ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2η ΣΕΙΡΑ ΓΡΑΠΤΩΝ ΑΣΚΗΣΕΩΝ

Ειρήνη Δόντη
ΑΜ 03119839

7ο εξάμηνο

Αθήνα 2022

ΑΣΚΗΣΗ 2.1:

Η κλάση $Hrec^n$ των παραλλήλων στους άξονες υπερ-παραλληλογράμμων του R^n είναι PAC εκπαιδεύσιμη, καθώς θεωρούμε ότι για μία περιοχή ri με i∈[1,2n]: $Pr_D[ri] = \frac{\varepsilon}{2n}$ και θεωρώντας ότι Pr_D[R - R'] > ε, ε > 0, η πιθανότητα κανένα από τα m να ταυτίζεται με την περιοχή ri είναι $(1-\frac{\varepsilon}{2n})^m$.

Οπότε, Pr[error(R') > ε] ≤ $2n(1-\frac{\varepsilon}{2n})^m \leq 2ne^{-\frac{\varepsilon}{2n}m}$ ή $m \geq \frac{2n}{\varepsilon}\log(\frac{2n}{\delta})$ με δ>0 πιθανότητα το λιγότερο 1 - δ, $error_D$(R') ≤ ε.

ΑΣΚΗΣΗ 2.2:

(a) Χρησιμοποιούμε τη μέθοδο των πολλαπλασιαστών Lagrange:

$$L(u,\theta,\lambda) = J(u,\theta) + \sum_i \lambda i\left(\sum_{j=1}^{m} uij - 1\right)$$

Πρέπει: $\frac{\partial L}{\partial uij} = 0$ ή $quij^{q-1}d(xi,\theta j)^2 - \lambda i = 0$ (1)

$\frac{\partial L}{\partial \theta j} = 0$ ή $\frac{\partial J}{\partial \theta j} = 0$ ή $-2\sum_i^N uij^q(xi-\theta j) = 0$ ή $\sum_i^N uij^q xi = \sum_i^N uij^q \theta j$ (2)

Οπότε από (2): $\theta j = \frac{\sum_{i=1}^N uij^q xi}{\sum_{i=1}^N uij^q}$

Από τη σχέση (1): $uij = \left(\frac{\lambda i}{qd(xi,\theta j)^2}\right)^{\frac{1}{q-1}} = \left(\frac{\lambda i}{q}\right)^{\frac{1}{q-1}}\left(\frac{1}{d(xi,\theta j)}\right)^{\frac{2}{q-1}}$

Επειδή $\sum_{k=1}^m uik = 1$: $\sum_{k=1}^{m}\left(\frac{\lambda i}{qd(xi,\theta k)^2}\right)^{\frac{1}{q-1}} = 1$ ή

$$\left(\frac{\lambda i}{q}\right)^{\frac{1}{q-1}} = \frac{1}{\sum_{j=1}^{m}\left(\frac{1}{d(xi,\theta k)}\right)^{\frac{2}{q-1}}}$$

Οπότε: $uij = \frac{1}{\sum_{k=1}^{m}\left(\frac{d(xi,\theta j)}{d(xi,\theta k)}\right)^{\frac{1}{q-1}}}$

(b) Παρατίθεται κώδικας στο .ipynb αρχείο

(c) Παρατίθεται κώδικας στο .ipynb αρχείο

(d) Παρατίθεται κώδικας στο .ipynb αρχείο

(e) Παρατίθεται σχόλια στο .ipynb αρχείο

ΑΣΚΗΣΗ 2.3:

(a) Για να αποδείξουμε ότι η d_n είναι μετρική, πρέπει να ικανοποιεί τα εξής:

- d_n(x,y) ≥ 0 για κάθε x,y και d_n(x,y) = 0 αν και μόνο αν x=y.

  Εφόσον υπάρχουν προσθέσεις με απόλυτα, προφανώς d_n(x,y) ≥ 0 για κάθε x,y. Αν x=y, τότε xi = yi και xj = yj, οπότε d_n(x,y) = 0. Αν d_n(x,y) = 0, τότε |xi – yi| = 0 και |xj – yj| = 0 ή xi = yi και xj = yj ή x=y

- d_n(x,y) = d_n(y,x) για κάθε x,y

  Προφανώς ισχύει, γιατί έχουμε απόλυτο.

