ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

# Τεχνητή Νοημοσύνη

*8ο φροντιστήριο (2024-25)*

Αν Α και Β είναι δύο ενδεχόμενα ενός πειράματος και P(B) > 0, τότε ο λόγος

$$\frac{\mathrm{P}(A \cap B)}{P(B)}$$

Λέγεται **δεσμευμένη πιθανότητα** του Α με δεδομένο το Β και συμβολίζεται με $P(A|B)$. Δηλαδή:

$$P(A|B) = \frac{\mathrm{P}(A \cap B)}{P(B)}$$

Δύο ενδεχόμενα $A$ και $B$ με $P(A) > 0$ και $P(B) > 0$ λέγονται **ανεξάρτητα**, αν και μόνον αν $P(A|B) = P(A)$ και $P(B|A) = P(B)$.

Δύο ενδεχόμενα $A$ και $B$ λέγονται **ανεξάρτητα**, αν

$$P(A \cap B) = P(A) \cdot P(B)$$

**Θεώρημα Bayes**

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

**Ανεξαρτησία και δεσμευμένη πιθανότητα**

$$P(A_1 \mid A_2 \cap B) = P(A_1 \mid B)$$

Ισοδύναμα

$$P(A_1 \cap A_2|B) = P(A_1|B) \cdot P(A_2|B).$$

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

| $X_1$ | $X_2$ | $X_3$ | C |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

**Αφελείς ταξινομητές Bays (Naïve Bayes)**
**Θεώρημα Bayes**

$$\boldsymbol{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

όπου Α, Β δύο ενδεχόμενα.
Ως Α θα έχουμε μια τιμή της μεταβλητής $C$, δηλαδή μια συγκεκριμένη κατάταξη c.
Ως Β θα έχουμε ένα μόνο αντικείμενο προς κατάταξη $\vec{X} =< X_1, X_2, \dots, X_m >$.
Οπότε, για ένα συγκεκριμένο αντικείμενο προς κατάταξη, στόχος είναι να συγκρίνουμε τις πιθανότητες $P(C = c \mid X)$ για όλες τις τιμές της μεταβλητής $c$. Το αντικείμενο $\vec{X}$ θα καταταχθεί στην κατάταξη με τη μεγαλύτερη πιθανότητα.

$$\boldsymbol{P\left(C = c \mid \vec{X}\right) = \frac{P\left(\vec{X} \mid C = c\right)P(C = c)}{P\left(\vec{X}\right)}}$$

Η πιθανότητα $P\left(\vec{X}\right)$ δεν χρειάζεται να υπολογιστεί γιατί απλά θα επιλέξουμε τον μεγαλύτερο αριθμητή.
Η πιθανότητα $P(C = c)$ υπολογίζεται από τα παραδείγματα
**Παραδοχή των αφελών ταξινομητών Bayes:** Οι τιμές των $x_i$ είναι ανεξάρτητες δεδομένης της τιμής της C.
Επομένως η πιθανότητα $P\left(\vec{X} \mid C = c\right)$ υπολογίζεται ως εξής:

$$\boldsymbol{P\left(\vec{X} \mid C = c\right) = P\left(\vec{X}_1 = x_1 \cap \dots \cap \vec{X}_m = x_m | C = c\right) = \prod_{i=1}^{m} P(X_i = x_i \mid C = c)}$$

Ανεξαρτησία και δεσμευμένη πιθανότητα
$P(A_1 \mid A_2 \cap B) = P(A_1 \mid B)$
Ισοδύναμα
$P(A_1 \cap A_2|B) = P(A_1|B) \cdot P(A_2|B).$

**Αφελείς ταξινομητές Bays (Naïve Bayes)**

Σε περίπτωση που για κάποιο $i$ είναι $\boldsymbol{P(X_i = x_i \mid C = c) = 0}$ μηδενίζεται το γινόμενο

$$\prod_{i=1}^{m} \boldsymbol{P(X_i = x_i \mid C = c)}$$

Ένας τρόπος εξομάλυνσης: **εκτιμήτρια Laplace.**
Θεωρούμε κατά την εκτίμηση της ότι υπάρχουν δύο ακόμη ψευτο-μηνύματα εκπαίδευσης κατηγορίας c:
ένα που περιέχει τη λέξη $X_i$
ένα που δεν την περιέχει.
Επομένως +1 στον αριθμητή της εκτίμησης, +2 στον παρονομαστή.
Γενικότερα, κατά την εκτίμηση τυχαίας μεταβλητής $\boldsymbol{X_i}$ με k δυνατές τιμές, +1 στον αριθμητή και +k στον παρονομαστή.

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

$$P(X_1 = 0|C = 1) = \frac{P(X_1 = 0 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

| $X_1$ | $X_2$ | $X_3$ | C |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| 0 | | | 1 |
| 1 | | | 1 |

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

$$P(X_1 = 0|C = 1) = \frac{P(X_1 = 0 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

$$P(X_2 = 1|C = 1) = \frac{P(X_2 = 1 \cap C = 1)}{P(C = 1)} = \frac{1+1}{2+2} = \mathbf{\frac{2}{4}}$$

| $X_1$ | $X_2$ | $X_3$ | C |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| | 0 | | 1 |
| | 1 | | 1 |

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

$$P(X_1 = 0|C = 1) = \frac{P(X_1 = 0 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

$$P(X_2 = 1|C = 1) = \frac{P(X_2 = 1 \cap C = 1)}{P(C = 1)} = \frac{1+1}{2+2} = \mathbf{\frac{2}{4}}$$

$$P(X_3 = 0|C = 1) = \frac{P(X_3 = 1 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

| $X_1$ | $X_2$ | $X_3$ | C |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| | | 0 | 1 |
| | | 1 | 1 |

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

| $X_1$ | $X_2$ | $X_3$ | C |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

$$P(X_1 = 0|C = 1) = \frac{P(X_1 = 0 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

$$P(X_2 = 1|C = 1) = \frac{P(X_2 = 1 \cap C = 1)}{P(C = 1)} = \frac{1+1}{2+2} = \mathbf{\frac{2}{4}}$$

$$P(X_3 = 0|C = 1) = \frac{P(X_3 = 1 \cap C = 1)}{P(C = 1)} = \frac{0+1}{2+2} = \mathbf{\frac{1}{4}}$$

Επιπλέον $P(C = 1) = \frac{1}{2}$. Άρα

$$P\left(C = 1 \mid \vec{X} = <0,1,0>\right) = \frac{1}{P\left(\vec{X} = <0,1,0>\right)} P(C = 1) \prod_{i=1}^{m} P(X_i = x_i \mid C = 1) =$$

$$= \frac{1}{P\left(\vec{X} = <0,1,0>\right)} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{2}{4} \cdot \frac{1}{4} = \frac{1}{P\left(\vec{X} = <0,1,0>\right)} \cdot \mathbf{\frac{1}{4^3}}$$

**Άσκηση 16.1.**

Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $< X_1, X_2, X_3 > = < 0,1,0 >$. Σε ποια από τις δύο κατηγορίες (C = 0 ή C = 1) θα το κατατάξει ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) που έχει στη διάθεσή του τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμήτρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Όλες οι ιδιότητες $X_i$ είναι δυαδικές (έχουν ως πεδίο τιμών το {0, 1}).

*Όμοια βρίσκουμε ότι*

$$P\left( C = 0 \mid \vec{X} = <0,1,0> \right) = \frac{1}{P\left(\vec{X} = <0,1,0>\right)} P(C = 0) \prod_{i=1}^{m} P( X_i = x_i \mid C = 0 ) =$$

$$= \frac{1}{P\left(\vec{X} = <0,1,0>\right)} \cdot \frac{1}{2} \cdot \frac{2}{4} \cdot \frac{2}{4} \cdot \frac{2}{4} = \frac{1}{P\left(\vec{X} = <0,1,0>\right)} \cdot \frac{\mathbf{1}}{\mathbf{4^2}}$$

Επομένως θα το κατατάξει στην C = 0.

**Άσκηση 16.3.**

Χρησιμοποιούμε μια παραλλαγή του αφελούς ταξινομητή Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα), με δύο κατηγορίες (C = 0 και C = 1) και m δυαδικές ιδιότητες $X_1, \dots, X_m$, η οποία κατατάσσει στη C = 1 αν:

$$P(C = 1) \cdot \prod_{i=1}^{m} P(X_i = x_i | C = 1) \geq K$$

όπου K μια σταθερά, ενώ διαφορετικά κατατάσσει στη C = 0. Αποδείξτε ότι ο ταξινομητής αυτός είναι ένας γραμμικός διαχωριστής. Δείξτε αναλυτικά τους υπολογισμούς σας.

