

3

AM *

Αν δεν έχετε/γνωρίζετε τον AM σας βάλτε ένα email σας.

Enter your answer

4

Ιδιότητα *

ΔΠΜΣ τάδε , υ.δ. σχολής τάδε κλπ

Enter your answer

5

964 άνθρωποι ερωτήθηκαν αν πίνουν αλκοόλ. Ο πίνακας συγκεντρώνει τις απαντήσεις τους

Φύλο Ναι Όχι Σύνολο

Άνδρες 152 299 451

Γυναίκες 195 318 513

Σύνολο 347 617 964

Ποιες είναι οι πιθανότητες (odds) για έναν άνδρα να πίνει (ως προς το να μην πίνει) αλκοόλ; Ποιο είναι ο σχετικός κίνδυνος να πίνουν οι γυναίκες σε σχέση με τους άνδρες;

(4 Points)

Enter your answer

6

Έχετε ένα πρόβλημα δυαδικής ταξινόμησης. Κάθε δείγμα x_i ανήκει είτε στην κλάση $c=1$ είτε στην κλάση $c=0$. Η διαδικασία συλλογής δεδομένων εκπαίδευσης είναι ωστόσο ατελής και κάποια παραδείγματα δεν λαμβάνουν σωστές ετικέτες. Για κάθε δείγμα x_i αντί να έχουμε την αληθινή ετικέτα του, c_i , έχουμε την πιθανότητα π_i που αντιστοιχεί στο να είναι $c_i=1$. Πώς εκφράζεται η λογαριθμική πιθανότητα (log likelihood) για ένα πιθανοτικό μοντέλο $p(c=1|\theta)$ αν αντί του c_i βάλουμε το π_i ;

(5 Points)

Enter your answer

7

Έστω ένα δέντρο αποφάσεων. Έχετε δύο επιλογές: α) να ματατρέψετε το δέντρο σε κανόνες και να κλαδέψετε τους κανόνες, β) να κλαδέψετε το δέντρο και στη συνέχεια να μετατρέψετε το δέντρο σε κανόνες. Ποια είναι τα πλεονεκτήματα και μειονεκτήματα της κάθε μεθόδου (εν συντομία);

(4 Points)

Enter your answer

8

Έχουμε έξι δείγματα δύο χαρακτηριστικών το καθένα: [4 1], [6 6], [9 5], [1 2], [7 3], [5 4]. Οι ετικέτες τους είναι [1 0 1 0 1 0]. Για την ταξινόμηση φτιάχνουμε ένα δέντρο αποφάσεων βάθους δύο. Οι διχοτομήσεις προκύπτουν με κανόνες που θέτουν ένα χαρακτηριστικό μεγαλύτερο ίσο από ένα κατώφλι. α) ποια είναι η εντροπία της ρίζας του δέντρου; β) ποιος είναι ο κανόνας της δεύτερης διχοτόμησης;

(5 Points)

Enter your answer

9

Ένα λαχείο σύμφωνα με τους εκδότες του έχει πιθανότητα να κερδίζει 0.3. Ποια είναι η πιθανότητα να χρειαστεί κάποιος το λιγότερο τρία λαχεία για να κερδίσει;

(3 Points)

Enter your answer

10

Χρησιμοποιήστε τις μεθόδους (a) κανονικοποίηση min-max θέτοντας min=0 και max=1 και (b) κανονικοποίηση z-score για να ομαλοποιήσετε τις ακόλουθες τιμές δεδομένων: 200, 300, 400, 600, 1000

(3 Points)

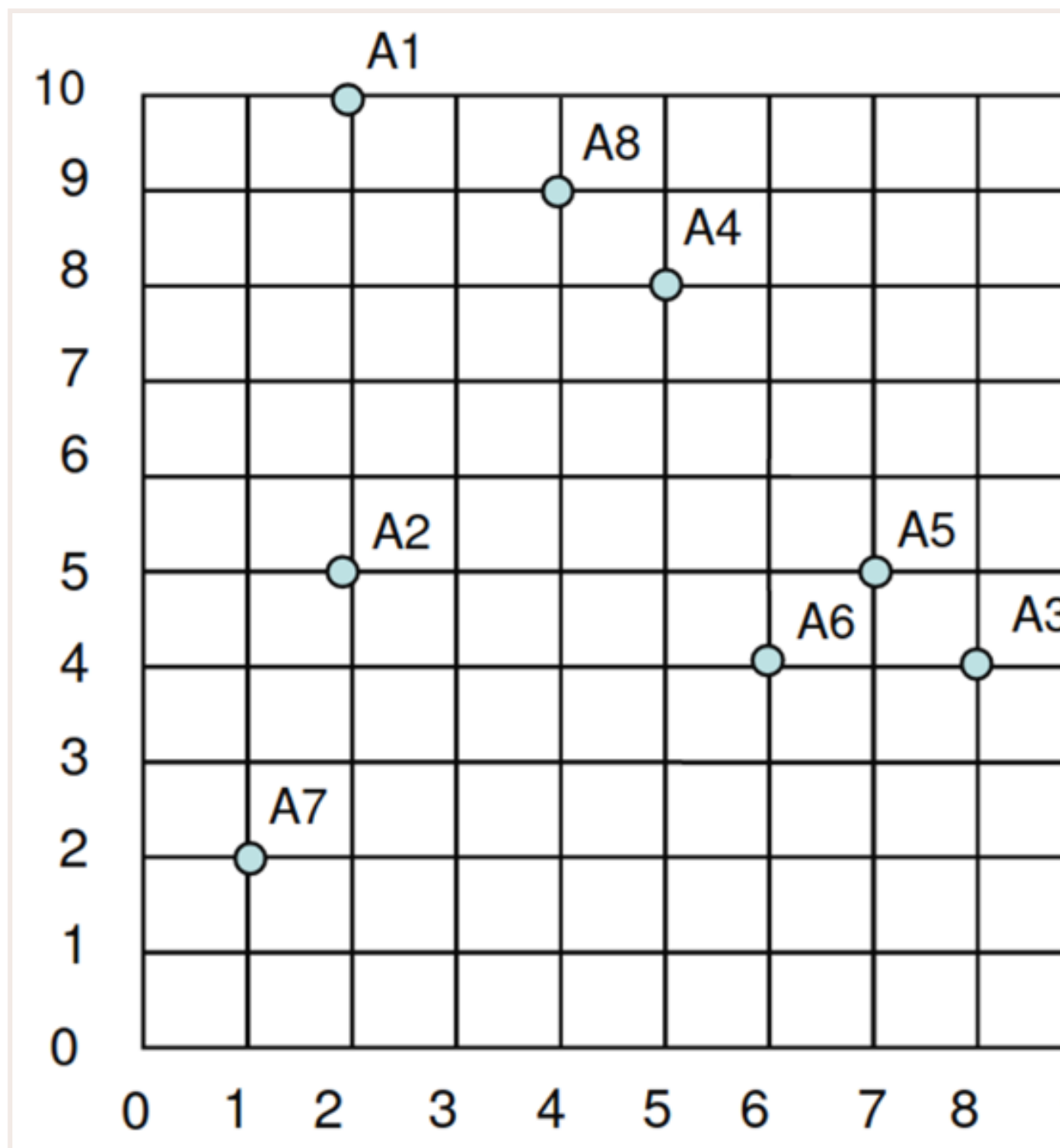
Enter your answer

11

Δίνονται τα 8 σημεία (A1 ως A8) στον δισδιάστατο χώρο, τα οποία τα συσταδοποιούμε με τη χρήση του αλγορίθμου DBSCAN με MinPts=2 και $Eps=\sqrt{5}$. Να γράψετε ποιες συστάδες σχηματίζονται, αν υπάρχουν οριακά σημεία (border points), ποια είναι και σε ποια συστάδα είναι οριακά καθώς και αν υπάρχουν σημεία θορύβου (noise points) και ποια είναι αυτά.

(5 Points)

Αν η εικόνα δεν εμφανίζεται, δείτε την στο <https://imgur.com/b4TgyJy.png>.



