Αλγοριθμική Επιστήμη Δεδομένων
1η Σειρά Ασκήσεων

Ακαδημαϊκό έτος 2023-24

Άσκηση 1

α) Θέτουμε $x = 0$ και $y = m$ (το τελευταίο υφίσταται διότι $U = [m^k]$, $k \geqslant 2$). Τότε:

- $h_{\alpha,\beta}(x = 0) = (\alpha \cdot 0 + \beta) \mod m \equiv \beta \mod m$
- $h_{\alpha,\beta}(y = m) = (\alpha \cdot m + \beta) \mod m \equiv (\alpha \cdot m) \mod m + \beta \mod m \equiv \beta \mod m$

Όλες οι $h_{\alpha,\beta}$ της $\mathcal{H}$ για $x = 0$ και $y = m$ δίνουν την ίδια τιμή άρα βρήκαμε $x \neq y$ τ.ω. $h_{\alpha,\beta}(x) = h_{\alpha,\beta}(y)$ $\forall h_{\alpha,\beta} \in \mathcal{H}$, οπότε $\mathbb{P}_{h\in\mathcal{H}}[h(x) = h(y)] = 1$ και άρα δεν ισχύει η ιδιότητα της καθολικότητας.
β) Θέτουμε την οικογένεια συναρτήσεων $\mathcal{H} = \{h_{\alpha,\beta} : \alpha \in [p] \setminus \{0\}, \beta \in [p]\}$ όπου

$$h_{\alpha,\beta}(x) = [(\alpha x + \beta) \mod p] \mod m, \;\; x \in \mathcal{U} = [p] \tag{1}$$

Θα δείξουμε ότι η $\mathcal{H}$ είναι καθολική.
Για $x \neq y$ θα υπολογίσουμε το πλήθος των $h_{\alpha,\beta} \in \mathcal{H}$ ώστε $h_{\alpha,\beta}(x) = h_{\alpha,\beta}(y)$.

$$h_{\alpha,\beta}(x) = h_{\alpha,\beta}(y) \implies$$
$$[(\alpha x + \beta) \mod p] \mod m \equiv [(\alpha y + \beta) \mod p] \mod m \implies$$

$$[(\alpha x+\beta) \mod p] \equiv [(\alpha y+\beta) \mod p] + im \;\mu\varepsilon\; i \in [-\left[\frac{p-1}{m}\right], \left[\frac{p-1}{m}\right]] \implies$$

$$\alpha(x-y) \equiv_p im \text{ και ο } p \text{ είναι πρώτος άρα αν } x \neq 0 \text{ τότε υπάρχει ο αντίστροφός}$$

$$\alpha \equiv_p im(x-y)^{-1} \; (*)$$

Υποθέτουμε Χ.Β.Τ.Γ. ότι $x > y \Rightarrow x - y > 0$ άρα υπάρχει ο $(x-y)^{-1} \not\equiv_p 0$ διότι π είναι πρώτος.

Πρέπει να ισχύει ότι $i > 0$ άρα υπάρχουν το πολύ $\left[\frac{p-1}{m}\right]$ τιμές του ι για τις οποίες θα ισχύει η $(*)$. Εφόσον λοιπόν $\alpha \in [p] \setminus \{0\}$, από τις $p-1$ επιλογές της τιμής του α, μόνο για τις $\left[\frac{p-1}{m}\right]$ το πλήθος θα ισχύει ότι $h_{\alpha,\beta}(x) = h_{\alpha,\beta}(y)$ (οι διάφορες τιμές του β δεν επηρεάζουν το αποτέλεσμα). Οπότε

$$\mathbb{P}_{h\in\mathcal{H}}[h_{\alpha,\beta}(x) = h_{\alpha.\beta}(y)] = \frac{\left[\frac{p-1}{m}\right]}{p-1} \leqslant \frac{p-1}{m(p-1)} = \frac{1}{m}$$

