Αλγοριθμική Επιστήμη Δεδομένων

1η Σειρά Ασκήσεων

Στοιχεία φοιτητή

Ονοματεπώνυμο: Κωνσταντίνος Τσόπελας

Αριθμός Μητρώου: 03400198

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Εθνικό Μετσόβιο Πολυτεχνείο

ΔΠΜΣ ΕΔΕΜΜ

3 Απριλίου 2024

MMDS, Exercise 6.3.1

Ερώτημα (a)

Baskets:

1. {1, 2, 3}
2. {2, 3, 4}
3. {3, 4, 5}
4. {4, 5, 6}
5. {1, 3, 5}
6. {2, 4, 6}
7. {1, 3, 4}
8. {2, 4, 5}
9. {3, 5, 6}
10. {1, 2, 4}
11. {2, 3, 5}
12. {3, 4, 6}

Μετράμε τα καλάθια στα οποία ανήκει το κάθε ένα από τα 6 αντικείμενα:

- Item 1: baskets 1, 5, 7, 10
- Item 2: baskets 1, 2, 6, 8, 10, 11
- Item 3: baskets 1, 2, 3, 5, 7, 9, 11, 12
- Item 4: baskets 2, 3, 4, 6, 7, 8, 10, 12
- Item 5: baskets 3, 4, 5, 8, 9, 11
- Item 6: baskets 4, 6, 9, 12

Και άρα:

- support$(\{1\}) = 4$
- support$(\{2\}) = 6$
- support$(\{3\}) = 8$
- support$(\{4\}) = 8$
- support$(\{5\}) = 6$
- support$(\{6\}) = 4$

Για τα support των δισυνόλων, θα εφαρμόσουμε την ίδια βασική λογική, απλά αντί να ξεκινήσουμε από τα καλάθια τα ίδια, θα ξεκινήσουμε από τις απαριθμήσεις όλων των ζευγών αντικειμένων που υπάρχουν σε κάθε καλάθι:

1. $\{1, 2, 3\} \rightarrow \{ \{1, 2\}, \{1, 3\}, \{2, 3\} \}$
2. $\{2, 3, 4\} \rightarrow \{ \{2, 3\}, \{2, 4\}, \{3, 4\} \}$
3. $\{3, 4, 5\} \rightarrow \{ \{3, 4\}, \{3, 5\}, \{4, 5\} \}$
4. $\{4, 5, 6\} \rightarrow \{ \{4, 5\}, \{4, 6\}, \{5, 6\} \}$
5. $\{1, 3, 5\} \rightarrow \{ \{1, 3\}, \{1, 5\}, \{3, 5\} \}$
6. $\{2, 4, 6\} \rightarrow \{ \{2, 4\}, \{2, 6\}, \{4, 6\} \}$
7. $\{1, 3, 4\} \rightarrow \{ \{1, 3\}, \{1, 4\}, \{3, 4\} \}$
8. $\{2, 4, 5\} \rightarrow \{ \{2, 4\}, \{2, 5\}, \{4, 5\} \}$
9. $\{3, 5, 6\} \rightarrow \{ \{3, 5\}, \{3, 6\}, \{5, 6\} \}$
10. $\{1, 2, 4\} \rightarrow \{ \{1, 2\}, \{1, 4\}, \{2, 4\} \}$
11. $\{2, 3, 5\} \rightarrow \{ \{2, 3\}, \{2, 5\}, \{3, 5\} \}$
12. $\{3, 4, 6\} \rightarrow \{ \{3, 4\}, \{3, 6\}, \{4, 6\} \}$

Συγκεντρώνοντας και πάλι τα καλάθια στα οποία ανήκει το κάθε δισύνολο:

- Pair $\{1, 2\}$: baskets 1, 10
- Pair $\{1, 3\}$: baskets 1, 5, 7
- Pair $\{2, 3\}$: baskets 1, 2, 11
- Pair $\{2, 4\}$: baskets 2, 6, 8, 10
- Pair $\{3, 4\}$: baskets 2, 3, 7, 12
- Pair $\{3, 5\}$: baskets 3, 5, 9, 11
- Pair $\{4, 5\}$: baskets 3, 4, 8
- Pair $\{4, 6\}$: baskets 4, 6, 12
- Pair $\{5, 6\}$: baskets 4, 9
- Pair $\{1, 5\}$: baskets 5
- Pair $\{2, 6\}$: baskets 6
- Pair $\{1, 4\}$: baskets 7, 10
- Pair $\{2, 5\}$: baskets 8, 11
- Pair $\{3, 6\}$: baskets 9, 12

