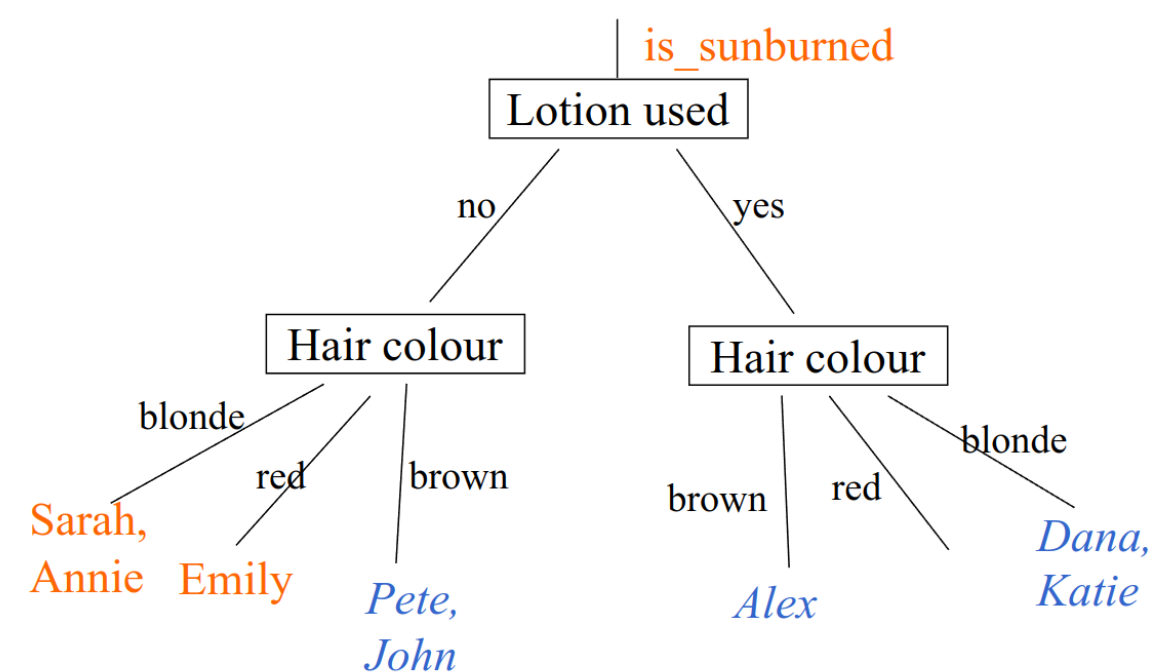
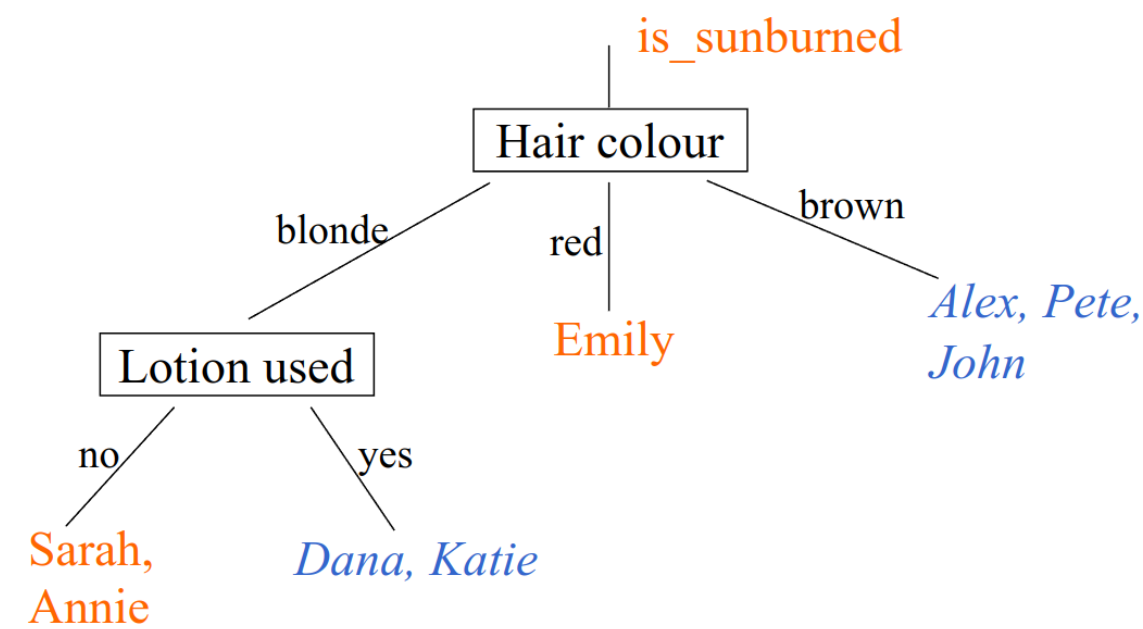
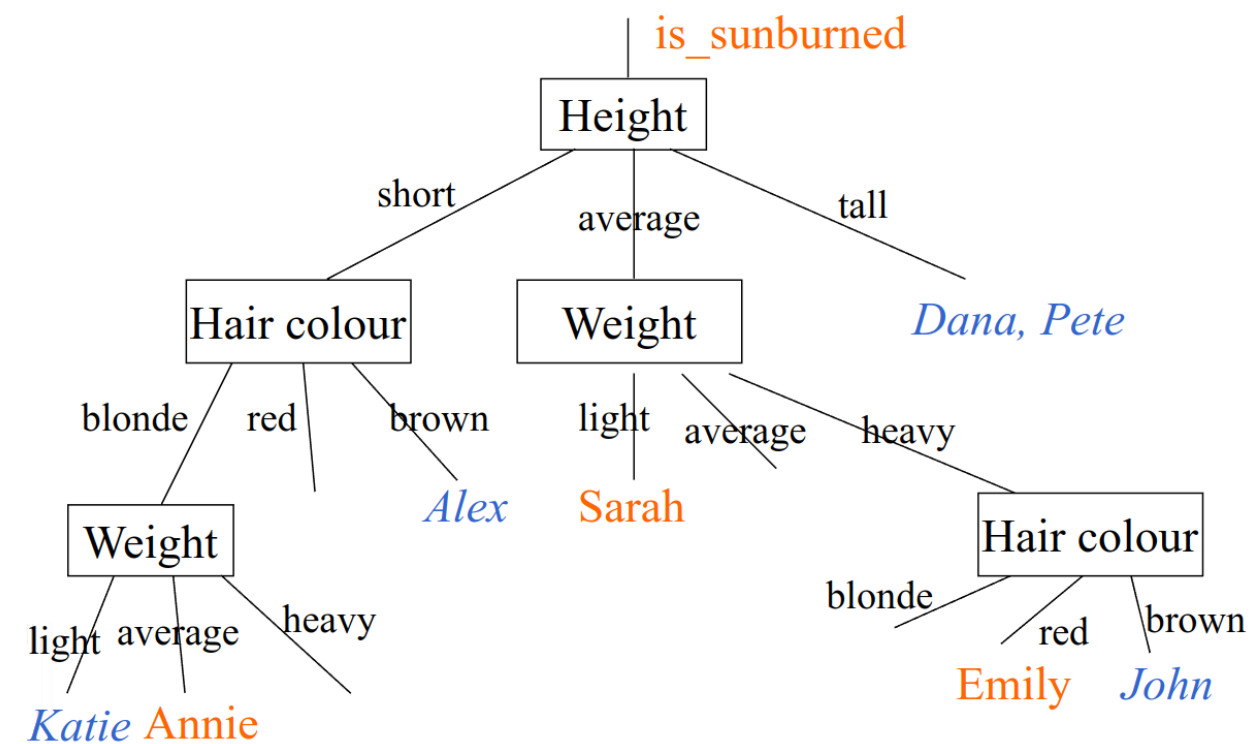
Δεν μπορούμε να πούμε  
πολλά από αυτό







## Πολλές δυνατότητες



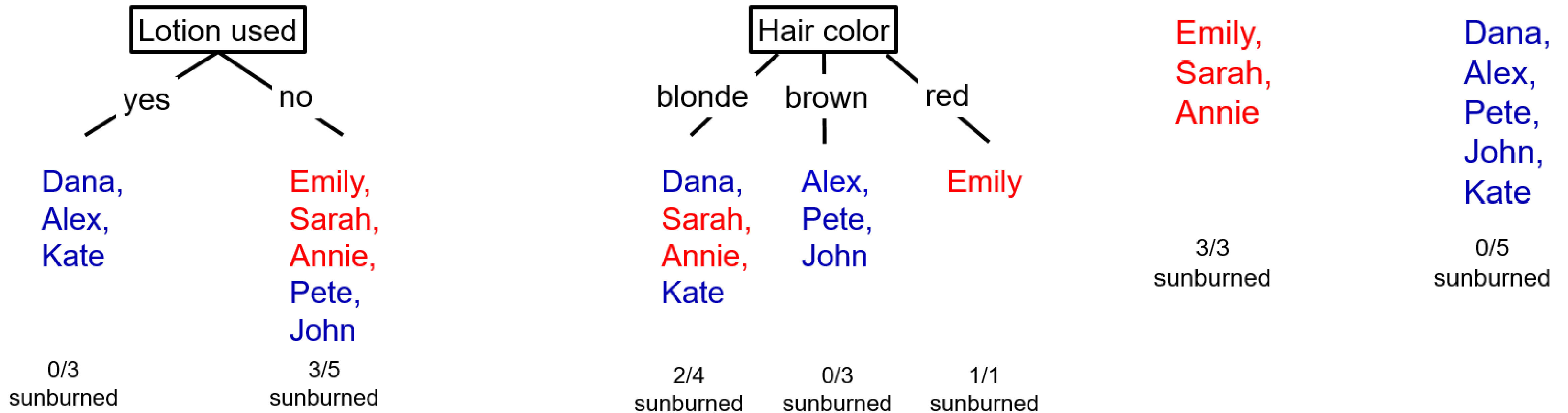
- Μεταξύ αυτών των δέντρων αποφάσεων, μερικοί τα πάνε καλύτερα από άλλα για την κατάρτιση, την επικύρωση και τα σύνολα δοκιμών.
- Ο αλγόριθμος εκμάθησης πρέπει να επιλέξει μία από όλες τις δυνατότητες για να τα πάει καλά στο σύνολο της εκπαίδευσης και ιδανικά για να γενικευτεί καλά στα σύνολα επικύρωσης και δοκιμών.