Δείξαμε ότι μια οικογένεια συναρτήσεων είναι καθολική για το $\mathcal{U} = [p]$ οπότε για κάθε $\mathcal{U}' \subseteq \mathcal{U}$ αν $x \neq y$ και $x, y \in \mathcal{U}' \Rightarrow x, y \in \mathcal{U} \Rightarrow$ υπάρχουν το πολύ $\frac{|\mathcal{H}|}{m}$ το πλήθος $h_{\alpha,\beta} \in \mathcal{H} : h_{\alpha,\beta}(x) = h_{\alpha.\beta}(y)$. Άρα, για κάθε υποσύνολο του $\mathcal{U}$ ισχύει η ιδιότητα της καθολικότητας, συνεπω΄ς και για το $\mathcal{U}" = [m^k]$, $k \geqslant 2$ με $m_k < p$ διότι τότε $[m^k] \subseteq [p]$.

γ) Με βάση τα παραπάνω η ιδιότητα της καθολικότητας ισχύει αν $\mathcal{U} = [p]$.

Άσκηση 2

α) Δεδομένου ότι η ανοιχτή διευθυνσιοδότηση δουλεύει με γραμμική διερεύνηση, ο χρόνος που απαιτείται για την εισαγωγή του ι-οστού στοιχείου είναι ο ίδιος με τον χρόνο που απαιτείται για την επιτυχή αναζήτησή του. Συνεπώς και ο μέσος χρόνος διεκπεραίωσης των δύο διαδικασιών είναι ίδιος.

β) Έστω Ω_i το πλήθος των βημάτων επιτυχούς αναζήτησης/εισαγωγής (είναι ίσα από το (α)) του i-οστού στοιχείου. Θα εξετάσουμε το ενδεχόμενο $\Omega_i \geqslant k$. Προκειμένου να χρειαστούν k βήματα για να εισάγουμε ένα στοιχείο, προϋποτίθεται η θέση h_i (όπου h συνάρτηση κατακερματισμού) καθώς και οι επόμενες $k-1$ θέσεις να είναι κατειλημμένες. Από την υπόθεση της ομοιομορφίας αυτό έχει πιθανότητα να συμβεί

$$\mathbb{P}(\Omega \geqslant k) = \frac{i-1}{m} \cdot \frac{i-2}{m} \cdot \ldots \cdot \frac{i-k}{m-k+1} \leqslant \alpha_{i-1}^k$$

όπου $\alpha_{i-1} = \frac{i-1}{m}$. Εμάς μας ενδιαφέρει να φράξουμε τη μέση τιμή της $\Omega = \frac{\Omega_1+\Omega_2+\ldots+\Omega_n}{n}$. Έχουμε

$$
\begin{aligned}
\mathbb{E}[\Omega] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Omega_i] \\
&= \frac{1}{n}\sum_{i=1}^{n}(\sum_{k=1}^{\infty}\mathbb{P}[\Omega_i \leqslant k]) \\
&\leqslant \frac{1}{n}\sum_{i=1}^{n}(\sum_{k=1}^{\infty}\alpha_{i-1}^{k}) \\
&\leqslant \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{1-\alpha_{i-1}}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{m}{m-(i-1)} \\
&= \frac{m}{n}\left(\frac{1}{m-n+1}+\ldots+\frac{1}{m}\right) \\
&\leqslant \frac{m}{n}\int_{m-n}^{m}\frac{1}{x}\,dx \\
&= \frac{m}{n}\cdot ln\frac{m}{m-n} \\
&= \frac{1}{\alpha}\cdot ln\frac{1}{1-\alpha},
\end{aligned}
\tag{2}
$$

δηλαδή το ζητούμενο.

Άσκηση 3

Άσκηση 6.3.1 από το βιβλίο *MMDS*:

Μας δίνεται *support threshold*=4 και *hash table* με 11 *buckets*.

