

Ασκήσεις μελέτης της 16^{ης} διάλεξης

16.1. (α) Έστω ένα αντικείμενο προς κατάταξη το οποίο παριστάνεται με το διάνυσμα $\langle X_1, X_2, X_3 \rangle = \langle 0, 1, 0 \rangle$. 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 εκτιμητήρια 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

Απάντηση:

$$\begin{aligned}
 P(C = 1|X_1 = 0, X_2 = 1, X_3 = 0) &\cong \frac{P(C = 1)}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot P(X_1 = 0|C = 1) \cdot P(X_2 = 1|C = 1) \cdot P(X_3 = 0|C = 1) \\
 &\cong \frac{1/2}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot \frac{0+1}{2+2} \cdot \frac{1+1}{2+2} \cdot \frac{0+1}{2+2} \\
 &= \frac{1/2}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot \frac{1}{4} \cdot \frac{2}{4} \cdot \frac{1}{4}
 \end{aligned}$$

$$\begin{aligned}
 P(C = 0|X_1 = 0, X_2 = 1, X_3 = 0) &\cong \frac{P(C = 0)}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot P(X_1 = 0|C = 0) \cdot P(X_2 = 1|C = 0) \cdot P(X_3 = 0|C = 0) \\
 &\cong \frac{1/2}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot \frac{1+1}{2+2} \cdot \frac{1+1}{2+2} \cdot \frac{1+1}{2+2} \\
 &= \frac{1/2}{P(X_1 = 0, X_2 = 1, X_3 = 0)} \cdot \frac{2}{4} \cdot \frac{2}{4} \cdot \frac{2}{4}
 \end{aligned}$$

Επομένως θα το κατατάξει στην $C = 0$.

16.2. Έστω ένα αντικείμενο με διάνυσμα $\langle X_1, X_2, X_3, X_4 \rangle = \langle b, d, b, a \rangle$. Σε ποια από τις τρεις κατηγορίες θα κατατάξει το αντικείμενο ένας αφελής ταξινομητής Bayes (της πολυ-μεταβλητής μορφής Bernoulli που συναντήσαμε στο μάθημα) ο οποίος διαθέτει τα δεδομένα εκπαίδευσης του πίνακα; Γράψτε αναλυτικά τους υπολογισμούς σας. Χρησιμοποιήστε εκτιμητήρια Laplace κατά τις εκτιμήσεις των πιθανοτήτων $P(X_i|C)$. Θεωρήστε ότι κάθε τυχαία μεταβλητή X_i έχει τέσσερις δυνατές τιμές: a, b, c, d.

X_1	X_2	X_3	X_4	C
a	b	b	a	1
b	a	a	b	1
b	b	a	b	1
a	a	b	b	2
a	b	b	b	2
b	a	b	a	2
c	d	d	c	3
d	c	c	d	3
d	d	c	d	3

Απάντηση:

Έχουμε: $P(C = 1) = P(C = 2) = P(C = 3) = 3/9$. Επομένως, οι a priori πιθανότητες δεν επηρεάζουν την απόφαση.

Έχουμε επίσης:

$$P(X_1 = b|C = 1) = \frac{2+1}{3+4} = \frac{3}{7} \quad P(X_2 = d|C = 1) = \frac{0+1}{3+4} = \frac{1}{7}$$

$$P(X_3 = b|C = 1) = \frac{1+1}{3+4} = \frac{2}{7} \quad P(X_4 = a|C = 1) = \frac{1+1}{3+4} = \frac{2}{7}$$

και:

$$P(X_1 = b|C = 2) = \frac{1+1}{3+4} = \frac{2}{7} \quad P(X_2 = d|C = 2) = \frac{0+1}{3+4} = \frac{1}{7}$$

$$P(X_3 = b|C = 2) = \frac{3+1}{3+4} = \frac{4}{7} \quad P(X_4 = a|C = 2) = \frac{1+1}{3+4} = \frac{2}{7}$$

και:

$$P(X_1 = b|C = 3) = \frac{0+1}{3+4} = \frac{1}{7} \quad P(X_2 = d|C = 3) = \frac{2+1}{3+4} = \frac{3}{7}$$

$$P(X_3 = b|C = 3) = \frac{0+1}{3+4} = \frac{1}{7} \quad P(X_4 = a|C = 3) = \frac{0+1}{3+4} = \frac{1}{7}$$

Άρα:

$$P(X_1 = b|C = 1) \cdot P(X_2 = d|C = 1) \cdot P(X_3 = b|C = 1) \cdot P(X_4 = a|C = 1) = \frac{3 \cdot 1 \cdot 2 \cdot 2}{7 \cdot 7 \cdot 7 \cdot 7}$$

$$P(X_1 = b|C = 2) \cdot P(X_2 = d|C = 2) \cdot P(X_3 = b|C = 2) \cdot P(X_4 = a|C = 2) = \frac{2 \cdot 1 \cdot 4 \cdot 2}{7 \cdot 7 \cdot 7 \cdot 7}$$

$$P(X_1 = b|C = 3) \cdot P(X_2 = d|C = 3) \cdot P(X_3 = b|C = 3) \cdot P(X_4 = a|C = 3) = \frac{1 \cdot 3 \cdot 1 \cdot 1}{7 \cdot 7 \cdot 7 \cdot 7}$$

και επομένως το αντικείμενο κατατάσσεται στην κατηγορία $C = 2$.