Enter your answer

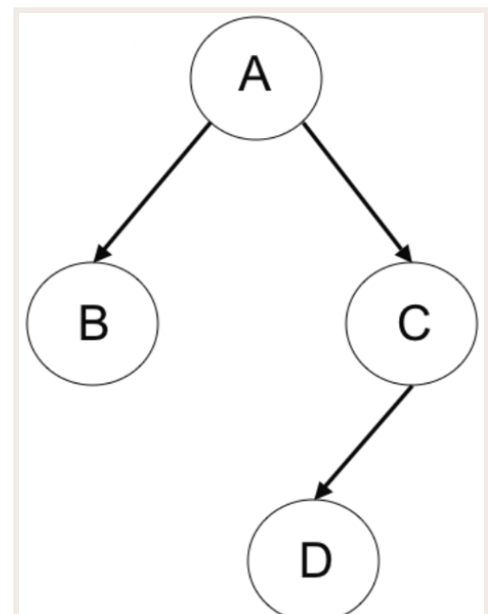
12

Εκτελούμε τον αλγόριθμο Expectation-Maximization για ένα μοντέλο μίξης δύο γκαουσιανών κατανομών με ίδια βάρη για ένα σύνολο 3 δειγμάτων δεδομένων και σε κάποιο βήμα της επανάληψης έχουμε $p(x_1 = -2|\theta_1) = A$, $p(x_1 = -2|\theta_2) = B$, $p(x_2 = 1|\theta_1) = C$, $p(x_2 = 1|\theta_2) = D$, $p(x_3 = 3|\theta_1) = E$, $p(x_3 = 3|\theta_2) = F$. Έστω $A=0.3$, $B=0.7$, $C=0.45$, $D=0.55$, $E=0.2$, και $F=0.8$. Να υπολογίσετε την πιθανότητα το κάθε δείγμα δεδομένων να προέρχεται από κάθε μια από τις δύο κατανομές καθώς και τις μέσες τιμές των δύο κατανομών μετά την ολοκλήρωση αυτού του βήματος (στο διάστημα $[0,1]$, με ακρίβεια 2 δεκαδικών ψηφίων)

(5 Points)

Enter your answer

13



Δίνεται το Μπεϋζιανό δίκτυο πεποίθησης του σχήματος και οι εξείς πιθανότητες: $P(A)=0.3$, $P(B|A)=0.7$, $P(B|\sim A)=0.5$, $P(C|A)=0.2$, $P(C|\sim A)=0.6$, $P(D|C)=0.7$, $P(D|\sim C)=0.7$.

Υπολογίστε την πιθανότητα $P(A|B)$

(5 Points)

Αν η εικόνα δεν εμφανίζεται δείτε την στο <https://imgur.com/zag4eDk>

Enter your answer

14

Δίνεται ένα σύνολο κανόνων με 200 θετικά δείγματα και 300 αρνητικά. Για καθένα από τους ακόλουθους κανόνες

R1 : A -> + (καλύπτει 5 θετικά και 2 αρνητικά δείγματα),

R2 : B -> + (καλύπτει 40 θετικά και 20 αρνητικά δείγματα),

R3 : C -> + (καλύπτει 80 θετικά 80 αρνητικά δείγματα),

υπολογίστε α) την ορθότητά του (accuracy) και β) το κέρδος πληροφορίας (information gain) FOIL του.

(4 Points)

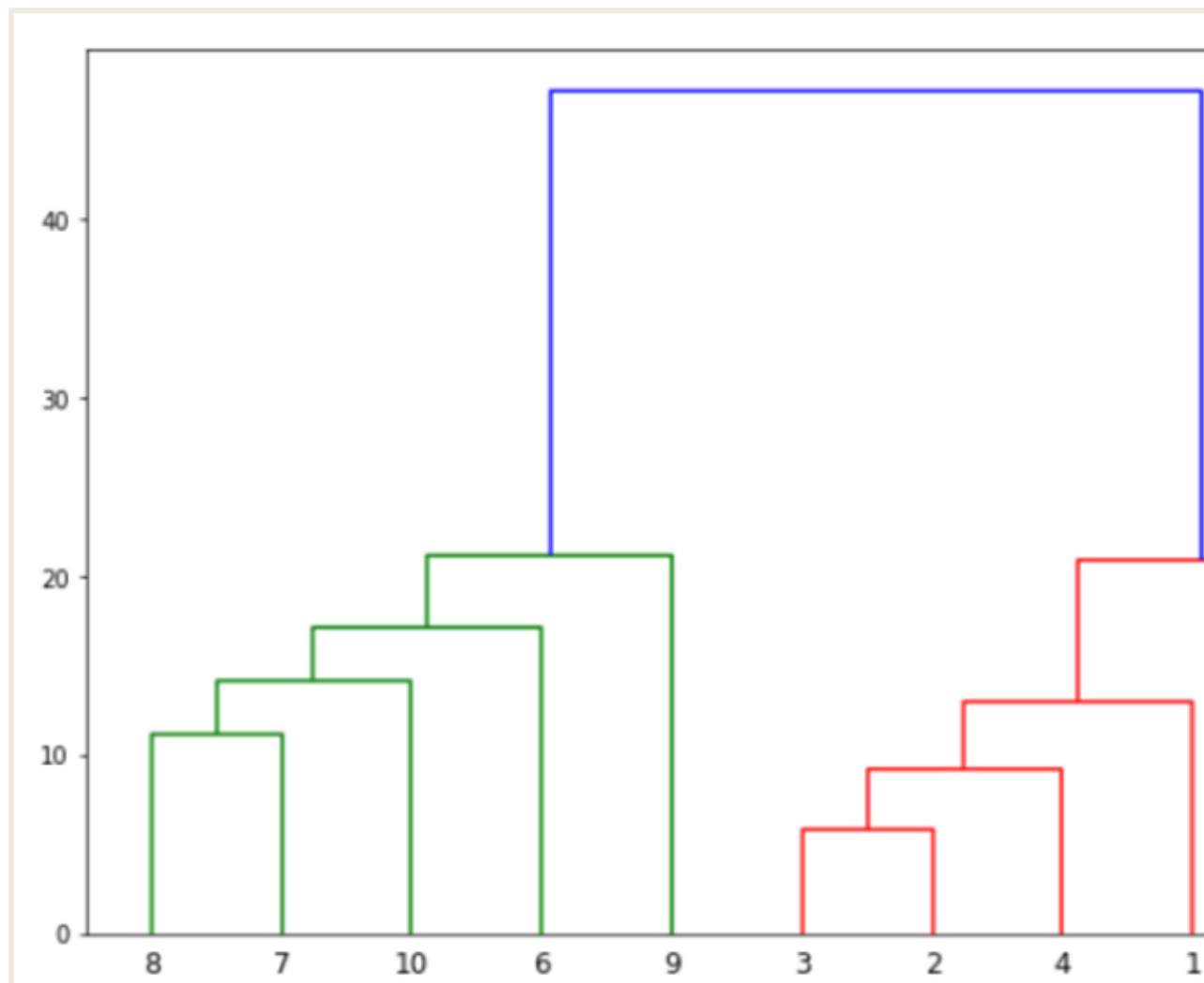
Enter your answer

15

Δίνεται το δενδρόγραμμα ενός συσσωρευτικού ιεραρχικού αλγορίθμου συσταδοποίησης (agglomerative hierarchical clustering) δέκα σημείων. Ποια από τις παρακάτω συστάδες σχηματίζεται πρώτη;

(5 Points)

Αν δεν εμφανίζεται η εικόνα, δείτε την στο <https://i.imgur.com/YfoqshN.png>.



- ☐ Η συστάδα των σημείων 3,2,4,1
- ☐ Δεν απαντώ
- ☐ Η συστάδα των σημείων 8,7
- ☐ Η συστάδα των σημείων 3,2,4
- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Η συστάδα των σημείων 8,7,10,6

Submit

3

AM *

Αν δεν έχετε/γνωρίζετε τον AM σας βάλτε ένα email σας.