α) Το *support* κάθε *itemset* ισούται με τον αριθμό των φορών που εμφανίζεται ως υποσύνολο σε κάθε ένα από τα *baskets*.

Item	Support
1	4
2	6
3	8
4	8
5	6
6	4

Table 1: Support of every item in baskets

Itemset	Support
{1,2}	2
{1,3}	3
{1,4}	2
{1,5}	1
{1,6}	0
{2,3}	3
{2,4}	4
{2,5}	2
{2,6}	1
{3,4}	4
{3,5}	4
{3,6}	2
{4,5}	3
{4,6}	3
{5,6}	2

Table 2: Support of every itemset in baskets

β) Κάθε ζεύγος i, j κατακερματίζεται στο καλάθι με τιμή $(i \times j) \mod 11$.

Itemset	Bucket
{1,2}	2
{1,3}	3
{1,4}	4
{1,5}	5
{1,6}	6
{2,3}	6
{2,4}	8
{2,5}	10
{2,6}	1
{3,4}	1
{3,5}	4
{3,6}	7
{4,5}	9
{4,6}	2
{5,6}	8

Table 3: Buckets of all the Pairs.

γ) Υπολογίζουμε το *support* του κάθε *bucket* αθροίζοντας τα *supports* των δυσυνόλων που αντιστοιχίζονται σε αυτό το *bucket*. Ορίζουμε *frequent buckets* αυτά τα οποία έχουν $support \geq 4$.

Bucket	Support
0	0
1	5
2	5
3	3
4	6
5	1
6	3
7	2
8	6
9	3
10	2

Table 4: Supports for each bucket.

Συμπεραίνουμε ότι τα $frequent\ buckets$ είναι τα 1, 2, 4 και 8.

δ)Στην δεύτερη φάση του PCY θα περάσουν τα 2,6 με 3,4 για το $bucket$ 1, τα 1,2 με 4,6 για το $bucket$ 2, τα 1,4 με 3,5 για το $bucket$ 4 και τα 2,4 με 3,4 για το $bucket$ 8.

Άσκηση 6.4.2 από το βιβλίο $MMDS$:

Άσκηση 4

α) Ο αλγόριθμος $A{-}priori$ έχει ως στόχο να εντοπίσει τα πιο $frequent\ itemsets$ τα οποία καθορίζονται από ένα κατώφλι s (αλλιώς $support$). Ένα $itemset$ χαρακτηρίζεται $frequent$ εαν συναντάται σε τουλάχιστον s το πλήθος $market\ baskets$. Δουλεύει αποκλείοντας τα περισσότερα πολυπληθή υποψήφια σύνολα εξετάζοντας πρώτα αν τα υποσύνολά τους ξεπερνούν το κατώφλι, λαμβάνοντας υπόψιν ότι ένα μεγάλο σύνολο δεν μπορεί να είναι $frequent$ αν δεν είναι πρώτα όλα τα υποσύνολά του (αντι-μονοτονικότητα). Διασφαλίζεται έτσι η ορθότητα του αλγορίθμου. Παρόλα αυτά είναι χρονοβόρος, καθώς το πέρασμα στη βάση δεδομέων μπορεί να συμβεί κ φορές, όσο και ο μεγαλύτερος από τους πληθάριθμους των $frequent\ itemsets$.

β) Η χρονική πολυπλοκότητα του Αλγορίθμου $A - priori$ είναι πολυωνυμική ($\mathcal{O}(m \cdot n^2)$) ως προς την έξοδο, όπου m ο αριθμός των συχνών $itemsets$ και n ο συνολικός αριθμός των $items$. Όπως αναφέρουμε και στο ερώτημα (α) ο χρόνος εκτέλεσης ισούται με τον μεγαλύτερο από τους πληθάριθμους των $frequent\ itemsets$, όπου αυτός ο αριθμός είναι συνήθως πολύ μικρότερος από το πλήθος των $items$.