Και άρα:

- support($\{1,2\}$) $= 2$
- support($\{1,3\}$) $= 3$
- support($\{2,3\}$) $= 3$
- support($\{2,4\}$) $= 4$
- support($\{3,4\}$) $= 4$
- support($\{3,5\}$) $= 4$
- support($\{4,5\}$) $= 3$
- support($\{4,6\}$) $= 3$
- support($\{5,6\}$) $= 2$
- support($\{1,5\}$) $= 1$
- support($\{2,6\}$) $= 1$
- support($\{1,4\}$) $= 2$
- support($\{2,5\}$) $= 2$
- support($\{3,6\}$) $= 2$

Τα υπόλοιπα ζεύγη δεν εμφανίζονται σε κανένα καλάθι, και άρα το support τους είναι 0.

Ερώτημα (b)

Ασχολούμαστε, φυσικά, μόνο με τα ζευγάρια που εμφανίζονται στην παραπάνω λίστα, δηλαδή που έχουν support θετικό (διότι αυτά είναι και τα μόνα ζευγάρια που θα δει ο PCY, αφού ο τρόπος που μετράει ζευγάρια είναι με διπλή επανάληψη στα αντικείμενα κάθε καλαθιού).

Όπως μας λέει η εκφώνηση, έχουμε 11 buckets και η hash function μας είναι η $h(\{i,j\}) = i \cdot j \mod 11$. Κατόπιν τούτου, μπορούμε να υπολογίσουμε τα hashes των ζευγαριών:

- $h(\{1,2\}) = 2 \mod 11 = 2$
- $h(\{1,3\}) = 3 \mod 11 = 3$
- $h(\{2,3\}) = 6 \mod 11 = 6$
- $h(\{2,4\}) = 8 \mod 11 = 8$
- $h(\{3,4\}) = 12 \mod 11 = 1$
- $h(\{3,5\}) = 15 \mod 11 = 4$

- $h(\{4,5\}) = 20 \mod 11 = 9$
- $h(\{4,6\}) = 24 \mod 11 = 2$
- $h(\{5,6\}) = 30 \mod 11 = 8$
- $h(\{1,5\}) = 5 \mod 11 = 5$
- $h(\{2,6\}) = 12 \mod 11 = 1$
- $h(\{1,4\}) = 4 \mod 11 = 4$
- $h(\{2,5\}) = 10 \mod 11 = 10$
- $h(\{3,6\}) = 18 \mod 11 = 7$

Συγκεντρώνοντας, για καλύτερη εποπτεία, τα ζευγάρια που χασάρονται στο κάθε bucket:

- Bucket 1: { {3, 4}, {2, 6} }
- Bucket 2: { {1, 2}, {4, 6} }
- Bucket 3: { {1, 3} }
- Bucket 4: { {3, 5}, {1, 4} }
- Bucket 5: { {1, 5} }
- Bucket 6: { {2, 3} }
- Bucket 7: { {3, 6} }
- Bucket 8: { {2, 4}, {5, 6} }
- Bucket 9: { {4, 5} }
- Bucket 10: { {2, 5} }
- Bucket 11: { }

Ερώτημα (c)

Για να δούμε ποια buckets είναι frequent, υπολογίζουμε το άθροισμα των support των itemsets που αντιστοιχούν σε κάθε bucket. Φυσικά, σημειώνουμε ότι ο αλγόριθμος δεν κρατάει όλη αυτή την πληροφορία, αλλά απλά απαριθμεί ζευγάρια, χασάρει και αυξάνει απευθείας το αντίστοιχο bucket. Απλά εδώ το κάνουμε επειδή είναι πιο καθαρό και explicit.