- d_n(x,z) ≤ d_n(x,y) + d_n(y,z)

  d_n(x,z) = |xi – zi| + $\frac{1}{\frac{l}{2}+1} \cdot \sum_{i=1,j\neq 1} |xj - zj|$

  d_n(x,y) = |xi – yi| + $\frac{1}{\frac{l}{2}+1} \cdot \sum_{i=1,j\neq 1} |xj - yj|$

  d_n(y,z) = |yi – zi| + $\frac{1}{\frac{l}{2}+1} \cdot \sum_{i=1,j\neq 1} |yj - zj|$

  Ισχύει ότι: |α+β| ≤ |α| + |β|. Οπότε, |xi – zi| ≤ |xi – yi| + |yi – zi| και |xj – zj| ≤ |xj – yj| + |yj – zj|.

  Οπότε, d_n(x,z) = |xi – zi| + $\frac{1}{\frac{l}{2}+1} \cdot \sum_{i=1,j\neq 1} |xj - zj|$ ≤ |xi – yi| + |yi – zi| + $\frac{1}{\frac{l}{2}+1} \cdot \sum_{i=1,j\neq 1}(|\mathrm{xj} - \mathrm{yj}| + |\mathrm{yj} - \mathrm{zj}|)$ ≤ d_n(x,y) + d_n(y,z).

(b) Για τον πίνακα προτύπων, χρησιμοποιούμε το σύνολο προτύπων X

$$D(X) = \begin{bmatrix} 0 & 3 \\ 1.4 & 2.6 \\ -1.5 & 3.4 \\ -0.2 & -0.4 \\ 1 & -1 \\ 2 & -1.5 \\ 2.6 & -1.8 \\ 3 & -2 \end{bmatrix}$$ και για το πίνακα εγγύτητας, χρησιμοποιούμε τη

μετρική P(X) = $$\begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 & 5.5 & 6.1 & 6.5 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 & 4.4 & 5 & 5.4 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 & 6.65 & 7.25 & 7.65 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 & 2.75 & 3.5 & 4 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 & 1.25 & 2 & 2.5 \\ 5.5 & 4.4 & 6.65 & 2.75 & 1.25 & 0 & 0.75 & 1.25 \\ 6.1 & 5 & 7.25 & 3.5 & 2 & 0.75 & 0 & 0.5 \\ 6.5 & 5.4 & 7.65 & 4 & 2.5 & 1.25 & 0.5 & 0 \end{bmatrix}$$

(c) Με βάση τον αλγόριθμο ιεραρχικού δεσμού, οι διαδοχικές ομαδοποιήσεις και το αντίστοιχο δεντρόγραμμα εγγύτητας προκύπτουν ως εξής:

{{x1}, {x2}, {x3}, {x4}, {x5}, {x6}, {x7}, {x8}}

- {{x1}, {x2}, {x3}, {x4}, {x5}, {x6}, {x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 & 5.5 & 6.1 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 & 4.4 & 5 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 & 6.65 & 7.25 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 & 2.75 & 3.5 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 & 1.25 & 2 \\ 5.5 & 4.4 & 6.65 & 2.75 & 1.25 & 0 & 0.75 \\ 6.1 & 5 & 7.25 & 3.5 & 2 & 0.75 & 0 \end{bmatrix}$$

- {{x1}, {x2}, {x3}, {x4}, {x5}, {x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 & 5.5 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 & 4.4 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 & 6.65 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 & 2.75 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 & 1.25 \\ 5.5 & 4.4 & 6.65 & 2.75 & 1.25 & 0 \end{bmatrix}$$

- {{x1}, {x2}, {x3}, {x4}, {x5, x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 \end{bmatrix}$$

- {{x1}, {x2}, {x3}, {x4, x5, x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 \\ 1.6 & 0 & 3.3 & 3.8 \\ 1.7 & 3.3 & 0 & 4.45 \\ 3.5 & 3.8 & 4.45 & 0 \end{bmatrix}$$