γ)Ο αλγόριθμος ΦΠ-Γροωτη βασίζεται σε μια δομή που ονομάζεται ΦΠ-τρεε. Η διαφορά του με τον Α-πριορι είναι πως δεν προσπαθεί να παράγει υποψήφια ιτεμσετς αλλά να ανακαλύψει συσχετίσεις. Υπάρχει όμως ομοιότητα στην πολυπλοκότητα εξόδου μεταξύ των δύο αλγορίθμων, η οποία είναι επίσης πολυωνυμική.

δ) Το $support\ threshold\ s = 5$. Καταγράφουμε για αρχή όλα τα μονοσύνολα και το $support$ τους στον παρακάτω πίνακα, δηλαδή το πλήθος των συνόλων στα οποία είναι υποσύνολα.

Item	Support
{a}	11
{b}	13
{c}	11
{d}	14

Εφόσον όλα τους έχουν $support \geqslant 5$ θα προχωρήσουν όλα στο επόμενο βήμα στο οποίο θα πάρουμε όλα τα πιθανά δυσυνόλα και θα μετρήσουμε το $support$ τους.

Item	Support
{a, b}	4
{a, c}	7
{a, d}	8
{b, c}	6
{b, d}	8
{c, d}	6

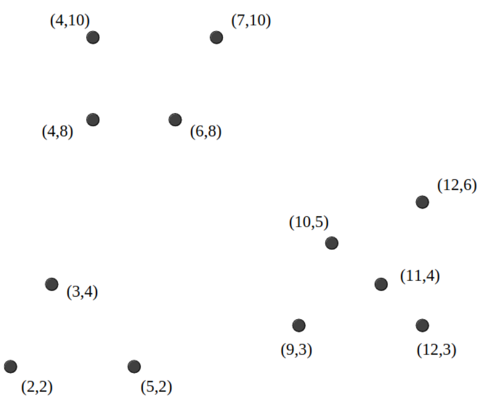
Το $\{a, b\}$ έχει $support = \ 4 < s$ οπότε δεν θα το συμπεριλάβουμε στις υπόλοιπες αναζητήσεις.

Item	Support
{a, c, d}	4
{b, c, d}	3

Παρατηρούμε ότι κανένα από τα σύνολα με τρία *items* δεν είναι $frequent$.
ε)

Άσκηση 5

Άσκηση 7.2.2 από το βιβλίο $MMDS$:



α) Εαν ορίζαμε ως απόσταση των δύο *clusters* την μικρότερη απόσταση δύο οποιωνδήποτε σημείων αυτών των δύο *clusters* , τότε η συσταδοποίση θα άλλαζε ως εξής:
Τα κοντινότερα ζευγάρια σημείων θα είναι τα (10, 5) με (11, 4) ή τα (11, 4) με (12, 3) και τα δύο με απόσταση $\sqrt{2}$. Συγχωνεύουμε αυθαίρετα τις (11, 4) και (12, 3). Το (10, 5) απέχει από το (11, 4) πάλι την μικρότερη απόσταση ($\sqrt{2}$), οπότε δημιουργείται το *cluster* { (10, 5), (11, 4), (12, 3) }.
Τώρα τα κοντινότερα *clusters* είναι η {(4, 8)} με την {(4, 10)} και η {(4, 8)} με την {(6,8)} έχοντας απόσταση 2. Άρα καταλήγουμε στον σχηματισμό του {(4, 8), (4, 10), (6, 8)}.
Τώρα υπάρχουν πολλά ζευγάρια με απόσταση $\sqrt{5}$:

- (2, 2) με (3, 4)
- (9, 3) με (10, 5) $\in$ {(10, 5), (11, 4), (12, 3)}
- (12, 6) με (11, 4) $\in$ {(10, 5), (11, 4), (12, 3)}
- (7, 12) με (6, 8) $\in$ {(4, 8), (4, 10), (6, 8)}