## Εκπαίδευση Δέντρων Αποφάσεων

**Απόφαση 1:** Πώς να επιλέξετε ποια λειτουργία θα χωρίσετε σε κάθε κόμβο  
 μεγιστοποίηση της τάξης (ή ελαχιστοποίηση της διαταραχής)



Slide βασισμένο στο [Andrew Ng's Machine Learning course - Coursera](#)





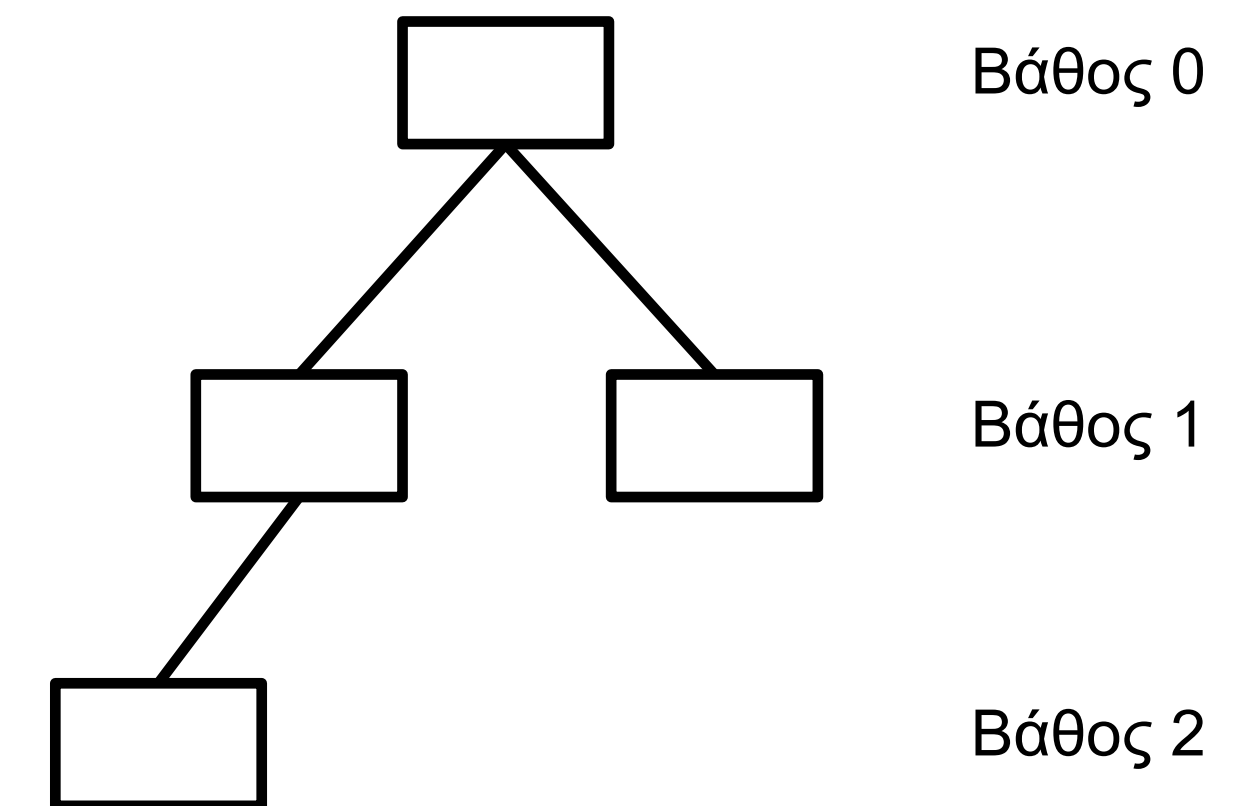


# Εκπαίδευση Δέντρων Αποφάσεων

## Απόφαση 2: Πότε σταματάς να χωρίζεις;

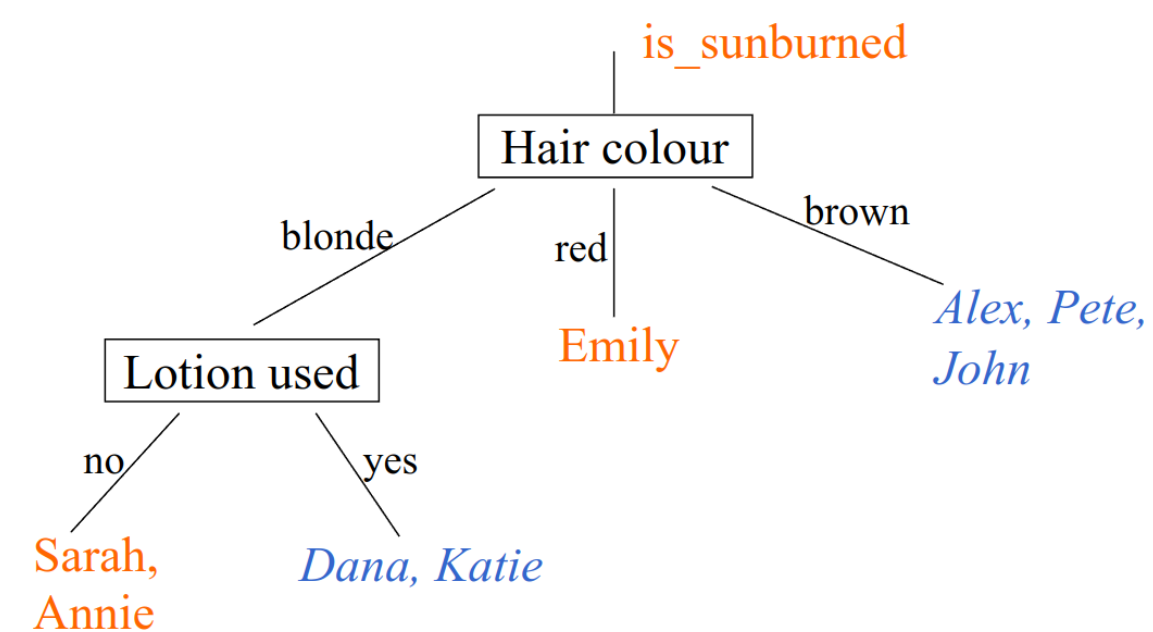
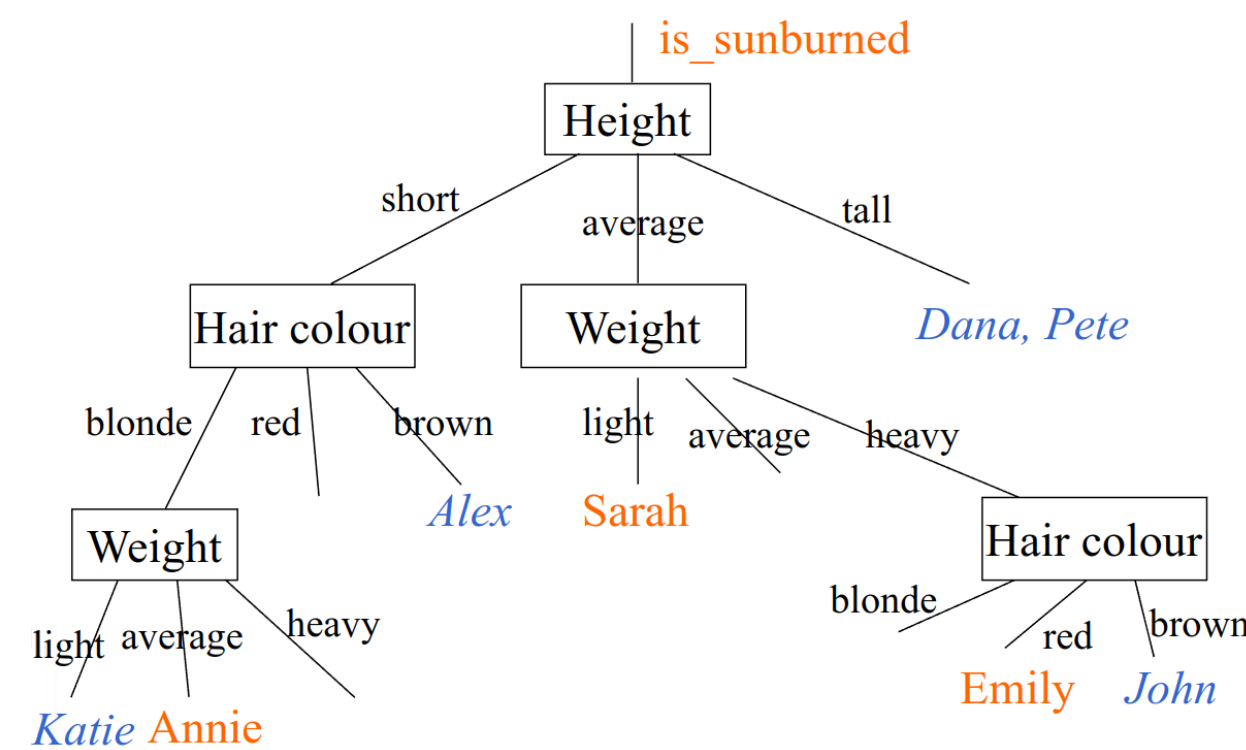
- Όταν ένας κόμβος είναι 100 % μία κατηγορία
- Κατά τη διάσπαση ενός κόμβου θα έχει ως αποτέλεσμα το δέντρο να υπερβαίνει ένα μέγιστο βάθος
- Όταν οι βελτιώσεις στη βαθμολογία «παραγγελίας» είναι κάτω από ένα κατώτατο όριο
- Όταν ο αριθμός των παραδειγμάτων στον κόμβο είναι κάτω από ένα κατώτατο όριο
- Όταν το σφάλμα στο σύνολο επικύρωσης αρχίσει να αυξάνεται

Slide βασισμένο στο [Andrew Ng's Machine Learning course - Coursera](#)





## Απόφαση 1: Επιλέγοντας το χαρακτηριστικό για να χωρίσετε



- **Άσχετα** χαρακτηριστικά δεν ταξινομούν τα δεδομένα καλά
- Η χρήση άσχετων χαρακτηριστικών προκαλεί έτσι μεγαλύτερα δέντρα αποφάσεων
- Ένας αλγόριθμος εκμάθησης θα μπορούσε να ψάξει για απλούστερα δέντρα αποφάσεων
- Ε: Πώς;

Slide βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.







## Α: Πώς το κάναμε;

- Ε: Ποιο είναι το καλύτερο χαρακτηριστικό για τον διαχωρισμό των δεδομένων;
- Α: Αυτό που είναι **πιο ενημερωτικό** για την ταξινόμηση που θέλουμε να πάρουμε
  
- Ε: Τι σημαίνει «περισσότερες πληροφορίες»;
- Α: Το χαρακτηριστικό που **μειώνει καλύτερα τη διαταραχή**
  
- Ε: Πώς μπορούμε να μετρήσουμε κάτι τέτοιο;
- Α: Δείτε τις επόμενες διαφάνειες

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





## Διαταραχή

- Χρειαζόμαστε μια ποσότητα για να μετρήσουμε τη διαταραχή σε ένα σύνολο παραδειγμάτων

$$S = \{s_1, s_2, \dots, s_m\}$$

όπου  $s_1 = \text{"Sarah"}$ ,  $s_2 = \text{"Dana"}$ , ...

