Τεχνικές Μείωσης Διαστάσεων

Ειδικά θέματα ψηφιακής επεξεργασίας σήματος και εικόνας

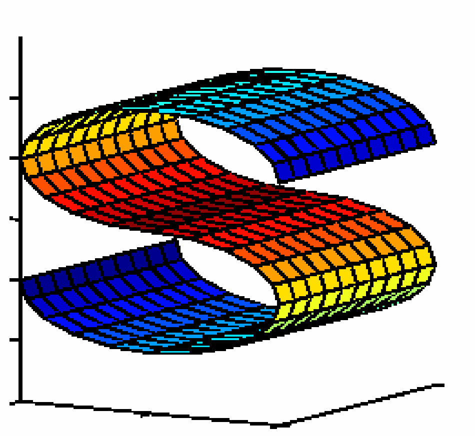
Σ. Φωτόπουλος- Α. Μακεδόνας

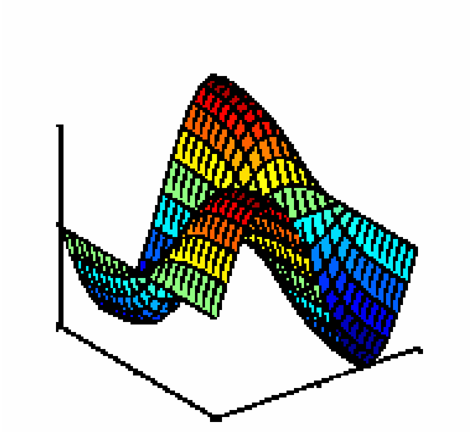
Εισαγωγή

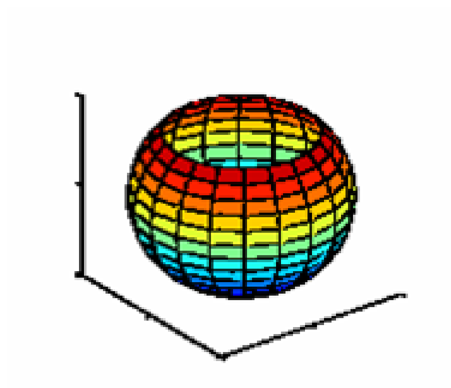
- Το μεγαλύτερο μέρος των δεδομένων που καλούμαστε να επεξεργαστούμε είναι πολυδιάστατα.

Εισαγωγή

- Το μεγαλύτερο μέρος των δεδομένων που καλούμαστε να επεξεργαστούμε είναι πολυδιάστατα.
- Τα δεδομένα στον πολυδιάστατο χώρο συχνά παρουσιάζουν κάποια συγκεκριμένη δομή.







Τεχνικές Μείωσης Διαστάσεων

- Τι είναι η μείωση διαστάσεων (**D**imensionality **R**eduction – **DR**);
  - Μια μεθοδολογία που προσπαθεί να προβάλει ένα σύνολο από διανύσματα υψηλής διάστασης σε ένα χώρο χαμηλότερης διάστασης

Τεχνικές Μείωσης Διαστάσεων

- Τι είναι η μείωση διαστάσεων (**D**imensionality **R**eduction – **DR**);
  - Μια μεθοδολογία που προσπαθεί να προβάλει ένα σύνολο από διανύσματα υψηλής διάστασης σε ένα χώρο χαμηλότερης διάστασης
- Τι μας παρέχει;
  - Δίνει μια λύση στο πρόβλημα διαχείρισης δεδομένων πολλών διαστάσεων, αναζητώντας δομή χαμηλής διάστασης στα πολυδιάστατα δεδομένα

Τεχνικές Μείωσης Διαστάσεων

- Τι είναι η μείωση διαστάσεων (**D**imensionality **R**eduction – **DR**);
  - Μια μεθοδολογία που προσπαθεί να προβάλει ένα σύνολο από διανύσματα υψηλής διάστασης σε ένα χώρο χαμηλότερης διάστασης
- Τι μας παρέχει;
  - Δίνει μια λύση στο πρόβλημα διαχείρισης δεδομένων πολλών διαστάσεων, αναζητώντας δομή χαμηλής διάστασης στα πολυδιάστατα δεδομένα
- Γιατί είναι απαραίτητη;
  - Οι αποστάσεις μεταξύ των δεδομένων στον ελαττωμένο χώρο υπολογίζονται πιο γρήγορα
  - Το μέγεθος του συνόλου δεδομένων μειώνεται
  - Αποκαλύπτεται η δομή των δεδομένων η οποία παραμένει κρυμμένη στον αρχικό πολυδιάστατο χώρο
  - Βελτιώνεται η αποδοτικότητα των τεχνικών εξόρυξης δεδομένων

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Ανελαστικές – Ελαστικές τεχνικές (Hard vs. Soft DR)
  - Ο διαχωρισμός αυτός έχει να κάνει με τον λόγο του πλήθους των αρχικών διαστάσεων ως προς το πλήθος των τελικών.
  - Οι «ανελαστικές» τεχνικές μείωσης διαστάσεων αναφέρονται σε προβλήματα που έχουν να κάνουν με δείγματα πολύ μεγάλης διαστατικότητας τα οποία πρέπει να αναπαρασταθούν σε χώρους πολύ μικρότερων διαστάσεων.
  - Αντίθετα οι «ελαστικές» τεχνικές χρησιμεύουν για προβλήματα όπου οι αρχικές διαστάσεις των δειγμάτων με τις τελικές τους διαφέρουν κατά λίγο (μία τάξη μεγέθους).

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Ανελαστικές – Ελαστικές τεχνικές (Hard vs. Soft DR)
  - Ο διαχωρισμός αυτός έχει να κάνει με τον λόγο του πλήθους των αρχικών διαστάσεων ως προς το πλήθος των τελικών.
  - Οι «ανελαστικές» τεχνικές μείωσης διαστάσεων αναφέρονται σε προβλήματα που έχουν να κάνουν με δείγματα πολύ μεγάλης διαστατικότητας τα οποία πρέπει να αναπαρασταθούν σε χώρους πολύ μικρότερων διαστάσεων.
  - Αντίθετα οι «ελαστικές» τεχνικές χρησιμεύουν για προβλήματα όπου οι αρχικές διαστάσεις των δειγμάτων με τις τελικές τους διαφέρουν κατά λίγο (μία τάξη μεγέθους).
- Παραδοσιακά – Παραγωγικά μοντέλα (Traditional vs. Generative model)
  - Τα παραδοσιακά μοντέλα προσπαθούν να δημιουργήσουν τα δείγματα στον ελαττωμένο χώρο με βάση τις παρατηρήσεις στον αρχικό.
  - Τα παραγωγικά μοντέλα προσπαθούν από τυχαίες μεταβλητές στον ελαττωμένο χώρο να αναδομήσουν τα δείγματα στον αρχικό χώρο.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Συνεχή – Διακριτά μοντέλα (Continuous vs. Discrete model)
  - Στα συνεχή μοντέλα η μεταφορά των σημείων από τον αρχικό στον ελαττωμένο χώρο γίνεται με τη χρήση μιας παραμετροποιημένης συνάρτησης ανάμεσα στους δύο αυτούς χώρους. Τα νέα δείγματα τοποθετούνται στον ελαττωμένο χώρο χωρίς να χρειάζεται ο εξαρχής προσδιορισμός της συνάρτησης αυτής.
  - Στα διακριτά μοντέλα τα νέα σημεία δεν είναι εύκολο να τοποθετηθούν στον ελαττωμένο χώρο και τεχνικές παρεμβολής (interpolation procedures) είναι απαραίτητες για το σκοπό αυτό.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Συνεχή – Διακριτά μοντέλα (Continuous vs. Discrete model)
  - Στα συνεχή μοντέλα η μεταφορά των σημείων από τον αρχικό στον ελαττωμένο χώρο γίνεται με τη χρήση μιας παραμετροποιημένης συνάρτησης ανάμεσα στους δύο αυτούς χώρους. Τα νέα δείγματα τοποθετούνται στον ελαττωμένο χώρο χωρίς να χρειάζεται ο εξαρχής προσδιορισμός της συνάρτησης αυτής.
  - Στα διακριτά μοντέλα τα νέα σημεία δεν είναι εύκολο να τοποθετηθούν στον ελαττωμένο χώρο και τεχνικές παρεμβολής (interpolation procedures) είναι απαραίτητες για το σκοπό αυτό.
- Υπονοούμενη – Ρητή χαρτογράφηση (Implicit vs. Explicit mapping)
  - Στη ρητή χαρτογράφηση υπάρχει άμεση ανάθεση κάθε αναπαράστασης στον ελαττωμένο χώρο με τα δείγματα. Είναι δύσκολη, όπως και στα διακριτά μοντέλα η τοποθέτηση νέων σημείων.
  - Στην υπονοούμενη χαρτογράφηση δεν υπάρχει άμεση συσχέτιση ανάμεσα στις παραμέτρους που χρησιμοποιούνται για την ελάττωση των διαστάσεων και τις συντεταγμένες των σημείων στον αρχικό χώρο.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Ενσωματωμένη – Εξωτερική εκτίμηση της διαστατικότητας (Integrated vs. External estimation of the dimensionality)
  - Οι περισσότερες μέθοδοι δεν έχουν ενσωματωμένο εκτιμητή του βέλτιστου αριθμού των διαστάσεων στις οποίες πρέπει να πέσει το σύστημά μας, ώστε να έχουμε την καλύτερη αναπαράσταση. Έτσι αυτή η δουλειά γίνεται συνήθως εξωτερικά από τον χρήστη.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Ενσωματωμένη – Εξωτερική εκτίμηση της διαστατικότητας (Integrated vs. External estimation of the dimensionality)
  - Οι περισσότερες μέθοδοι δεν έχουν ενσωματωμένο εκτιμητή του βέλτιστου αριθμού των διαστάσεων στις οποίες πρέπει να πέσει το σύστημά μας, ώστε να έχουμε την καλύτερη αναπαράσταση. Έτσι αυτή η δουλειά γίνεται συνήθως εξωτερικά από τον χρήστη.
- Σταδιακή – Αυτόνομη ενσωμάτωση (Layered vs. Standalone embeddings)
  - Στις μεθόδους της σταδιακής ενσωμάτωσης η προσθήκη ή η αφαίρεση μιας διάστασης των δεδομένων δεν απαιτεί τον επαναπροσδιορισμό των συντεταγμένων στις υπόλοιπες χρησιμοποιούμενες διαστάσεις. Όλες οι μέθοδοι, οι οποίες προσεγγίζουν το πρόβλημα της μείωσης της διαστατικότητας με τη λύση ενός προβλήματος ιδιοτιμών ανήκουν σε αυτή την κατηγορία και καλούνται φασματικές μέθοδοι (spectral methods).
  - Αντίθετα στην αυτόνομη ενσωμάτωση αν αλλάξει ο αριθμός των διαστάσεων πρέπει να υπολογιστούν εξαρχής οι συντεταγμένες.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