Υπόδειξη: Αν παραστήσουμε με $t_i$ το ενδεχόμενο που παριστάνεται με $X_i = 1$ (π.χ. την εμφάνιση μιας συγκεκριμένης λέξης), τότε:

$$P(X_i = x_i | C = 1) = P(t_i | C = 1)^{x_i} \cdot [1 - P(t_i | C = 1)]^{1 - x_i}$$

Υπενθυμίζεται, επίσης, ότι $\log(a \cdot b) = \log a + \log b$ και $\log a^b = b \cdot \log a$.

$$P(C = 1) \cdot \prod_{i=1}^{m} P(X_i = x_i | C = 1) \geq K$$

$$P(C = 1) \cdot \prod_{i=1}^{m} (P(t_i | C = 1)^{x_i} \cdot [1 - P(t_i | C = 1)]^{1 - x_i}) \geq K$$

$$\log\left( P(C = 1) \cdot \prod_{i=1}^{m} (P(t_i | C = 1)^{x_i} \cdot [1 - P(t_i | C = 1)]^{1 - x_i}) \right) \geq \log K$$

$\log(a \cdot b) = \log a + \log b$

**Άσκηση 16.3.**

$$\log P(C=1) + \log \prod_{i=1}^{m} (P(t_i|C=1)^{x_i} \cdot [1 - P(t_i|C=1)]^{1-x_i}) \geq \log K$$

$\log(a \cdot b) = \log a + \log b$

$$\log P(C=1) + \sum_{i=1}^{m} \log(P(t_i|C=1)^{x_i} \cdot [1 - P(t_i|C=1)]^{1-x_i}) \geq \log K$$

$$\log P(C=1) + \sum_{i=1}^{m} \log P(t_i|C=1)^{x_i} + \sum_{i=1}^{m} \log[1 - P(t_i|C=1)]^{1-x_i} \geq \log K$$

$\log a^b = b \cdot \log a$

$$\log P(C=1) + \sum_{i=1}^{m} \mathrm{x_i} \log P(t_i|C=1) + (1-x_i) \sum_{i=1}^{m} \log[1 - P(t_i|C=1)] \geq \log K$$

$$\log P(C=1) + \sum_{i=1}^{m} \log[1 - P(t_i|C=1)] - \log K + \left( \sum_{i=1}^{m} \log P(t_i|C=1) - \sum_{i=1}^{m} \log[1 - P(t_i|C=1)] \right) x_i \geq 0$$

$$w_0 + \sum_{i=1}^{m} w_i x_i \geq 0$$

Επομένως πρόκειται για γραμμικό ταξινομητή.

**Άσκηση 16.4.**

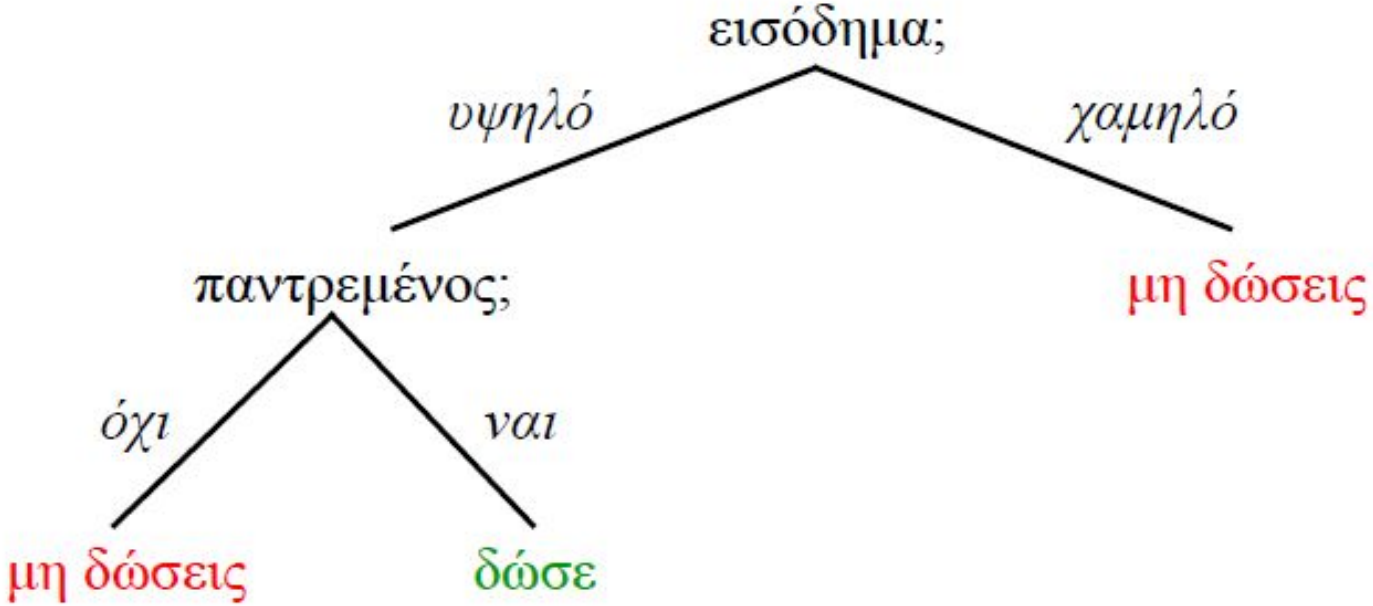
A) Βάσει των παραδειγμάτων εκπαίδευσης του πίνακα, πόση είναι η εντροπία H(C) της κατηγορίας C και γιατί;

Απάντηση:

P(C=μαύρο) = P(C=άσπρο) $= \frac{1}{2}$ . Έχουμε δύο ισοπίθανα ενδεχόμενα C = μαύρο και C = άσπρο και άρα μέγιστη εντροπία (αβεβαιότητα), που για δύο ενδεχόμενα είναι ίση με 1. Το ίδιο συμπέρασμα προκύπτει από τον ορισμό της εντροπίας, με αριθμητικούς υπολογισμούς.

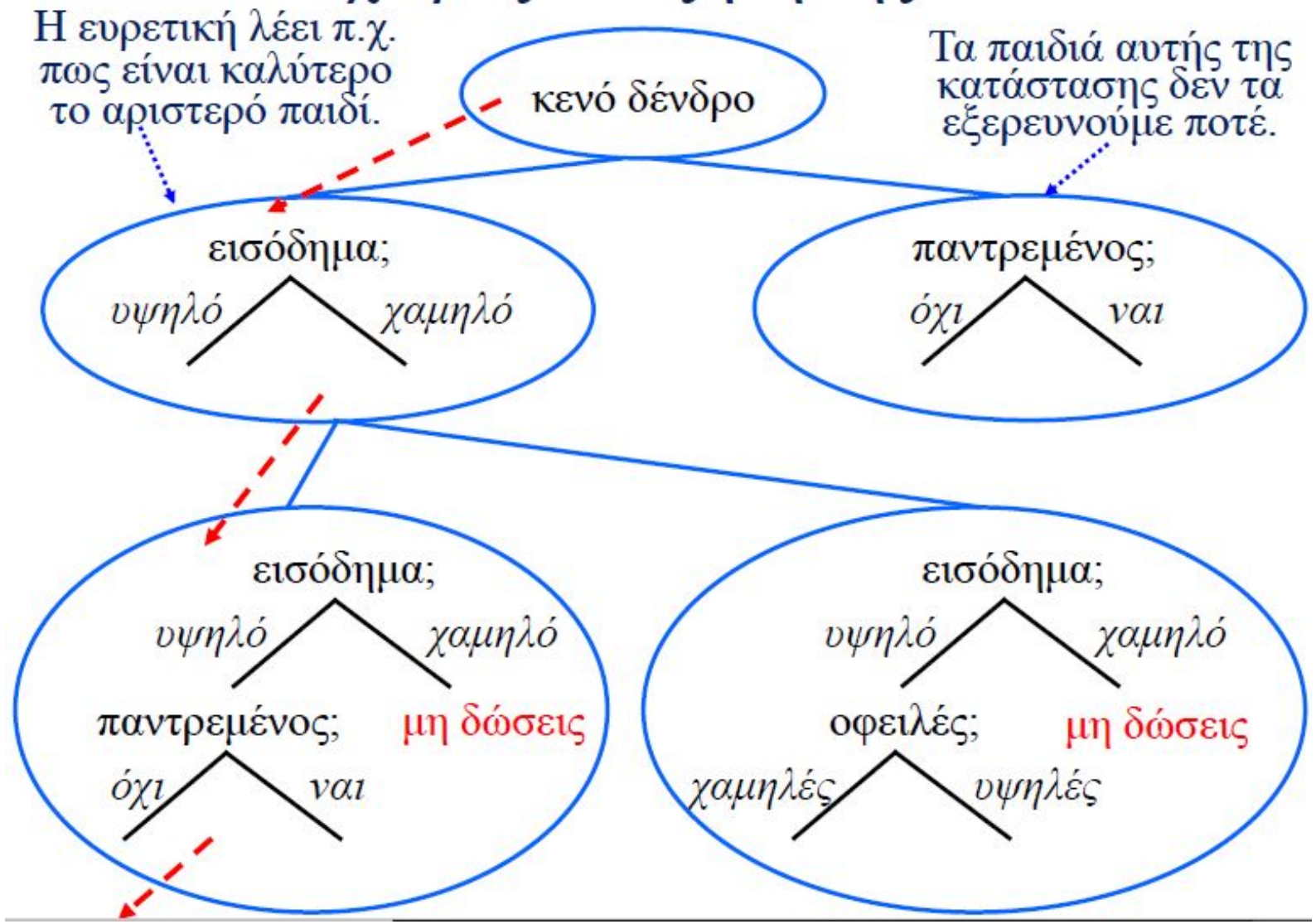
| ID | X | Y | Z | C |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | Μαύρο |
| 2 | 1 | 0 | 1 | Μαύρο |
| 3 | 0 | 0 | 1 | Μαύρο |
| 4 | 1 | 1 | 0 | Μαύρο |
| 5 | 0 | 1 | 1 | Άσπρο |
| 6 | 1 | 0 | 0 | Άσπρο |
| 7 | 0 | 0 | 0 | Άσπρο |
| 8 | 1 | 0 | 0 | Άσπρο |