- Ποιες ιδιότητες πρέπει να έχει η Διαταραχή (D);
  - Όταν  $D(S)=0 \rightarrow$  όλα τα πρώηνδείγματα στο S έχουν την ίδια κατηγορία
  - Όταν  $D(S)=1$  τα  $\rightarrow$  μισά από τα παραδείγματα στο S είναι μιας κατηγορίας και τα μισά είναι η αντίθετη κατηγορία
  - Χρειαζόμαστε μια λειτουργία που να λαμβάνει ως εισαγωγή το **ποσοστό** των θετικών παραδειγμάτων
    - Καθώς το ποσοστό αυξάνεται από 0 σε 0,5, D πρέπει να πάει από 0 σε 1
    - Καθώς το ποσοστό αυξάνεται από 0,5 σε 1, D πρέπει να πάει από 1 σε 0

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.







## Παραδείγματα

- $D(\{\text{'Dana'}, \text{'Pete'}\}) = 0$
- $D(\{\text{'Sarah'}, \text{'Annie'}, \text{'Emily'}\}) = 0$
- $D(\{\text{'Sarah'}, \text{'Emily'}, \text{'Alex'}, \text{'John'}\}) = 1$
- $D(\{\text{'Sarah'}, \text{'Emily'}, \text{'Alex'}\}) = ?$

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.



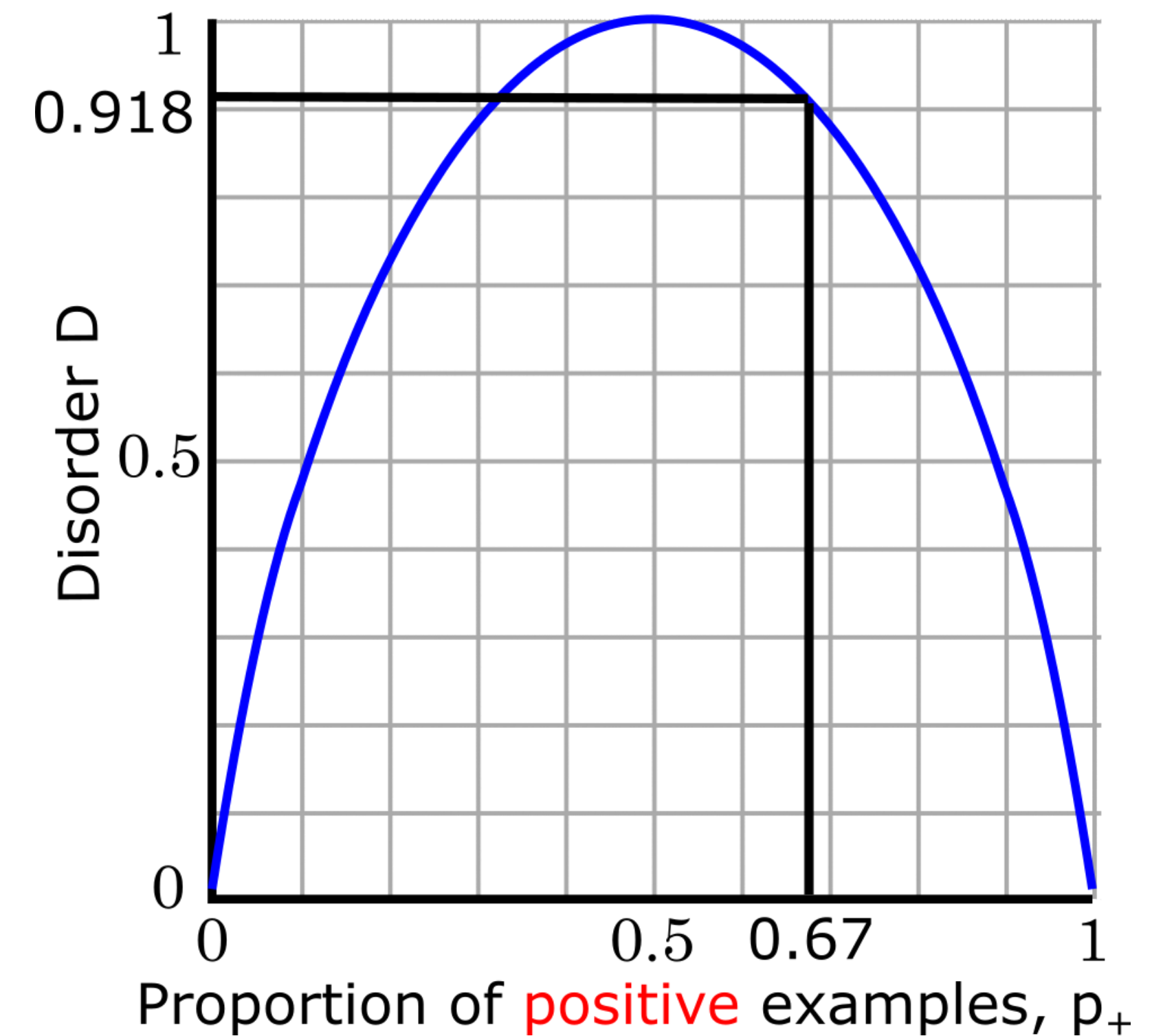


# Διαίσθηση συνάρτησης διαταραχής

$$D(\{\text{'Sarah'}, \text{'Emily'}, \text{'Alex'}\}) = 0.918$$

Χρειαζόμαστε μια λειτουργία που να λαμβάνει ως εισαγωγή το **ποσοστό** των **θετικών** παραδειγμάτων

- Καθώς το ποσοστό αυξάνεται από 0 σε 0,5, D πρέπει να πάει από 0 σε 1
- Καθώς το ποσοστό αυξάνεται από 0,5 σε 1, D πρέπει να πάει από 1 σε 0





## Η εντροπία ως μέτρο διαταραχής

Η **εντροπία** μετρά τη διαταραχή ενός συνόλου  $S$  που περιέχει συνολικά  $m$  παραδείγματα του οποίου  $m_+$   $m_-$  είναι θετικά και αρνητικά και δίνεται από

$$H(p_+) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

όπου

$$p_+ = m_+/m \quad \text{and} \quad p_- = m_-/m = 1 - p_+$$

- $H(0) = 0$
- $H(1) = 0$
- $H(0.5) = 1$







## Επιστροφή στην παραλία

$$D(\{\text{"Sara"}, \text{"Dana"}, \text{"Alex"}, \text{"Annie"}, \text{"Emily"}, \text{"Pete"}, \text{"John"}, \text{"Catie"}\}) = H(3/8)$$

$$H(3/8) = -\frac{3}{8} \log_2 \frac{3}{8} - \frac{5}{8} \log_2 \frac{5}{8}$$

$$= 0.954$$

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





# Ελαχιστοποίηση της διαταραχής

- Τι ξέρουμε:
  - Πώς να μετρήσετε τη διαταραχή ενός συνόλου
- Ό,τι έχει απομείνει:
  - Θέλουμε να μετρήσουμε πόσο θα μείωνε η διαταραχή ενός συνόλου γνωρίζοντας την αξία ενός συγκεκριμένου χαρακτηριστικού.
  - Αυτό θα μας βοηθήσει να καταλάβουμε ποια λειτουργία θα χρησιμοποιήσετε για τον διαχωρισμό των δεδομένων

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





# Κέρδος πληροφοριών

Το **κέρδος πληροφοριών** μετρά την αναμενόμενη **μείωση** της εντροπίας λόγω της διάσπασης σε ένα χαρακτηριστικό  $A$

$$Gain(S, A) = D(S) - \underbrace{\sum_{v \in Values(A)} \frac{|S_v|}{|S|} D(S_v)}$$

Η μέση διαταραχή είναι μόνο το σταθμισμένο άθροισμα των διαταραχών στα κλαδιά (υποσύνολα) που δημιουργούνται από τις τιμές του  $A$

Θέλουμε το μεγαλύτερο κέρδος

➤ Το ίδιο με τη μικρότερη μέση διαταραχή

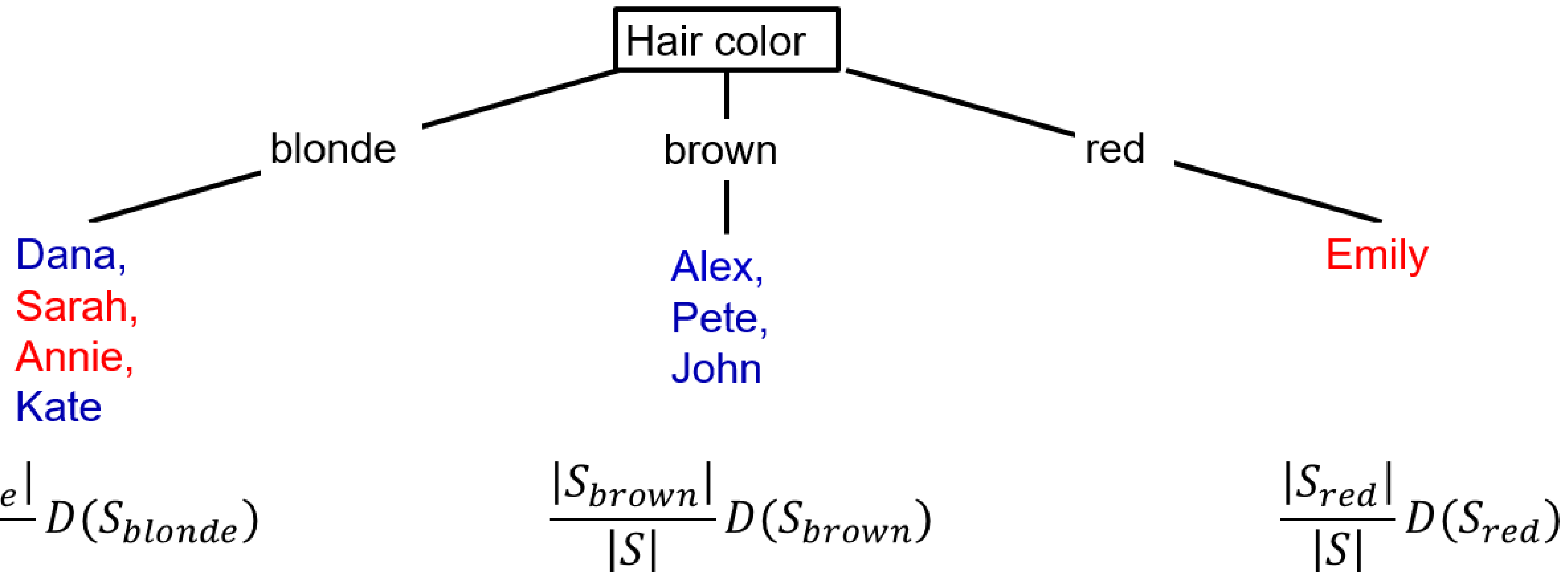
[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.