- Γραμμικά – μη Γραμμικά μοντέλα (Linear vs. Nonlinear model)
  - Τα μη γραμμικά μοντέλα είναι πιο ισχυρά από τα αντίστοιχα γραμμικά, γιατί μπορούν να δώσουν καλύτερη αναπαράσταση της δομής που υπάρχει στον αρχικό χώρο.
  - Για την υλοποίησή τους όμως είναι απαραίτητος ο υπολογισμός πολλών παραμέτρων, οι οποίες με τη σειρά τους για να υπολογιστούν προϋποθέτουν την ύπαρξη πολλών δεδομένων.

Κατηγοριοποίηση Τεχνικών Μείωσης Διαστάσεων

Τεχνικές Μείωσης Διαστατικότητας

Γραμμικές

Principal **C**omponent **A**nalysis

Multi-**D**imensional **S**caling

Μη γραμμικές

Locally **L**inear **E**mbedding

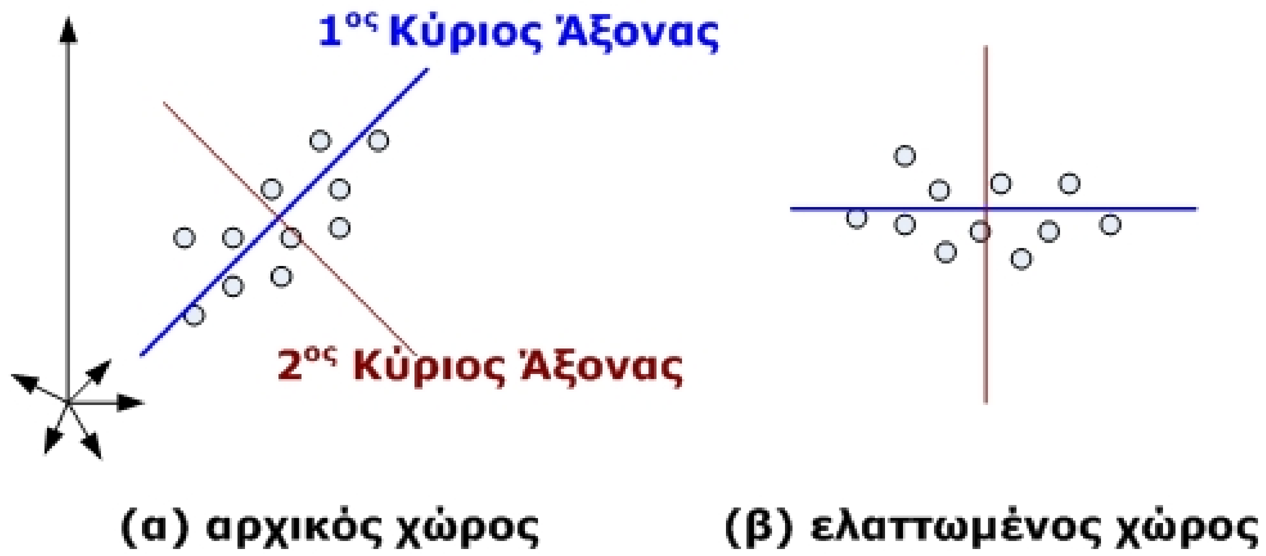
ISOMAP

Γραμμικές τεχνικές μείωσης διαστάσεων

- Principal Component Analysis (PCA)
- Multidimensional Scaling (MDS)

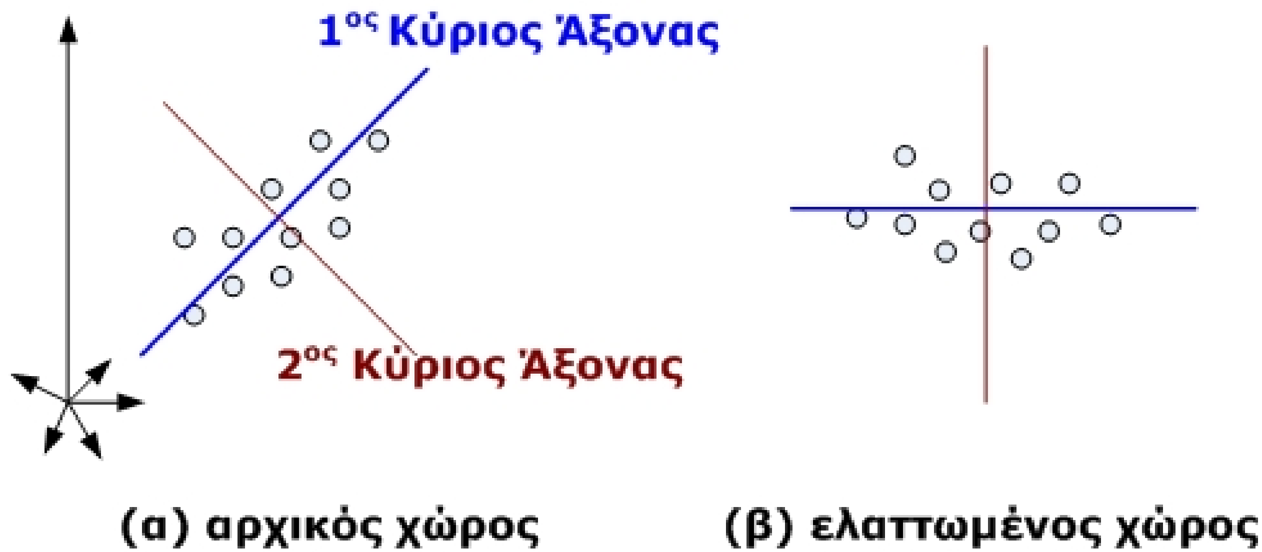
Principal Component Analysis

- Η Ανάλυση Κύριων Συνιστωσών (**P**rincipal **C**omponent **A**nalysis – PCA) προσπαθεί να υπολογίσει τους άξονες εκείνους στους οποίους παρατηρείται η μέγιστη διασπορά των δεδομένων.



Principal Component Analysis

- **Με μαθηματική διατύπωση:**
  - Βρες τα *ιδιοδιανύσματα του διαγωνοποιημένου πίνακα* συνδιασποράς *επί των αρχικών συντεταγμένων. Αυτά αποτελούν τους άξονες του μετασχηματισμένου χώρου και οι ιδιοτιμές αποδίδουν τη διασπορά κατά μήκος των αξόνων.*



Principal Component Analysis

Αλγόριθμος:

Για τα δεδομένα $\{X_1, X_2, \ldots, X_n\} \in R^D$

Principal Component Analysis

Αλγόριθμος:

Για τα δεδομένα $\{X_1, X_2, \ldots, X_n\} \in R^D$

- Γίνεται υπολογισμός του πίνακα συνδιασποράς $S = X \cdot X^T$

Principal Component Analysis

Αλγόριθμος:

Για τα δεδομένα $\{X_1, X_{2,} \dots, X_n\} \in R^D$

- Γίνεται υπολογισμός του πίνακα συνδιασποράς $S = X \cdot X^T$
- Υπολογίζεται η μέση τιμή τους μ

Principal Component Analysis

Αλγόριθμος:

Για τα δεδομένα $\{X_1, X_{2,} \ldots, X_n\} \in R^D$

- Γίνεται υπολογισμός του πίνακα συνδιασποράς $S = X \cdot X^T$
- Υπολογίζεται η μέση τιμή τους μ
- Υπολογίζονται οι ιδιοτιμές I_i και τα ιδιοδιανύσματα e_i μέσω της διαδικασίας ιδιοανάλυσης του S,

$$I_i \cdot e_i = S \cdot e_i$$

Principal Component Analysis

Αλγόριθμος:

Για τα δεδομένα $\{X_1, X_2, \ldots, X_n\} \in R^D$

- Γίνεται υπολογισμός του πίνακα συνδιασποράς $S = X \cdot X^T$
- Υπολογίζεται η μέση τιμή τους μ
- Υπολογίζονται οι ιδιοτιμές I_i και τα ιδιοδιανύσματα e_i μέσω της διαδικασίας ιδιοανάλυσης του S,

$$I_i \cdot e_i = S \cdot e_i$$

- Επιλέγονται τα d μεγαλύτερα ιδιοδιανύσματα και βάση αυτών υπολογίζονται οι νέες μεταβλητές,

$$Y_i = [e_1, e_2, \ldots, e_d]^T \times (X_i - \mu)$$

Principal Component Analysis

- Εφαρμογές:
  - Βήμα προεπεξεργασίας που προηγείται της εφαρμογής αλγορίθμων εξόρυξης δεδομένων (όπως clustering)
  - Οπτικοποίηση δεδομένων (Data Visualization)
  - Μείωση θορύβου (Noise Reduction)

