



Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

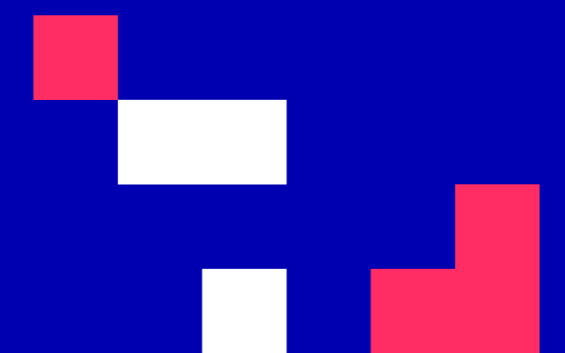
Διάλεξη 13: Μείωση των διαστάσεων

Βασίλης Βασιλειάδης, PhD

Χειμερινό Εξάμηνο 2022/23



CYENS
CENTRE OF EXCELLENCE



MAI4CAREU

Master programmes in Artificial
Intelligence 4 Careers in Europe



Επανάληψη





Εισαγωγή στη βαθιά μάθηση

- Η βαθιά μάθηση αφορά την εκμάθηση διαδοχικών επιπέδων αναπαραστάσεων
 - Χρησιμοποιώντας NNs με περισσότερα από 2 κρυμμένα στρώματα
- Το έχουμε τώρα κυρίως λόγω των εξελίξεων του υλικού (GPUs) και της αφθονίας των δεδομένων
- Όταν θέλουμε να εντοπίσουμε μια συγκεκριμένη ιδέα που θα μπορούσε να βρίσκεται σε διαφορετικά σημεία της εισόδου, χρησιμοποιούμε την κατανομή βάρους, δηλαδή, κατασκευάζουμε έναν ενιαίο ανιχνευτή χαρακτηριστικών για αυτή την έννοια, εκπαιδεύοντας τα βάρη αυτών των εισόδων από κοινού.
 - Αυτό βοηθά στη γενίκευση
 - Παράδειγμα 1: δημιουργία ενός ταξινομητή εικόνας σκυλιών: ένας σκύλος μπορεί να εμφανιστεί παντού σε μια εικόνα
 - Παράδειγμα 2: δίκτυο ολοκλήρωσης κειμένου: θέλουμε το μέρος του NN που μαθαίνει τι είναι ένας σκύλος να επαναχρησιμοποιείται κάθε φορά που το NN βλέπει τη λέξη σκύλος.





Εισαγωγή στη βαθιά μάθηση

- Τα συνελευτικά δίκτυα είναι πιο κατάλληλα για δεδομένα εικόνας
- Χρησιμοποιούν τοπικά δεκτικά πεδία (φίλτρα) και τα μετατοπίζουν (συγκέντρωση) πάνω από το χάρτη ενεργοποίησης του προηγούμενου στρώματος για να δημιουργήσουν το χάρτη ενεργοποίησης του τρέχοντος στρώματος
 - Αυτό μειώνει τον αριθμό των παραμέτρων σε σύγκριση με τα πλήρως συνδεδεμένα δίκτυα τροφοδοσίας
 - Κάθε φίλτρο γίνεται ανιχνευτής χαρακτηριστικών πάνω από διαφορετικά μέρη της εισόδου (μετάφραση αναλλοίωτη):
 - Στρώμα 1: ανιχνευτές ακρών
 - Στρώμα 2: ανιχνευτές γωνιών
 - Στρώμα 3: ανιχνευτές μερών των αντικειμένων
 - Στρώμα 4: ανιχνευτές πλήρων αντικειμένων





Εισαγωγή στη βαθιά μάθηση

- Διαδοχικά δεδομένα: δεδομένα διατεταγμένα σε ακολουθίες
 - Τυπικά: δεδομένα χρονολογικών σειρών
- Ένας τρόπος χειρισμού των διαδοχικών δεδομένων χρησιμοποιώντας ΝΔ είναι χρησιμοποιώντας συνδέσεις ανατροφοδότησης (καθυστερημένες): Επαναλαμβανόμενα ΝΔ (ΕΝΔ)
 - Ένα επαναλαμβανόμενο ΝΔ δημιουργεί τη δική του εσωτερική αναπαράσταση του χρόνου
- Μπορούμε να εκπαιδεύσουμε ΕΝΔ χρησιμοποιώντας οπισθοπολλαπλασιασμό μέσα στο χρόνο
 - Ξετυλίγουμε το ΕΝΔ με την πάροδο του χρόνου και προπαγανδίζουμε τα λάθη από το τελευταίο βήμα στο πρώτο
 - Συσσωρεύουμε τις κλίσεις και εφαρμόζουμε κάθοδο κλίσης
 - Τα unrolled ΕΝΔ μπορούν να γίνουν πολύ βαθιά δίκτυα
 - Όταν το κάνουμε αυτό μπορεί να έχουμε τα προβλήματα εξαφάνισης ή έκρηξης gradient