# Επιστροφή στην παραλία: υπολογίστε τη μέση διαταραχή που σχετίζεται με το χρώμα των μαλλιών

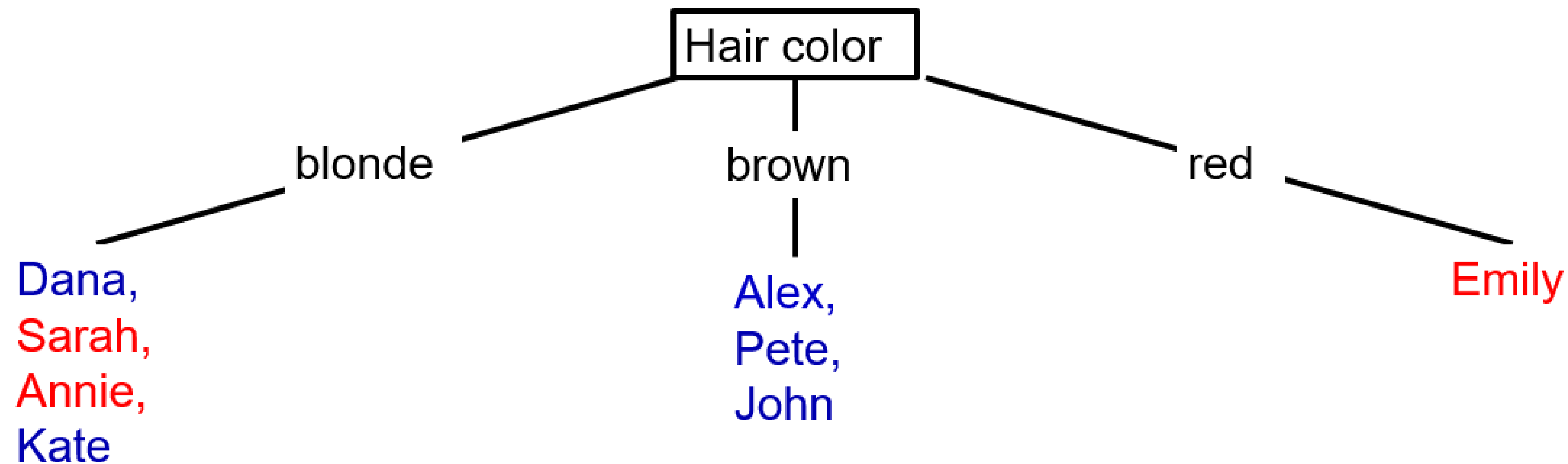


[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





# Επιστροφή στην παραλία: υπολογίστε τη μέση διαταραχή που σχετίζεται με το χρώμα των μαλλιών



$$\frac{4}{8} H(2/4) = 0.5 * 1$$

$$\frac{|S_{brown}|}{|S|} D(S_{brown})$$

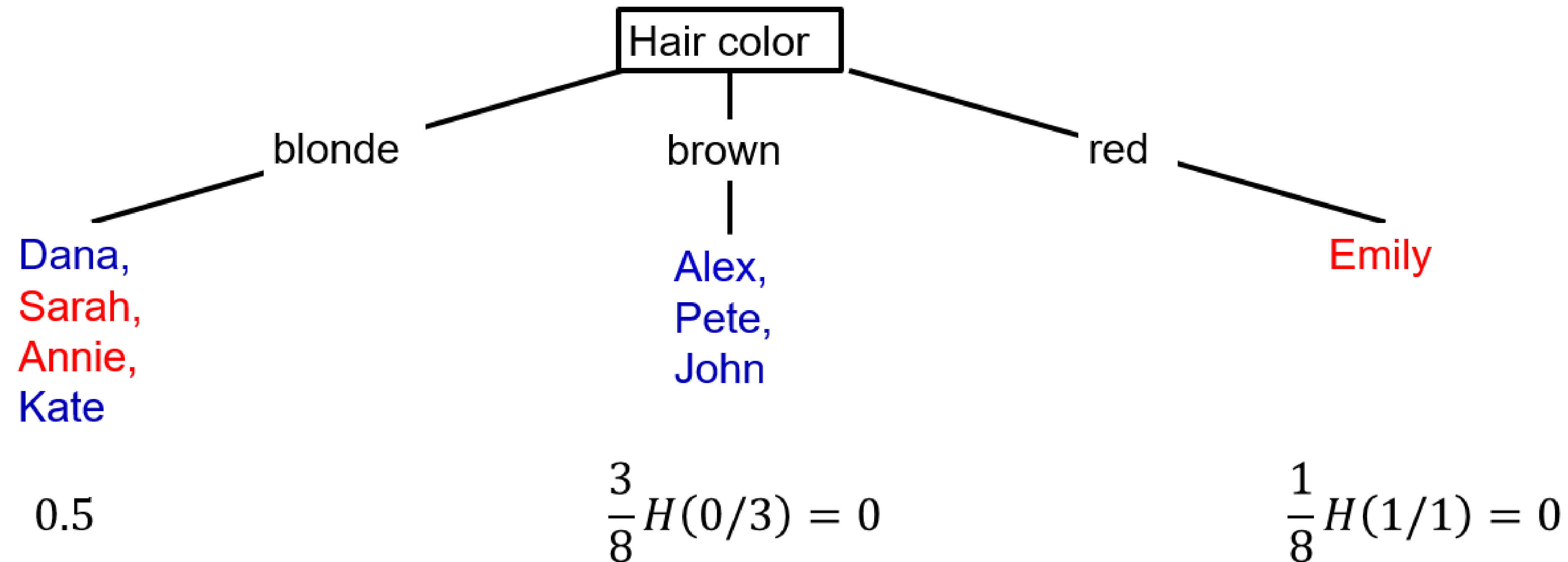
$$\frac{|S_{red}|}{|S|} D(S_{red})$$

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





# Επιστροφή στην παραλία: υπολογίστε τη μέση διαταραχή που σχετίζεται με το χρώμα των μαλλιών



Έτσι, η μέση διαταραχή που δημιουργείται κατά τη διάσπαση στο χρώμα των μαλλιών είναι  $0,5 + 0 + 0 = 0,5$

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.







## Ποια μεταβλητή απόφασης ελαχιστοποιεί τη διαταραχή;

Δοκιμή	Μέση διαταραχή	Κέρδος πληροφοριών
Χρώμα μαλλιών	0.5	$0.954 - 0,5 = 0,454$
Ύψος	0.69	$0.954 - 0,69 = 0.264$
Βάρος	0.94	$0.954 - 0,94 = 0,014$
Αντλιακό	0.61	$0.954 - 0,61 = 0,344$

Αυτό είναι ό, τι έχουμε μόλις υπολογίσει

Αυτές είναι οι μέσες διαταραχές των άλλων χαρακτηριστικών, υπολογιζόμενες με τον ίδιο τρόπο

- Ποια μεταβλητή απόφασης μεγιστοποιεί το κέρδος πληροφοριών στη συνέχεια;
- Αυτό που ελαχιστοποιεί τη μέση διαταραχή

### Χρώμα μαλλιών

[Slide](#) βασισμένο στο Ata Kaban's Machine Learning course - University of Birmingham.





## Διαχωρισμός σε συνεχή μεταβλητή

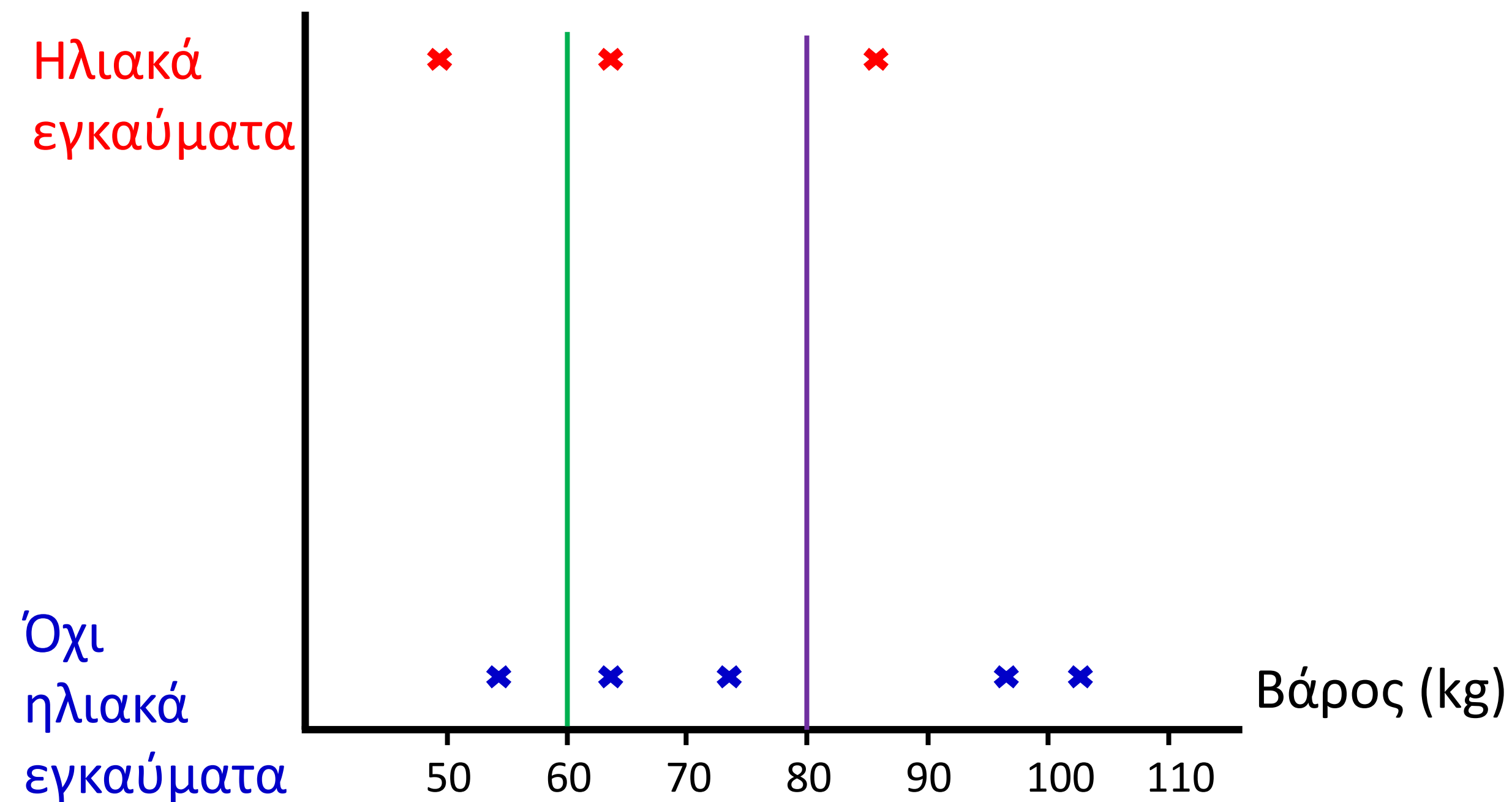
- Ας υποθέσουμε ότι ένα χαρακτηριστικό (π.χ., βάρος) αντιπροσωπεύεται με συνεχείς τιμές
- Πώς μπορούμε να χωρίσουμε αυτό το χαρακτηριστικό;