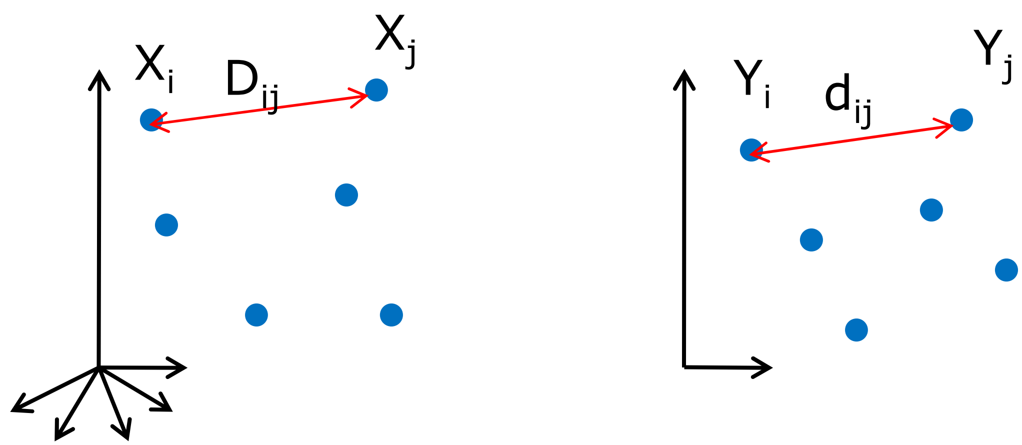
Γραμμικές τεχνικές μείωσης διαστάσεων

- Principal Component Analysis (PCA)

- Multidimensional Scaling (MDS)

Multidimensional Scaling

- Με την τεχνική αυτή απεικονίζονται τα δεδομένα σε ένα χώρο χαμηλών διαστάσεων με τέτοιο τρόπο ώστε οι γειτνιάσεις των δεδομένων να διατηρούνται με τον καλύτερο δυνατό τρόπο.



Multidimensional Scaling

Αλγόριθμος:

- Αρχικά τα δεδομένα $\{X_1, X_2, \ldots, X_n\} \in R^D$ απεικονίζονται σε χώρο k-διαστάσεων.

Multidimensional Scaling

Αλγόριθμος:

- Αρχικά τα δεδομένα $\{X_1, X_2, \ldots, X_n\} \in R^D$ απεικονίζονται σε χώρο k-διαστάσεων.
- Γίνεται ελαχιστοποίηση της συνάρτησης

$$stress = \sqrt{\frac{\sum_{ij}\left(D(X_i, X_j) - d(Y_i, Y_j)\right)^2}{\sum_{ij}\left(D(X_i, X_j)\right)^2}}$$

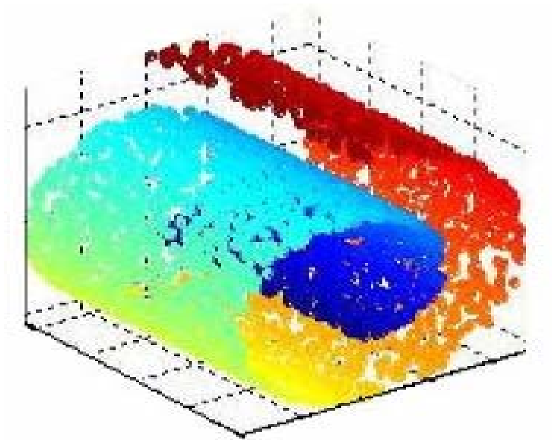
με τη μετακίνηση των σημείων στον ελαττωμένο χώρο. Συγκεκριμένα για κάθε σημείο ρυθμίζουμε τη θέση των υπολοίπων ώστε να ελαχιστοποιείται η συνάρτηση αυτή.

Γραμμικές τεχνικές μείωσης διαστάσεων

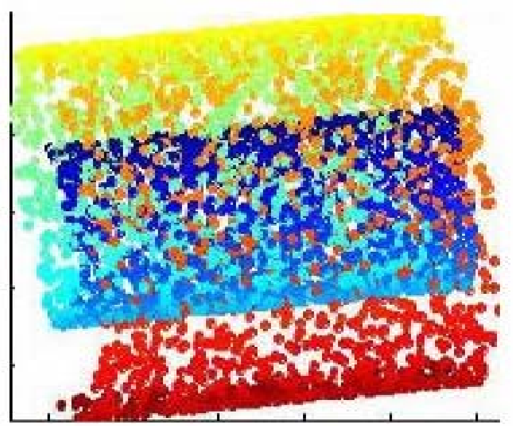
- Οι PCA και MDS είναι τεχνικές απλές στην υλοποίηση και αποδοτικές όταν ο αρχικός χώρος είναι γραμμικός ή σχεδόν γραμμικός.

Γραμμικές τεχνικές μείωσης διαστάσεων

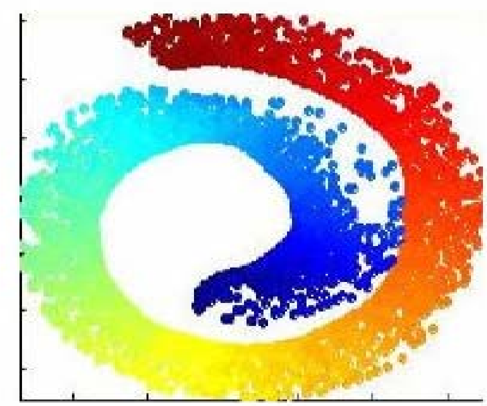
- Οι PCA και MDS είναι τεχνικές απλές στην υλοποίηση και αποδοτικές όταν ο αρχικός χώρος είναι γραμμικός ή σχεδόν γραμμικός.
- Αποτυγχάνουν όμως να ανακαλύψουν μη γραμμικές δομές στα δεδομένα



Swiss roll



PCA



MDS

Γραμμικές τεχνικές μείωσης διαστάσεων

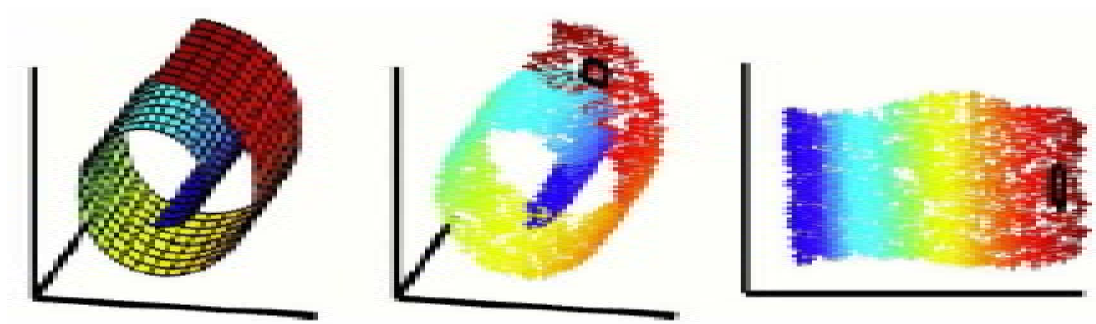
- Locally Linear Embedding (LLE)
- ISOMAP

Locally Linear Embedding

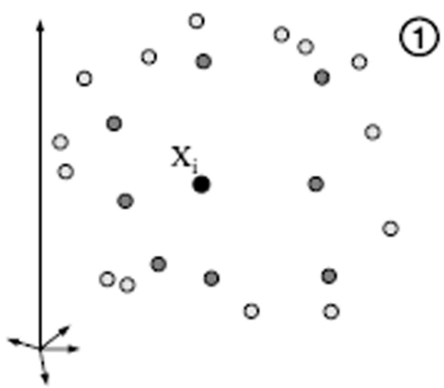
- Ο LLE είναι ένας μη γραμμικός αλγόριθμος ενσωμάτωσης σε χώρο χαμηλών διαστάσεων, όπου διατηρεί τη γενική μη γραμμικότητα της υπερεπιφάνειας εκμεταλλευόμενος την τοπικά γραμμική δομή.

Locally Linear Embedding

- Ο LLE είναι ένας μη γραμμικός αλγόριθμος ενσωμάτωσης σε χώρο χαμηλών διαστάσεων, όπου διατηρεί τη γενική μη γραμμικότητα της υπερεπιφάνειας εκμεταλλευόμενος την τοπικά γραμμική δομή.
- Θεωρεί ότι μια υπερεπιφάνεια είναι κατά προσέγγιση «γραμμική» αν αναφερθούμε μεμονωμένα σε ένα μικρό τμήμα της.