# Μάθηση δέντρων απόφασης

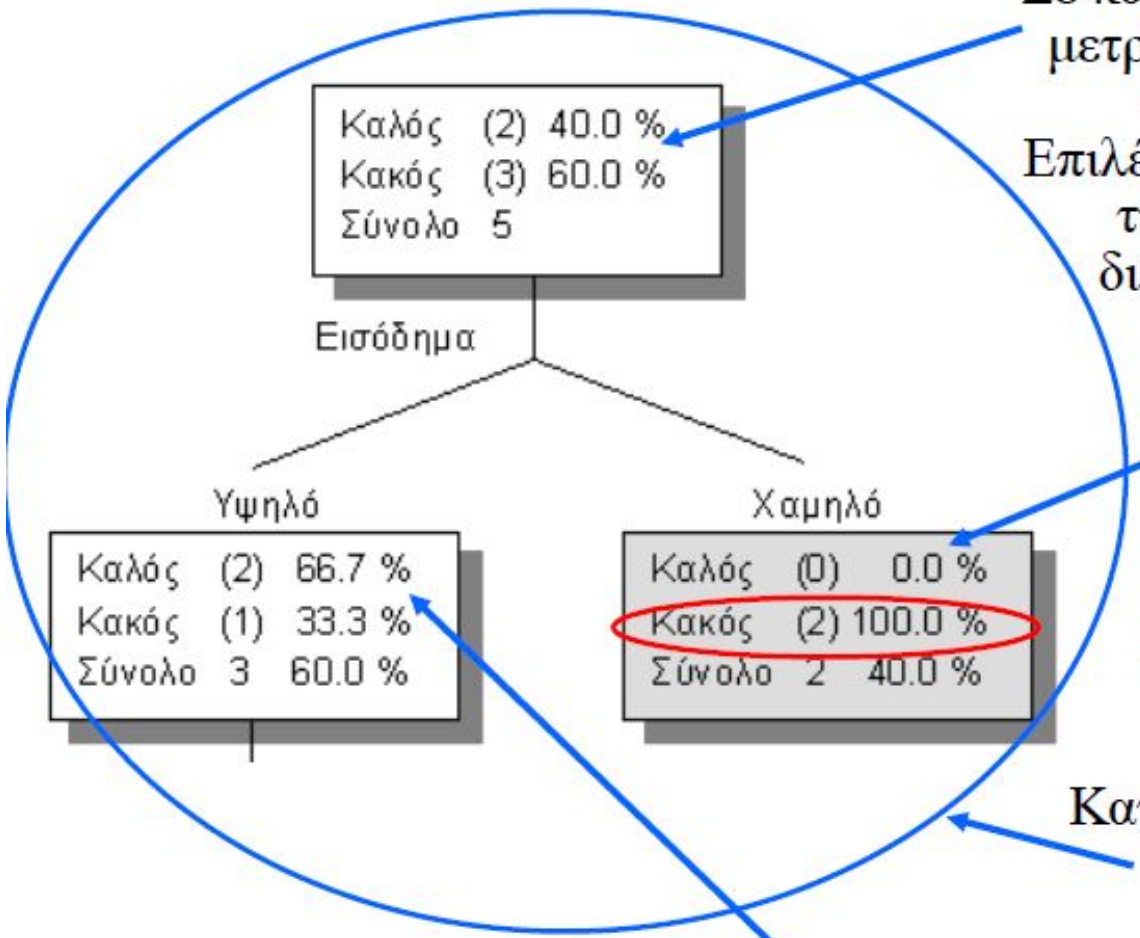



- Ο **αλγόριθμος ID3** κατασκευάζει δέντρα αυτής της μορφής από τα παραδείγματα εκπαίδευσης.
- Σε κάθε **εσωτερικό κόμβο** ελέγχουμε την **τιμή μιας ιδιότητας**.
- Τα **φύλλα** αντιστοιχούν σε **αποφάσεις**.

# Ο υποχώρος αναζήτησης του ID3



# Καταστάσεις: περισσότερες λεπτομέρειες



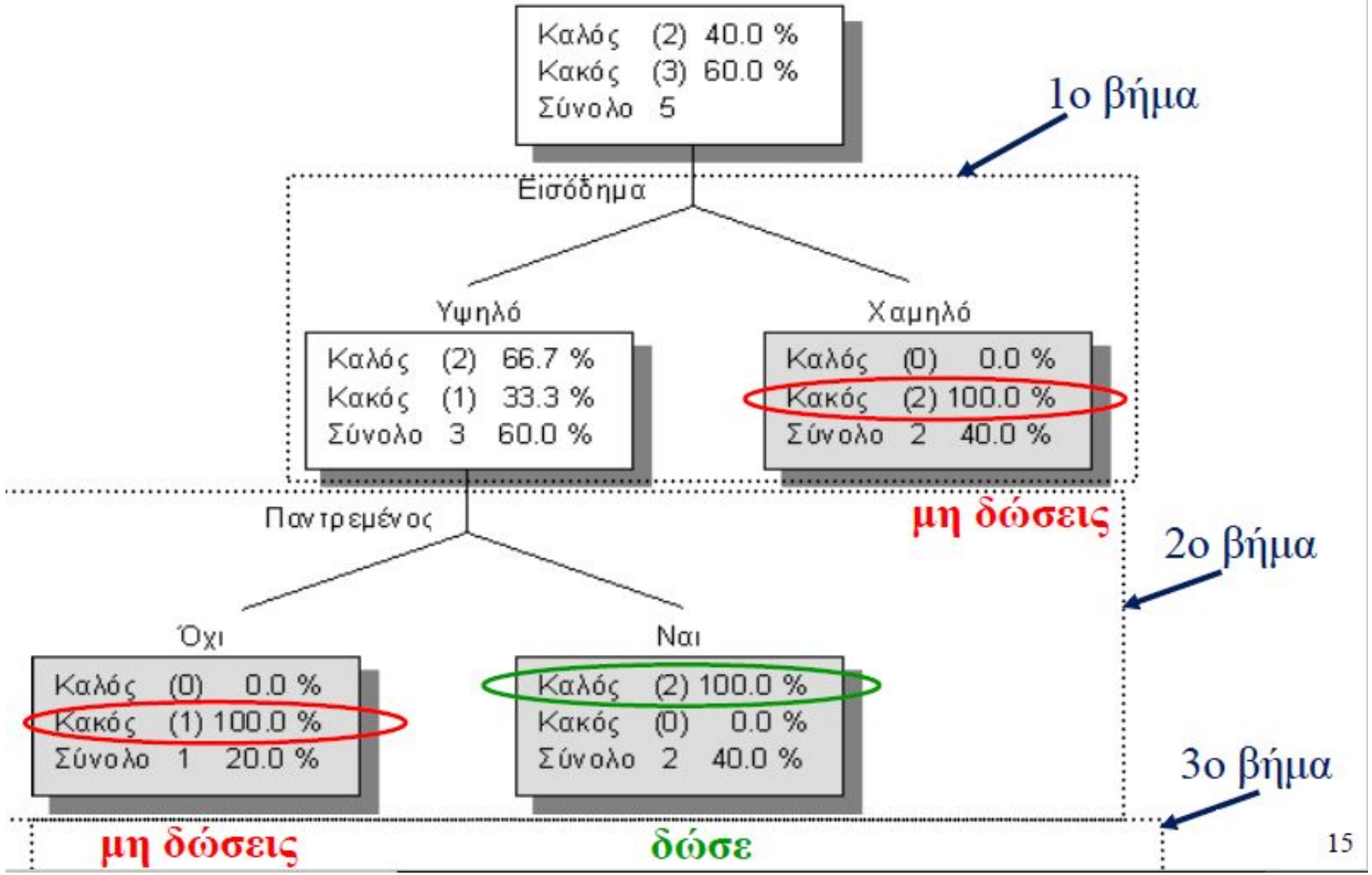

Σε κάθε κόμβο του **δένδρου** μετράμε τα παραδείγματα κάθε κατηγορίας. Επιλέγουμε μέσω ευρετικής την ιδιότητα που τα διαχωρίζει καλύτερα.

Όλα τα παραδείγματα εκπαίδευσης που έχουν μείνει ανήκουν σε **μία κατηγορία**. Απόφαση: **μη δώσεις**.