Έτσι προκύπτουν τα εξής *clusters*:

- {(2, 2), (3, 4)}
- {(10, 5), (11, 4), (12, 3), (9, 3), (12, 6)}
- {(6, 8), (7, 10), (4, 8), (4, 10)}

Τέλος έχουμε απόσταση $\sqrt{8}$ μεταξύ των

- (5, 2) με (3, 4) $\in$ {(2, 2), (3, 4)}

Καταλήγουμε στα *clusters*:

- {(2, 2), (3, 4), (5, 2)}
- {(6, 8), (7, 10), (4, 8), (4, 10)}
- {(10, 5), (11, 4), (12, 3), (9, 3), (12, 6)}

β) Θεωρώντας τώρα ως απόσταση τον μέσο όλων των αποστάσεων μεταξύ των σημείων του ενός με τα σημεία του άλλου *cluster* έχουμε ότι τα κοντινότερα ζευγάρια σημείων είναι τα (10, 5) με (11, 4) και (11, 4) με (12, 3) με απόσταση $\sqrt{2}$. Συγχωνεύουμε αυθαίρετα ένα ζευγάρι και παίρνουμε το *cluster* {(11, 4), (12, 3)}. Η αμέσως επόμενη μικρότερη απόσταση είναι αυτή των (4, 8) με (4, 10) και των (4, 8) με (6, 8) μετρώντας απόσταση 2. Δημιουργούμε λοιπόν το *cluster* {(4, 8), (4, 10)} όπως και στο Παράδειγμα 7.2. Έπειτα η επόμενη μικρότερη απόσταση είναι αυτή των {(4, 8), (4, 10)} και {(6, 8)} με απόσταση $\frac{2+\sqrt{8}}{2}$ οπότε τα συγχωνεύουμε. Στη συνέχεια η απόσταση του (10, 5) από το (11, 4) είναι $\sqrt{2}$ και του (10, 5) από το (12, 3) είναι $\sqrt{8}$ άρα $dist\Big(\{(10,5)\},\{(11,4),(12,3)\}\Big) = \frac{\sqrt{2}+\sqrt{8}}{2}$.
Έτσι προκύπτουν τα *clusters*:

- {(10, 5), (11, 4), (12, 3)}
- {(4, 8), (4, 10), (6, 8)}

Συνεχίζουμε με $dist\Big((9,3),\{(10,5),(11,4),(12,3)\}\Big) = \sqrt{5}$ οπότε σχηματίζεται καινούργιο *cluster* {(9, 3), (10, 5), (11, 4), (12, 3)}. Τέλος, έχουμε

- $dist\Big((5,2),\{(2,2)\,(3,4)\}\Big) = \frac{3+\sqrt{8}}{2} \approx 2.91$
- $dist\Big((12,6),\{(10,5),(11,4),(12,3),(9,3)\}\Big) = \frac{3+\sqrt{5}+\sqrt{5}+\sqrt{18}}{4} \approx 2.92$
- $dist\Big((7,10),\{(4,8),(4,10),(6,8)\}\Big) = \frac{3+\sqrt{13}+\sqrt{5}}{3} \approx 2.94$

Με αυτή τη σειρά λοιπόν οδηγούμαστε στα *clusters*

- {(2, 2) (3, 4), (5, 2)}
- {(9, 3), (10, 5), (11, 4), (12, 3), (12, 6)}
- {(4, 8), (4, 10), (6, 8), (7, 10)}

Άσκηση 7.2.3 από το βιβλίο *MMDS*:

α) Το κριτήριο συνένωσης δύο *clusters* είναι να προκύψει η μικρότερη δυνατή ακτίνα. Ως ακτίνα ορίζεται η μέγιστη απόσταση ενός σημείου του *cluster* από το *centroid*.