Εισαγωγή στη βαθιά μάθηση

- Echo state networks :
 - χρησιμοποιήστε ένα μεγάλο αραιά συνδεδεμένο κρυφό στρώμα που δεν είναι εκπαιδευμένο
 - δεν χρησιμοποιούν backprop μέσα στο χρόνο
 - δεν έχουν τα προβλήματα εξαφάνισης ή έκρηξης gradient
 - μπορούμε να υπολογίσουμε την αναλυτική λύση για ένα πρόβλημα παλινδρόμησης (παρόμοια με τη γραμμική παλινδρόμηση)
 - καλή απόδοση σε εργασίες που απαιτούν γρήγορη, προσαρμοστική εκπαίδευση
 - μη καλή απόδοση σε εργασίες με πολλές μεταβλητές και μακροπρόθεσμες εξαρτήσεις





Εισαγωγή στη βαθιά μάθηση

- Long short-term memory networks:
 - Χρησιμοποιήστε μηχανισμούς περιτύλιξης που επιτρέπουν στο δίκτυο να μάθει τι να ξεχάσει, τι να αποθηκεύσει στη μνήμη και τι να εξάγει
 - Gating: σιγμοειδές πολλαπλασιασμένο με το σήμα: το σιγμοειδές ρυθμίζει πόσο από το σήμα επιτρέπεται να περάσει
 - Μπορεί να εφαρμοστεί σε εργασίες που απαιτούν μακροχρόνιες εξαρτήσεις
- Δεδομένα κειμένου:
 - One-hot word representation: αραιή, υψηλή διάσταση, δεν γενικεύει καλά
 - Η λέξη ενσωματώνει: μαθημένη, αριθμητική διανυσματική αναπαράσταση
 - Προσπαθήστε να αποτυπώσετε το νόημα των λέξεων με βάση τη χρήση τους σε προτάσεις
 - Λέξεις με παρόμοια σημασία έχουν παρόμοιες αναπαραστάσεις διανυσμάτων
 - Βασιλιάς - Άνδρας + Γυναίκα = Βασίλισσα





Clustering

- Το clustering είναι το πρόβλημα του clustering δεδομένων με παρόμοια χαρακτηριστικά
 - Δεν έχουμε ετικέτες που καθορίζουν τις σωστές εξόδους
- Το K-means clustering χρησιμοποιεί έναν προκαθορισμένο αριθμό (k) των συστιάδων
 - Τυχαία αρχικοποιεί τα centroid συστιάδας k
 - Εναλλάσσεται μεταξύ της εκχώρησης κάθε σημείου σε ένα centroid και της επικαιροποίησης των centroid
- Μπορούμε να αποφύγουμε την τοπική optima με την εκτέλεση του k-means πολλές φορές και επιλέγοντας την ομαδοποίηση με το χαμηλότερο κόστος
- Επιλέγουμε το k χρησιμοποιώντας τη γνώση τομέα, τη μέθοδο αγκώνα ή το σκορ Silhouette
- Η ομαδοποίηση μπορεί να βοηθήσει την εποπτευόμενη μάθηση, π.χ.
 - Εύρεση των αρχικών κέντρων των δικτύων RBF
 - Δυνατότητα χρήσης τόσο των δεδομένων με επισήμανση όσο και των μη επισημασμένων δεδομένων για τη βελτίωση της γενίκευσης





Διάλεξη 13: Μείωση των διαστάσεων

Μαθησιακά αποτελέσματα

Θα μάθετε για:

