

Ερώτηση 1 (1 Μονάδα)

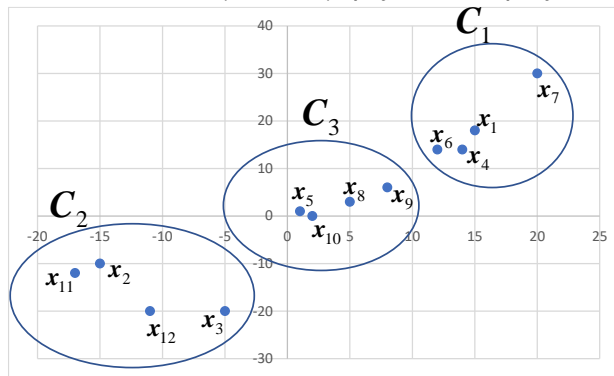
Στον διπλανό πίνακα δίνονται έξι δυσδιάστατα διανυσματικά δεδομένα. Ποιο από τα παρακάτω δυσδιάστατα διανύσματα είναι ο μέσος όρος των παραπάνω έξι διανυσμάτων;

- (1) [3 3]
- (2) [5 6]
- (3) [-3 -2]
- (4) [10 7]

5	6
2	5
10	2
5	5
3	7
9	11
1	12

Ερώτηση 2 (2 Μονάδες)

Στο παρακάτω σχήμα 12 δυσδιάστατα διανύσματα διαμερίζονται σε τρεις συστάδες.



Ποιος από τους παρακάτω πίνακες είναι ο πίνακας συναρτήσεων συμμετοχής του παραπάνω διαμερισμού;

	(1)	(2)	(3)	(4)
	C_1 C_2 C_3	C_1 C_2 C_3	C_1 C_2 C_3	C_1 C_2 C_3
x_1	1 0 0	1 0 0	1 0 0	0 0 1
x_2	0 1 0	0 1 0	0 1 0	1 0 0
x_3	0 1 0	0 1 0	0 1 0	0 1 0
x_4	1 0 0	0 0 1	1 0 0	0 1 0
x_5	0 0 1	0 0 1	0 0 1	0 0 1
x_6	1 0 0	1 0 0	1 1 0	0 0 1
x_7	1 0 0	1 0 0	1 0 0	1 0 0
x_8	0 0 1	0 1 0	0 0 1	0 0 1
x_9	0 0 1	0 1 0	1 0 1	0 0 1
x_{10}	0 0 1	0 0 1	0 0 1	1 0 0
x_{11}	0 1 0	0 1 0	0 1 0	1 0 0
x_{12}	0 1 0	0 1 0	0 1 1	0 1 0

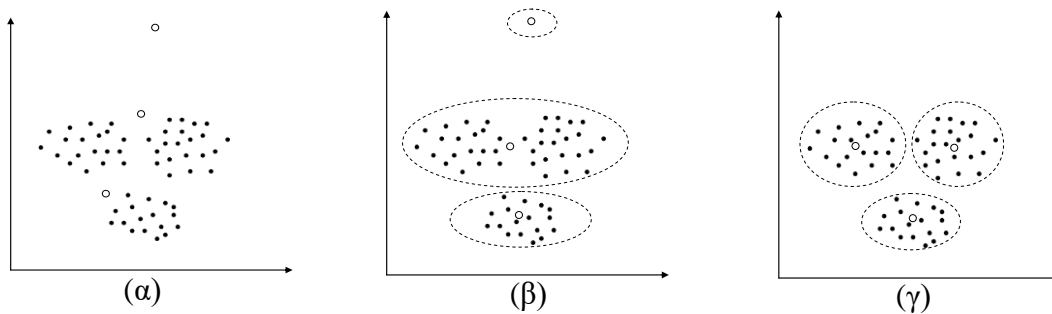
Ερώτηση 3 (0.5 Μονάδα)

Βασικό βήμα κατά την προεπεξεργασία δεδομένων είναι:

- (1) Η δημιουργία κατάλληλης υπολογιστικής εφαρμογής
- (2) Η δημιουργία κατάλληλης βάσης δεδομένων με καλά ορισμένο σχήμα βάσης
- (3) Η δημιουργία χρονοδιαγράμματος υλοποίησης της όλης επεξεργασίας
- (4) Η δημιουργία εξειδικευμένων αλγόριθμων που θα εφαρμοστούν σε τυχαίο δείγμα των δεδομένων

Ερώτηση 4 (1.5 Μονάδα)

Στο παρακάτω σχήμα (α) δίνεται ένα σύνολο δεδομένων στον δυσδιάστατο χώρο και οι **αρχικές θέσεις** των κέντρων τριών συστάδων. Τα δεδομένα συμβολίζονται με τα συμπαγή σημεία, ενώ τα τρία κέντρα με τους κύκλους. Στα σχήματα (β) και (γ) φαίνονται δύο πιθανές τελικές διαμερίσεις σε τρεις ομάδες.