Name	Hair	Height	Weight	Lotion	Result
<b>Sarah</b>	Blonde	Average	49,3 kg	No	<b>Sunburned</b>
<i>Dana</i>	Blonde	Tall	63,6 kg	Yes	<i>None</i>
<i>Alex</i>	Brown	Short	72,1 kg	Yes	<i>None</i>
<b>Annie</b>	Blonde	Short	65,8 kg	No	<b>Sunburned</b>
<b>Emily</b>	Red	Average	86,7 kg	No	<b>Sunburned</b>
<i>Pete</i>	Brown	Tall	97,3 kg	No	<i>None</i>
<i>John</i>	Brown	Average	102,2 kg	No	<i>None</i>
<i>Kate</i>	Blonde	Short	53,4 kg	Yes	<i>None</i>





## Διαχωρισμός σε συνεχή μεταβλητή



Βάρος < 60 kg

$$\frac{2}{8}H(1/2) + \frac{6}{8}H(2/6) = 0.94$$

Βάρος < 80 kg

$$\frac{5}{8}H(2/5) + \frac{3}{8}H(1/3) = 0.95$$

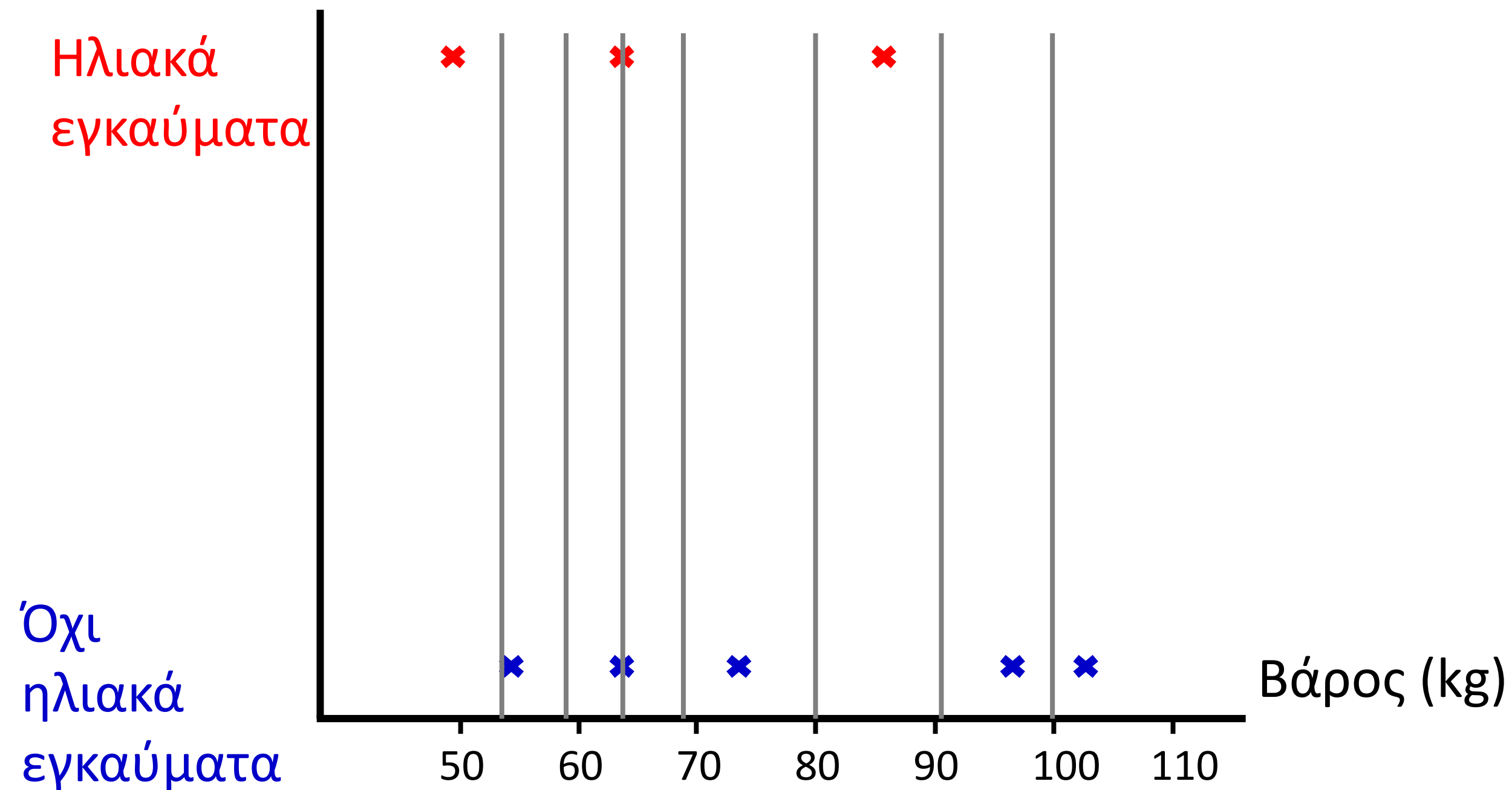
**Με βάση αυτές τις δύο διαιρέσεις, ποια θα επέλεγες;**







## Διαχωρισμός σε συνεχή μεταβλητή



### Διαδικασία

- Αποκτήστε ζεύγη ταξινομημένων σημείων και επιλέξτε το μεσαίο σημείο ως την τιμή για διαίρεση
- Υπολογίστε το κέρδος πληροφοριών για όλες τις διαιρέσεις
- Επιλέξτε το κατώτατο όριο με το υψηλότερο κέρδος πληροφοριών





## Δέντρα παλινδρόμησης

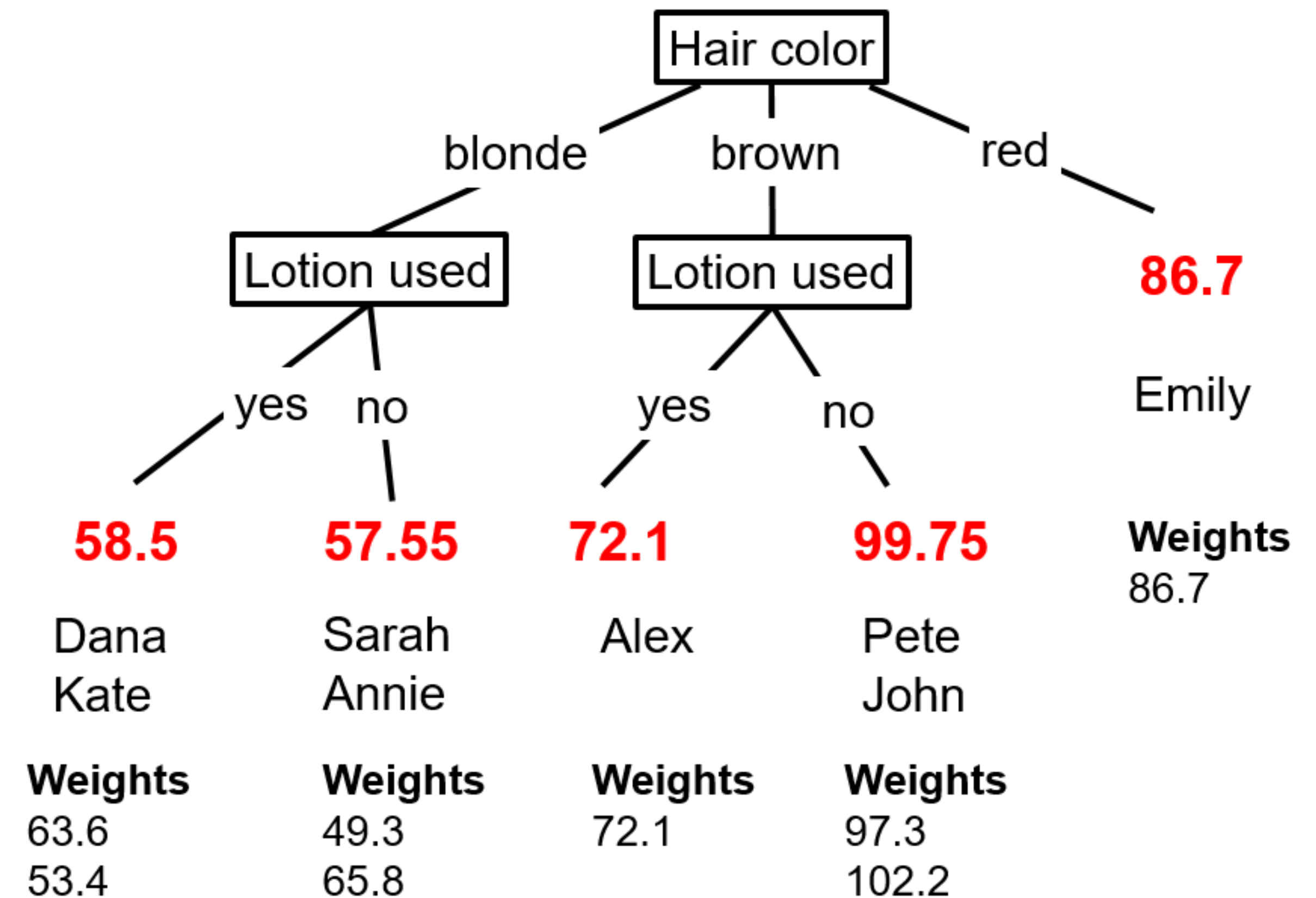
Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	49,3 kg	No	Sunburned
Dana	Blonde	Tall	63,6 kg	Yes	None
Alex	Brown	Short	72,1 kg	Yes	None
Annie	Blonde	Short	65,8 kg	No	Sunburned
Emily	Red	Average	86,7 kg	No	Sunburned
Pete	Brown	Tall	97,3 kg	No	None
John	Brown	Average	102,2 kg	No	None
Kate	Blonde	Short	53,4 kg	Yes	None