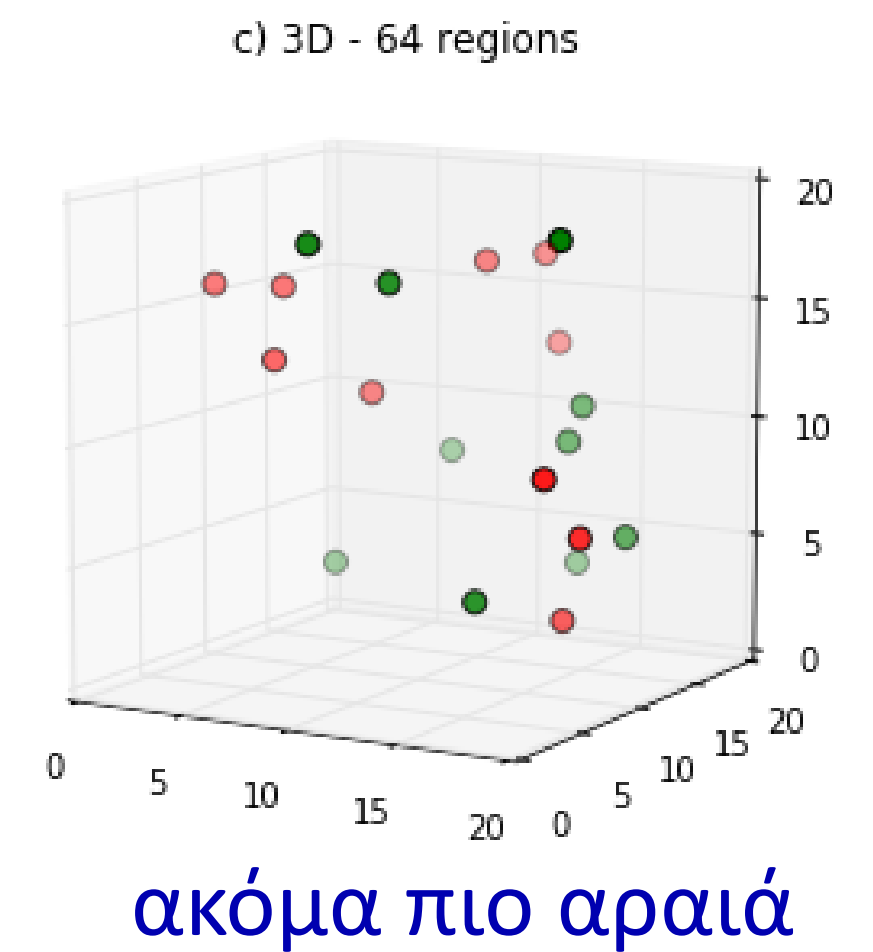
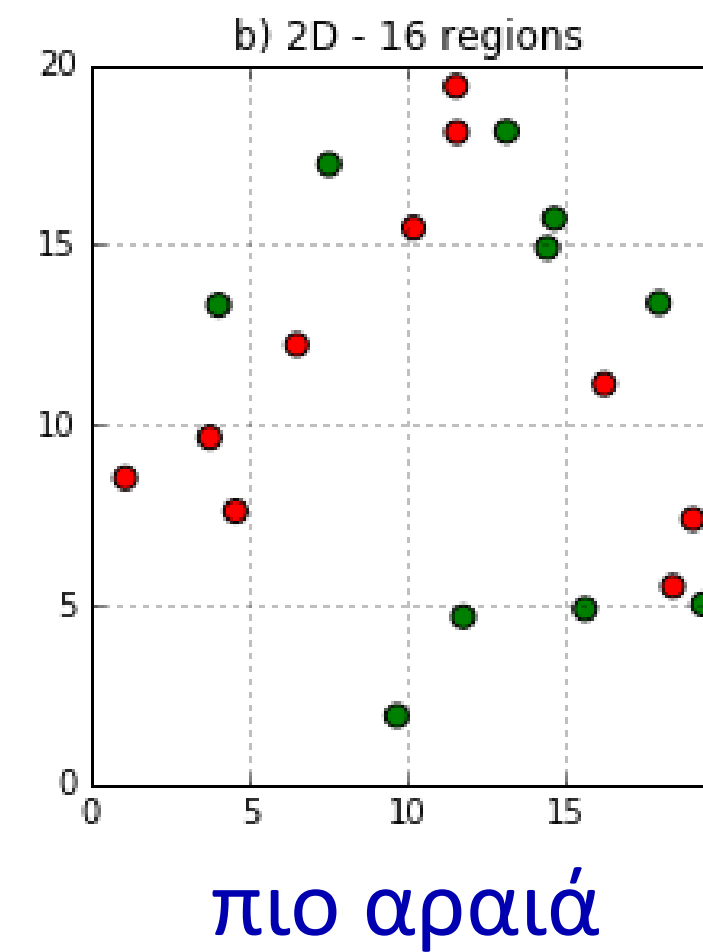
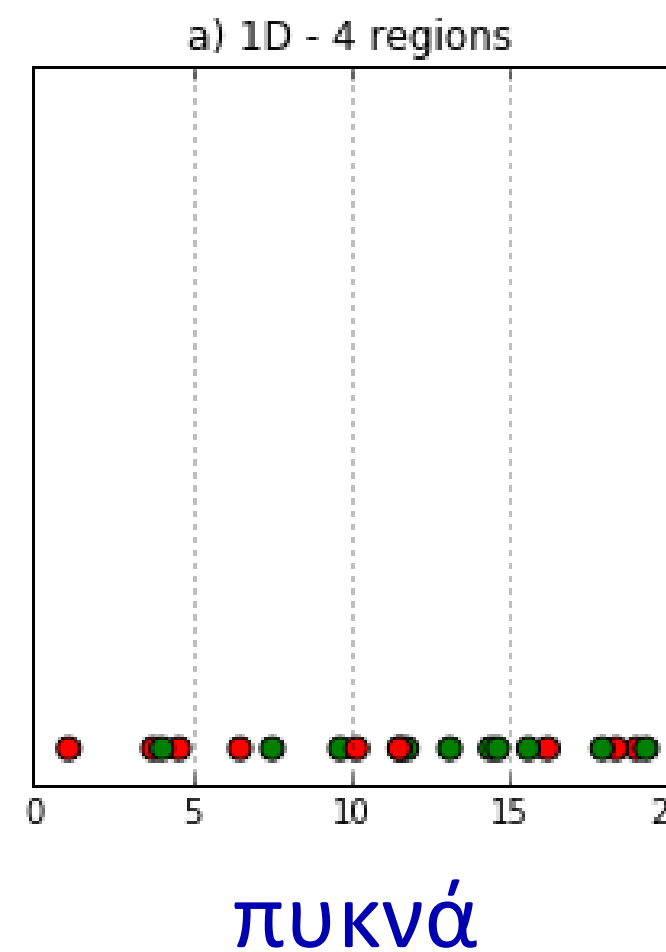
1. Το πρόβλημα της μείωσης των διαστάσεων
2. Ο αλγόριθμος Principal Components Analysis (PCA)
3. Μη γραμμική μείωση της διάστασης με τη χρήση του πυρήνα PCA και Autoencoders
4. Πολλαπλές προσεγγίσεις μάθησης
5. Το t-distributed stochastic neighbor embedding (t-SNE)
για την απεικόνιση των υψηλών διαστάσεων δεδομένων