- Bucket 1: support($\{3,4\}$) + support($\{2,6\}$) $= 4 + 1 = 5$
- Bucket 2: support($\{1,2\}$) + support($\{4,6\}$) $= 2 + 3 = 5$
- Bucket 3: support($\{1,3\}$) $= 3$

- Bucket 4: support($\{3,5\}$) + support($\{1,4\}$) $= 4+2 = 6$
- Bucket 5: support($\{1,5\}$) $= 1$
- Bucket 6: support($\{2,3\}$) $= 3$
- Bucket 7: support($\{3,6\}$) $= 2$
- Bucket 8: support($\{2,4\}$) + support($\{5,6\}$) $= 4+2 = 6$
- Bucket 9: support($\{4,5\}$) $= 3$
- Bucket 10: support($\{2,5\}$) $= 2$
- Bucket 11: 0

Αφού, λοιπόν, έχουμε support threshold 4, frequent buckets θα είναι όσοι έχουν άθροισμα τουλάχιστον 4, δηλαδή οι 1, 2, 4, 8.

Ερώτημα (d)

Τα υποψήφια ζεύγη C_2, για τα οποία ο αλγόριθμος PCY θα πάει μετά να υπολογίσει τα support τους για να ελέγξει αν είναι όντως frequent, είναι αυτά που ικανοποιούν τις 2 συνθήκες:

1. και τα δύο αντικείμενά του ζεύγους είναι frequent.
2. το ζεύγος χασάρεται σε frequent bucket.

Στην συγκεκριμένη περίπτωση, όλα τα μονοσύνολα έχουν support ≥ 4, άρα η πρώτη συνθήκη ικανοποιείται για κάθε ζεύγος και αρκεί να ελέγξουμε τη 2η.

Δεδομένου, λοιπόν, ότι οι frequent buckets είναι οι 1, 2, 4, 8, τα υποψήφια ζεύγη C_2 θα είναι ακριβώς αυτά που χασάρονται στους συγκεκριμένους buckets, δηλαδή τα { {3, 4}, {2, 6}, {1, 2}, {4, 6}, {3, 5}, {1, 4}, {2, 4}, {5, 6} }.

MMDS, Exercise 6.4.1

Καταρχάς, τα αντικείμενα G και H δεν ανήκουν σε maximal frequent itemset, άρα δεν είναι frequent, και άρα τα μονοσύνολα $\{G\}, \{H\}$ είναι στο negative border. Επίσης, έπεται ότι κανένα άλλο itemset του negative border δεν θα περιέχει τα G, H, αφού τότε θα ήταν υπερσύνολο άλλου στοιχείου του negative border, που δεν γίνεται.

Όλα τα άλλα αντικείμενα είναι frequent μόνα τους, γιατί ανήκουν σε κάποιο από τα δοσμένα maximal frequent itemsets. Οπότε, από μονοσύνολα τελειώσαμε.

Τα δισύνολα που ανήκουν στο negative border είναι ακριβώς αυτά δεν είναι frequent, αλλά και τα δύο στοιχεία τους από μόνα τους είναι frequent. Δεδομένου ότι τα frequent μονοσύνολα είναι τα A, B, C, D, E, F, μπορούμε να δούμε ότι από όλους τους συνδυασμούς τους, αυτοί που δεν είναι frequent, και άρα είναι στο negative border, είναι τα εξής:

- {A, E}
- {A, F}

- {B, D}
- {B, E}
- {B, F}
- {C, D}
- {C, E}
- {C, F}
- {D, E}
- {D, F}
- {E, F}

Ολοκληρώνουμε με τα itemsets 3 στοιχείων, μια και κανένα δεν είναι frequent, και άρα αποκλείεται κάποιο itemset 4 στοιχείων να είναι στο negative border.

Για τα 3 στοιχείων, πρέπει και τα 3 υποσύνολά τους με 2 στοιχεία να είναι frequent. Δεδομένου ότι τα μόνα frequent 2 στοιχείων είναι τα {A, B}, {B, C}, {A, C}, {A, D}, βλέπουμε ότι τα μόνα που "ταιριάζουν" και σχηματίζουν ένα itemset 3 στοιχείων είναι τα 3 πρώτα, και άρα το τελευταίο itemset του negative border είναι το {A, B, C}.

MMDS, Exercise 5.2.2

(a) Χρησιμοποιώντας την αναπαράσταση της ενότητας 5.2.1 του βιβλίου, ο πίνακας μετάβασης για το γράφο 5.4 θα είναι ο Πίνακας 1. Ας σημειωθεί ότι ο κόμβος E αποτελεί dead end, για αυτό και δεν του αντιστοιχεί γραμμή στον πίνακα.