$x_1$

$x_2$

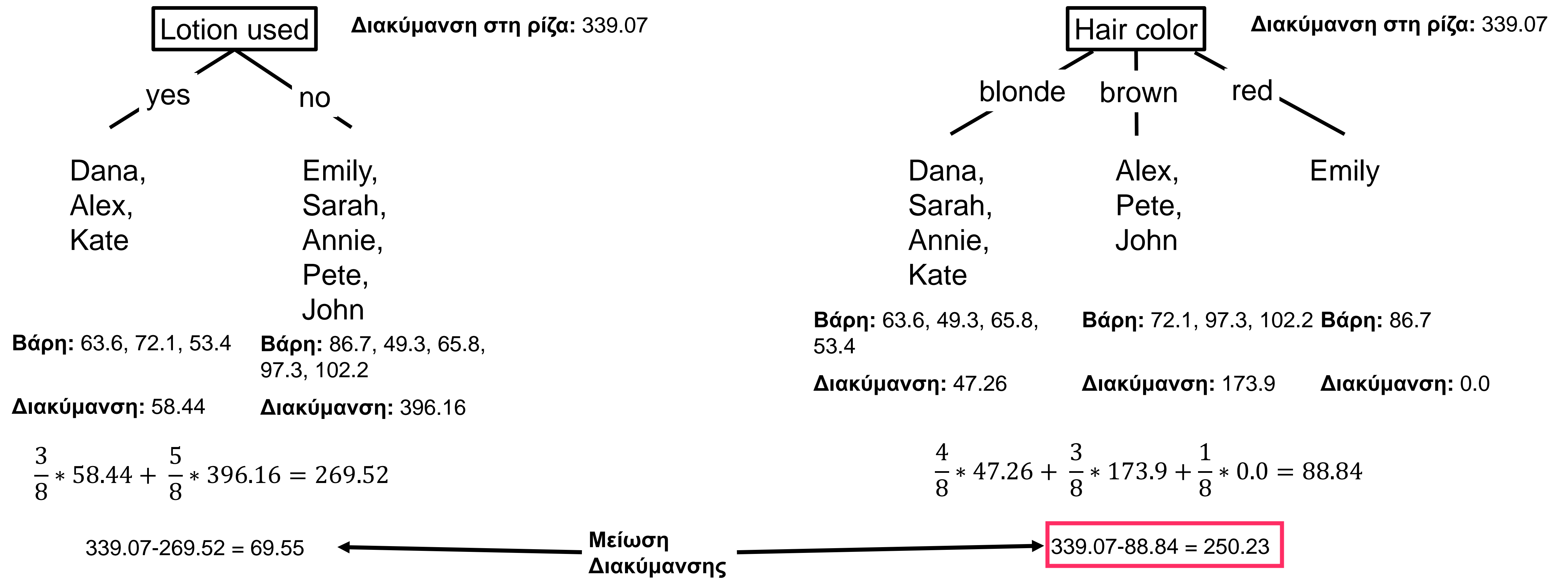
$y$

$x_3$





# Δέντρα παλινδρόμησης: επιλέγοντας μια διαίρεση





**MAI4CAREU**

Μεταπτυχιακά προγράμματα στην  
Τεχνητή Πληροφορία 4 Σταδιοδρομίες  
στην Ευρώπη



# Ensemble Learning





## Ensemble Learning

«Δύο (ή περισσότερα) κεφάλια είναι καλύτερα από ένα»

«Η σοφία του πλήθους»

- Η κρίση του κάθε ατόμου έχει θόρυβο
- Η συλλογική γνώμη (μέσω του μέσου όρου των απαντήσεων) είναι συχνά ανώτερη από εκείνη οποιουδήποτε ατόμου, καθώς ακυρώνει τις επιπτώσεις αυτού του θορύβου.

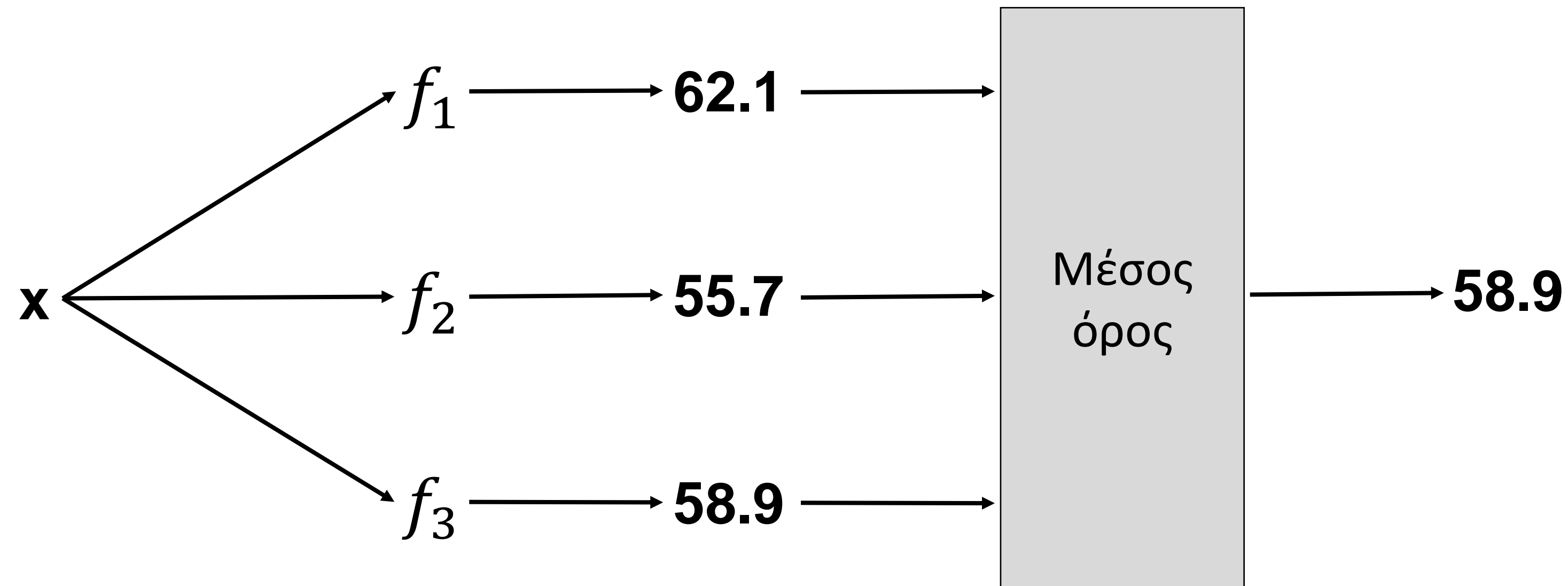
### Μέθοδοι συνόλων:

- Εκμάθηση πολλαπλών (συνήθως αδύναμων) μοντέλων αντί για ένα
- κάθε μοντέλο θα παράγει διαφορετικά σφάλματα σε διαφορετικές περιπτώσεις
- η διαφορετικότητα συμβάλλει στην προώθηση της γενίκευσης και στην οικοδόμηση ενός ισχυρού μοντέλου
- χρειάζεστε έναν τρόπο για να **συνδυάσετε/αθροίσετε** την παραγωγή από διαφορετικά μοντέλα:
  - για παλινδρόμηση: πάρτε τον μέσο όρο
  - για την ταξινόμηση: ψηφίστε την πλειοψηφία





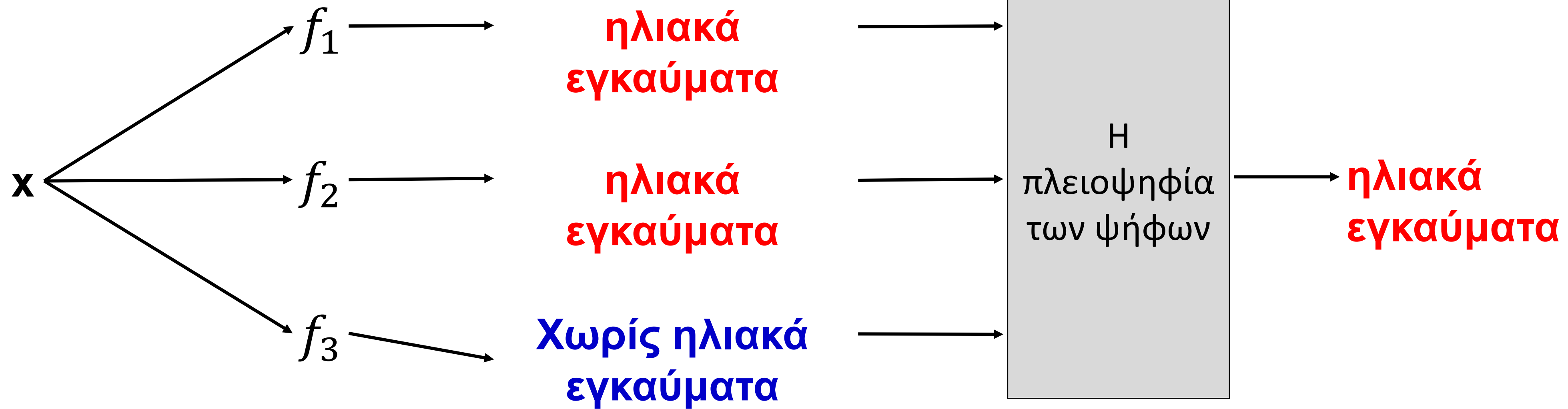
# Παράδειγμα παλινδρόμησης







## Παράδειγμα ταξινόμησης





# Αλγόριθμοι Ensemble Learning

- Τα σύνολα τείνουν να παράγουν καλύτερα αποτελέσματα όταν υπάρχει σημαντική ποικιλομορφία μεταξύ των μοντέλων
- Οι αλγόριθμοι **Ensemble επιδιώκουν να προωθήσουν την ποικιλομορφία** μεταξύ των μοντέλων που συνδυάζουν

## Κατηγορίες αλγορίθμων:

- Παράλληλες μέθοδοι (π.χ. Bagging)
- Διαδοχικές μέθοδοι (π.χ., Boosting)
- Stacking