Η κατάρα των διαστάσεων

- Δεδομένου του ίδιου αριθμού δειγμάτων, καθώς αυξάνεται η διάσταση των δεδομένων, γίνονται **εκθετικά αραιά**
- Κάνει τα δεδομένα να είναι πιο **εύκολα διαχωρίσιμα**



- **Ωστόσο,:**
 - **Πιο εύκολο να** υπερπροσαρμοστεί σε σύνολα δεδομένων υψηλής διάστασης
 - Ανάγκη **εκθετικά μεγάλου αριθμού δειγμάτων** για να καλύψει το διάστημα και να αποφύγει την υπερπροσαρμογή
 - **Οι μετρήσεις απόστασης δεν έχουν νόημα** σε υψηλές διαστάσεις (π.χ. αλγόριθμοι όπως K-means clustering ή K-Nearest Neighbors δεν μπορούν να αντιμετωπίσουν δεδομένα υψηλής διάστασης)





Κίνητρο μείωσης των διαστάσεων

- Βρείτε σημαντικές δομές χαμηλών διαστάσεων κρυμμένες σε παρατηρήσεις υψηλών διαστάσεων
 - Ο ανθρώπινος εγκέφαλος αντιμετωπίζει το ίδιο πρόβλημα στην αντίληψη:
 - 30,000 ακουστικές νευρικές ίνες
 - 10^6 οπτικές ίνες νεύρων
 - Εξαγωγή μικρού αριθμού σχετικών με την αντίληψη χαρακτηριστικών
- Δύσκολο να απεικονιστούν δεδομένα σε διαστάσεις μεγαλύτερες από τρεις