β) Διάμετρος ενός *cluster* ορίζεται η μεγαλύτερη απόσταση μεταξύ δύο σημείων του *cluster*. Για αρχή, επιλέγουμε να συνενώσουμε τα *clusters* $\{(11,\ 4)\}$ με $\{(12,\ 3)\}$ γιατί θα προκύψει η μικρότερη δυνατή διάμετρος ($\sqrt{2}$). Έπειτα ενώνω τις $\{(3,\ 4)\}$ με $\{(2,\ 2)\}$, τις $\{(6,\ 8)\}$ με $\{(7,\ 10)\}$ και τις $\{(10,\ 5)\}$ με $\{(12,\ 6)\}$ όλες με διάμετρο $\sqrt{5}$. Στη συνέχεια τα παρακάτω *clusters* έχουν διάμετρο 3 και θα συγχωνευτούν:

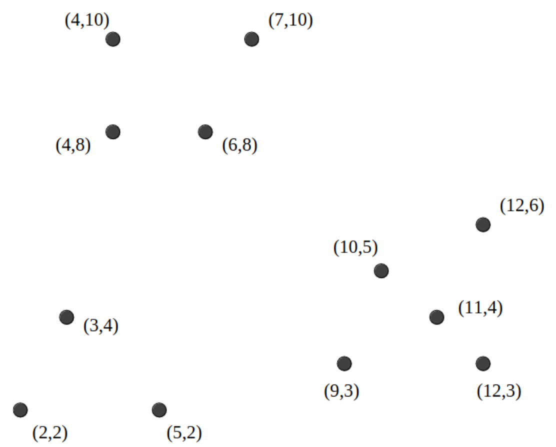
- $\{(4, 10)\}$ με $\{(6, 8), (7, 10)\}$
- $\{(5, 2)\}$ με $\{(3, 4), (2, 2)\}$
- $\{(10, 5), (12, 6)\}$ με $\{(11, 4), (12, 3)\}$

Μένει να συνενώσουμε τα

- $\{(4, 8)\}$ και $\{(4, 10), (6, 8), (7, 10)\}$ με διάμετρο $\sqrt{13}$
- $\{(9, 3)\}$ και $\{(10, 5), (12, 6), (11, 4), (12, 3)\}$ με διάμετρο $\sqrt{18}$

Άσκηση 7.2.5 από το βιβλίο *MMDS*:

Θεωρούμε τα 3 *natural clusters* της παρακάτω εικόνας.



Το *clustroid* για το καθένα, θεωρώντας πως το κριτήριο επιλογής *clustroid* είναι το σημείο με το ελάχιστο άθροισμα αποστάσεων από τα υπόλοιπα σημεία στο *cluster*.
Τα 3 *clusters* είναι:

- *Cluster* 1: $\{(3,4),(2,2),(5,2)\}$
- *Cluster* 2: $\{(4,10),(7,10),(4,8),(6,8)\}$
- *Cluster* 3: $\{(10,5),(12,6),(11,4),(9,3),(12,3)\}$

Τα *clustroids* είναι:

- Το (3, 4) για το *Cluster* 1
- Το (6, 8) για το *Cluster* 2
- Το (11, 4) για το *Cluster* 3

Άσκηση 7.3.2 από το βιβλίο *MMDS*:
Θεώρημα
Έστω κ κλάσεις $C_1, C_2, .., C_\kappa$ με την ιδιότητα

$$\max diam(C_i) < min\_dist(C_i, C_j)\ \forall i, j \in \{1, .., \kappa\}$$

όπου $min\_dist(C_i, C_j)$ (*minimum intercluster distance*) είναι η μικρότερη απόσταση μεταξύ δύο σημείων από δύο διαφορετικές κλάσεις. Τότε ανεξάρτητα από το πρώτο σημείο που επιλέξαμε, αν εφαρμόσουμε τον Αλγόρισθμο του 7.3.2 για κ *starting points*, αυτά θα ανήκουν σε διαφορετιές κλάσεις μεταξύ τους.