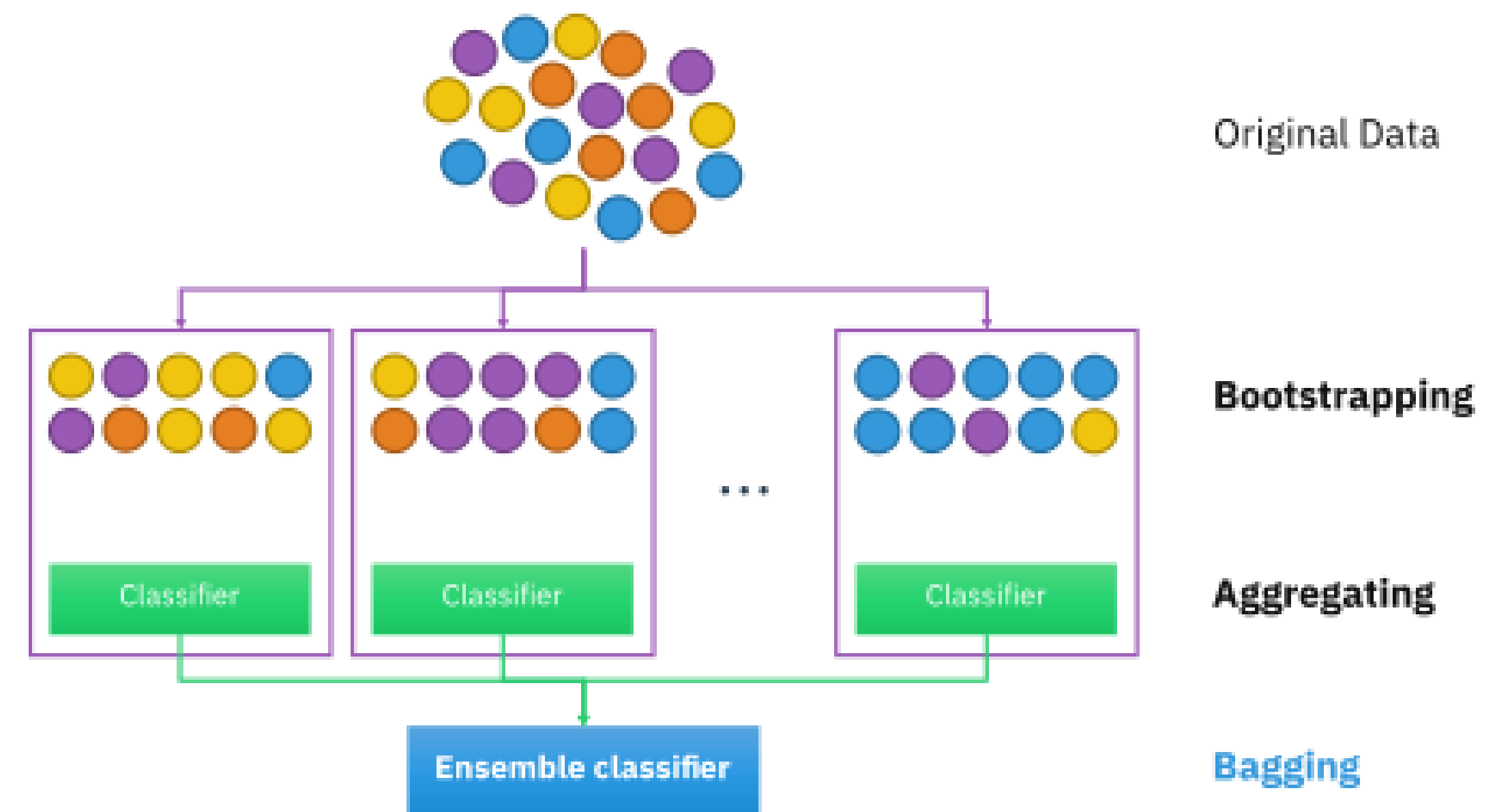
Η εκμάθηση συνόλων μπορεί να χρησιμοποιηθεί με οποιοδήποτε μοντέλο





# Bagging (Bootstrap aggregating)

- Επιδιώκει την ποικιλομορφία των μελών του συνόλου, **μεταβάλλοντας τα δεδομένα της εκπαίδευσής τους**
- **Bootstrapping:** Δεδομένου ότι ένα σύνολο εκμάθησης  $D$  από  $m$  περιπτώσεις δημιουργούν  $B$  νέα σύνολα εκμάθησης  $D_i$  καθένα μεγέθους  $m'$  με **δειγματοληψία από το  $D$  ομοιόμορφα με αντικατάσταση**
  - Αυτό εξασφαλίζει ότι τα δείγματα είναι **ανεξάρτητα**
- **Συνάθροιση:** Τοποθετήστε τα μοντέλα  $B$  στα αντίστοιχα σύνολα εκμάθησης  $D_i$  και συνδυάζουν τα αποτελέσματα με μέσο όρο ή ψηφοφορία



[ΠΗΓΗ](#)







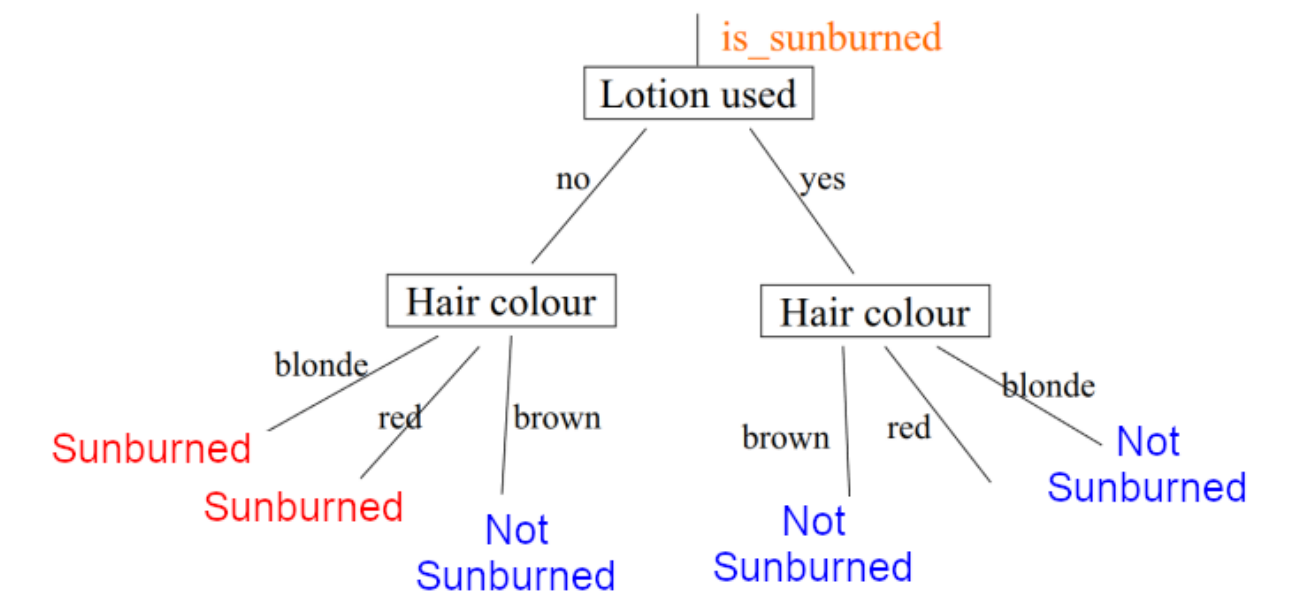
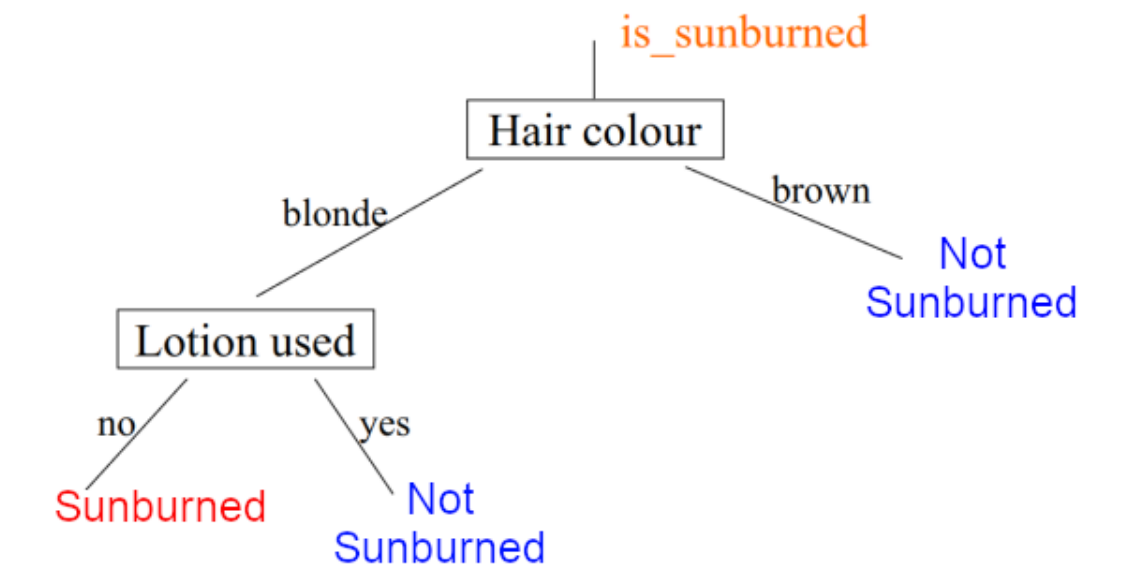
## Bagged decision tree

### Διαδικασία

- Σύνολο εκπαίδευσης  $D$  μεγέθους  $m$
- Για  $b = 1$  έως  $B$ 
  - Δημιουργία νέου συνόλου κατάρτισης  $D_b$  μεγέθους  $m$  χρησιμοποιώντας δειγματοληψία με αντικατάσταση
  - Εκπαιδεύστε ένα δέντρο αποφάσεων στο  $D_b$
- Ομαδοποίηση των αποτελεσμάτων πάνω από τα δέντρα  $B$

Η ρύθμιση του  $B$  σε μεγάλο αριθμό (π.χ., 100) έχει ως αποτέλεσμα καλύτερη απόδοση, ωστόσο, μια πολύ μεγάλη τιμή (π.χ., 1000) επιβραδύνει τον υπολογισμό χωρίς να προσφέρει μεγάλη βελτίωση.

Όνομα	Μαλλιά	Ύψος	Βάρος	Λοσιόν	Αποτέλεσμα
Η Άννι	Ξανθιά	Κοντός	Μέσος όρος	Όχι	Ηλιακά εγκαύματα
Dana	Ξανθιά	Ψηλός	Μέσος όρος	Ναι	Κανένα
Ο Ιωάννης	Καφέ	Μέσος όρος	Βαρύς	Όχι	Κανένα
Dana	Ξανθιά	Ψηλός	Μέσος όρος	Ναι	Κανένα
Η Σάρα	Ξανθιά	Μέσος όρος	Φως	Όχι	Ηλιακά εγκαύματα
Ο Πιτ	Καφέ	Ψηλός	Βαρύς	Όχι	Κανένα
Όνομα	Μαλλιά	Ύψος	Βάρος	Λοσιόν	Αποτέλεσμα
Ο Πιτ	Καφέ	Ψηλός	Βαρύς	Όχι	Κανένα
Η Κέιτ	Ξανθιά	Κοντός	Φως	Ναι	Κανένα
Η Κέιτ	Ξανθιά	Κοντός	Φως	Ναι	Κανένα
Η Σάρα	Ξανθιά	Μέσος όρος	Φως	Όχι	Ηλιακά εγκαύματα
Η Άννι	Ξανθιά	Κοντός	Μέσος όρος	Όχι	Ηλιακά εγκαύματα
Η Έμιλι	Κόκκινο	Μέσος όρος	Βαρύς	Όχι	Ηλιακά εγκαύματα
Η Έμιλι	Κόκκινο	Μέσος όρος	Βαρύς	Όχι	Ηλιακά εγκαύματα
Ο Αλέξης	Καφέ	Κοντός	Μέσος όρος	Ναι	Κανένα
Η Έμιλι	Κόκκινο	Μέσος όρος	Βαρύς	Όχι	Ηλιακά εγκαύματα





# Τυχαία δάση

## Επιπρόσθετο βήμα:

Τυχαιοποιήστε την επιλογή του χαρακτηριστικού

Σε κάθε κόμβο, κατά την επιλογή ενός χαρακτηριστικού για διαχωρισμό, εάν υπάρχουν  $n$  διαθέσιμα χαρακτηριστικά, επιλέξτε ένα τυχαίο υποσύνολο  $k < n$  χαρακτηριστικών και επιτρέψτε στον αλγόριθμο να επιλέξει μόνο από αυτό το υποσύνολο