Κατάσταση του χώρου αναζήτησης.

Τα παραδείγματα εκπαίδευσης **δεν** ανήκουν όλα σε μία κατηγορία. **Επέκτεινε** το δέντρο κάτω από αυτόν τον κόμβο.

# ID3: επέκταση δένδρου απόφασης



# ID3: Αναρρίχηση λόφου

- **Σε κάθε κατάσταση** (ημιτελές δέντρο) του χώρου αναζήτησης, επιλέγει να **επεκτείνει** κάθε κόμβο του δέντρου απόφασης (που πρέπει να επεκταθεί) με την **ιδιότητα** που **εκτιμά** πως είναι η **χρησιμότερη**.
  - Χρησιμοποιεί ως ευρετική συνάρτηση αξιολόγησης των ιδιοτήτων το **κέρδος πληροφορίας** (IG).
- Προκύπτει έτσι μια **μοναδική κατάσταση-παιδί** (νέο δέντρο απόφασης) στην οποία μεταβαίνουμε.
  - Το **μέτωπο** περιέχει πάντα **μία μόνο κατάσταση**. Δεν υπάρχει δυνατότητα εξέτασης εναλλακτικών μονοπατιών.

# Αλγόριθμος ID3

**συνάρτηση** ID3(*παραδείγματα*, *ιδιότητες*, *προεπιλεγμένη*)

**είσοδοι**: *παραδείγματα*: σύνολο παραδειγμάτων εκπαίδευσης

*ιδιότητες*: σύνολο διαθέσιμων ιδιοτήτων

*προεπιλεγμένη*: προεπιλεγμένη κατηγορία

**αν** *παραδείγματα* = {} **τότε επίστρεψε** *προεπιλεγμένη κατηγορία*

**διαφορετικά αν** όλα τα *παραδείγματα* ανήκουν στην ίδια κατηγορία **τότε επίστρεψε** αυτή την κατηγορία

**διαφορετικά αν** *ιδιότητες* = {} **τότε επίστρεψε** την κατηγορία που είναι συχνότερη στα *παραδείγματα*

**διαφορετικά** ...

Επιλέγουμε την ιδιότητα που παρέχει το μεγαλύτερο κέρδος πληροφορίας.

**διαφορετικά**

*καλύτερη* ← επιλογή-ιδιότητας(*ιδιότητες*, *παραδείγματα*)

*δέντρο* ← νέο δέντρο που στη ρίζα του ελέγχει την *καλύτερη*

έστω *m* η συχνότερη κατηγορία μεταξύ των *παραδειγμάτων*

**για κάθε** δυνατή τιμή $v_i$ της *καλύτερης*

$παραδείγματα_i \leftarrow \{\pi \in παραδείγματα \mid \pi.καλύτερη = v_i\}$

*υποδέντρο* ← ID3($παραδείγματα_i$, *ιδιότητες* – *καλύτερη*, *m*)

πρόσθεσε κλαδί με ετικέτα $v_i$ στο *δέντρο* που να οδηγεί από τη ρίζα στο *υποδέντρο*

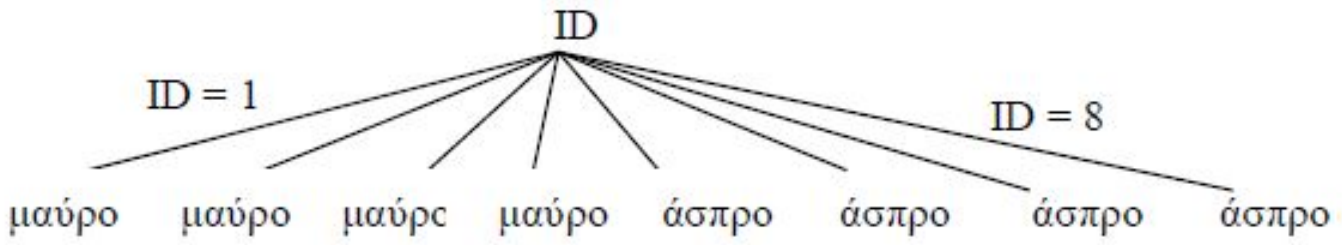
**επίστρεψε** το *δέντρο*

**Άσκηση 16.4.**

Β) Σχεδιάστε το δέντρο απόφασης που θα κατασκευάσει ο ID3 (Iterative Dichotomiser 3), αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες τις ID, X, Y, Z (το δέντρο προβλέπει την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.

| ID | X | Y | Z | C |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | Μαύρο |
| 2 | 1 | 0 | 1 | Μαύρο |
| 3 | 0 | 0 | 1 | Μαύρο |
| 4 | 1 | 1 | 0 | Μαύρο |
| 5 | 0 | 1 | 1 | Άσπρο |
| 6 | 1 | 0 | 0 | Άσπρο |
| 7 | 0 | 0 | 0 | Άσπρο |
| 8 | 1 | 0 | 0 | Άσπρο |

**Απάντηση**: Αν ξέρουμε την τιμή της ιδιότητας ID, τότε ξέρουμε με βεβαιότητα την τιμή της κατηγορίας C όλων των παραδειγμάτων. Βάσει, δηλαδή, των παραδειγμάτων του πίνακα, $H(C \mid ID) = 0$ και επομένως $IG(ID, C) = H(C) - H(C \mid ID) = 1$. Αντίθετα, καμία από τις άλλες ιδιότητες (X, Y, Z) δεν προβλέπει με απόλυτη βεβαιότητα την κατηγορία όλων των παραδειγμάτων, επομένως το κέρδος πληροφορίας που παρέχουν είναι μικρότερο από 1. Επομένως ο ID3 θα προτιμήσει να τοποθετήσει στη ρίζα του δέντρου απόφασης την ερώτηση για την ιδιότητα ID. Θα υπάρχουν 8 κλαδιά κάτω από τη ρίζα, ένα για κάθε δυνατή τιμή της ID που εμφανίζεται στα παραδείγματα εκπαίδευσης. Στο υποδέντρο κάτω από κάθε κλαδί θα καταλήξει ακριβώς ένα από τα παραδείγματα, εκείνο με την αντίστοιχη τιμή ID. Επομένως τα παραδείγματα (είναι μόνο ένα) κάθε υποδέντρου θα ανήκουν σε μία μόνο (ανά υποδέντρο) κατηγορία. Άρα ο ID3 θα σταματήσει και θα επιστρέψει το παρακάτω δέντρο, που δεν χρησιμοποιεί τις άλλες ιδιότητες.



**Άσκηση 16.4.**

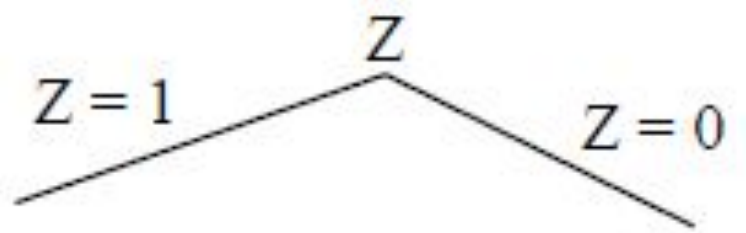
Γ) Σχεδιάστε τώρα το δέντρο απόφασης που θα κατασκευάσει ο ID3, αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες μόνο τις X, Y, Z (το δέντρο προβλέπει πάλι την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.

| ID | X | Y | Z | C |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | Μαύρο |
| 2 | 1 | 0 | 1 | Μαύρο |
| 3 | 0 | 0 | 1 | Μαύρο |
| 4 | 1 | 1 | 0 | Μαύρο |
| 5 | 0 | 1 | 1 | Άσπρο |
| 6 | 1 | 0 | 0 | Άσπρο |
| 7 | 0 | 0 | 0 | Άσπρο |
| 8 | 1 | 0 | 0 | Άσπρο |

**Απάντηση**: Τα δεδομένα εκπαίδευσης δείχνουν ότι αν μάθουμε πως Z = 1, είναι πολύ πιθανό (πιθανότητα ¾) ότι C = μαύρο· και αν μάθουμε πως Z = 0, είναι πολύ πιθανό (πιθανότητα ¾) ότι C = άσπρο. Επομένως, η γνώση της τιμής της ιδιότητας Z μειώνει την εντροπία (αβεβαιότητα για την τιμή) της C, εντροπία που αρχικά ήταν μέγιστη, δηλαδή IG(C, Z) > 0. Το ίδιο συμπέρασμα προκύπτει υπολογίζοντας αναλυτικά το IG(C, Z). Αντίθετα, τα δεδομένα εκπαίδευσης δείχνουν ότι αν μάθουμε πως Y = 1, η πιθανότητα να έχουμε C = μαύρο παραμένει ½ και ίση με την πιθανότητα να έχουμε C = άσπρο· ομοίως, αν μάθουμε ότι X = 0, η πιθανότητα να έχουμε C = μαύρο παραμένει ½ και ίση με την πιθανότητα να έχουμε C = άσπρο. Επομένως, όποια κι αν είναι η τιμή της Y, εξακολουθούμε να έχουμε δύο ισοπίθανα ενδεχόμενα C = μαύρο και C = άσπρο και, επομένως, μέγιστη εντροπία (αβεβαιότητα) H(C) = 1. Άρα η γνώση της τιμής της Y δεν μειώνει καθόλου την εντροπία (αβεβαιότητα για την τιμή της) C, δηλαδή IG(C, Y) = 0. Το ίδιο συμπέρασμα προκύπτει υπολογίζοντας αναλυτικά το IG(C, Y). Ομοίως IG(C, X) = 0. Επομένως ο ID3 θα επιλέξει να τοποθετήσει στην κορυφή του δέντρο απόφασης την ερώτηση για την ιδιότητα Z.