Τι είναι η Μείωση των διαστάσεων

- Μετατροπή των δεδομένων από έναν ορατό χώρο υψηλών διαστάσεων $x \in \mathbb{R}^n$, σε ένα χαμηλής διάστασης λανθάνον χώρο $z \in \mathbb{R}^k$, $k < n$
- Η χαμηλής διάστασης αναπαράσταση διατηρεί κάποιες σημαντικές ιδιότητες των αρχικών δεδομένων
- Η χαρτογράφηση μπορεί να είναι:
 - **Μη** παραμετρικό, όπου υπολογίζουμε μια **ενσωμάτωση** $z^{(i)}$ για κάθε είσοδο $x^{(i)}$ στο σύνολο δεδομένων, αλλά όχι για οποιαδήποτε άλλα σημεία.
 - Αυτό χρησιμοποιείται κυρίως για την οπτικοποίηση και την **ερμηνευσιμότητα** ($k = 2$ or $k = 3$) **των δεδομένων**
 - Ένα **παραμετρικό** μοντέλο $z = f_{\theta}(x)$, το οποίο μπορεί να εφαρμοστεί σε οποιαδήποτε είσοδο. Μπορέστε να χρησιμοποιηθείτε:
 - Για **συμπίεση δεδομένων** ($1 \leq k < n$)
 - Ως **στάδιο προεπεξεργασίας** για άλλους αλγόριθμους μάθησης (π.χ. εποπτευόμενη μάθηση ή ομαδοποίηση)
 - Για **οπτικοποίηση δεδομένων**
- Όταν μιλάμε για αμυδρή μείωση, συνήθως εννοούμε **εξαγωγή χαρακτηριστικών** (όχι επιλογή)





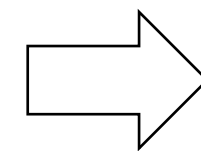
Οπτικοποίηση δεδομένων

Γιατί;

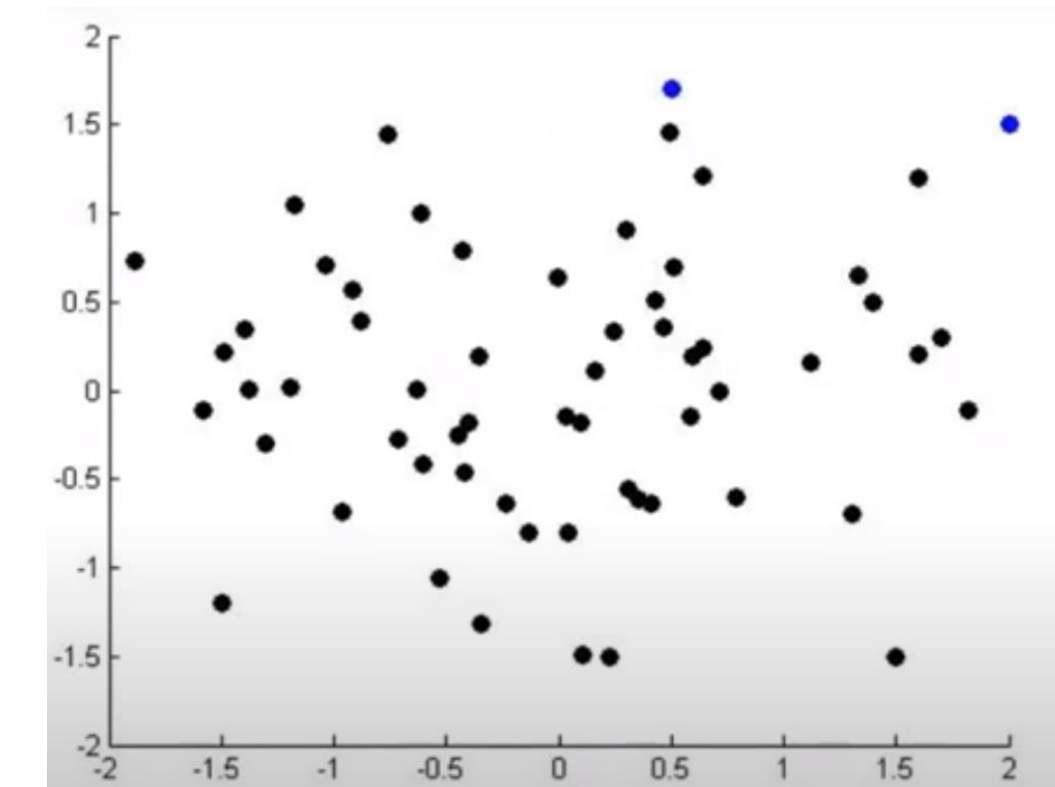
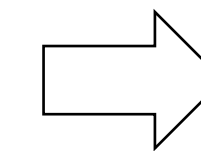
- Κατανόηση της σχέσης μεταξύ των δεδομένων (π.χ., εάν υπάρχουν συστάδες)
- Δυνατότητα ερμηνείας: μπορεί να είναι μια απαίτηση σε ορισμένους τομείς (π.χ., ιατρική, κυβερνητική)