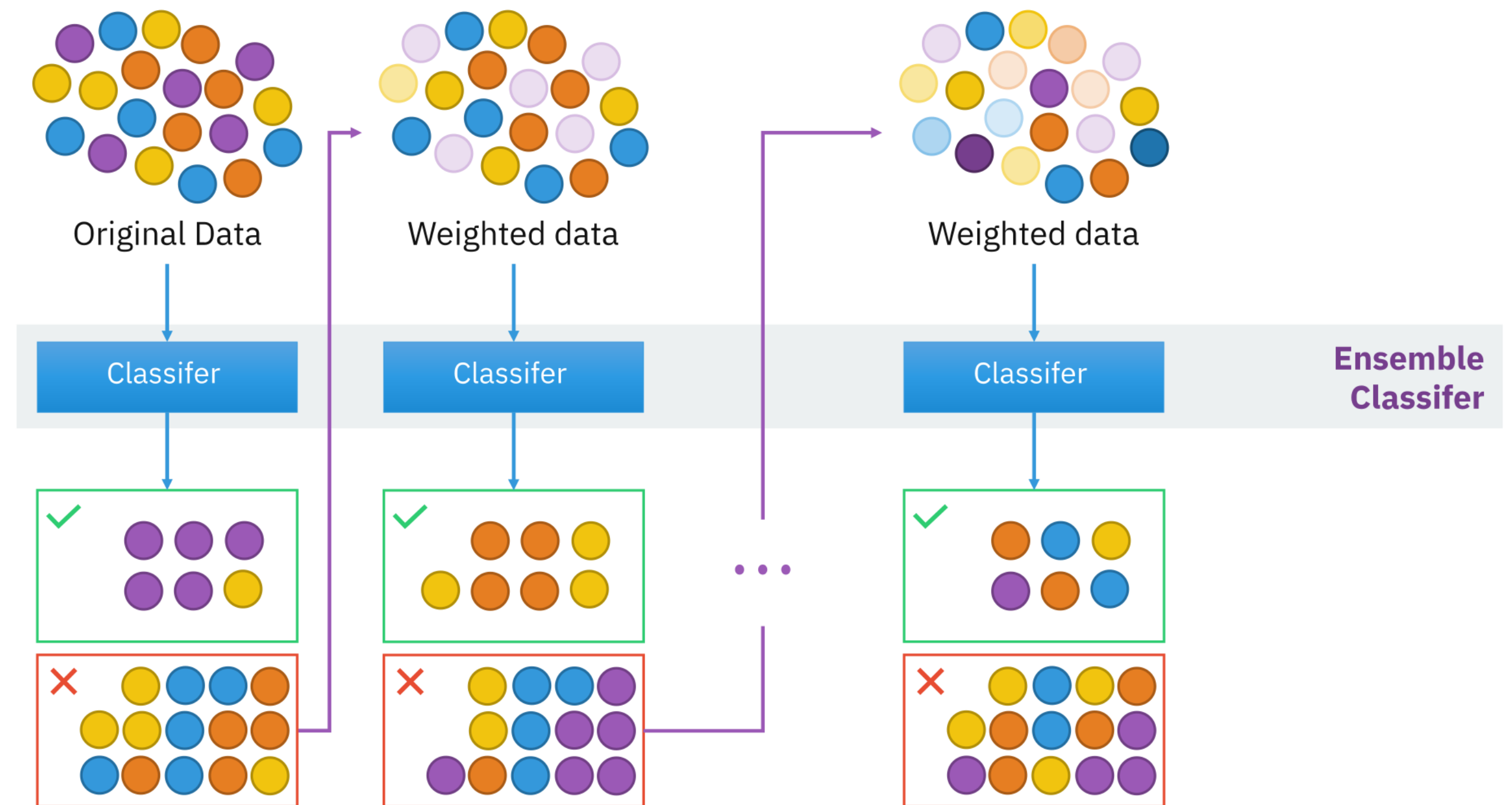
Τυπικά:  $k = \sqrt{n}$





## Boosting

- **Δημιουργήστε σταδιακά το σύνολο** όπου κάθε νέο μοντέλο επικεντρώνεται σε προηγούμενα λανθασμένα παραδείγματα
- Αρχικά, όλες οι περιπτώσεις εκμάθησης (D1) δίνεται το ίδιο βάρος
- Τα δεδομένα εκμάθησης (D1) δίνονται σε ένα βασικό μοντέλο (L1)
- Οι λανθασμένες περιπτώσεις κατά L1 έχουν πλέον **μεγαλύτερο βάρος** από τις σωστά ταξινομημένες περιπτώσεις
- Τα ενισχυμένα δεδομένα (D2) δίνονται σε ένα δεύτερο μοντέλο (L2) και ούτω καθεξής
- Τα αποτελέσματα συγκεντρώνονται από όλα τα μοντέλα με τη χρήση **σταθμισμένου μέσου όρου**



[ΠΗΓΗ](#)







## XGBoost (eXtreme Gradient Boosting)

- Βιβλιοθήκη λογισμικού ανοιχτού κώδικα των ενισχυμένων δέντρων
- Αποτελεσματική, επεκτάσιμη και φορητή εφαρμογή
- Ενσωματωμένος στην τακτοποίηση για την αποφυγή υπερπροσαρμογής
- Τρέχει σε ένα μόνο μηχάνημα ή σε κατανεμημένα περιβάλλοντα
- Πολύ επιτυχημένος και ιδιαίτερα ανταγωνιστικός σε διαγωνισμούς μηχανικής μάθησης (π.χ., Kaggle)





# Παραδείγματα XGBoost

## Ταξινόμηση

```
from xgboost import XGBClassifier  
model = XGBClassifier()  
model.fit(X_train, y_train)  
y_pred = model.predict(X_test)
```

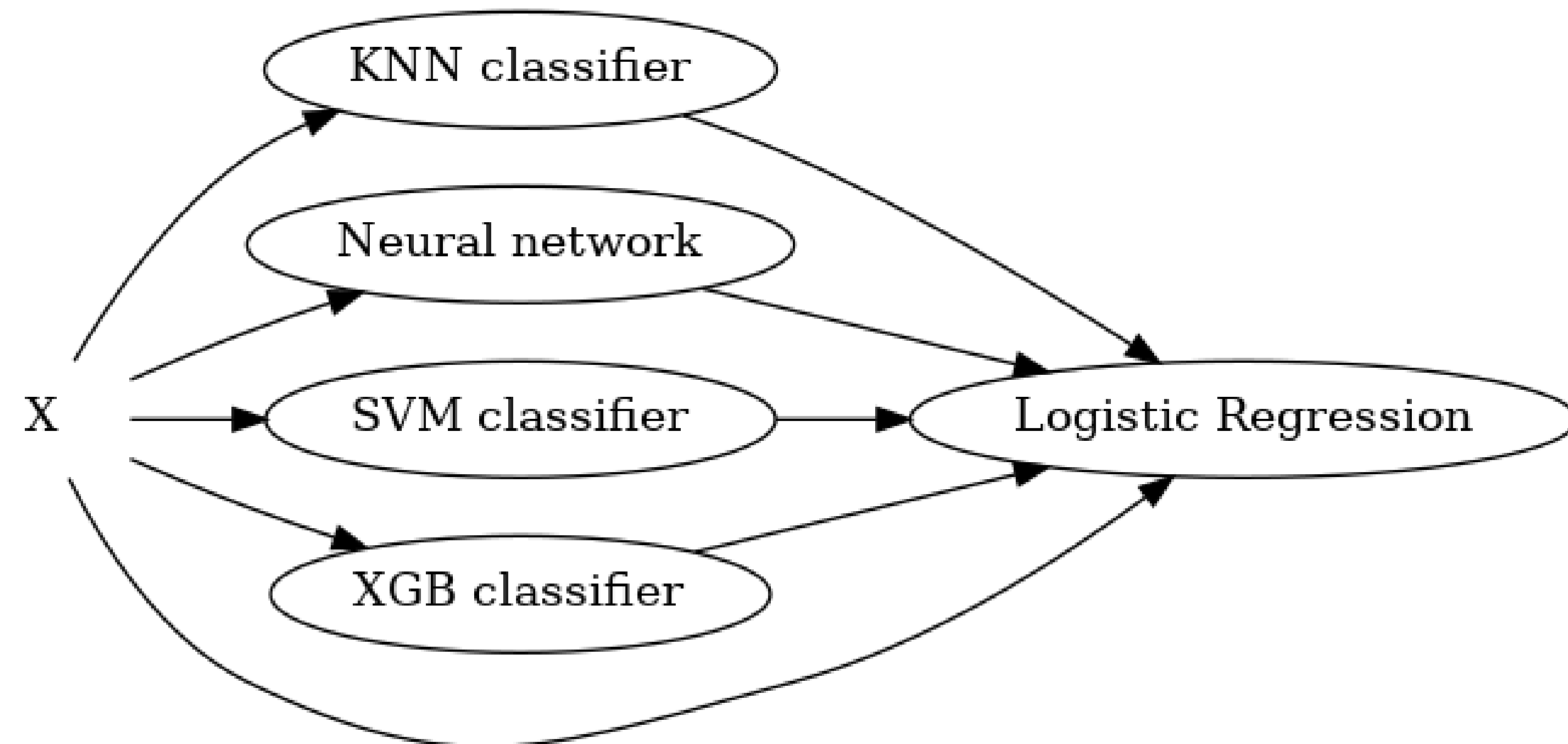
## Οπισθοδρόμηση

```
from xgboost import XGBRegressor  
model = XGBRegressor()  
model.fit(X_train, y_train)  
y_pred = model.predict(X_test)
```





## Stacking



[ΠΗΓΗ](#)

- Εκπαίδευση πολλαπλών μοντέλων στα δεδομένα της εκπαίδευσης
- Οι προβλέψεις των μοντέλων γίνονται τα δεδομένα εκμάθησης ενός 2<sup>ου</sup> επιπέδου (συνήθως logistic regression) που μαθαίνει πώς να τα συνδυάσει





## Κουίζ

Ε: Ποια είναι η κύρια διαφορά μεταξύ του bagging και του boosting;

Α: Το bagging εκπαιδεύει τα μοντέλα παράλληλα, ενώ το boosting τους εκπαιδεύει διαδοχικά

Ε: Πού πηγαίνει ένας μηχανικός μηχανικής εκμάθησης για κάμπινγκ;

Α: Σε ένα τυχαίο δάσος







## Επόμενη Διάλεξη

- Μέθοδοι με βάση τον πυρήνα



**MAI4CAREU**

Μεταπτυχιακά προγράμματα στην  
Τεχνητή Πληροφορία 4 Σταδιοδρομίες  
στην Ευρώπη



# Σας ευχαριστούμε