Source	Degree	Destinations
A	3	B, C, D
B	2	A, D
C	1	E
D	2	B, C

Πίνακας 1: Πίνακας Μετάβασης για γράφο 5.4

(b) Ομοίως, για το γράφο της εικόνας 5.7, ο πίνακας μετάβασης είναι ο Πίνακας 2

Source	Degree	Destinations
a	3	a, b, c
b	2	a, c
c	2	b, c

Πίνακας 2: Πίνακας Μετάβασης για γράφο 5.7

MMDS, Exercise 5.2.3

Έχουμε έναν γράφο 4 κόμβων, άρα ο πίνακας μετάβασής του είναι 4×4. Άρα, για πλευρά block 2, θα έχουμε ακριβώς 4 blocks, για μεταβάσεις μεταξύ των ομάδων κόμβων {A, B} και {C, D}.

Μάλιστα, η αναπαράσταση εδώ θα μοιάζει πολύ με αυτήν της εικόνας 5.14 του βιβλίου. Η μόνη διαφορά στο γράφημα είναι ότι εδώ δεν υπάρχει ακμή από το C στο A. Οπότε, εκτός από το block C, D προς C, D, δεν θα υπάρχει γραμμή για τον C ούτε στο block C, D προς A, B.

Τα blocks παρουσιάζονται στον Πίνακα 3.

Source	Degree	Destinations
A	3	B
B	2	A

(a) A, B προς A, B

Source	Degree	Destinations
D	2	B

(b) C, D προς A, B

Source	Degree	Destinations
A	3	C, D
B	2	D

(c) A, B προς C, D

Source	Degree	Destinations
D	2	C

(d) C, D προς C, D

Πίνακας 3: Πίνακας Μετάβασης σε blocks για γράφο 5.3

MMDS, Exercise 5.2.4

Καταρχάς, είναι εμφανές ότι στο γράφημα της εικόνας 5.9, όλοι οι κόμβοι θα έχουν βαθμό 1, με μοναδική εξαίρεση τον πρώτο (με βαθμό 2) και τον τελευταίο (με βαθμό 0).

Από εκεί και πέρα, με πλευρά block k, αυτό που θα συμβεί, ουσιαστικά, είναι ότι χωρίζουμε τους κόμβους μας σε ομάδες μεγέθους k. Υποθέτοντας ότι οι κόμβοι διατρέχονται με την σειρά, από αριστερά προς τα δεξιά, οι ομάδες αυτές θα είναι οι κόμβοι υπ' αριθμόν 0 έως k-1, μετά οι k έως $2k - 1$, κ.ο.κ.

Κατόπιν τούτου, όπως έχουμε δει στη θεωρία και περιγράφει και το βιβλίο, θα δημιουργηθούν, καταρχήν, ακριβώς $(\frac{n}{k})^2$ blocks στον πίνακα μετάβασης, όπου το block M_{ij} θα έχει όλες τις μεταβάσεις από το block i προς το block j.

Στην περίπτωσή μας, παρατηρούμε ότι κάθε ομαδούλα από συνεχόμενους κόμβους έχει, προφανώς, συνδέσεις προς τους ίδιους τους κόμβους της ομάδας (ο καθένας προς τον επόμενο), καθώς και μία σύνδεση προς τους κόμβους της επόμενης ομάδας. Δεν υπάρχει καμία άλλη σύνδεση.

Οπότε, τελικά, συνειδητοποιούμε ότι θα έχουμε ακριβώς $2 * \frac{n}{k} - 1$ μη κενά blocks (τα υπόλοιπα δεν έχουν καθόλου συνδέσεις, οπότε δεν χρειάζεται και να τα κρατάμε). Συγκεκριμένα, αδρά, για κάθε i από 0 έως $\frac{n}{k} - 1$, θα έχουμε ακριβώς 2 blocks:

1. Ένα block με $k - 1$ γραμμές, όπου απλά ο κάθε κόμβος της ομάδας i θα έχει ακμή προς τον επόμενό του (εκτός του τελευταίου της ομάδας και του πρώτου πρώτου κόμβου που έχει δύο ακμές).