Για παράδειγμα: Σύνολο δεδομένων με στατιστικά στοιχεία και στοιχεία για διαφορετικές χώρες στον κόσμο

x_1 ΑΕΠ
 x_2 Κατά κεφαλήν ΑΕΠ
 x_3 Προσδόκιμο ζωής
 ...
 x_{50} Μέσο εισόδημα νοικοκυριού



z_1 Μέγεθος χώρας/ΑΕΠ
 z_2 Κατά κεφαλήν ΑΕΠ





Συμπύεση δεδομένων

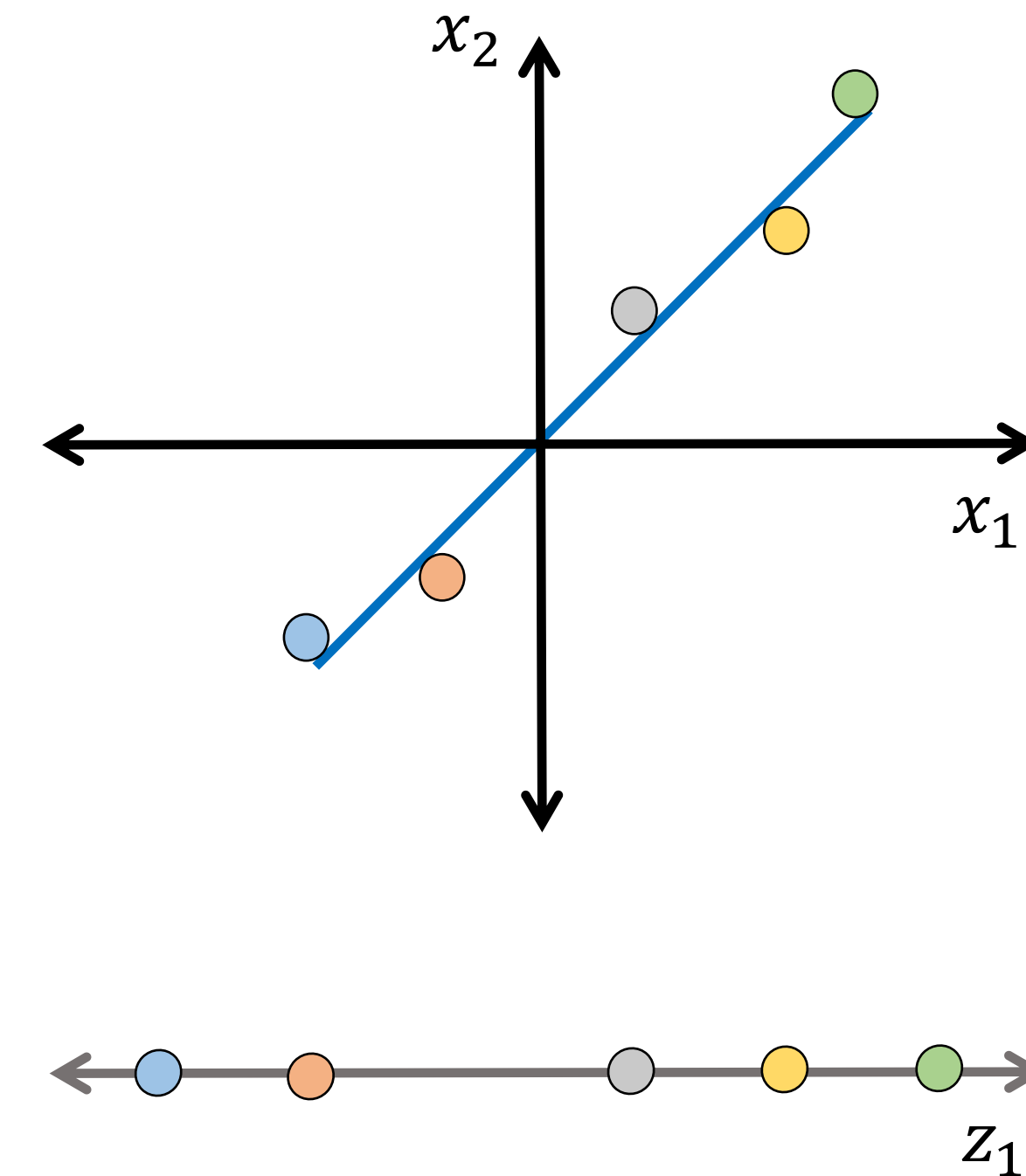
Πλεονεκτήματα

- Μειώστε τα περιττά χαρακτηριστικά και τον θόρυβο
- Μειώστε τη μνήμη που απαιτείται για την αποθήκευση των δεδομένων
- Μειώστε την υπερπροσαρμογή
- Κάνουν τους αλγόριθμους MM υπολογιστικά αποδοτικούς (π.χ., SVMs)

Μειονεκτήματα

- Μπορεί να οδηγήσει σε κάποια απώλεια δεδομένων
- Μπορεί να οδηγήσει σε μειωμένη ακρίβεια