Enter your answer

4

Ιδιότητα *

ΔΠΜΣ τάδε , υ.δ. σχολής τάδε κλπ

Enter your answer

5

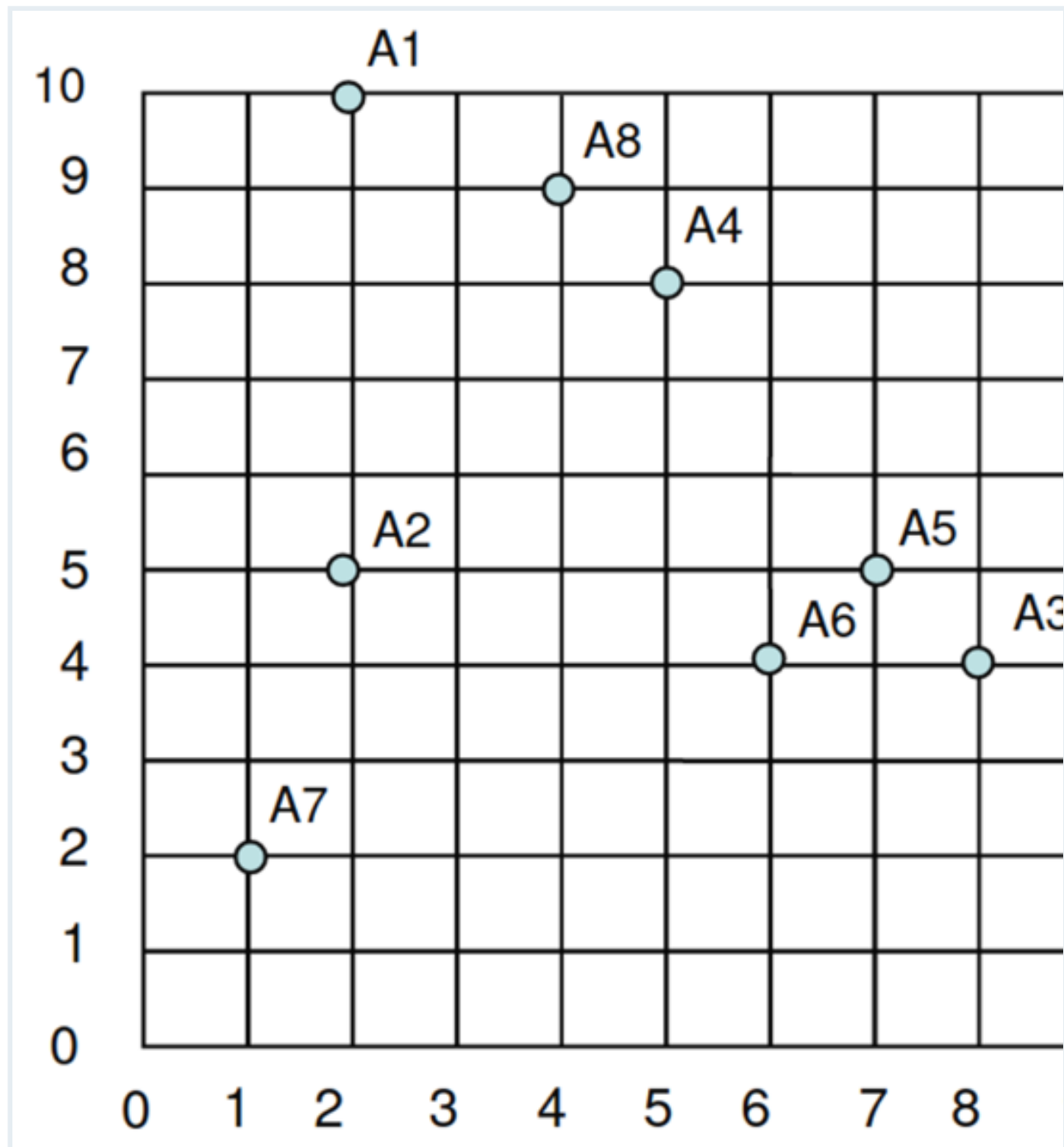
Ποιες από τις επόμενες ιδιότητες είναι χαρακτηριστικές των δέντρων αποφάσεων;
(4 Points)

- ☐ Μη ομαλότητα στις επιφάνειες απόφασης
- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Υψηλή διακύμανση
- ☐ Μη φραγμένο σύνολο παραμέτρων
- ☐ Υψηλή πόλωση

6

Δίνονται τα 8 σημεία (A1 ως A8) στον δισδιάστατο χώρο, τα οποία τα συσταδοποιούμε με τη χρήση του αλγορίθμου DBSCAN με $\text{MinPts}=2$ και $\text{Eps}=\sqrt{10}$. Να γράψετε ποιες συστάδες σχηματίζονται, αν υπάρχουν οριακά σημεία (border points), ποια είναι και σε ποια συστάδα είναι οριακά καθώς και αν υπάρχουν σημεία θορύβου (noise points) και ποια είναι αυτά.
(5 Points)

Αν η εικόνα δεν εμφανίζεται, δείτε την στο <https://imgur.com/b4TqyJy.png>.



Enter your answer

7

Ένας επενδυτής εξετάζει δύο μετοχές, την A και την B. Οι αναλυτές λένε ότι η πιθανότητα να φέρει κέρδη η A σε ένα τρίμηνο είναι 60% και η B 50%. Εκτιμούν επίσης ότι η πιθανότητα το να φέρουν κέρδη και οι δύο σε ένα τρίμηνο είναι 30%. Ποια είναι η πιθανότητα ο επενδυτής να έχει κέρδη στο τρίμηνο από την A, την B ή και τις δύο μετοχές
(3 Points)

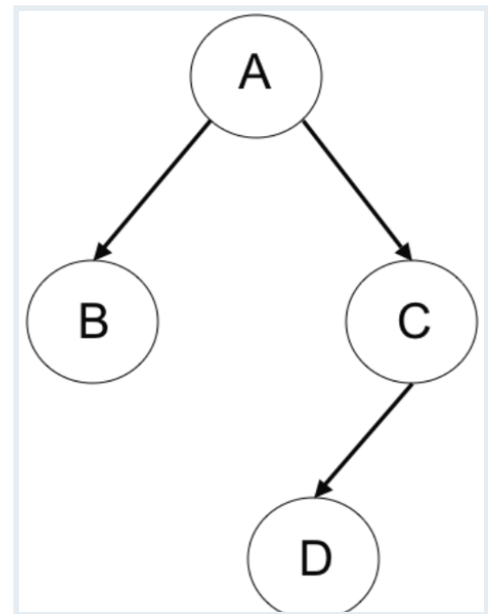
Enter your answer

8

Δίνεται ένα σύνολο κανόνων με 200 θετικά δείγματα και 300 αρνητικά. Για καθένα από τους ακόλουθους κανόνες
R1 : A -> + (καλύπτει 5 θετικά και 2 αρνητικά δείγματα),
R2 : B -> + (καλύπτει 40 θετικά και 20 αρνητικά δείγματα),
R3 : C -> + (καλύπτει 80 θετικά 80 αρνητικά δείγματα),
υπολογίστε α) την ορθότητά του (accuracy) και β) το κέρδος πληροφορίας (information gain) FOIL του.
(4 Points)

Enter your answer

9



Δίνεται το Μπεϋζιανό δίκτυο πεποίθησης του σχήματος και οι εξείς πιθανότητες: $P(A)=0.3$, $P(B|A)=0.7$, $P(B|\sim A)=0.5$, $P(C|A)=0.2$, $P(C|\sim A)=0.6$, $P(D|C)=0.7$, $P(D|\sim C)=0.7$.

Υπολογίστε την πιθανότητα $P(B|D)$

(5 Points)

Αν η εικόνα δεν εμφανίζεται δείτε την στο <https://imgur.com/zag4eDk>

Enter your answer

10

Έστω ένα μοντέλο λογιστικής παλινδρόμησης και ένα γραμμικά διαχωρίσιμο πρόβλημα. Αν έχουμε μια απεικόνιση $g(x)$ των χαρακτηριστικών x , και βάρη w , εξηγήστε ποιοτικά πώς το μοντέλο μέγιστης πιθανότητας προκύπτει αν βρούμε την ευθεία με $w^T g(x) = 0$ και στη συνέχεια θέσουμε τη νόρμα του w στο άπειρο.