**Άσκηση 16.4.**

Γ) Σχεδιάστε τώρα το δέντρο απόφασης που θα κατασκευάσει ο ID3, αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες μόνο τις X, Y, Z (το δέντρο προβλέπει πάλι την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.




*Στο υποδέντρο για Z = 1, θα καταλήξουν τα παραδείγματα με ID = 1, 2, 3, 5, ενώ στο υποδέντρο για Z =* 0 τα παραδείγματα με *ID = 4, 6, 7, 8.*
*Και στα δύο υποδέντρα, αν μάθουμε κατόπιν την τιμή της ιδιότητας Y, πετυχαίνουμε πλήρη διαχωρισμό (πρόβλεψη) των κατηγοριών, ενώ αντίθετα δεν συμβαίνει το ίδιο αν μάθουμε την τιμή της ιδιότητας X.*
*Επομένως και στα δύο υποδέντρα η ιδιότητα Y παρέχει μεγαλύτερο κέρδος πληροφορίας απ΄ ό,τι η X.*

| ID | X | Y | Z | C |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | Μαύρο |
| 2 | 1 | 0 | 1 | Μαύρο |
| 3 | 0 | 0 | 1 | Μαύρο |
| 5 | 0 | 1 | 1 | Άσπρο |

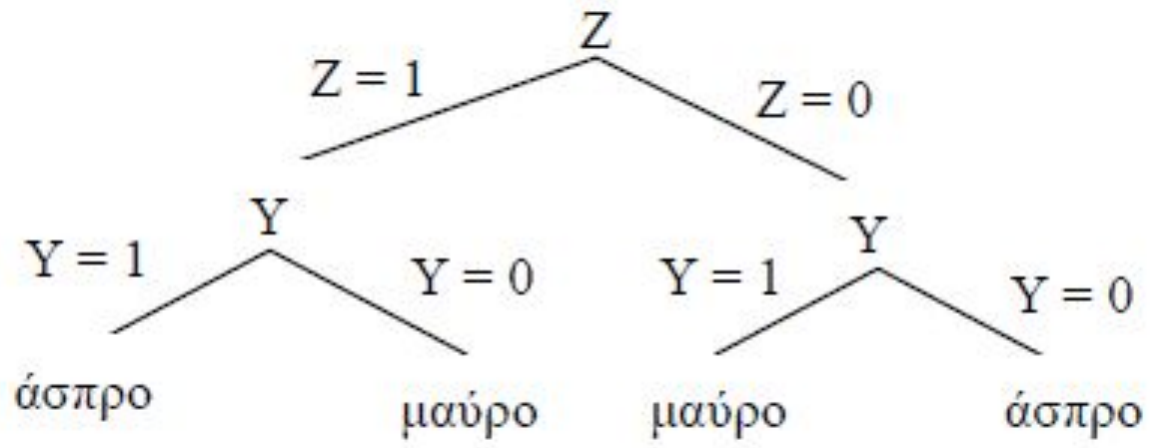
| ID | X | Y | Z | C |
|---|---|---|---|---|
| 4 | 1 | 1 | 0 | Μαύρο |
| 6 | 1 | 0 | 0 | Άσπρο |
| 7 | 0 | 0 | 0 | Άσπρο |
| 8 | 1 | 0 | 0 | Άσπρο |

**Άσκηση 16.4.**

Γ) Σχεδιάστε τώρα το δέντρο απόφασης που θα κατασκευάσει ο ID3, αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες μόνο τις X, Y, Z (το δέντρο προβλέπει πάλι την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.

| ID | X | Y | Z | C |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | Μαύρο |
| 2 | 1 | 0 | 1 | Μαύρο |
| 3 | 0 | 0 | 1 | Μαύρο |
| 5 | 0 | 1 | 1 | Άσπρο |

*Άρα ο ID3 θα προτιμήσει να προσθέσει και στα δύο υποδέντρα της ερωτήσεις για την ιδιότητα Y.*
*Σε κάθε κλαδί κάτω από τις δύο ερωτήσεις Y, καταλήγουν παραδείγματα μίας μόνο κατηγορίας.*
*Επομένως ο ID3 θα σταματήσει και θα επιστρέψει το παρακάτω δέντρο, που δεν χρησιμοποιεί την ιδιότητα X.*

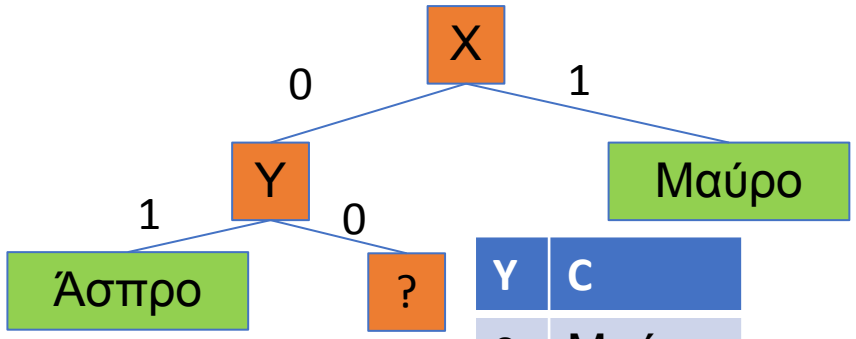



| ID | X | Y | Z | C |
|---|---|---|---|---|
| 4 | 1 | 1 | 0 | Μαύρο |
| 6 | 1 | 0 | 0 | Άσπρο |
| 7 | 0 | 0 | 0 | Άσπρο |
| 8 | 1 | 0 | 0 | Άσπρο |

**Άσκηση 16.5.**
Αν αξιολογήσουμε έναν ταξινομητή ID3 (χωρίς πριόνισμα) στο ίδιο σύνολο διανυσμάτων στο οποίο τον εκπαιδεύσαμε, αλλά στα διανύσματα εκπαίδευσης περιλαμβάνονται και ασυνεπή παραδείγματα, το ποσοστό ορθότητας του ταξινομητή θα βρεθεί να είναι:

**Απάντηση**: Τα ασυνεπή διανύσματα εκπαίδευσης δεν είναι δυνατόν να διαχωριστούν, αφού έχουν τις ίδιες τιμές σε όλες τις ιδιότητες, οπότε καταλήγουν στα ίδια φύλλα και (αφού ανήκουν σε διαφορετικές κατηγορίες) δεν συμφωνούν οι κατηγορίες όλων τους με τις κατηγορίες των φύλλων στα οποία κατέληξαν. Αφού αξιολογούμε χρησιμοποιώντας τα ίδια διανύσματα που χρησιμοποιήθηκαν κατά την εκπαίδευση, κάθε διάνυσμα αξιολόγησης καταλήγει στο ίδιο φύλλο όπου κατέληξε και κατά την εκπαίδευση και κάποια από τα ασυνεπή διανύσματα αξιολόγησης (και εκπαίδευσης) καταλήγουν πάλι σε φύλλα των οποίων οι κατηγορίες είναι διαφορετικές από εκείνες των ασυνεπών διανυσμάτων. Επομένως, θα υπάρχουν σίγουρα λάθη κατάταξης και άρα το ποσοστό ορθότητας θα είναι σίγουρα μικρότερο του 100%.