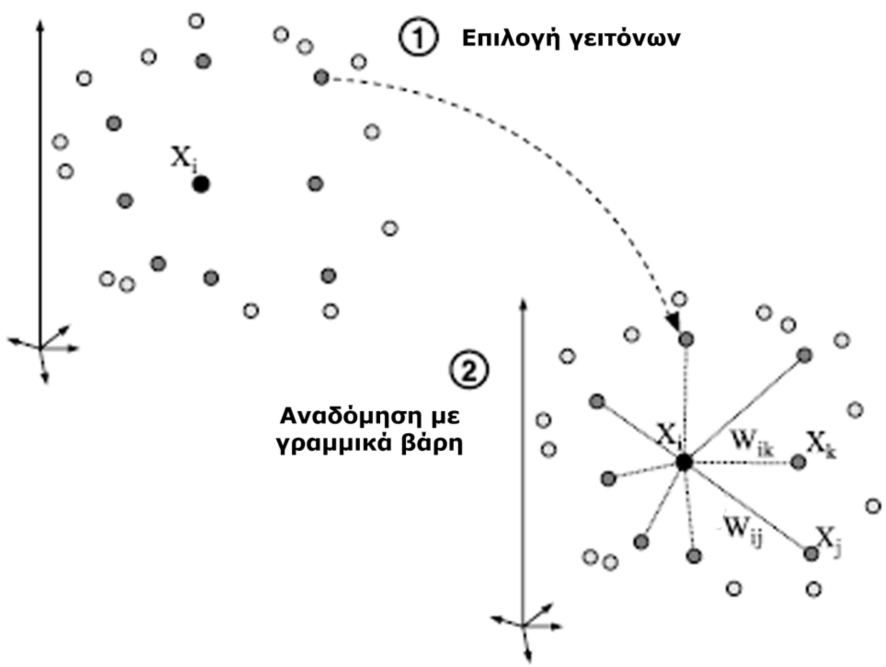
Locally Linear Embedding



Αλγόριθμος:

1.Εύρεση των k κοντινότερων γειτόνων κάθε σημείου X_i, στον χώρο R^D μέσω ευκλείδειων αποστάσεων.

Locally Linear Embedding



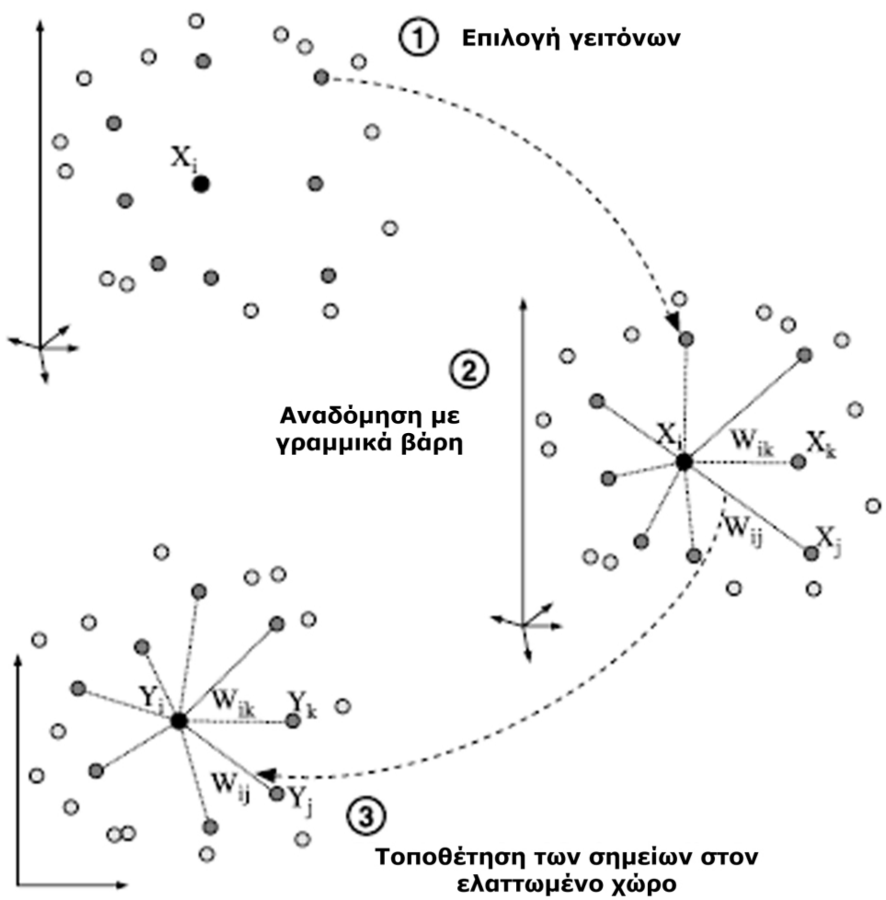
Αλγόριθμος:

1.Εύρεση των k κοντινότερων γειτόνων κάθε σημείου X_i, στον χώρο R^D μέσω ευκλείδειων αποστάσεων.

2.Υπολογισμός των βαρών W_{ij} που ανακατασκευάζουν κάθε σημείο X_i καλύτερα από τους γείτονές του.

$$X_i \approx \sum_j W_{ij} X_j$$

Locally Linear Embedding



Αλγόριθμος:

1.Εύρεση των k κοντινότερων γειτόνων κάθε σημείου X_i, στον χώρο R^D μέσω ευκλείδειων αποστάσεων.

2.Υπολογισμός των βαρών W_{ij} που ανακατασκευάζουν κάθε σημείο X_i καλύτερα από τους γείτονές του.

$$X_i \approx \sum_j W_{ij} X_j$$

3.Προσδιορισμός των σημείων Y_i, στον ελαττωμένο R^d χώρο, που ανακατασκευάζονται καλύτερα από τα βάρη W_{ij}

$$Y_i \approx \sum_j W_{ij} Y_j$$

Locally Linear Embedding

- Χαρακτηριστικά των βαρών ανακατασκευής W_{ij}:
  - Ανεξάρτητα περιστροφής, κλιμάκωσης και μετασχηματισμού (λόγω της συνθήκης

$$\sum_j W_{ij} = 1$$

Locally Linear Embedding

- Χαρακτηριστικά των βαρών ανακατασκευής W_{ij}:
  - Ανεξάρτητα περιστροφής, κλιμάκωσης και μετασχηματισμού (λόγω της συνθήκης

$$\sum_j W_{ij} = 1$$

  - Τα βάρη που υπολογίζονται στις αρχικές διαστάσεις, ανακατασκευάζουν τα σημεία και στον ελαττωμένο χώρο ενσωμάτωσης k-διαστάσεων.

Locally Linear Embedding

- Χαρακτηριστικά των βαρών ανακατασκευής W_{ij}:
  - Ανεξάρτητα περιστροφής, κλιμάκωσης και μετασχηματισμού (λόγω της συνθήκης

$$\sum_j W_{ij} = 1$$

  - Τα βάρη που υπολογίζονται στις αρχικές διαστάσεις, ανακατασκευάζουν τα σημεία και στον ελαττωμένο χώρο ενσωμάτωσης k-διαστάσεων.
  - Χαρακτηρίζουν τις εγγενείς γεωμετρικές ιδιότητες κάθε γειτονιάς σημείων.

Locally Linear Embedding

- Χαρακτηριστικά των βαρών ανακατασκευής W_{ij}:
  - Ανεξάρτητα περιστροφής, κλιμάκωσης και μετασχηματισμού (λόγω της συνθήκης

$$\sum_j W_{ij} = 1$$

  - Τα βάρη που υπολογίζονται στις αρχικές διαστάσεις, ανακατασκευάζουν τα σημεία και στον ελαττωμένο χώρο ενσωμάτωσης k-διαστάσεων.
  - Χαρακτηρίζουν τις εγγενείς γεωμετρικές ιδιότητες κάθε γειτονιάς σημείων.
- Τα βέλτιστα βάρη υπολογίζονται μέσω της ελαχιστοποίησης του σφάλματος ανακατασκευής

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2$$

Locally Linear Embedding

- Συνθήκες
  - $W_{ij}=0$ αν το X_j δεν είναι γείτονας του X_i
  - $\sum_j W_{ij} = 1$

Locally Linear Embedding

- Συνθήκες
  - $W_{ij}=0$ αν το X_j δεν είναι γείτονας του X_i
  - $\sum_j W_{ij} = 1$
- Η ελαχιστοποίηση του ε(W) και οι περιορισμοί, συνθέτουν ένα πρόβλημα «ελαχίστων τετραγώνων».

Locally Linear Embedding

- Συνθήκες
  - $W_{ij}=0$ αν το X_j δεν είναι γείτονας του X_i
  - $\sum_j W_{ij} = 1$
- Η ελαχιστοποίηση του ε(W) και οι περιορισμοί, συνθέτουν ένα πρόβλημα «ελαχίστων τετραγώνων».
- Οι συντεταγμένες στις d-διαστάσεις κάθε σημείου Y_i υπολογίζονται ελαχιστοποιώντας τις αντίστοιχες συναρτήσεις κόστους για τα δεδομένα βάρη

$$\Phi(W) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$$

Locally Linear Embedding

- Περιορισμοί
  - $\sum_i \vec{Y}_i = 0$ (μέση τιμή)
  - $\frac{1}{N}\sum_i \vec{Y}\vec{Y}^T = 1$ (διασπορά)