2. Ένα block με 1 μόνο γραμμή, που θα αναπαριστά την ακμή του τελευταίου κόμβου της ομάδας i προς τον πρώτο κόμβο της ομάδας $i+1$.

Μοναδική εξαίρεση είναι η τελευταία ομάδα, που δεν έχει επόμενη, και άρα δεν έχει το 2ο τύπο block, εξ ου και το -1 στον συνολικό αριθμό των blocks.

MMDS, Exercise 5.3.1

Ας ξεκινήσουμε γράφοντας τον πίνακα μετάβασης του γραφήματος της εικόνας 5.15. Αυτός είναι ο εξής:

$$M = \begin{pmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

Θεωρούμε και εμείς, όπως στην αντίστοιχη ενότητα του βιβλίου, ότι $\beta = 0.8$

(a) Αφού το teleportation set είναι μόνο το A, η επαναληπτική εξίσωση που θα μας δώσει το topic-sensitive pagerank είναι:

$$\begin{aligned} v' &= \beta M v + (1-\beta) e_{\{A\}}/|\{A\}| \\ &= \begin{pmatrix} 0 & 2/5 & 4/5 & 0 \\ 4/15 & 0 & 0 & 2/5 \\ 4/15 & 0 & 0 & 2/5 \\ 4/15 & 2/5 & 0 & 0 \end{pmatrix} v + \begin{pmatrix} 1/5 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Ξεκινώντας, όπως και σε αντίστοιχα παραδείγματα στο βιβλίο, από αρχική κατανομή στο teleportation set (μήπως και βοηθήσει να επιταχύνει λίγο τη σύγκλιση), λαμβάνουμε την ακολουθία διανυσμάτων:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 \\ 0.2666 \\ 0.2666 \\ 0.2666 \end{pmatrix}, \begin{pmatrix} 0.52 \\ 0.16 \\ 0.16 \\ 0.16 \end{pmatrix}, \begin{pmatrix} 0.392 \\ 0.2026 \\ 0.2026 \\ 0.2026 \end{pmatrix}, \begin{pmatrix} 0.4432 \\ 0.1856 \\ 0.1856 \\ 0.1856 \end{pmatrix}, \begin{pmatrix} 0.4227 \\ 0.1924 \\ 0.1924 \\ 0.1924 \end{pmatrix}, \ldots, \begin{pmatrix} 0.4285 \\ 0.1904 \\ 0.1904 \\ 0.1904 \end{pmatrix}$$

(b) Με ακριβώς όμοιο τρόπο:

$$\begin{aligned} v' &= \beta M v + (1-\beta) e_{\{A\}}/|\{A\}| \\ &= \begin{pmatrix} 0 & 2/5 & 4/5 & 0 \\ 4/15 & 0 & 0 & 2/5 \\ 4/15 & 0 & 0 & 2/5 \\ 4/15 & 2/5 & 0 & 0 \end{pmatrix} v + \begin{pmatrix} 1/10 \\ 0 \\ 1/10 \\ 0 \end{pmatrix} \end{aligned}$$

Παρατηρείστε ότι το μόνο κομμάτι που χρειάζεται να αλλάξει είναι ο σταθερός όρος, αφού ο πίνακας δεν εξαρτάται από το teleportation set.

Επαναλαμβάνοντας τη διαδικασία, παίρνουμε την ακολουθία:

$$\begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.1333 \\ 0.2333 \\ 0.1333 \end{pmatrix}, \begin{pmatrix} 0.34 \\ 0.1867 \\ 0.2867 \\ 0.1867 \end{pmatrix}, \begin{pmatrix} 0.404 \\ 0.1653 \\ 0.2653 \\ 0.1653 \end{pmatrix}, \begin{pmatrix} 0.3784 \\ 0.1739 \\ 0.2739 \\ 0.1739 \end{pmatrix}, \begin{pmatrix} 0.3886 \\ 0.1705 \\ 0.2705 \\ 0.1705 \end{pmatrix}, \ldots, \begin{pmatrix} 0.3857 \\ 0.1714 \\ 0.2714 \\ 0.1714 \end{pmatrix}$$