Αλγόριθμος 7.3.2

```
Διαλέγουμε τυχαίο σημείο
while πλήθος starting points do
    Πρόσθεσε το σημείο του οποίου η mindist από τα επιλεγμένα σημεία
    είναι όσο το δυνατόν μεγαλύτερη γίνεται
end while
```

Απόδειξη

Έστω ότι έχουμε επιλέξει $X = \{x_1, .., x_s\}$ *starting points* όπου $s < \kappa$ καθένα από τα οποία ανληκει σε διαφορετική κλάσημε βάση τον παραπάνω αλγόριθμο. Καλούμαστε να επιλέξουμε σημείο x_{s+1}. Εφόσον $s < \kappa$ υπάρχει κάποια κλάση που να μην έχει κανένα από τα s σημεία που έχουμε επιλέξει. Άρα αν επιλέξουμε το x_{s+1} από μία καινούργια κλάση και αν επιλέξουμε σημείο y σε οποιαδήποτε από τις υπάρχουσες κλάσεις τότε:

$$dist(y, X) \leqslant maxdiam_{1 \leqslant i \leqslant \kappa}(C_i) < min\_dist_{1 \leqslant i,j \leqslant \kappa,\, i \neq j}(C_i, C_j) \leqslant dist(x_{s+1}, X)$$

Οπότε το σημείο που επιτυγχάνει την μεγιστική απόσταση από το X είναι αναγκατικά έξω από τις προηγούμενες κλάσεις και έπεται το ζητούμενο.

Άσκηση 7.4.1 από το βιβλίο $MMDS$:

Στον Αλγόριθμο $CURE$ τα *representative points* έχουν επιλεχθεί έτσι ώστε να βρίσκονται στα όρια των δύο κλάσεων. Ονομάζουμε C_1 την κλάση στο δαχτυλίδι η οποία έχει τα *representative points* της πάνω στον κύκλο ακτίνας o και κέντρου (0, 0) και στον ομόκεντρο κύκλο ακτίνας i. Ονομάζουμε C_2 την κλάση κέντρου (0, 0) και ακτίνας c.

Εφόσον όλα τα *representative points* μετακινήθηκαν κατά 20% εγκύτερα προς τα κεντροϊδές τους (αυτό είναι το (0, 0) για όλα), τα εξωτερικά *representative points* της C_1 βρίσκονται πλεόν σε κύκλο ακτίνας $0.8{\cdot}o$ με κέντρο (0, 0), τα εσωτερικά *representative points* της C_1 σε ακτίνα $0.8 \cdot i$ στον ομόκεντρο κύκλο και αυτά της C_2 απέχουν $0.8 \cdot c$ από το (0, 0).

Σύμφωνα με την εκφώνηση, δύο κλάσεις συγχωνεύονται αν έπειτα από την αναμετάθεση των σημείων τους υπάρχουν *representative points* και από τις δύο κλάσεις απόστασης μικρότερης ή ίσης του d. Μαθηματικά αυτό μεταφράζεται ως:

$$\|x_1 - x_2\|_2 \leqslant d,$$

όπου x_1 *representative point* της C_1, x_2 *representative point* της C_2. Αν επιλέξουμε τα εγκύτερα δυνατά *representative points* των δύο κλάσεων, το ένα θα βρίσκεται στο *inner circle* της C_1 με απόσταση $0.8 \cdot i$ από το κέντρο και αυτό της C_2 στην ίδια ακτίνα με απόσταση $0.8 \cdot c$ από το κέντρο. Η απόστασή τους είναι

$$(0.8 \cdot i - 0.8 \cdot c) = 0.8(i - c)$$

Τελικά οι κλάσεις C_1 και C_2 δεν πρόκειται να συγχωνευθούν αν

$$d < 0.8(i - c)$$