Locally Linear Embedding

- Περιορισμοί
  - $\sum_i \vec{Y}_i = 0$ (μέση τιμή)
  - $\frac{1}{N}\sum_i \vec{Y}\vec{Y}^T = 1$ (διασπορά)
- Αναλύοντας τη συνάρτηση κόστους έχουμε

$$\Phi(W) = \sum_{ij} M_{ij}\left(\vec{Y}_i \cdot \vec{Y}_j\right)$$

όπου

$$M = (I - W)^T (I - W)$$

Locally Linear Embedding

- Περιορισμοί
  - $\sum_i \vec{Y}_i = 0$ (μέση τιμή)
  - $\frac{1}{N}\sum_i \vec{Y}\vec{Y}^T = 1$ (διασπορά)
- Αναλύοντας τη συνάρτηση κόστους έχουμε

$$\Phi(W) = \sum_{ij} M_{ij}\left(\vec{Y}_i \cdot \vec{Y}_j\right)$$

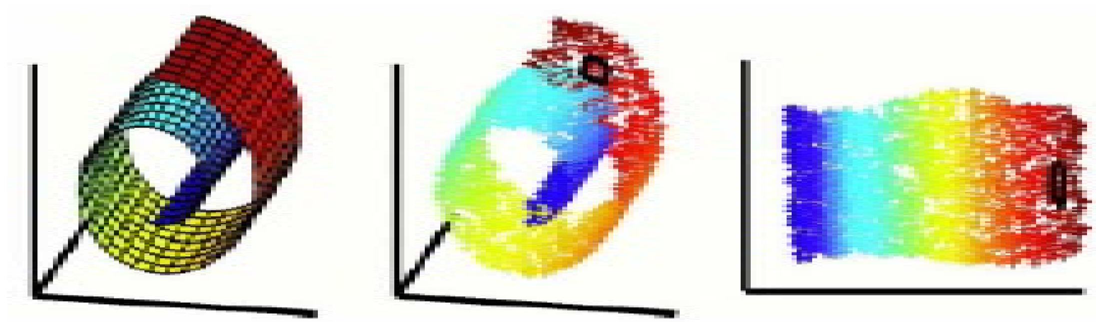
όπου

$$M = (I - W)^T (I - W)$$

- Το πρόβλημα ανάγεται σε πρόβλημα ιδιοτιμών του M
  - Η επιλογή των μικρότερων ιδιοτιμών ελαχιστοποιεί τη συνάρτηση κόστους.
  - Αγνοώντας το 1ο ιδιοδιανύσματα επιλέγονται τα επόμενα d

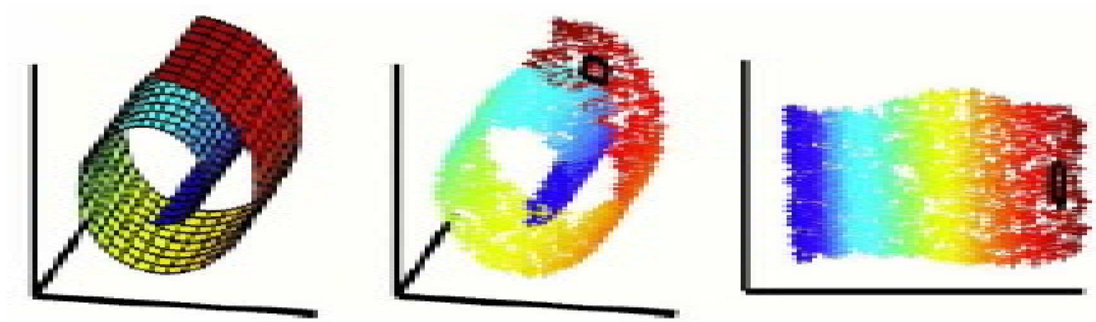
Locally Linear Embedding

- Πλεονεκτήματα
  - Διατήρηση των τοπικών γειτνιάσεων
  - Ικανότητα ανακάλυψης μη γραμμικών υπερεπιφανειών
  - Μη επαναληπτικός αλγόριθμος



Locally Linear Embedding

- Πλεονεκτήματα
  - Διατήρηση των τοπικών γειτνιάσεων
  - Ικανότητα ανακάλυψης μη γραμμικών υπερεπιφανειών
  - Μη επαναληπτικός αλγόριθμος
- Μειονεκτήματα
  - Απαιτεί ομαλές, μη κλειστές, πυκνά δειγματοληπτημένες υπερεπιφάνειες
  - Επιλογή γειτόνων
  - Ευαίσθητο σε απομακρυσμένα σημεία (outliers)



Γραμμικές τεχνικές μείωσης διαστάσεων

- Locally Linear Embedding (LLE)

- ISOMAP

ISOMAP

- Ο ISOMAP υπολογίζει τον χώρο ενσωμάτωσης Rd, διατηρώντας τις γεωδαιτικές αποστάσεις μεταξύ των σημείων της υπερεπιφάνειας RD

ISOMAP

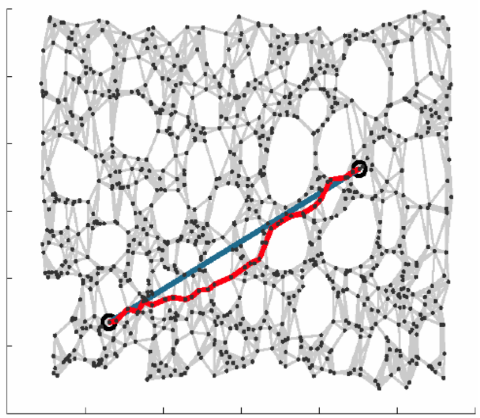
- Ο ISOMAP υπολογίζει τον χώρο ενσωμάτωσης Rd, διατηρώντας τις γεωδαιτικές αποστάσεις μεταξύ των σημείων της υπερεπιφάνειας RD
- Στον πίνακα γεωδαιτικών αποστάσεων, ο οποίος υπολογίζεται μεταξύ όλων των σημείων εφαρμόζεται ο κλασικός MDS αλγόριθμος

ISOMAP

- Ο ISOMAP υπολογίζει τον χώρο ενσωμάτωσης Rd, διατηρώντας τις γεωδαιτικές αποστάσεις μεταξύ των σημείων της υπερεπιφάνειας RD
- Στον πίνακα γεωδαιτικών αποστάσεων, ο οποίος υπολογίζεται μεταξύ όλων των σημείων εφαρμόζεται ο κλασικός MDS αλγόριθμος
- Διατηρεί την εγγενή γεωμετρία των δεδομένων

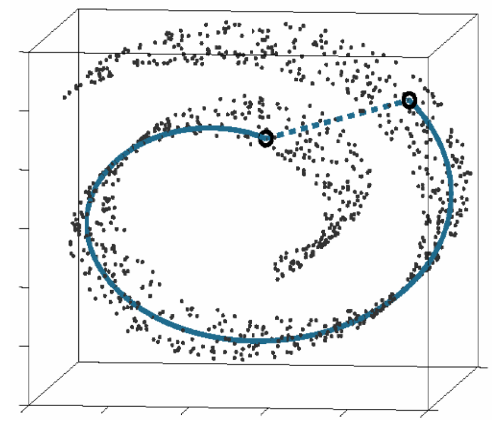
ISOMAP

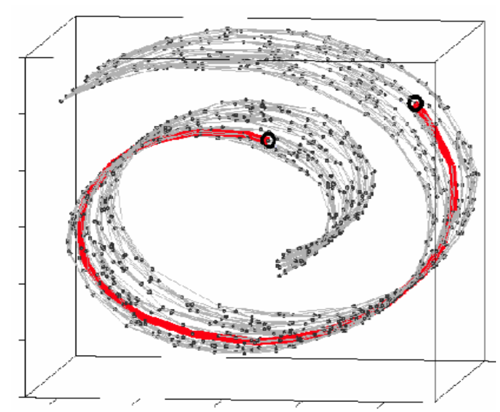
- Γεωδαιτική απόσταση:
  - Είναι το μήκος της μικρότερης καμπύλης που ενώνει δύο απομακρυσμένα σημεία μιας υπερεπιφάνειας
  - Για γειτονικά σημεία, η γεωδαιτική τους απόσταση ταυτίζεται ικανοποιητικά με την ευκλείδεια απόστασή τους.



ISOMAP

- Γεωδαιτική απόσταση:
  - Είναι το μήκος της μικρότερης καμπύλης που ενώνει δύο απομακρυσμένα σημεία μιας υπερεπιφάνειας
  - Για γειτονικά σημεία, η γεωδαιτική τους απόσταση ταυτίζεται ικανοποιητικά με την ευκλείδεια απόστασή τους.
  - Για απομακρυσμένα σημεία, η απόσταση προσδιορίζεται από μια ακολουθία μικρών βημάτων μεταξύ γειτονικών σημείων. Δημιουργείται από ένωση των ακμών μεταξύ των γειτονικών σημείων.





ISOMAP

Αλγόριθμος:

- Προσδιορισμός των γειτόνων κάθε σημείου Z_i
  - Όλα τα σημεία Z_j εντός σφαίρας ακτίνας ε
  - k κοντινότεροι γείτονες

ISOMAP

Αλγόριθμος:

- Προσδιορισμός των γειτόνων κάθε σημείου Z_i
  - Όλα τα σημεία Z_j εντός σφαίρας ακτίνας ε
  - k κοντινότεροι γείτονες
- Κατασκευή του γράφου γειτονίας G
  - Κάθε σημείο ενώνεται με ευκλείδεια ακμή $d_x(i, j)$ με τα γειτονικά του σημεία
  - Δημιουργία του πίνακα αποστάσεων $D_x=\{d_x(i, j)\}$

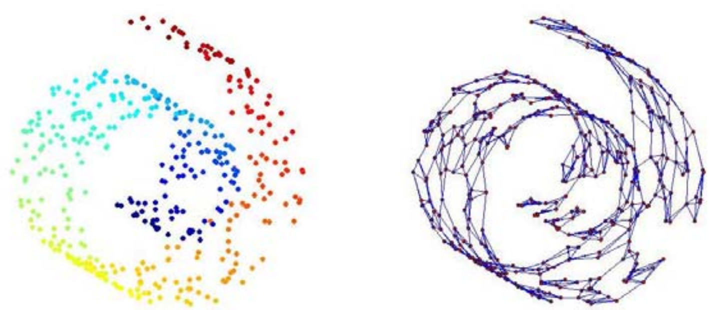
ISOMAP

Αλγόριθμος:

- Προσδιορισμός των γειτόνων κάθε σημείου Z_i
  - Όλα τα σημεία Z_j εντός σφαίρας ακτίνας ε
  - k κοντινότεροι γείτονες
- Κατασκευή του γράφου γειτονίας G
  - Κάθε σημείο ενώνεται με ευκλείδεια ακμή $d_x(i, j)$ με τα γειτονικά του σημεία
  - Δημιουργία του πίνακα αποστάσεων $D_x=\{d_x(i, j)\}$
- Υπολογισμός των αποστάσεων μεταξύ όλων των σημείων πάνω στον γράφο G και εφαρμογή του κλασικού MDS.
  - Αλγόριθμος Dijkstra για τον υπολογισμό των $d_G(i, j)$ γεωδαιτικών αποστάσεων
  - Υπολογισμός του πίνακα γεωδαιτικών αποστάσεων $D_G=\{d_G(i, j)\}$
  - Εφαρμογή του κλασικού MDS στον πίνακα D_G
  - Με λύση του προβλήματος ιδιοτιμών που προκύπτει βρίσκεται η ενσωμάτωση των σημείων στον ελαττωμένο χώρο.