| Y | C |
|---|---|
| 0 | Μαύρο |
| 0 | Άσπρο |
| 1 | Άσπρο |



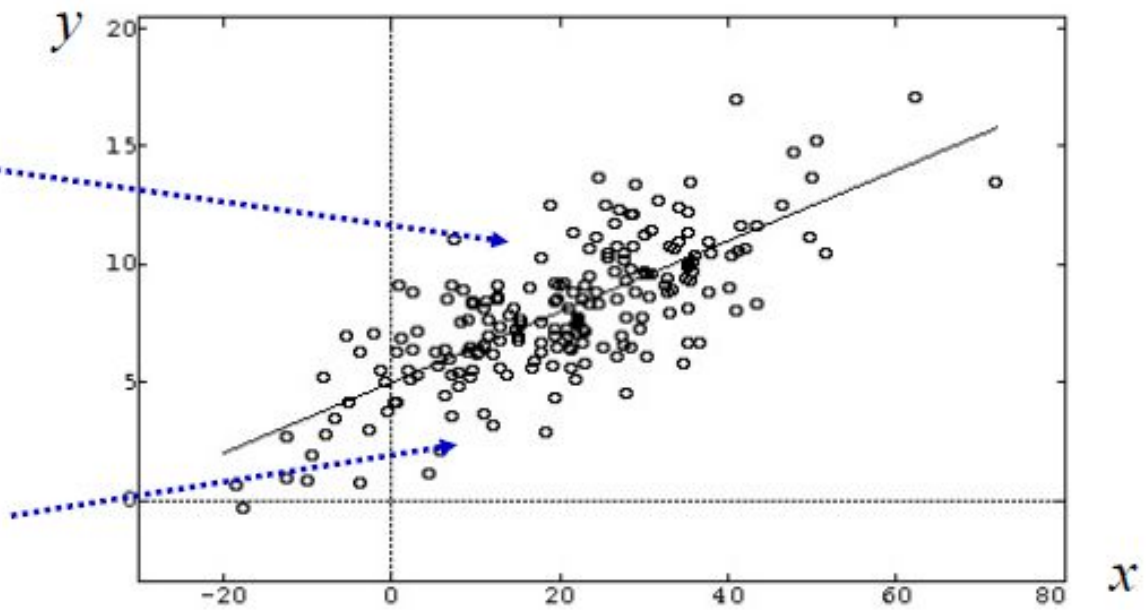

| Y | C |
|---|---|
| 0 | Μαύρο |
| 0 | Άσπρο |

| X | Y | C |
|---|---|---|
| 0 | 0 | Μαύρο |
| 1 | 0 | Μαύρο |
| 0 | 0 | Άσπρο |
| 0 | 1 | Άσπρο |

# Γραμμική παλινδρόμηση

Τα σημεία πάνω από τη γραμμή της $f(x)$ έχουν: $y > w_1 x + w_0$

Τα σημεία κάτω από τη γραμμή της $f(x)$ έχουν: $y < w_1 x + w_0$




Πηγή εικόνας: http://en.wikipedia.org/wiki/Linear_regression

- Θέλουμε να **μάθουμε την $f(x)$** από ένα **δείγμα** (τελείες).
- Περιοριζόμαστε σε **γραμμικές υποθέσεις** (συναρτήσεις):

$$y = f_{w_1, w_0}(x) = w_1 x + w_0$$

- Άρα ψάχνουμε τα **καλύτερα**: $w_1, w_0$

# Γραμμική παλινδρόμηση – συνέχεια

- Αν έχουμε **δύο ιδιότητες** $x_1, x_2$, οι γραμμικές μας υποθέσεις αντιστοιχούν σε **επίπεδα** του τριδιάστατου χώρου:

$$y = f_{w_2, w_1, w_0}(x_1, x_2) = w_2 x_2 + w_1 x_1 + w_0$$

- Γενικότερα, αν έχουμε ιδιότητες $x_1, x_2, \ldots, x_n$, οι γραμμικές μας υποθέσεις αντιστοιχούν σε **υπερ-επίπεδα** του $(n+1)$-διάστατου χώρου:

$$y = f_{w_n, \ldots, w_0}(x_1, \ldots, x_n) = w_n x_n + \ldots w_1 x_1 + w_0$$

$$= \sum_{l=0}^{n} w_l x_l = \langle w_0, w_1, \ldots, w_n \rangle \cdot \langle x_0, x_1, \ldots x_n \rangle$$

$$f_{\vec{w}}(\vec{x}) = \vec{w} \cdot \vec{x} = W^T X$$

Θεωρούμε ότι πάντα $x_0 = 1$.

Αν θεωρήσουμε κάθε διάνυσμα ως πίνακα μιας στήλης.

και ψάχνουμε το **καλύτερο** $\vec{w}$.

# Συνάρτηση αξιολόγησης

- Ο **χώρος αναζήτησης** περιλαμβάνει τα δυνατά $\vec{w}$.
- Για να **αξιολογήσουμε** κάθε **κατάσταση** $\vec{w}$, θα χρησιμοποιήσουμε τη συνάρτηση:

$$E(\vec{w}) = \frac{1}{2}\sum_{i=1}^{m}[f_{\vec{w}}(\vec{x}^{(i)}) - y^{(i)}]^2$$

όπου:

$(\vec{x}^{(i)}, y^{(i)})$ τα **παραδείγματα εκπαίδευσης** (δείγμα),

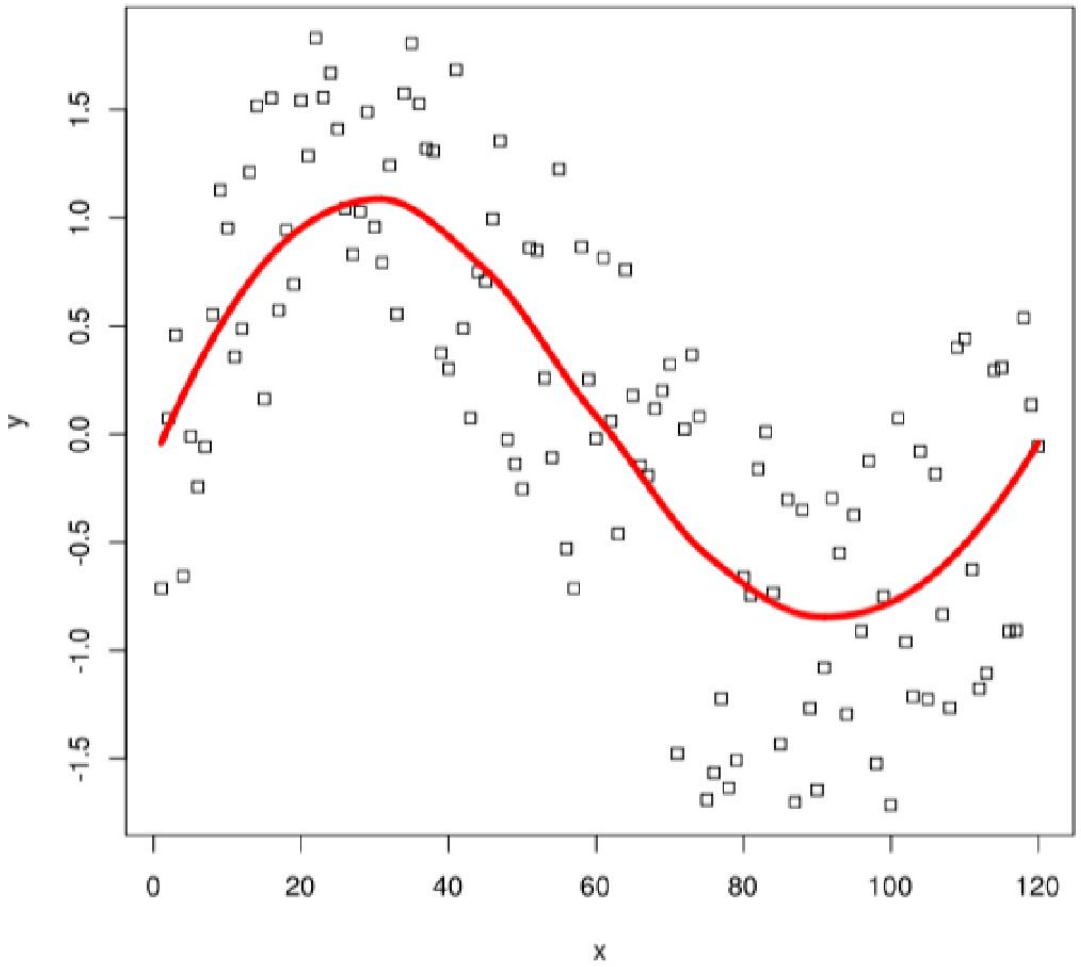
$y^{(i)}$ η **ορθή απόκριση** για είσοδο $\vec{x}^{(i)}$.

- «Γραμμική παλινδρόμηση **ελαχίστων τετραγώνων**».
  - Αξιολογούμε αθροίζοντας τα **τετράγωνα** των **διαφορών** των **αποκρίσεων** από τις **επιθυμητές** τιμές.

**Άσκηση 17.1.**

Οι τελείες του σχήματος στα δεξιά παριστάνουν έναν δείγμα που προέρχεται από πληθυσμό ο οποίος ακολουθεί στην πραγματικότητα την άγνωστη συνάρτηση y = f(x), της οποίας η γραφική παράσταση είναι η συνεχής καμπύλη[1]. Λόγω θορύβου κατά τις μετρήσεις της δειγματοληψίας, όμως, τα σημεία του δείγματος δεν βρίσκονται ακριβώς πάνω στη συνεχή καμπύλη. Θέλουμε να μάθουμε από το δείγμα μια συνάρτηση y = h(x), που να προσεγγίζει κατά το δυνατόν περισσότερο την f(x).

Α) Εξηγήστε γιατί η χρήση γραμμικής παλινδρόμησης ελαχίστων τετραγώνων δεν θα οδηγούσε σε ικανοποιητική h(x).