16.3. Χρησιμοποιούμε μια παραλλαγή του αφελούς ταξινομητή Bayes που συναντήσαμε στο μάθημα (δηλαδή μια παραλλαγή της πολυ-μεταβλητής μορφής Bernoulli), 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 $\log a^b = b \cdot \log a$.

Απάντηση: Ο ταξινομητής κατατάσσει στην κατηγορία $C = 1$ αν και μόνο αν (ανν):

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

ισοδύναμα:

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

ισοδύναμα:

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

ισοδύναμα:

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

Θέτοντας $K_1 = \log P(C = 1)$, $K_{2,i} = \log P(t_i|C = 1)$, $K_{3,i} = \log[1 - P(t_i|C = 1)]$, $K_4 = \log K$, η προηγούμενη σχέση γίνεται:

$$K_1 + \sum_{i=1}^m [x_i \cdot K_{2,i} + (1 - x_i) \cdot K_{3,i}] \geq K_4$$

ισοδύναμα:

$$K_1 + \sum_{i=1}^m K_{3,i} - K_4 + \sum_{i=1}^m x_i \cdot (K_{2,i} - K_{3,i}) \geq 0$$

Θέτοντας $w_0 = K_1 + \sum_{i=1}^m K_{3,i} - K_4$ και $w_i = (K_{2,i} - K_{3,i})$, η προηγούμενη σχέση γίνεται:

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

Επομένως πρόκειται για γραμμικό ταξινομητή.

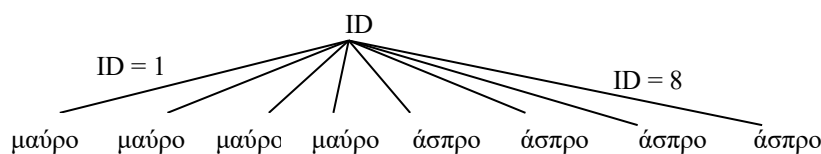
16.4. (α) Βάσει των παραδειγμάτων εκπαίδευσης του πίνακα, πόση είναι η εντροπία $H(C)$ της κατηγορίας C και γιατί;

Απάντηση: $P(C=\text{μαύρο}) = P(C=\text{άσπρο}) = 1/2$. Έχουμε δύο ισοπίθανα ενδεχόμενα $C = \text{μαύρο}$ και $C = \text{άσπρο}$ 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

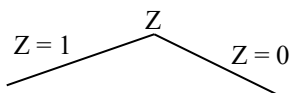
(β) Σχεδιάστε το δέντρο απόφασης που θα κατασκευάσει ο ID3, αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες τις ID, X, Y, Z (το δέντρο προβλέπει την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.

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, $H(C | ID) = 0$ και επομένως $IG(ID, C) = H(C) - H(C | ID) = 1$. 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.

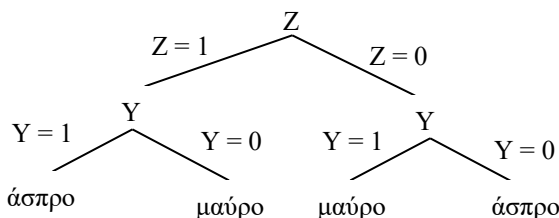


(γ) Σχεδιάστε τώρα το δέντρο απόφασης που θα κατασκευάσει ο ID3, αν του επιτρέψουμε να χρησιμοποιήσει ως ιδιότητες μόνο τις X, Y, Z (το δέντρο προβλέπει πάλι την κατηγορία C). Δεν χρειάζονται αριθμητικοί υπολογισμοί, μόνο προσεκτική παρατήρηση των δεδομένων. Εξηγήστε γιατί θα προκύψει το δέντρο που σχεδιάσατε.

Απάντηση: Τα δεδομένα εκπαίδευσης δείχνουν ότι αν μάθουμε πως $Z = 1$, είναι πολύ πιθανό (πιθανότητα $\frac{3}{4}$) ότι $C = \text{μαύρο}$ και αν μάθουμε πως $Z = 0$, είναι πολύ πιθανό (πιθανότητα $\frac{3}{4}$) ότι $C = \text{άσπρο}$. 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 $C = \text{μαύρο}$ παραμένει $\frac{1}{2}$ και ίση με την πιθανότητα να έχουμε $C = \text{άσπρο}$. Ομοίως, αν μάθουμε ότι $X = 0$, η πιθανότητα να έχουμε $C = \text{μαύρο}$ παραμένει $\frac{1}{2}$ και ίση με την πιθανότητα να έχουμε $C = \text{άσπρο}$. Επομένως, όποια κι αν είναι η τιμή της Y, εξακολουθούμε να έχουμε δύο ισοπίθανα ενδεχόμενα $C = \text{μαύρο}$ και $C = \text{άσπρο}$ 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.



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. 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.



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

σίγουρα 100%

σίγουρα μικρότερο του 100%

τίποτα από τα παραπάνω (το αποτέλεσμα εξαρτάται από τις τυχαιότητες των δεδομένων).

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%.