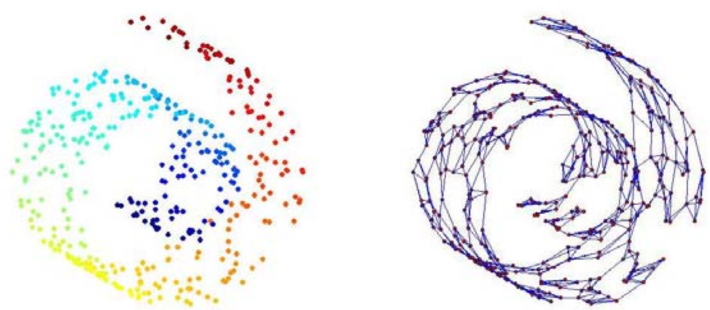
ISOMAP

- Πλεονεκτήματα:
  - Μη γραμμικός, μη επαναληπτικός αλγόριθμος
  - Υπολογιστική αποδοτικότητα

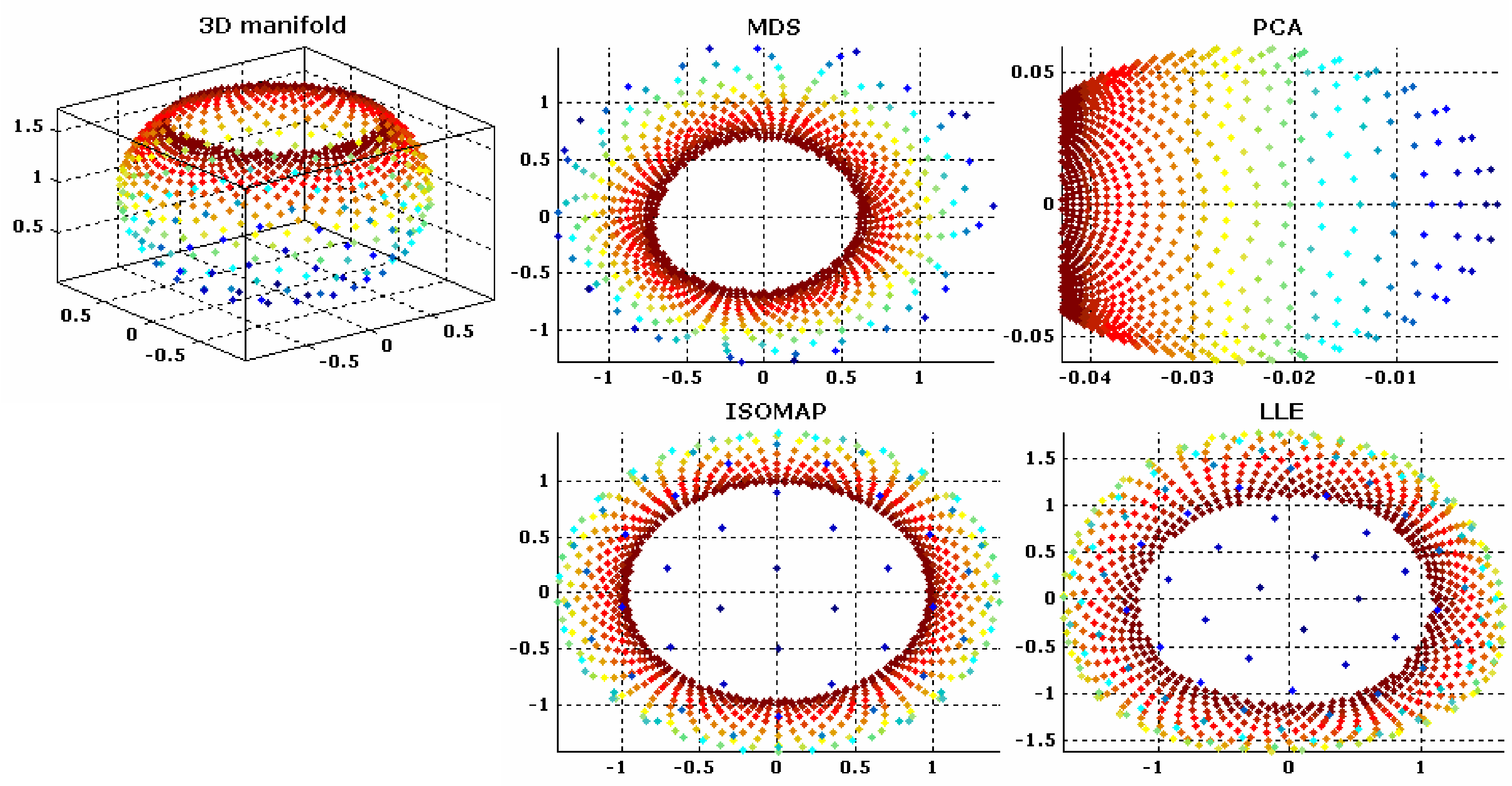


ISOMAP

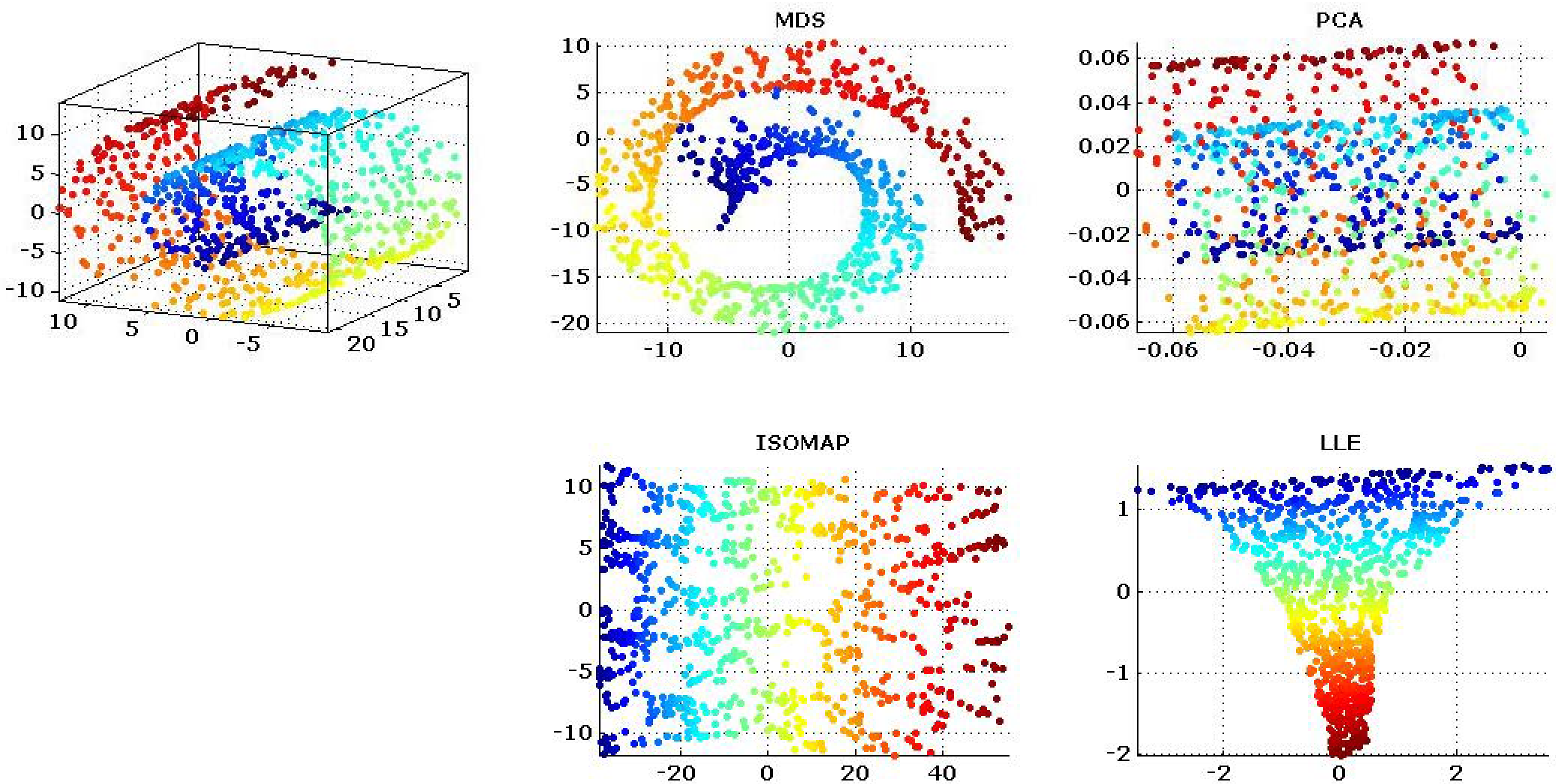
- Πλεονεκτήματα:
  - Μη γραμμικός, μη επαναληπτικός αλγόριθμος
  - Υπολογιστική αποδοτικότητα
- Μειονεκτήματα:
  - Μικρός αριθμός δειγμάτων οδηγεί σε ανακριβή υπολογισμό της γεωδαιτικής απόστασης
  - Μεγάλη καμπυλότητα της υπερεπιφάνειας, απαιτεί μεγάλο αριθμό γειτόνων για τον εντοπισμό της