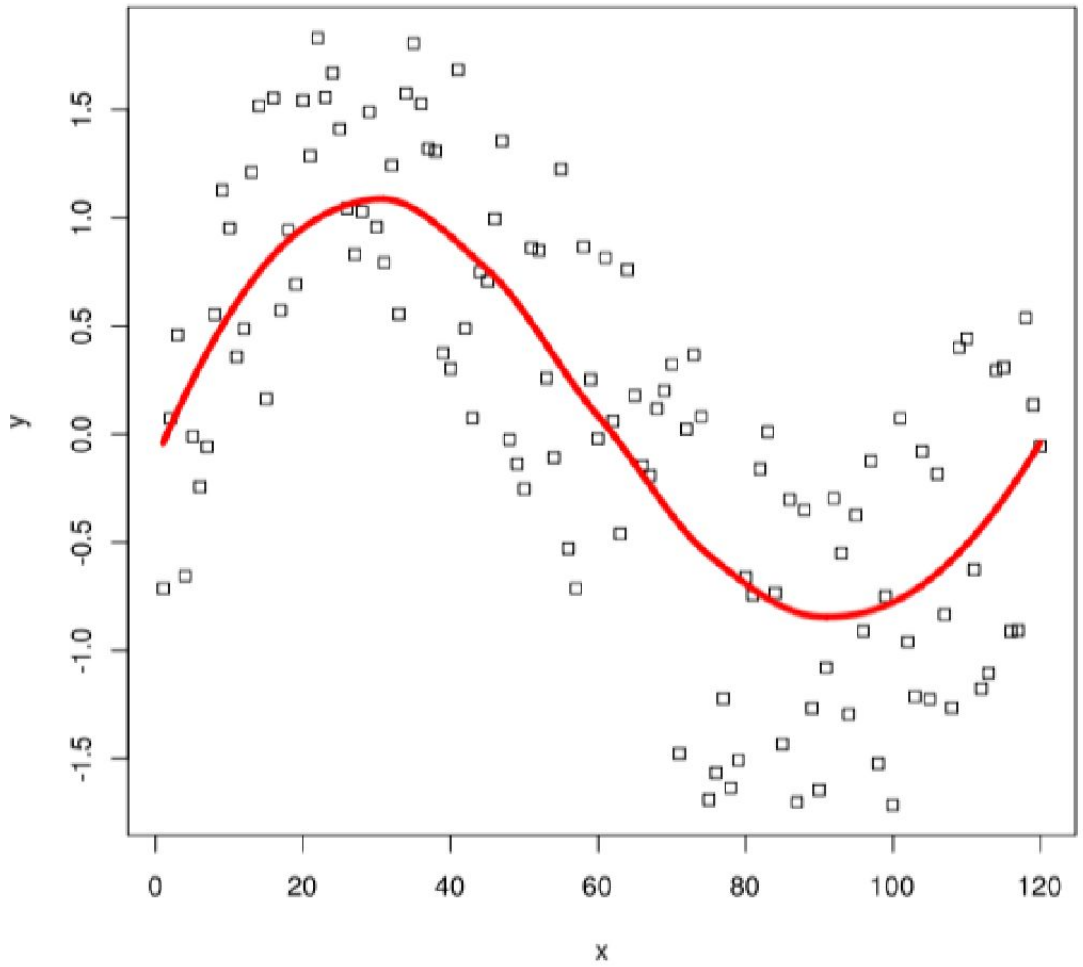



**Απάντηση**: Η γραμμική παλινδρόμηση ελαχίστων τετραγώνων μαθαίνει ευθείες γραμμές (ή γενικότερα επίπεδα ή υπερ-επίπεδα). Στη συγκεκριμένη περίπτωση η καμπύλη της y= f(x) προφανώς δεν μπορεί να προσεγγιστεί καλά από μια μόνο ευθεία γραμμή.

[1]Το σχήμα προέρχεται από την ιστοσελίδα http://en.wikipedia.org/wiki/Local_regression.

**Άσκηση 17.1.**

Β) Πώς θα μπορούσαμε να μάθουμε μια πιο ικανοποιητική h(x) χρησιμοποιώντας τον αλγόριθμο των k κοντινότερων γειτόνων (ή μια παραλλαγή του); Τι θα κάναμε κατά το στάδιο της εκπαίδευσης και τι όποτε (κατόπιν) μας δίνουν ένα x για το οποίο πρέπει να επιστρέψουμε το y = h(x);




**Απάντηση**:

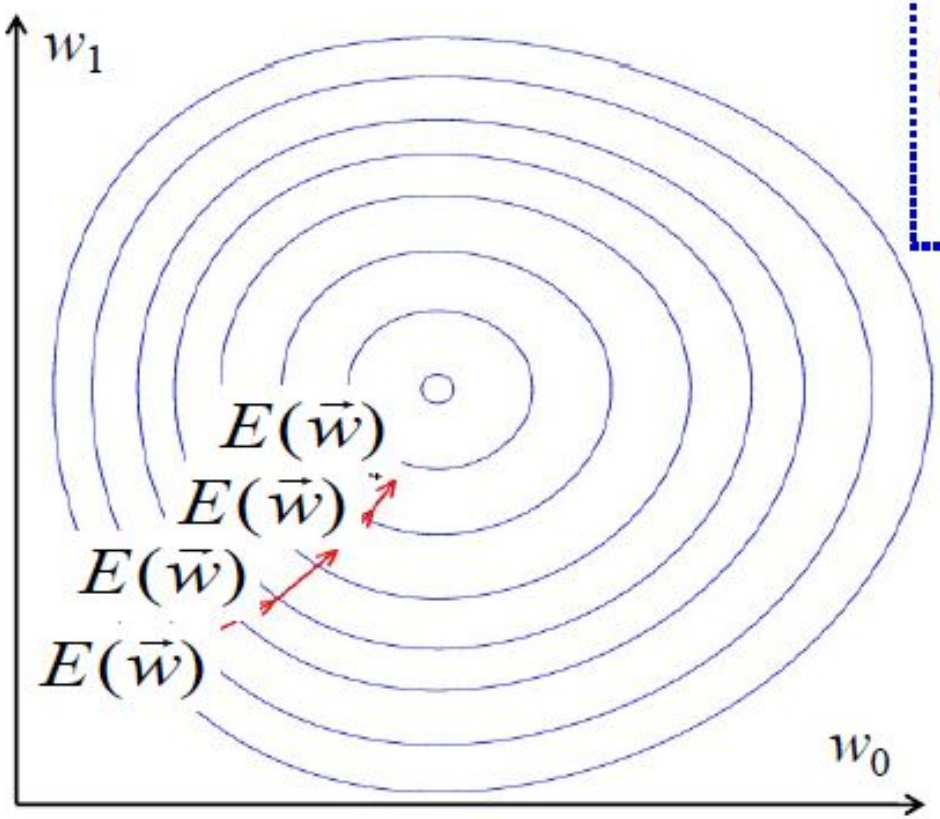
Κατά το στάδιο της εκπαίδευσης απλά αποθηκεύουμε όλες τις συντεταγμένες (x, y) των σημείων του δείγματος.

Κατόπιν, όποτε μας δίνουν ένα νέο x' και μας ζητούν το y' = h(x'), ανακτούμε τα k σημεία του δείγματος (γείτονες) των οποίων οι τιμές x βρίσκονται πιο κοντά στο x' και επιστρέφουμε το μέσο όρο των τιμών y αυτών των k σημείων (των γειτόνων). Μπορούμε επίσης να ζυγίζουμε κατά τον υπολογισμό του μέσου όρου τις y τιμές των γειτόνων, δίνοντας σε κάθε μία y τιμή γείτονα βάρος π.χ. αντιστρόφως ανάλογο της απόστασης της x τιμής του γείτονα από το x'.

# Κατάβαση κλίσης (gradient descent)

Ξεκινώ με τυχαία βάρη.
Μετράω σφάλμα $E(\vec{w})$ στα παραδείγματα εκπαίδευσης με τα τρέχοντα βάρη $\vec{w}$. Προς τα πού να μεταβάλω τα βάρη;

$$E(\vec{w}) = \frac{1}{2}\sum_{i=1}^{m}[f_{\vec{w}}(\vec{x}^{(i)}) - y^{(i)}]^2$$




Η κλίση $\nabla E(\vec{w})$ είναι ένα διάνυσμα που δείχνει προς την κατεύθυνση μεταβολής των βαρών που οδηγεί στη μεγαλύτερη **αύξηση** του $E(\vec{w})$. Το $-\nabla E(\vec{w})$ δείχνει προς την μεγαλύτερη **μείωση**.