(5 Points)

Enter your answer

11

964 άνθρωποι ερωτήθηκαν αν πίνουν αλκοόλ. Ο πίνακας συγκεντρώνει τις απαντήσεις τους

Φύλο Ναι Όχι Σύνολο

Άνδρες 152 299 451

Γυναίκες 195 318 513

Σύνολο 347 617 964

Ποιες είναι οι πιθανότητες (odds) για μια γυναίκα να πίνει (ως προς το να μην πίνει) αλκοόλ; Ποιο είναι ο σχετικός κίνδυνος να πίνουν οι άνδρες σε σχέση με τις γυναίκες;

(4 Points)

Enter your answer

12

Εκτελούμε τον αλγόριθμο Expectation-Maximization για ένα μοντέλο μίξης δύο γκαουσιανών κατανομών με ίδια βάρη για ένα σύνολο 3 δειγμάτων δεδομένων και σε κάποιο βήμα της επανάληψης έχουμε $p(x_1 = -2|\theta_1) = A$, $p(x_1 = -2|\theta_2) = B$, $p(x_2 = 1|\theta_1) = C$, $p(x_2 = 1|\theta_2) = D$, $p(x_3 = 3|\theta_1) = E$, $p(x_3 = 3|\theta_2) = F$. Έστω $A = 0.1$, $B = 0.9$, $C = 0.4$, $D = 0.6$, $E = 0.5$, και $F = 0.5$. Να υπολογίσετε την πιθανότητα το κάθε δείγμα δεδομένων να προέρχεται από κάθε μια από τις δύο κατανομές καθώς και τις μέσες τιμές των δύο κατανομών μετά την ολοκλήρωση αυτού του βήματος (στο διάστημα $[0,1]$, με ακρίβεια 2 δεκαδικών ψηφίων)

(5 Points)

Enter your answer

13

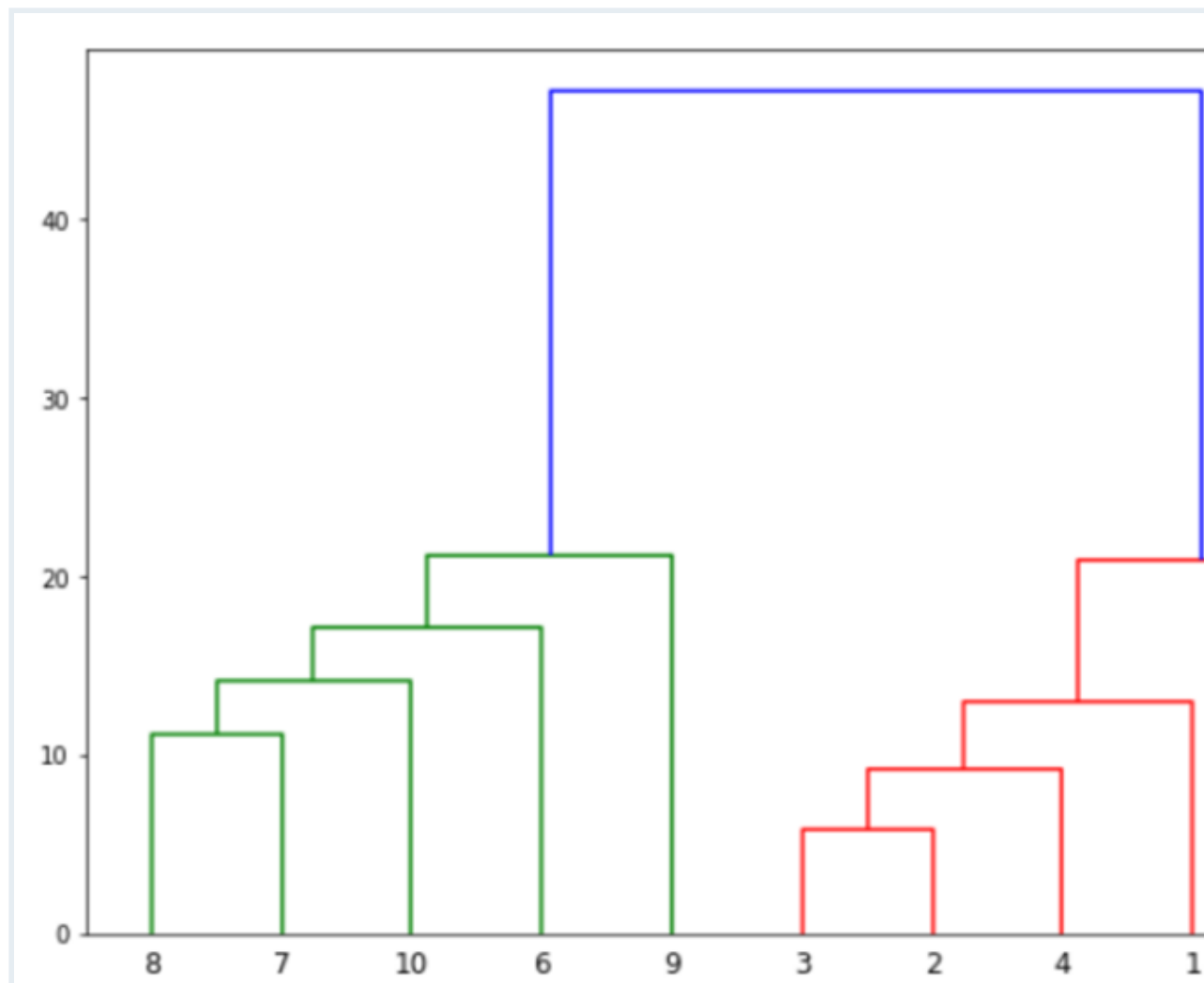
Ποιο ή ποιιά από τα ακόλουθα είναι παραδείγματα εξαγωγής χαρακτηριστικών;
(3 Points)

- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Κατασκευή bag of words vector από ένα email
- ☐ Εφαρμογή PCA σε δεδομένα υψηλής διαστατικότητας
- ☐ Αφαίρεση stopwords σε μια πρόταση

14

Δίνεται το δενδρόγραμμα ενός συσσωρευτικού ιεραρχικού αλγορίθμου συσταδοποίησης (agglomerative hierarchical clustering) δέκα σημείων. Ποια από τις παρακάτω συστάδες σχηματίζεται τελευταία;
(5 Points)

Αν δεν εμφανίζεται η εικόνα, δείτε την στο <https://i.imgur.com/YfoqshN.png>.



- ☐ Δεν απαντώ
- ☐ Η συστάδα των σημείων 3,2,4
- ☐ Η συστάδα των σημείων 8,7
- ☐ Η συστάδα των σημείων 3,2,4,1
- ☐ Η συστάδα των σημείων 8,7,10,6
- ☐ Καμία από τις υπόλοιπες απαντήσεις

15

Έχουμε έξι δείγματα δύο χαρακτηριστικών το καθένα: [4 1], [6 6], [9 5], [1 2], [7 3], [5 4]. Οι ετικέτες τους είναι [1 0 1 0 1 0]. Για την ταξινόμηση φτιάχνουμε ένα δέντρο αποφάσεων βάθους δύο. Οι διχοτομήσεις προκύπτουν με κανόνες που θέτουν ένα χαρακτηριστικό μεγαλύτερο ίσο από ένα κατώφλι. α) ποια είναι η εντροπία της ρίζας του δέντρου; β) ποιος είναι ο κανόνας της δεύτερης διχοτόμησης;
(5 Points)

Enter your answer

Submit

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Powered by Microsoft Forms | [Privacy and cookies](#) | [Terms of use](#)

3

AM *

Αν δεν έχετε/γνωρίζετε τον AM σας βάλτε ένα email σας.