- {{x1, x2}, {x3}, {x4, x5, x6, x7, x8}}

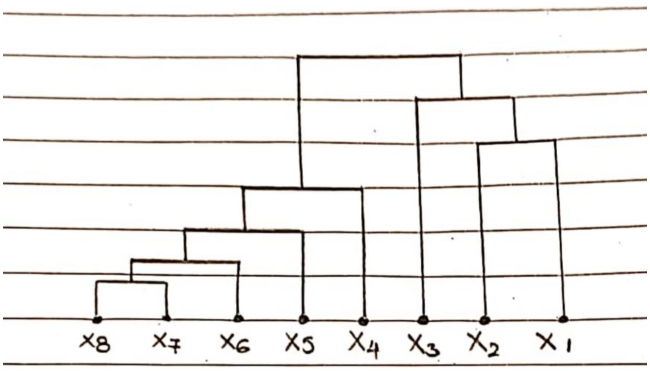
$$P(X) = \begin{bmatrix} 0 & 1.7 & 3.5 \\ 1.7 & 0 & 4.45 \\ 3.5 & 4.45 & 0 \end{bmatrix}$$

- {{x1, x2, x3}, {x4, x5, x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 3.5 \\ 3.5 & 0 \end{bmatrix}$$

{{x1, x2, x3, x4, x5, x6, x7, x8}}

Το δενδρόγραμμα παρουσιάζεται παρακάτω:



(d) Ιεραρχικός αλγόριθμος πλήρους δεσμού:

{{x1}, {x2}, {x3}, {x4}, {x5}, {x6}, {x7}, {x8}}

- {{x1}, {x2}, {x3}, {x4}, {x5}, {x6}, {x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 & 5.5 & 6.5 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 & 4.4 & 5.4 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 & 6.65 & 7.65 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 & 2.75 & 4 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 & 1.25 & 2.5 \\ 5.5 & 4.4 & 6.65 & 2.75 & 1.25 & 0 & 1.25 \\ 6.5 & 5.4 & 7.65 & 4 & 2.5 & 1.25 & 0 \end{bmatrix}$$

- {{x1}, {x2}, {x3}, {x4}, {x5}, {x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 3.5 & 4.5 & 6.5 \\ 1.6 & 0 & 3.3 & 3.8 & 3.8 & 5.4 \\ 1.7 & 3.3 & 0 & 4.45 & 5.65 & 7.65 \\ 3.5 & 3.8 & 4.45 & 0 & 1.5 & 4 \\ 4.5 & 3.8 & 5.65 & 1.5 & 0 & 2.5 \\ 6.5 & 5.4 & 7.65 & 4 & 2.5 & 0 \end{bmatrix}$$

- {{x1}, {x2}, {x3}, {x4, x5}, {x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 1.6 & 1.7 & 6.5 & 4.5 \\ 1.6 & 0 & 3.3 & 5.4 & 3.8 \\ 1.7 & 3.3 & 0 & 7.65 & 5.65 \\ 6.5 & 5.4 & 7.65 & 0 & 4 \\ 4.5 & 3.8 & 5.65 & 4 & 0 \end{bmatrix}$$

- {{x1, x2}, {x3}, {x4, x5}, {x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 3.3 & 6.5 & 4.5 \\ 3.3 & 0 & 7.65 & 5.65 \\ 6.5 & 7.65 & 0 & 4 \\ 4.5 & 5.65 & 4 & 0 \end{bmatrix}$$

- {{x1, x2, x3}, {x4, x5}, {x6, x7, x8}}

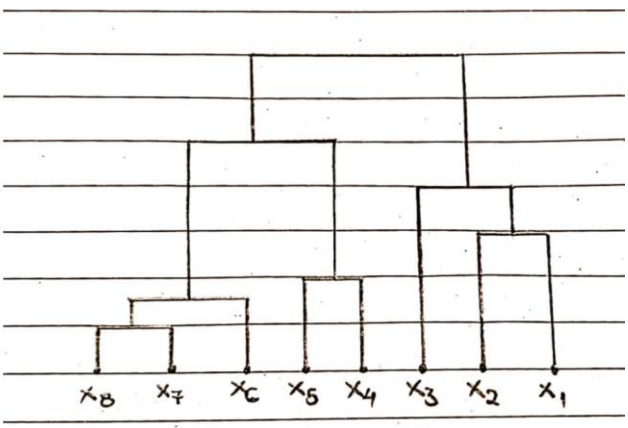
$$P(X) = \begin{bmatrix} 0 & 7.65 & 5.65 \\ 7.65 & 0 & 4 \\ 5.65 & 4 & 0 \end{bmatrix}$$