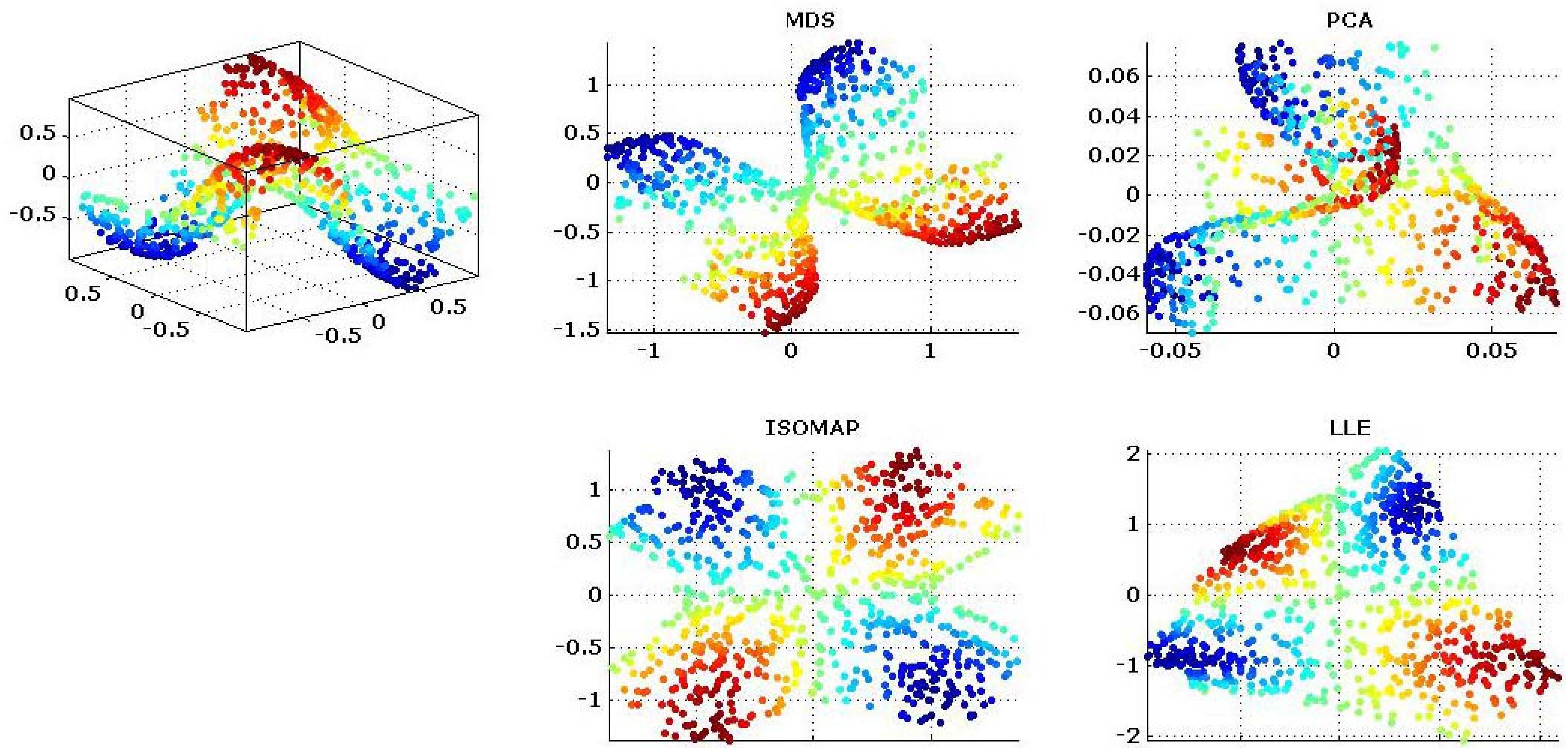
Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



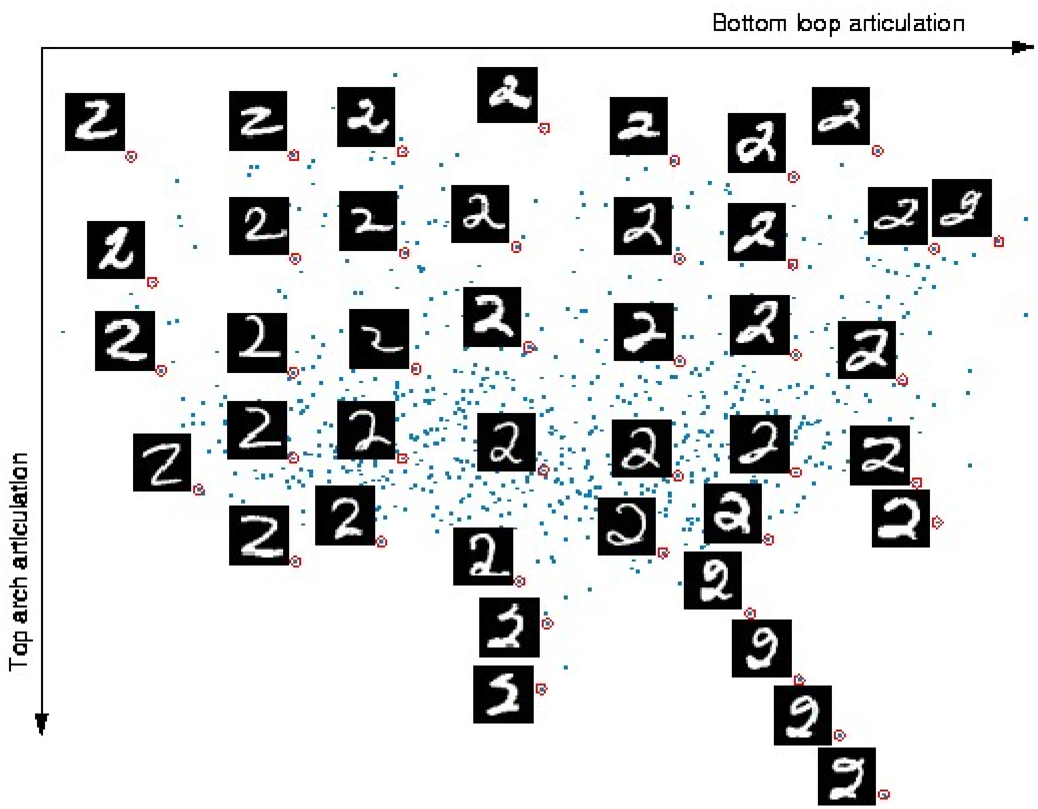
Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



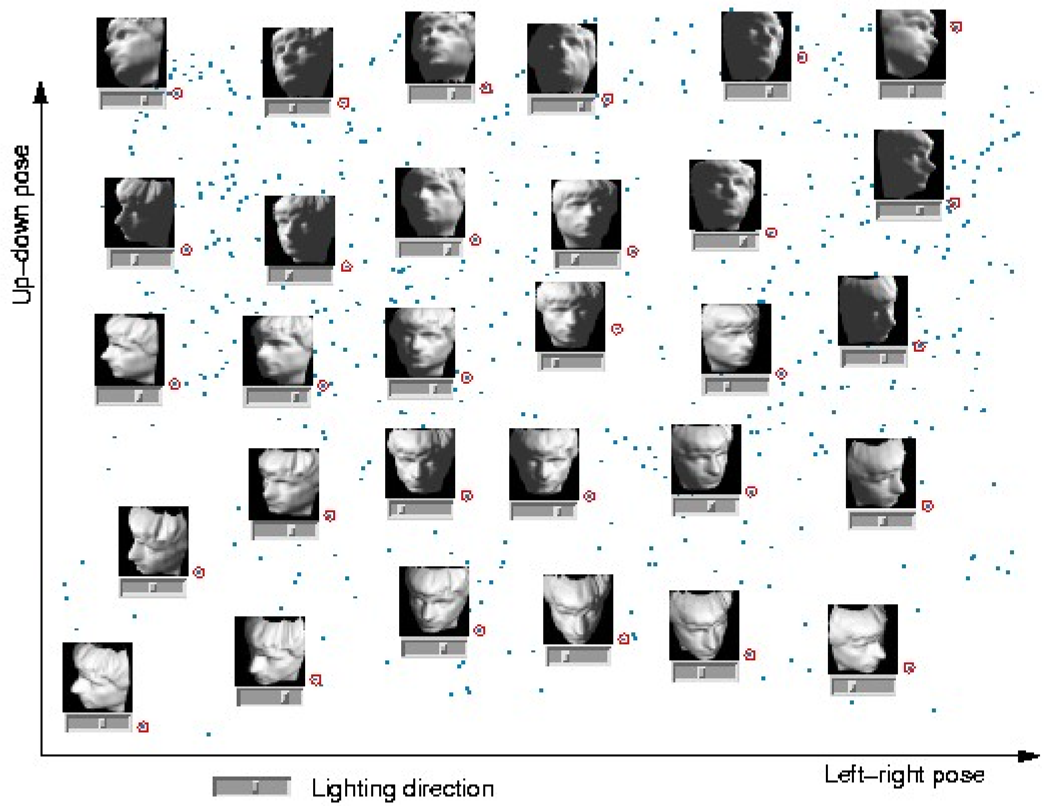
Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