Enter your answer

4

Ιδιότητα *

ΔΠΜΣ τάδε , υ.δ. σχολής τάδε κλπ

Enter your answer

5

964 άνθρωποι ερωτήθηκαν αν πίνουν αλκοόλ. Ο πίνακας συγκεντρώνει τις απαντήσεις τους

Φύλο Ναι Όχι Σύνολο

Άνδρες 152 299 451

Γυναίκες 195 318 513

Σύνολο 347 617 964

Ποιες είναι οι πιθανότητες (odds) για έναν άνδρα να πίνει (ως προς το να μην πίνει) αλκοόλ; Ποιο είναι ο σχετικός κίνδυνος να πίνουν οι γυναίκες σε σχέση με τους άνδρες;

(4 Points)

Enter your answer

6

Έχουμε έξι δείγματα δύο χαρακτηριστικών: [4 1], [6 6], [9 5], [1 2], [7 3], [5 4]. Οι ετικέτες τους είναι [1 0 1 0 1 0]. Για την ταξινόμηση φτιάχνουμε ένα δέντρο βάθους δύο. Οι διχοτομήσεις προκύπτουν με κανόνες που θέτουν ένα χαρακτηριστικό μεγαλύτερο ίσο από ένα κατώφλι.

α) στη ρίζα του δέντρου, ποιοι κανόνες δίνουν μηδενικό κέρδος πληροφορίας; β) ποιος είναι ο κανόνας της πρώτης διχοτόμησης;
(5 Points)

Enter your answer

7

Δίνεται ένα σύνολο κανόνων με 400 θετικά δείγματα και 200 αρνητικά. Για καθένα από τους ακόλουθους κανόνες

R1 : A -> + (καλύπτει 6 θετικά και 9 αρνητικά δείγματα),

R2 : B -> + (καλύπτει 40 θετικά και 70 αρνητικά δείγματα),

R3 : C -> + (καλύπτει 110 θετικά 130 αρνητικά δείγματα),

υπολογίστε α) την ορθότητά του (accuracy) και β) το κέρδος πληροφορίας (information gain) FOIL του.

(4 Points)

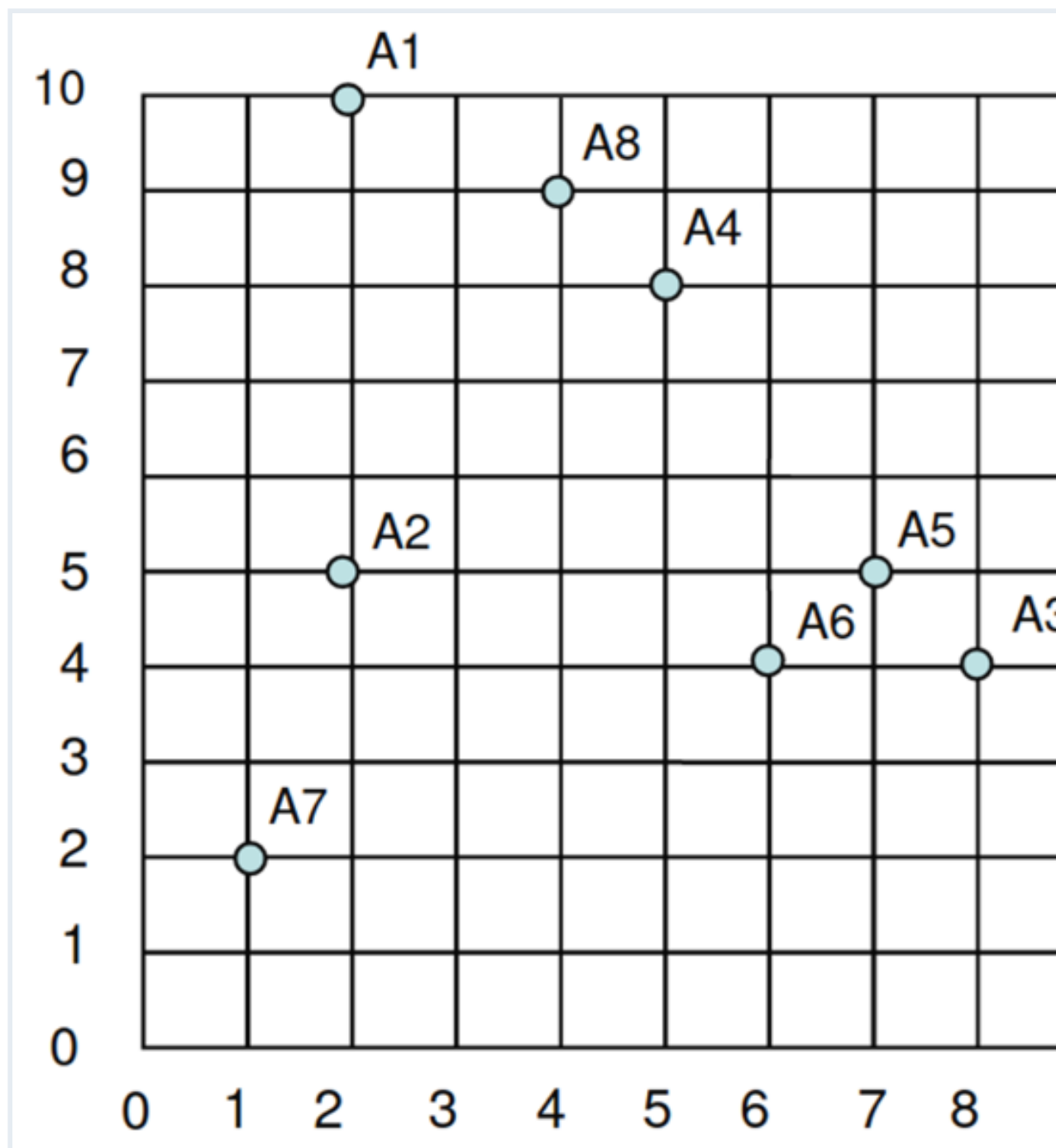
Enter your answer

8

Δίνονται τα 8 σημεία (A1 ως A8) στον δισδιάστατο χώρο, τα οποία τα συσταδοποιούμε με τη χρήση του αλγορίθμου DBSCAN με MinPts=3 και Eps= $\sqrt{5}$. Να γράψετε ποιες συστάδες σχηματίζονται, αν υπάρχουν οριακά σημεία (border points), ποια είναι και σε ποια συστάδα είναι οριακά καθώς και αν υπάρχουν σημεία θορύβου (noise points) και ποια είναι αυτά.

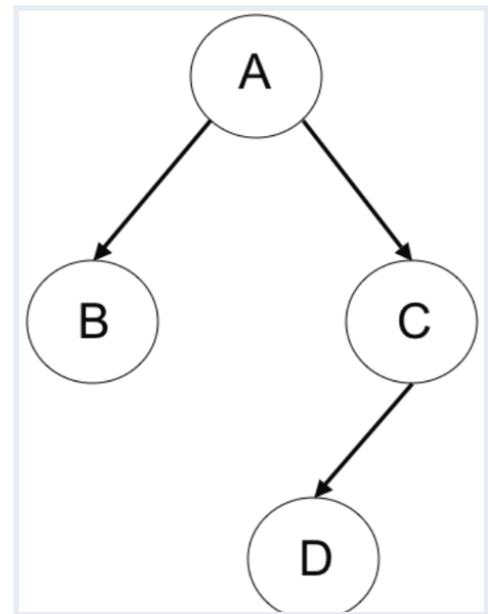
(5 Points)

Αν η εικόνα δεν εμφανίζεται, δείτε την στο <https://imgur.com/b4TgyJy.png>.



Enter your answer

9



Δίνεται το Μπεϋζιανό δίκτυο πεποίθησης του σχήματος και οι εξείς πιθανότητες: $P(A)=0.3$, $P(B|A)=0.7$, $P(B|\sim A)=0.5$, $P(C|A)=0.2$, $P(C|\sim A)=0.6$, $P(D|C)=0.7$, $P(D|\sim C)=0.7$.