Για παράδειγμα: Μειώστε τα δεδομένα από 2D σε 1D





Principal Component Analysis





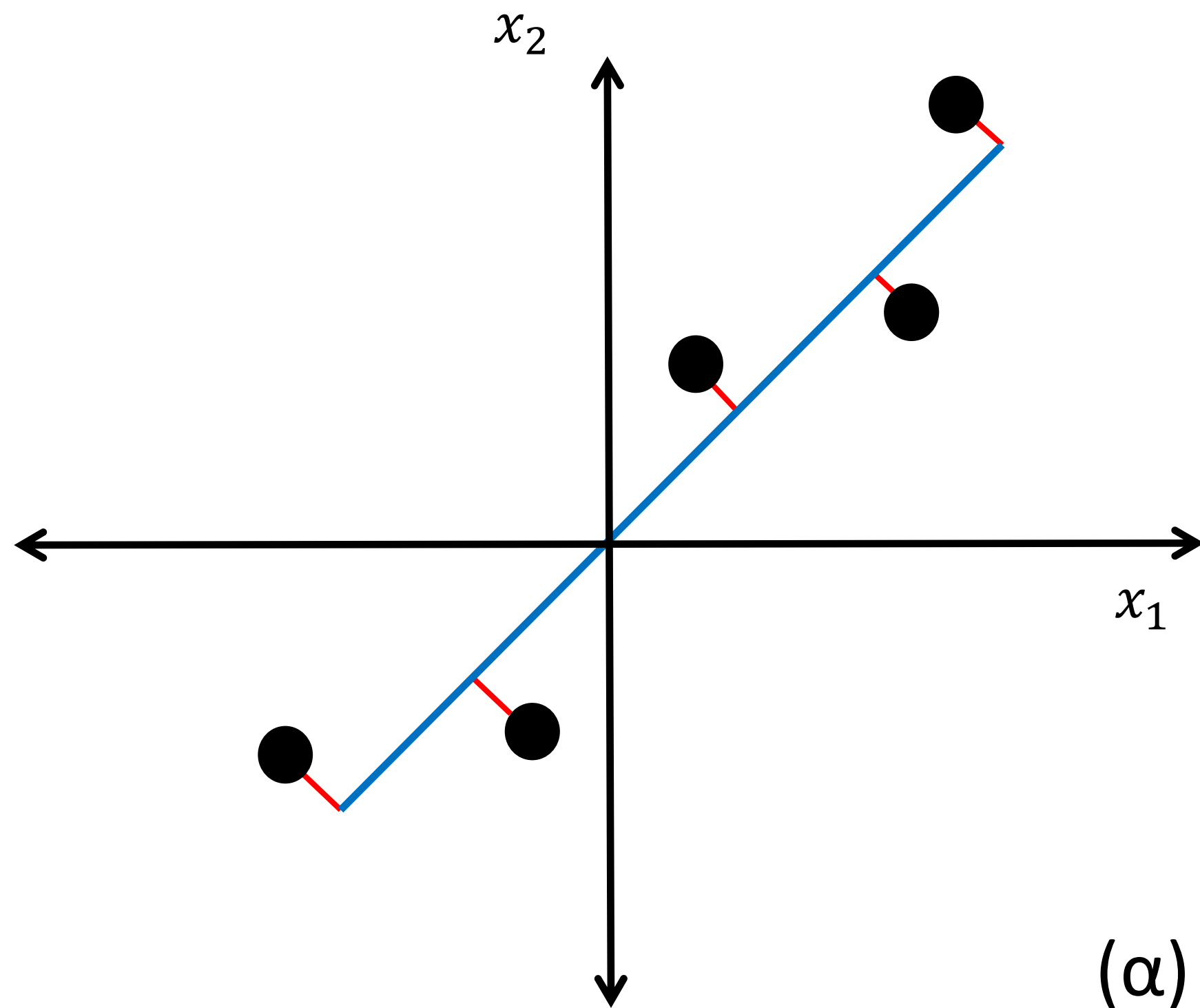
Principal Component Analysis (PCA)

- Η απλούστερη και πιο ευρέως χρησιμοποιούμενη μορφή εξαγωγής χαρακτηριστικών
- **Γραμμικός μετασχηματισμός** των δεδομένων σε ένα **νέο σύστημα συντεταγμένων** όπου το μεγαλύτερο μέρος της διακύμανσης των δεδομένων μπορεί να περιγραφεί με λιγότερες διαστάσεις από τα αρχικά δεδομένα
- Βρίσκει έναν γραμμικό συνδυασμό χαρακτηριστικών που **συνοψίζει καλύτερα τα δεδομένα**
- Ποια είναι η καλύτερη (γραμμική) «περίληψη»;
 - Ένα που επιτρέπει την ανακατασκευή των αρχικών δεδομένων όσο το δυνατόν καλύτερα
 - Ένα που μεγιστοποιεί τη διακύμανση των χαμηλών διαστάσεων δεδομένων
 - Η διακύμανση είναι πληροφορία!
 - Αυτά είναι **ισοδύναμα**, έτσι το PCA κάνει και τα δύο!