Κατά την γνώμη σας αν εφαρμόσουμε τον αλγόριθμο c-means στην τοπολογία του σχήματος (α), τότε θα πάρουμε ως αποτέλεσμα:

- (1) Το σχήμα (β)
- (2) Το σχήμα (γ)
- (3) Κανένα από τα δύο
- (4) Δεν ξέρω/δεν απαντώ

Ερώτηση 5 (0.5 Μονάδα)

Το βασικότερο πλεονέκτημα του αλγόριθμου c-means είναι:

- (1) Η απλότητα του αλγοριθμικού σχεδιασμού
- (2) Η εύκολη ενσωμάτωσή του σε άλλα αλγοριθμικά σχήματα
- (3) Το μικρό υπολογιστικό κόστος
- (4) Η σαφής εννοιολογική του υπόσταση

Ερώτηση 6 (1 Μονάδα)

Η πρώτη (αρχική) έκδοση του αλγόριθμου c-means είχε ως βασική παραδοχή ότι:

- (1) Οι ομάδες/συστάδες έχουν συγκεκριμένη αλληλοκάλυψη
- (2) Ο αριθμός των ομάδων/συστάδων μπορεί να προσαρμόζεται από μία επανάληψη στην επόμενη
- (3) Ο αριθμός των ομάδων/συστάδων είναι σταθερός και ορίζεται στην αρχή της επαναληπτικής διαδικασίας
- (4) Ο αριθμός των ομάδων/συστάδων είναι άγνωστος

Ερώτηση 7 (0.5 Μονάδα)

Η ανάπτυξη ενός συστήματος εξόρυξης δεδομένων βασίζεται

- (1) Στην συλλογή δεδομένων και στην επεξεργασία τους με αλγόριθμους επιβλεπόμενης, μη-επιβλεπόμενης, και/ή ημί-επιβλεπόμενης μάθησης
- (2) Στην συλλογή δεδομένων και στην επεξεργασία τους με μαθηματικά μοντέλα
- (3) Στην συλλογή δεδομένων και στην επεξεργασία τους με συλλογιστική λογική
- (4) Στην συλλογή δεδομένων και στην επεξεργασία τους με εξελικτικούς αλγόριθμους

Ερώτηση 8 (1 Μονάδα)

Ένα από μειονεκτήματα του αλγόριθμου c-means είναι ότι:

- (1) Η προεπεξεργασία των δεδομένων είναι πολύ μεγάλη σε υπολογιστικό κόστος
- (2) Έχει πολύ μεγάλο υπολογιστικό κόστος
- (3) Είναι πολύ ευαίσθητος στην επιλογή των αρχικών τιμών των κέντρων των συστάδων
- (4) Μπορεί να χρησιμοποιηθεί μόνο σε μικρό αριθμό εφαρμογών

Ερώτηση 9 (1 Μονάδα)

Μία βασική εφαρμογή του αλγόριθμου c-means είναι:

- (1) Η δημιουργία τρισδιάστατων γραφικών
- (2) Η συμπίεση σήματος και δεδομένων
- (3) Η δημιουργία ψηφιακών εικόνων μεγάλης ανάλυσης
- (4) Η σάρωση ψηφιακών εικόνων

Ερώτηση 10 (1 Μονάδα)

Μία στρατηγική για την βελτίωση της απόδοσης του αλγόριθμου c-means είναι:

- (1) Η επιλογή τυχαίων αρχικών τιμών για τα κέντρα των συστάδων
- (2) Οι αρχικές τιμές των κέντρων των συστάδων να είναι όσο γίνεται πιο μακριά η μία από την άλλη
- (3) Οι αρχικές τιμές των κέντρων των συστάδων να είναι όσο γίνεται πιο κοντά η μία από την άλλη
- (4) Η επιλογή μηδενικών αρχικών τιμών για τα κέντρα των συστάδων