Υπολογίστε την πιθανότητα $P(C|B)$

(5 Points)

Αν η εικόνα δεν εμφανίζεται δείτε την στο <https://imgur.com/zag4eDk>

Enter your answer

10

Εκτελούμε τον αλγόριθμο Expectation-Maximization για ένα μοντέλο μίξης δύο γκαουσιανών κατανομών με ίδια βάρη για ένα σύνολο 3 δειγμάτων δεδομένων και σε κάποιο βήμα της επανάληψης έχουμε $p(x_1 = -2|\theta_1) = A$, $p(x_1 = -2|\theta_2) = B$, $p(x_2 = 1|\theta_1) = C$, $p(x_2 = 1|\theta_2) = D$, $p(x_3 = 3|\theta_1) = E$, $p(x_3 = 3|\theta_2) = F$. Έστω $A=0.55$, $B=0.45$, $C=0.9$, $D=0.1$, $E=0.85$, και $F=0.15$. Να υπολογίσετε την πιθανότητα το κάθε δείγμα δεδομένων να προέρχεται από κάθε μια από τις δύο κατανομές καθώς και τις μέσες τιμές των δύο κατανομών μετά την ολοκλήρωση αυτού του βήματος (στο διάστημα $[0,1]$, με ακρίβεια 2 δεκαδικών ψηφίων)

(5 Points)

Enter your answer

11

Έστω ένα δέντρο αποφάσεων. Έχετε δύο επιλογές: α) να ματατρέψετε το δέντρο σε κανόνες και να κλαδέψετε τους κανόνες, β) να κλαδέψετε το δέντρο και στη συνέχεια να μετατρέψετε το δέντρο σε κανόνες. Ποια είναι τα πλεονεκτήματα και μειονεκτήματα της κάθε μεθόδου (εν συντομία);

(4 Points)

Enter your answer

12

Ποιος είναι ο σκοπός της διασταυρούμενης επικύρωσης;

(3 Points)



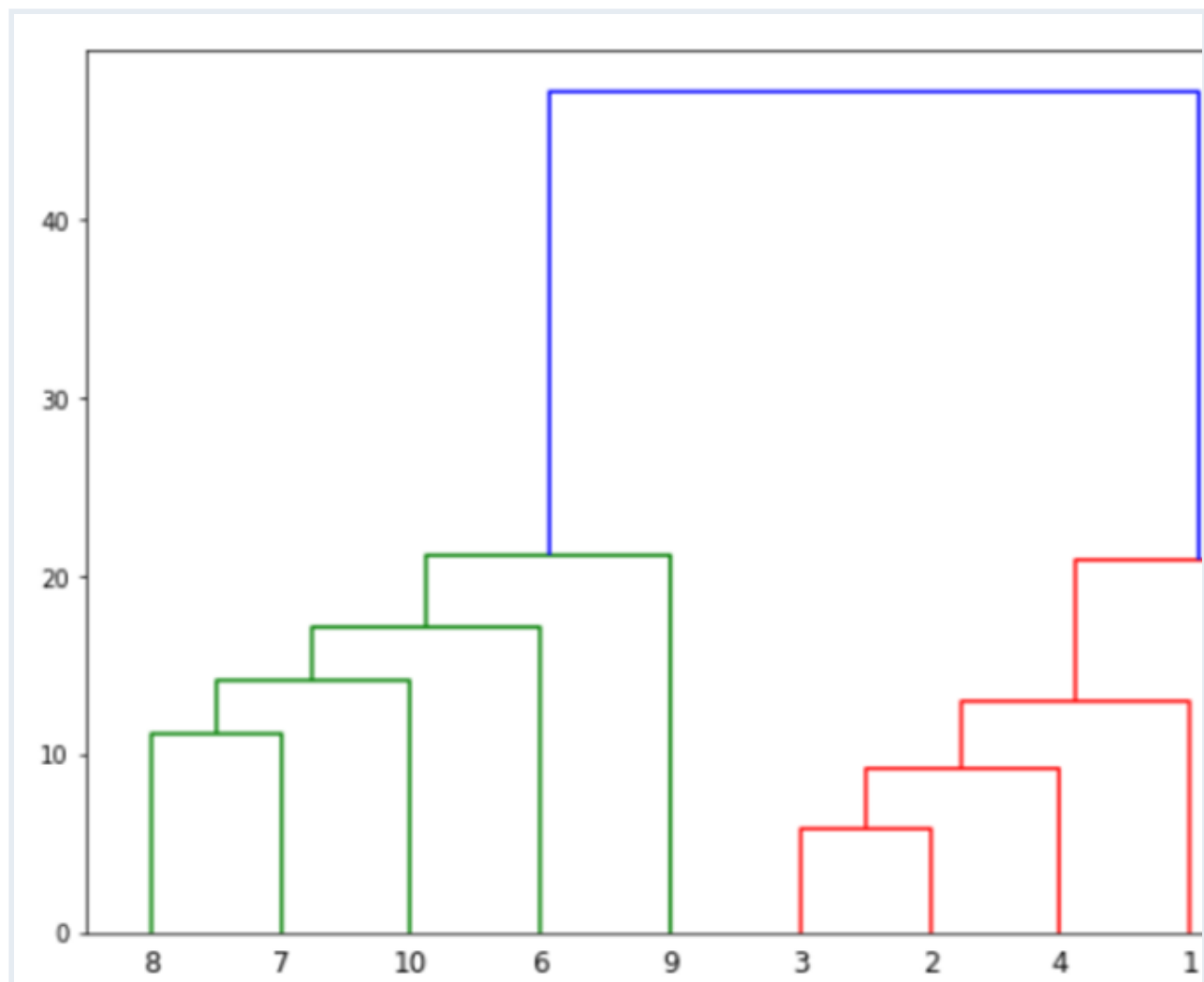
Η εκτίμηση της απόδοσης του μοντέλου εκτός του συνόλου δεδομένων εκπαίδευσης

- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Η εκτίμηση της προγνωστικής απόδοσης των μοντέλων
- ☐ Η εκτίμηση της απόδοσης του μοντέλου στα σύνολα εκπαίδευσης και επικύρωσης

13

Δίνεται το δενδρόγραμμα ενός διαχωριστικού ιεραρχικού αλγορίθμου συσταδοποίησης (divisive hierarchical clustering) δέκα σημείων. Ποια από τις παρακάτω συστάδες σχηματίζεται πρώτη;
(5 Points)

Αν δεν εμφανίζεται η εικόνα, δείτε την στο <https://i.imgur.com/YfoqshN.png>.



- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Η συστάδα των σημείων 8,7,10,6

- ☐ Δεν απαντώ
- ☐ Η συστάδα των σημείων 3,2,4
- ☐ Η συστάδα των σημείων 8,7
- ☐ Η συστάδα των σημείων 3,2,4,1

14

Έχετε ένα πρόβλημα δυαδικής ταξινόμησης. Κάθε δείγμα x_i ανήκει είτε στην κλάση $c=1$ είτε στην κλάση $c=0$. Η διαδικασία συλλογής δεδομένων εκπαίδευσης είναι ωστόσο ατελής και κάποια παραδείγματα δεν λαμβάνουν σωστές ετικέτες. Για κάθε δείγμα x_i αντί να έχουμε την αληθινή ετικέτα του, c_i , έχουμε την πιθανότητα π_i που αντιστοιχεί στο να είναι $c_i=1$. Πώς εκφράζεται η λογαριθμική πιθανότητα (log likelihood) για ένα πιθανοτικό μοντέλο $p(c=1|\theta)$ αν αντί του c_i βάλουμε το π_i ;

(5 Points)

Enter your answer

15

Ένας επενδυτής εξετάζει δύο μετοχές, την A και την B. Οι αναλυτές λένε ότι η πιθανότητα να φέρει κέρδη η A σε ένα τρίμηνο είναι 60% και η B 50%. Εκτιμούν επίσης ότι η πιθανότητα το να φέρουν κέρδη και οι δύο σε ένα τρίμηνο είναι 30%. Ποια είναι η πιθανότητα ο επενδυτής να έχει κέρδη στο τρίμηνο από την A, την B ή και τις δύο μετοχές

(3 Points)

Enter your answer

3

AM *

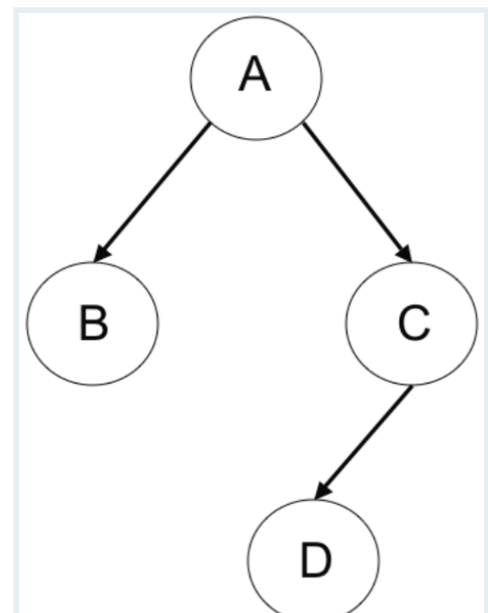
Αν δεν έχετε/γνωρίζετε τον AM σας βάλτε ένα email σας.