MMDS, Exercise 7.2.2

Το πρώτο βήμα σίγουρα δεν θα άλλαζε σε καμία από τις 2 περιπτώσεις, αφού μεταξύ δύο clusters που έχουν και τα δύο μόνο ένα σημείο, η απόσταση των centroids, η ελάχιστη απόσταση μεταξύ οποιωνδήποτε σημείων τους και η μέση απόσταση μεταξύ οποιωνδήποτε σημείων τους ταυτίζονται. Οπότε, και πάλι, στο πρώτο βήμα, τα ζευγάρια σημείων που είναι πιο κοντά μεταξύ τους είναι τα (10, 5), (11, 4) και (11, 4), (12, 3). Κάνουμε αυθαίρετα την ίδια επιλογή με το παράδειγμα 7.2, δηλαδή θα συγχωνεύσουμε τα σημεία (11, 4) και (12, 3).

(a) Ελάχιστη απόσταση μεταξύ σημείων των clusters Αφού πλέον πάμε σε αποστάσεις σημείων και όχι centroids, είναι εμφανές ότι το επόμενο βήμα θα είναι να ενωθεί το σημείο (10, 5) με το cluster που μόλις δημιουργήθηκε, αφού εδώ η απόσταση του cluster από το (10, 5) θα είναι ίση με $\sqrt{2}$ (ίση με του (11, 4) με το (10, 5)).

Σε αυτό το σημείο, η ελάχιστη απόσταση μεταξύ 2 cluster είναι 2, και θα επιλέξουμε, όπως και το αντίστοιχο παράδειγμα του βιβλίου, αυθαίρετα να συγχωνεύσουμε τα σημεία (4, 8) και (4, 10) σε ένα νέο cluster. Αν και εδώ, ανεξαρτήτως σειράς, τα σημεία (4, 8) και (6, 8) απέχουν και αυτά απόσταση 2, οπότε αμέσως μετά θα συγχωνεύσουμε το cluster που μόλις φτιάξαμε με το σημείο (6, 8).

Σε αυτή τη φάση, η ελάχιστη απόσταση είναι $\sqrt{5}$, και αυτή εμφανίζεται μεταξύ αρκετών σημείων. Συγκεκριμένα, η ελάχιστη απόσταση από το cluster (4, 10), (10, 8), (6, 8) προς το σημείο (7, 10) είναι $\sqrt{5}$, και το ίδιο και οι αποστάσεις των σημείων (9, 3) και (12, 6) από το cluster (10, 5), (11, 4), (12, 3).

Μπορούμε να πούμε, λοιπόν, κάπως αυθαίρετα, ότι πρώτα συγχωνεύεται το (7, 10) με το ένα cluster 3 σημείων, μετά συγχωνεύεται το (9, 3) με το άλλο cluster 3 σημείων και αμέσως μετά το (12, 6) με το cluster που μόλις δημιουργήθηκε, και τέλος συγχωνεύονται το (2, 2) με το (3, 4) σε ένα νέο cluster.

Ανακεφαλαιώνοντας, έχουμε: 1 cluster 4 σημείων πάνω αριστερά, 1 cluster 5 σημείων κάτω δεξιά, ένα cluster 2 σημείων κάτω αριστερά και ένα cluster 1 σημείου που είναι το (5, 2).

Το επόμενο βήμα είναι να συγχωνευτούν τα 2 τελευταία clusters που αναφέραμε (2 και 1 σημείων) σε ένα cluster 3 σημείων, αφού η απόσταση του (5, 2) απο το (3, 4) είναι $\sqrt{8}$, που είναι μικρότερο από κάθε άλλη απόσταση μεταξύ των clusters που έχουν σχηματιστεί.

Οι αποστάσεις ανά δύο των 3 εναπομείναντων clusters είναι ίσες με $\sqrt{17}$, $\sqrt{17}$ και 5, που είναι μεγαλύτερο. Οπότε, αυθαίρετα λέμε ότι θα συγχωνευτούν τα clusters των 3 και 4 σημείων, και στην συνέχεια το cluster που δημιοθργήθηκε με αυτό των 5 σημείων κάτω δεξιά που απέμεινε, για να σχηματιστεί εν τέλει ένα cluster (η ρίζα του δέντρου).

Συνοψίζουμε τα παραπάνω αποτελέσματα στ

MMDS, Exercise 7.2.3

MMDS, Exercise 7.3.2

MMDS, Exercise 7.4.1