- {{x1, x2, x3}, {x4, x5, x6, x7, x8}}

$$P(X) = \begin{bmatrix} 0 & 3.5 \\ 3.5 & 0 \end{bmatrix}$$

{{x1, x2, x3, x4, x5, x6, x7, x8}}

Το δενδρόγραμμα παρουσιάζεται παρακάτω:



(e) Οι αλγόριθμοι απλού δεσμού και πλήρους δεσμού διαφέρουν στη μέθοδο που χρησιμοποιείται για τον υπολογισμό της απόστασης μεταξύ των συστάδων.

Στην απλό δεσμό, η απόσταση μεταξύ δύο συστάδων ορίζεται ως η ελάχιστη απόσταση μεταξύ οποιωνδήποτε δύο σημείων στις δύο συστάδες, ενώ στον πλήρη δεσμό, η απόσταση μεταξύ δύο συστάδων ορίζεται ως η μέγιστη απόσταση μεταξύ οποιωνδήποτε δύο σημείων στις δύο συστάδες. Οπότε, το σχήμα των δενδρογραμμάτων κάθε αλγορίθμου είναι διαφορετικό. Στον απλό δεσμό, οι συστάδες τείνουν να σχηματίζονται από σημεία που είναι κοντά το ένα στο άλλο και τείνουν να σχηματίζουν περισσότερες συστάδες. Όπως και στον πλήρη δεσμό, οι συστάδες σχηματίζονται με σημεία που είναι πιο διακριτά μεταξύ τους και τείνουν να σχηματίζουν λιγότερα συμπλέγματα. Οπότε, η βέλτιστη ομαδοποίηση, βάσει των παραπάνω, επιτυγχάνεται με τον ιεραρχικό αλγόριθμο πλήρους δεσμού.

<u>ΑΣΚΗΣΗ 2.4:</u>

(a) J(α1, …, αn, e) = $\frac{1}{N}\sum_{n=1}^{N}\|x_n - a_n e\|^2$ =

= $\frac{1}{N}\sum_{n=1}^{N}({x_n}^2 - 2a_n x_n e + (a_n e)^2)$

Βρίσκουμε την τιμή που ελαχιστοποιεί το J: $\frac{\partial J}{\partial a_n}$ = -2${x_n}^T e + 2a_n e^2$ ή

$a_n = \frac{{x_n}^T e}{e^T e} = {x_n}^T e$ = <x_n, e>

(b) Αντικαθιστούμε το βέλτιστο a_n στο J.

J1(α1, …, αn, e) = $\frac{1}{N}\sum_{n=1}^{N}\left\|x_n - ({x_n}^T e)e\right\|^2$ =

= $\frac{1}{N}\sum_{n=1}^{N}({x_n}^2 - 2({x_n}^T e)x_n e + (({x_n}^T e)e)^2)$ =

= $\frac{1}{N}\sum_{n=1}^{N}({x_n}^2 - 2{x_n}^T e e x_n + {x_n}^T e e x_n)$ =

= $\frac{1}{N}\sum_{n=1}^{N}\|x_n\|^2 - e^T\frac{1}{N}\sum_{n=1}^{N}x_n {x_n}^T e$ =

= -$e^T\frac{1}{N}R_x e$ + $\frac{1}{N}\sum_{n=1}^{N}\|x_n\|^2$

(c) Χρησιμοποιούμε τον πολλαπλασιαστή Lagrange: L(e, λ) = J1(e) + λ1($e^T e$ -1)

Βρίσκουμε το e ώστε να ελαχιστοποιηθεί ο J1: $\frac{\partial L}{\partial e} = \frac{\partial(-e^T R_x e)}{\partial e} + \frac{\partial(\lambda 1 e^T e)}{\partial e}$ =

= -2$R_x e$ + 2λ1e = 0 ή $R_x e$ = λ1e. Οπότε, το βέλτιστο e που ελαχιστοποιεί το J1 είναι το ιδιοδιάνυσμα R_x που αντιστοιχεί στη μέγιστη ιδιοτιμή.