4

Ιδιότητα *

ΔΠΜΣ τάδε , υ.δ. σχολής τάδε κλπ

5



Δίνεται το Μπεϋζιανό δίκτυο πεποίθησης του σχήματος και οι εξείς πιθανότητες: $P(A)=0.3$, $P(B|A)=0.7$, $P(B|\sim A)=0.5$, $P(C|A)=0.2$, $P(C|\sim A)=0.6$, $P(D|C)=0.7$, $P(D|\sim C)=0.7$.

Υπολογίστε την πιθανότητα $P(A|B)$

(5 Points)

Αν η εικόνα δεν εμφανίζεται δείτε την στο <https://imgur.com/zag4eDk>

Enter your answer

6

Σε εκλογές, N υποψήφιοι ανταγωνίζονται ο ένας τον άλλον. Οι ψηφοφόροι ψηφίζουν για έναν από τους υποψηφίους και δεν επικοινωνούν μεταξύ τους την ώρα που ψηφίζουν. Ποια από τις παρακάτω μεθόδους συνόλων λειτουργεί παρόμοια με αυτή την εκλογική διαδικασία;

(3 Points)

- ☐ Bagging και Boosting
- ☐ Boosting
- ☐ Δεν απαντώ
- ☐ Καμία από τις υπόλοιπες απαντήσεις
- ☐ Bagging

7

Δίνεται ένα σύνολο κανόνων με 200 θετικά δείγματα και 300 αρνητικά. Για καθένα από τους ακόλουθους κανόνες

R1 : A \rightarrow + (καλύπτει 5 θετικά και 2 αρνητικά δείγματα),

R2 : B \rightarrow + (καλύπτει 40 θετικά και 20 αρνητικά δείγματα),

R3 : C \rightarrow + (καλύπτει 80 θετικά 80 αρνητικά δείγματα),

υπολογίστε α) την ορθότητά του (accuracy) και β) το κέρδος πληροφορίας (information gain) FOIL του.

(4 Points)

Enter your answer

8

Εκτελούμε τον αλγόριθμο Expectation-Maximization για ένα μοντέλο μίξης δύο γκαουσιανών κατανομών με ίδια βάρη για ένα σύνολο 3 δειγμάτων δεδομένων και σε κάποιο βήμα της επανάληψης έχουμε $p(x_1 = -2|\theta_1) = A$, $p(x_1 = -2|\theta_2) = B$, $p(x_2 = 1|\theta_1) = C$, $p(x_2 = 1|\theta_2) = D$, $p(x_3 = 3|\theta_1) = E$, $p(x_3 = 3|\theta_2) = F$. Έστω $A = 0.3$, $B = 0.7$, $C = 0.45$, $D = 0.55$, $E = 0.2$, και $F = 0.8$. Να υπολογίσετε την πιθανότητα το κάθε δείγμα δεδομένων να προέρχεται από κάθε μια από τις δύο κατανομές καθώς και τις μέσες τιμές των δύο κατανομών μετά την ολοκλήρωση αυτού του βήματος (στο διάστημα $[0,1]$, με ακρίβεια 2 δεκαδικών ψηφίων)

(5 Points)

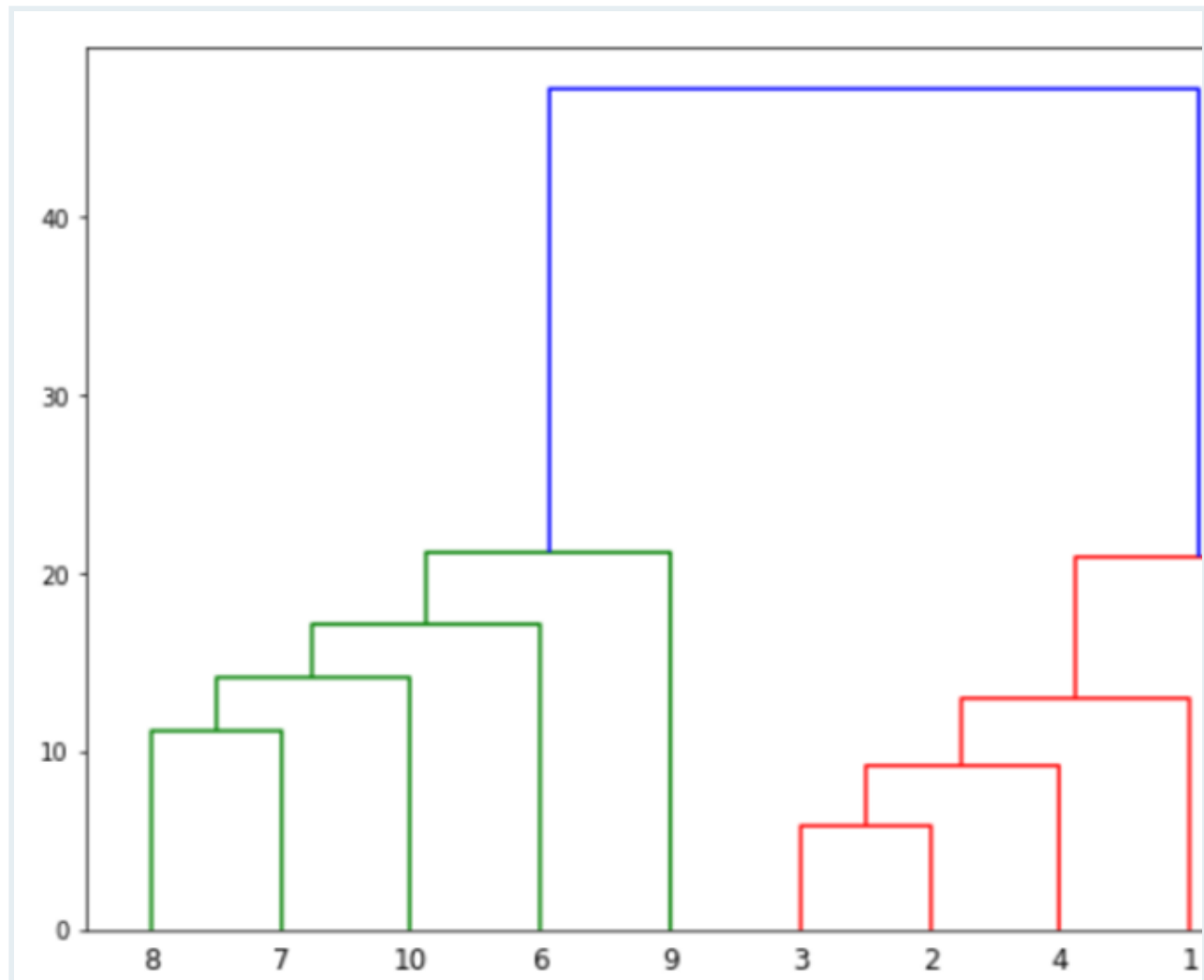
Enter your answer

9

Δίνεται το δενδρόγραμμα ενός συσσωρευτικού ιεραρχικού αλγορίθμου συσταδοποίησης (agglomerative hierarchical clustering) δέκα σημείων. Ποια από τις παρακάτω συστάδες σχηματίζεται τελευταία;

(5 Points)

Αν δεν εμφανίζεται η εικόνα, δείτε την στο <https://i.imgur.com/YfoqshN.png>.