Σε κάθε βήμα, τροποποιούμε το $\vec{w}$ κατά $\eta$ προς την κατεύθυνση που προκαλεί τη μεγαλύτερη μείωση του σφάλματος:

$$\vec{w} \leftarrow \vec{w} - \eta \cdot \nabla E(\vec{w})$$

**Κατάβαση λόφου** με συνάρτηση αξιολόγησης $E$.

# Στοχαστική κατάβαση κλίσης

1. Ξεκίνα με τυχαία βάρη $\vec{w}$.
2. Θέσε $i \leftarrow 1$ και $s \leftarrow 0$. Ανακάτεψε τα παραδείγματα.
3. Υπολόγισε το $E_i(\vec{w}) = \frac{1}{2}[f_{\vec{w}}(\vec{x}^{(i)}) - y^{(i)}]^2$ **μόνο στο <u>τρέχον</u>** ($i$-στό) παράδειγμα εκπαίδευσης.
4. $s \leftarrow s + E_i(\vec{w})$
5. Ενημέρωσε τα βάρη: $\vec{w} \leftarrow \vec{w} - \eta \cdot \nabla E_i(\vec{w})$ δηλαδή: $w_l \leftarrow w_l - \eta \cdot [f_{\vec{w}}(\vec{x}^{(i)}) - y^{(i)}] \cdot x_l^{(i)}$
6. Αν υπάρχει $(i+1)$-στό παράδειγμα, θέσε $i \leftarrow i+1$ και πήγαινε στο βήμα 3.
7. Αν το $s$ δεν έχει συγκλίνει και δεν υπερβήκαμε το μέγιστο αριθμό επαναλήψεων, πήγαινε στο βήμα 2.

Προκύπτει υπολογίζοντας τις μερικές παραγώγους...

**Άσκηση 17.2.**
Γράψτε τους υπολογισμούς με τους οποίους προκύπτει ο κανόνας ενημέρωσης βαρών της γραμμικής παλινδρόμησης ελαχίστων τετραγώνων, όταν χρησιμοποιείται **στοχαστική κατάβαση κλίσης.**

**Απάντηση**:
*Ο κανόνας ενημέρωσης βαρών της γραμμικής παλινδρόμησης ελαχίστων τετραγώνων, όταν χρησιμοποιείται στοχαστική κατάβαση κλίσης, δίνεται από τον τύπο:*

$\vec{w} \leftarrow \vec{w} - \eta \cdot \nabla_{\vec{w}} E_i(\vec{w})$, όπου $E_i(\vec{w}) = \frac{1}{2}\left[f_{\vec{w}}\left(\vec{x}^{(i)}\right) - y^{(i)}\right]^2$ και $f_{\vec{w}}\left(\vec{x}^{(i)}\right) = \mathrm{x_n w_n} + \cdots + \mathrm{x_1 w_1} + \mathrm{w_0}$

*Επίσης:*

$$\nabla_{\vec{w}} E_i(\vec{w}) = < \frac{\partial E_i(\vec{w})}{\partial w_0}, \frac{\partial E_i(\vec{w})}{\partial w_1}, \ldots, \frac{\partial E_i(\vec{w})}{\partial w_l}, \ldots, \frac{\partial E_i(\vec{w})}{\partial w_n} >$$

*Για $l \in \{0, \ldots, n\}$*:

$$\frac{\partial E_i(\vec{w})}{\partial w_l} = \left[\mathrm{f}_{\vec{w}}\left(\vec{x}^{(i)}\right) - y^{(i)}\right] \cdot x_l^{(i)}$$

*Άρα,*

$$\nabla_{\vec{w}} E_i(\vec{w}) = \left[\mathrm{f}_{\vec{w}}\left(\vec{x}^{(i)}\right) - y^{(i)}\right] < x_0^{(i)}, x_1^{(i)}, \ldots, x_l^{(i)}, \ldots, x_n^{(i)} > = \left[\mathrm{f}_{\vec{w}}\left(\vec{x}^{(i)}\right) - y^{(i)}\right] \cdot \vec{x}^{(i)}$$

*και ο τύπος ενημέρωσης βαρών γίνεται:*

$$\vec{w} \leftarrow \vec{w} - \eta \cdot \left[\mathrm{f}_{\vec{w}}\left(\vec{x}^{(i)}\right) - y^{(i)}\right] \cdot \vec{x}^{(i)}$$

**Άσκηση 17.3.**

Καταγράψαμε μεγάλο αριθμό παρτίδων (συνολικά M παρτίδες) μεταξύ παικτών Othello. Σε κάθε παρτίδα καταγράψαμε όλα τα στιγμιότυπα της σκακιέρας (τις θέσεις όλων των πιονιών), ένα στιγμιότυπο αμέσως μετά από κάθε κίνηση παίκτη. Τα στιγμιότυπα όλων των παρτίδων συνολικά ήταν K, ενώ τα διαφορετικά στιγμιότυπα (μετρώντας τα ίδια στιγμιότυπα μόνο μία φορά το καθένα) όλων των παρτίδων συνολικά ήταν L. Για κάθε στιγμιότυπο, καταγράψαμε τις τιμές των ιδιοτήτων $f_1(n)$, $f_2(n)$, $f_3(n)$, όπου τώρα **n** είναι το στιγμιότυπο· εμφανίστηκαν συνολικά N διαφορετικοί συνδυασμοί τιμών των $f_1(n)$, $f_2(n)$, $f_3(n)$ στις M παρτίδες. Για κάθε στιγμιότυπο, καταγράψαμε ακόμη αν η παρτίδα στην οποία εμφανίστηκε το στιγμιότυπο έληξε υπέρ του μαύρου παίκτη, υπέρ του άσπρου ή ισόπαλη.

Εξηγήστε πώς θα μπορούσαμε να εκμεταλλευτούμε τα καταγεγραμμένα στοιχεία των M παρτίδων, για να μάθουμε με γραμμική παλινδρόμηση ελαχίστων τετραγώνων τις καλύτερες δυνατές τιμές των $w_0$, $w_1$, $w_2$, $w_3$, ώστε για κάθε κόμβο n του δέντρου αναζήτησης, με MiniMax, η $h(n) = w_1 \times f_1(n) + w_2 \times f_2(n) + w_3 \times f_3(n) + w_0$ επιστρέφει μια κατά το δυνατόν ακριβέστερη εκτίμηση του *αναμενόμενου* οφέλους με το οποίο θα τελειώσει το παιχνίδι, αν βρεθεί στην κατάσταση του κόμβου n.

Θεωρήστε ότι το όφελος ενός παιχνιδιού είναι 1 όταν κερδίζει ο Max (μαύρος), –1 όταν κερδίζει ο Min (άσπρος) και 0 όταν το παιχνίδι λήγει ισόπαλο.

**Άσκηση 17.3.**

**Α)** Πόσα παραδείγματα (διανύσματα ιδιοτήτων) εκπαίδευσης θα είχαμε στη γραμμική παλινδρόμηση και ποιες θα ήταν οι μεταβλητές της συνάρτησης που θα μάθαινε η γραμμική παλινδρόμηση; Φροντίστε να μην υπάρχουν ασυνεπή παραδείγματα εκπαίδευσης.

**Απάντηση**:

*Θα είχαμε N παραδείγματα εκπαίδευσης, ένα για κάθε διαφορετικό συνδυασμό τιμών των $f_1(n)$, $f_2(n)$, $f_3(n)$ που παρατηρήθηκε στις M παρτίδες. Οι μεταβλητές θα ήταν οι ιδιότητες $f_1(n)$, $f_2(n)$, $f_3(n)$.*

**Άσκηση 17.3.**

**Β)** Πώς ακριβώς (δώστε μαθηματικό τύπο) θα υπολογίζαμε για κάθε παράδειγμα εκπαίδευσης την «ορθή» (επιθυμητή) απόκριση της συνάρτησης που θα θέλαμε να μάθει η γραμμική παλινδρόμηση;

**Απάντηση**:

*Κάθε παράδειγμα εκπαίδευσης παριστάνει έναν από τους Ν διαφορετικούς συνδυασμούς τιμών των τριών ιδιοτήτων που παρατηρήθηκε στις Μ παρτίδες. Έστω σ ένας από αυτούς τους διαφορετικούς συνδυασμούς και* $\sigma_1$ , $\sigma_2$ , ..., $\sigma_k$ *τα καταγεγραμμένα στιγμιότυπα (όχι αναγκαστικά διαφορετικά μεταξύ τους) που είχαν το συγκεκριμένο συνδυασμό τιμών του σ στις* $f_1(n)$, $f_2(n)$, $f_3(n)$. *Η επιθυμητή απόκριση για το σ θα ήταν:*

$$\frac{1}{k}\sum_{i=1}^{k} u(\sigma_i) = y^{(\sigma)}$$

όπου $u(\sigma_i)$ *είναι το όφελος με το οποίο τελείωσε η παρτίδα στην οποία καταγράφηκε το στιγμιότυπο .*

**Άσκηση 17.3.**

**Γ)** *Ποια θα ήταν η σχέση της συνάρτησης $f_w( x ) = w \times x$ που θα μάθαινε η γραμμική παλινδρόμηση με την h(n) που θα χρησιμοποιούσαμε τελικά;*

**Απάντηση**:

$$h(n) = f_{\vec{w}}(\vec{x}) \text{ , } \mu\varepsilon : \vec{x} = < f_1(n), f_2(n), f_3(n) >$$

$$f_{\vec{w}}(\vec{x}) = f_{w_0,w_1,w_2,w_3}(\langle f_1(n), f_2(n), f_3(n) \rangle) = w_1 \cdot f_1(n) + w_2 \cdot f_2(n) + w_3 \cdot f_3(n) + w_0 = h(n)$$