(d) Η παραπάνω ανάλυση σχετίζεται με την PCA, η οποία χρησιμοποιείται για τη μείωση διαστάσεων και επιδιώκει να βρίσκει κατευθύνσεις στα δεδομένα με τη μεγαλύτερη διακύμανση, προβάλλοντας τα προκειμένου να λάβουμε μία αναπαράσταση των δεδομένων σε χαμηλότερες διαστάσεις. Τα παραπάνω ερωτήματα, υπολογίζουν το πρώτο δεδομένο με τη μέγιστη διακύμανση. Στην τεχνική PCA, κεντράρουμε τα δεδομένα, αφαιρώντας τον μέσο όρο κάθε χαρακτηριστικού ώστε τα δεδομένα να έχουν μηδενικό μέσο όρο. Ο εμπειρικός πίνακας R_x είναι ο πίνακας συνδιακύμανσης δείγματος και τα ιδιοδιανύσματα του είναι κύρια συστατικά των δεδομένων. Το πρώτο κύριο συστατικό είναι το ιδιοδιάνυσμα με τη μέγιστη ιδιοτιμή. Οπότε, τα παραπάνω υποερωτήματα αποτελούν το πρώτο βήμα της PCA.

(e) Η SVD είναι μία παραγοντοποίηση ενός πίνακα X στο γινόμενο πινάκων X = UDV^T, όπου U και V είναι ορθογώνιοι πίνακες και D ένας διαγώνιος. Εάν στοιβάζουμε τα διανύσματα x_n ως γραμμές για να σχηματίσουμε τον πίνακα X, τα ιδιοδιανύσματα του πίνακα αυτοσυσχέτισης R_x ταυτίζονται με τον πίνακα V. Οι ιδιοτιμές του R_x είναι τα τετράγωνα των τιμών του X. Επίσης, η SVD σχετίζεται με τα παραπάνω υπερωτήματα, καθώς η ιδιοτιμή λ1 ταυτίζεται με το τετράγωνο του πρώτου στοιχείου του X.

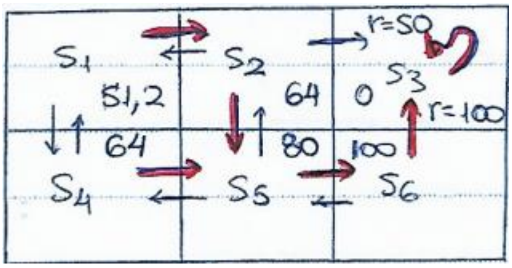
<u>ΑΣΚΗΣΗ 2.5:</u>

Παρατίθεται κώδικας στο .ipynb αρχείο.

ΑΣΚΗΣΗ 2.6:

(a) Ισχύει ότι u*(s3) = 0. Οπότε, u*(s6) = 100, u*(s5) = 100·0.8 = 80, u*(s2) = 80·0.8 = 64, u*(s1) = 64·0.8 = 51,2 και u*(s4) = 64.

(b) Σημειώνουμε στην εικόνα τα ζητούμενα βέλη μαζί με την τιμή u*(s) για κάθε κατάσταση s.



(c) Εξετάζουμε τις πλήρεις επαναλήψεις ως προς την αξία:

	S_1	S_2	S_3	S_4	S_5	S_6
0	0	0	0	0	0	0
1	0	50	0	0	0	100
2	40	50	0	0	80	100
3	40	64	0	64	80	100
4	51,2	64	0	64	80	100

Οπότε, χρειάζονται 4 πλήρεις επαναλήψεις για την εύρεση της βέλτιστης πολιτικής.

(d) Αν μεταβληθούν οι ανταμοιβές με τον ίδιο τρόπο, τότε η V* θα μεταβληθεί, αλλά η βέλτιστη πολιτική Π* θα παραμείνει αμετάβλητη.

Π.χ. Ισχύει ότι u*(s3) = 0. Οπότε, αυξάνοντας τις ανταμοιβές κατά 10 ισχύει ότι: u*(s6) = 110, u*(s5) = 110·0.8 = 88, u*(s2) = 88·0.8 = 70,4 , u*(s1) = 70,4·0.8 = 56,32 και u*(s4) = 70,4.

Συμπληρώνοντας το παραπάνω πίνακα με τις παραπάνω τιμές, θα προκύψουν πάλι 4 πλήρεις επαναλήψεις για την εύρεση της βέλτιστης πολιτικής.