- ☐ Η συστάδα των σημείων 3,2,4
- ☐ Δεν απαντώ
- ☐ Η συστάδα των σημείων 8,7,10,6
- ☐ Η συστάδα των σημείων 3,2,4,1
- ☐ Καμία από τις υπόλοιπες απαντήσεις

☐ Η συστάδα των σημείων 8,7

10

Έχουμε έξι δείγματα δύο χαρακτηριστικών: [4 1], [6 6], [9 5], [1 2], [7 3], [5 4]. Οι ετικέτες τους είναι [1 0 1 0 1 0]. Για την ταξινόμηση φτιάχνουμε ένα δέντρο βάθους δύο. Οι διχοτομήσεις προκύπτουν με κανόνες που θέτουν ένα χαρακτηριστικό μεγαλύτερο ίσο από ένα κατώφλι.

α) στη ρίζα του δέντρου, ποιοι κανόνες δίνουν μηδενικό κέρδος πληροφορίας; β) ποιος είναι ο κανόνας της πρώτης διχοτόμησης;

(5 Points)

Enter your answer

11

Ένας επενδυτής εξετάζει δύο μετοχές, την A και την B. Οι αναλυτές λένε ότι η πιθανότητα να φέρει κέρδη η A σε ένα τρίμηνο είναι 60% και η B 50%.

Εκτιμούν επίσης ότι η πιθανότητα το να φέρουν κέρδη και οι δύο σε ένα τρίμηνο είναι 30%. Ποια είναι η πιθανότητα ο επενδυτής να έχει κέρδη στο τρίμηνο από την A, την B ή και τις δύο μετοχές

(3 Points)

Enter your answer

12

Έστω ένα δέντρο αποφάσεων. Έχετε δύο επιλογές: α) να ματατρέψετε το δέντρο σε κανόνες και να κλαδέψετε τους κανόνες, β) να κλαδέψετε το δέντρο και στη συνέχεια να μετατρέψετε το δέντρο σε κανόνες. Ποια είναι τα πλεονεκτήματα και μειονεκτήματα της κάθε μεθόδου (εν συντομία);

(4 Points)

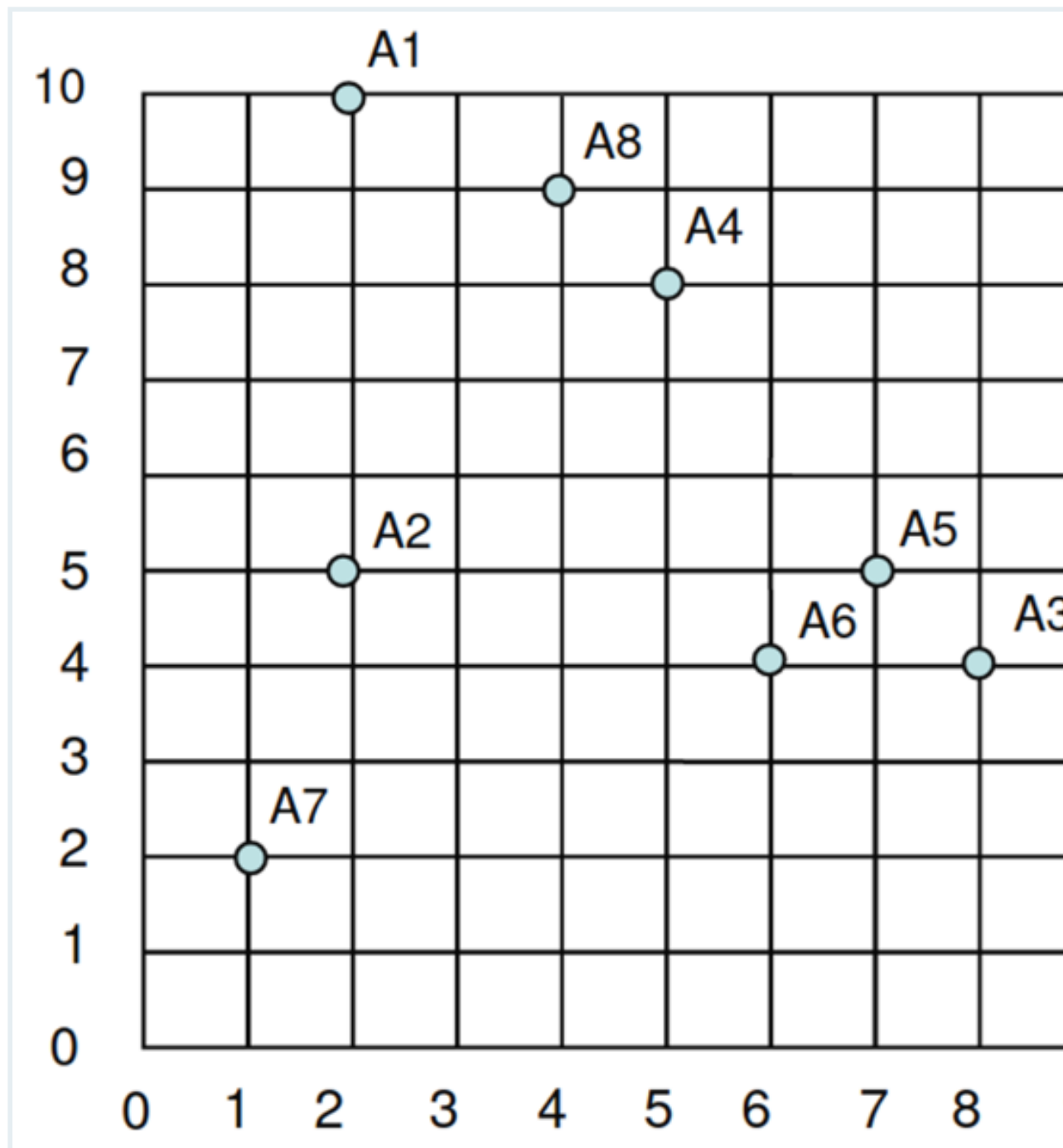
Enter your answer

13

Δίνονται τα 8 σημεία (A1 ως A8) στον δισδιάστατο χώρο, τα οποία τα συσταδοποιούμε με τη χρήση του αλγορίθμου DBSCAN με $\text{MinPts}=3$ και $\text{Eps}=\sqrt{5}$. Να γράψετε ποιες συστάδες σχηματίζονται, αν υπάρχουν οριακά σημεία (border points), ποια είναι και σε ποια συστάδα είναι οριακά καθώς και αν υπάρχουν σημεία θορύβου (noise points) και ποια είναι αυτά.

(5 Points)

Αν η εικόνα δεν εμφανίζεται, δείτε την στο <https://imgur.com/b4TgyJy.png>.



Enter your answer

14

Έστω ένα μοντέλο λογιστικής παλινδρόμησης και ένα γραμμικά διαχωρίσιμο πρόβλημα. Αν έχουμε μια απεικόνιση $g(x)$ των χαρακτηριστικών x , και βάρη w , εξηγήστε ποιοτικά πώς το μοντέλο μέγιστης πιθανότητας προκύπτει αν βρούμε την ευθεία με $w^T g(x) = 0$ και στη συνέχεια θέσουμε τη νόρμα του w στο άπειρο.

(5 Points)

Enter your answer

15

964 άνθρωποι ερωτήθηκαν αν πίνουν αλκοόλ. Ο πίνακας συγκεντρώνει τις απαντήσεις τους

Φύλο Ναι Όχι Σύνολο

Άνδρες 152 299 451

Γυναίκες 195 318 513

Σύνολό 347 617 964

Ποιες είναι οι πιθανότητες (odds) για μια γυναίκα να πίνει (ως προς το να μην πίνει) αλκοόλ; Ποιο είναι ο σχετικός κίνδυνος να πίνουν οι άνδρες σε σχέση με τις γυναίκες;

(4 Points)

Enter your answer

Submit