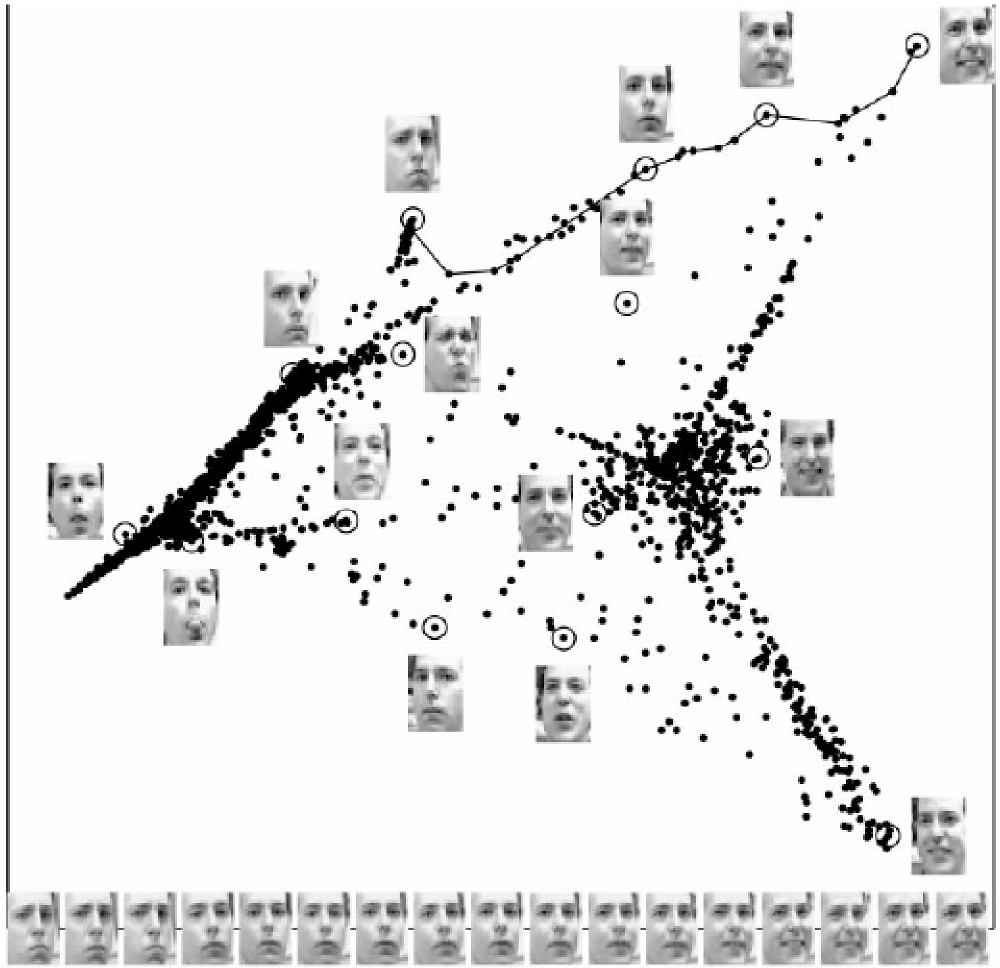
Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



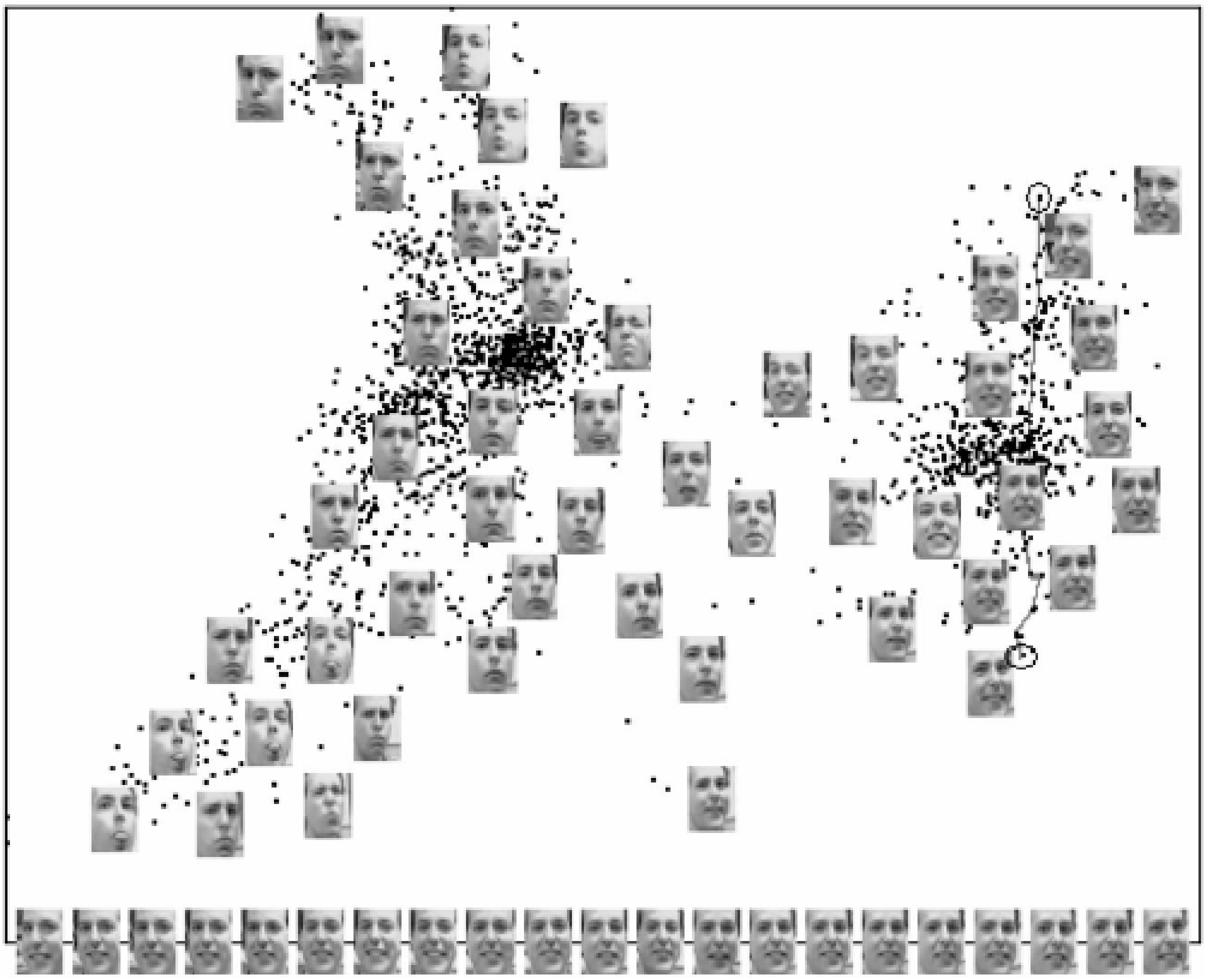
Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων

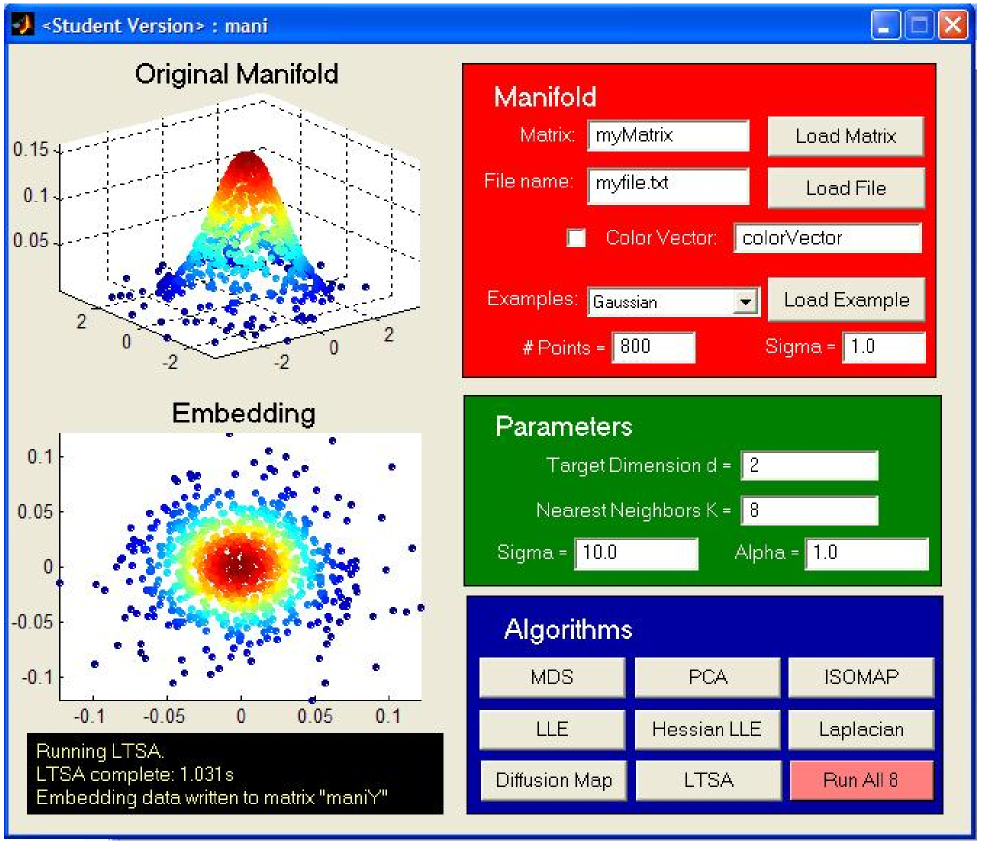


Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων



Παραδείγματα εφαρμογής τεχνικών μείωσης διαστάσεων

Εφαρμογής τεχνικών μείωσης διαστάσεων στο matlab



http://www.math.umn.edu/~wittman/mani/

Ευχαριστώ