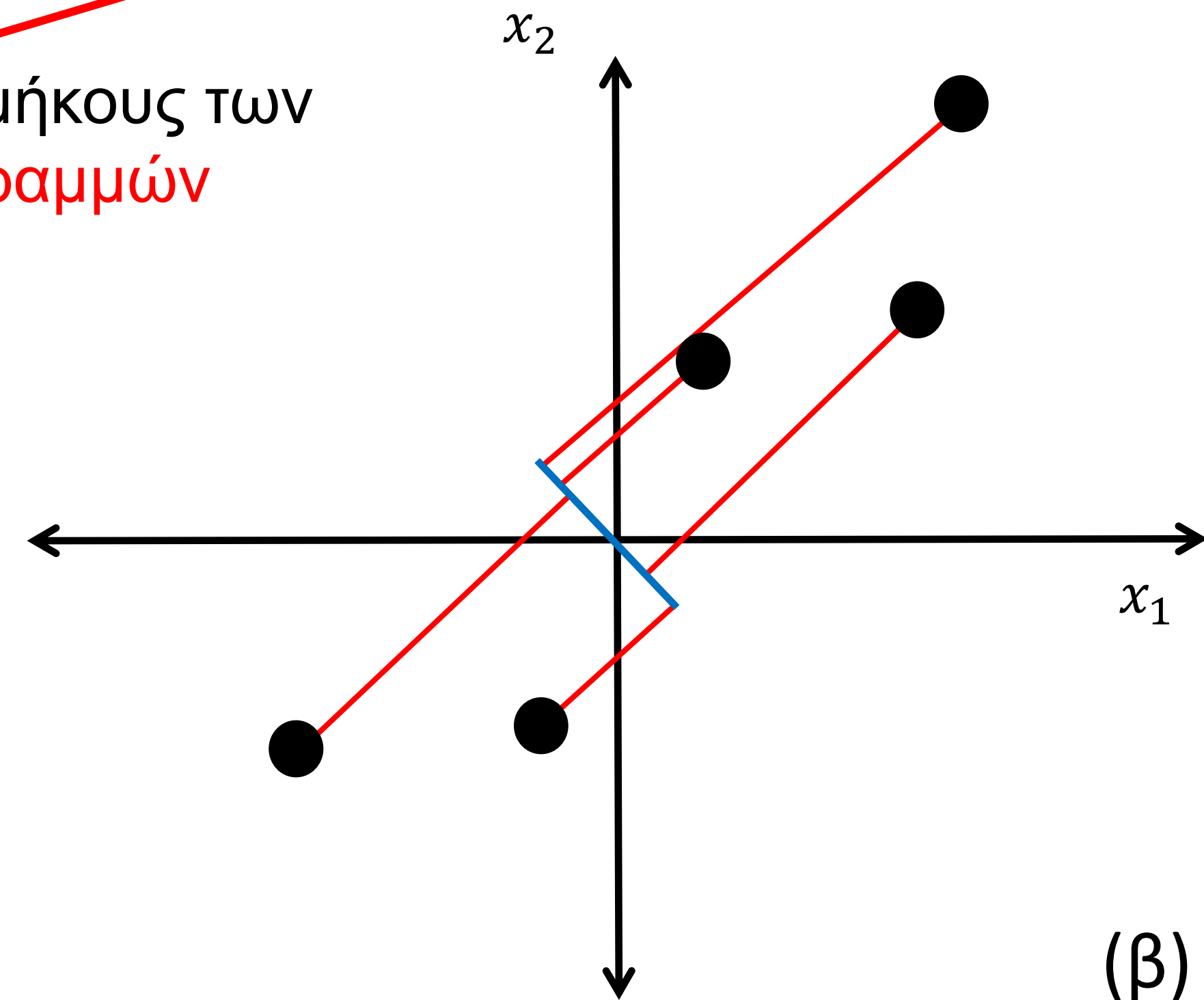


Ποια γραμμή έχει χαμηλότερο σφάλμα προβολής;



(α)

Άθροισμα του μήκους των κόκκινων γραμμών

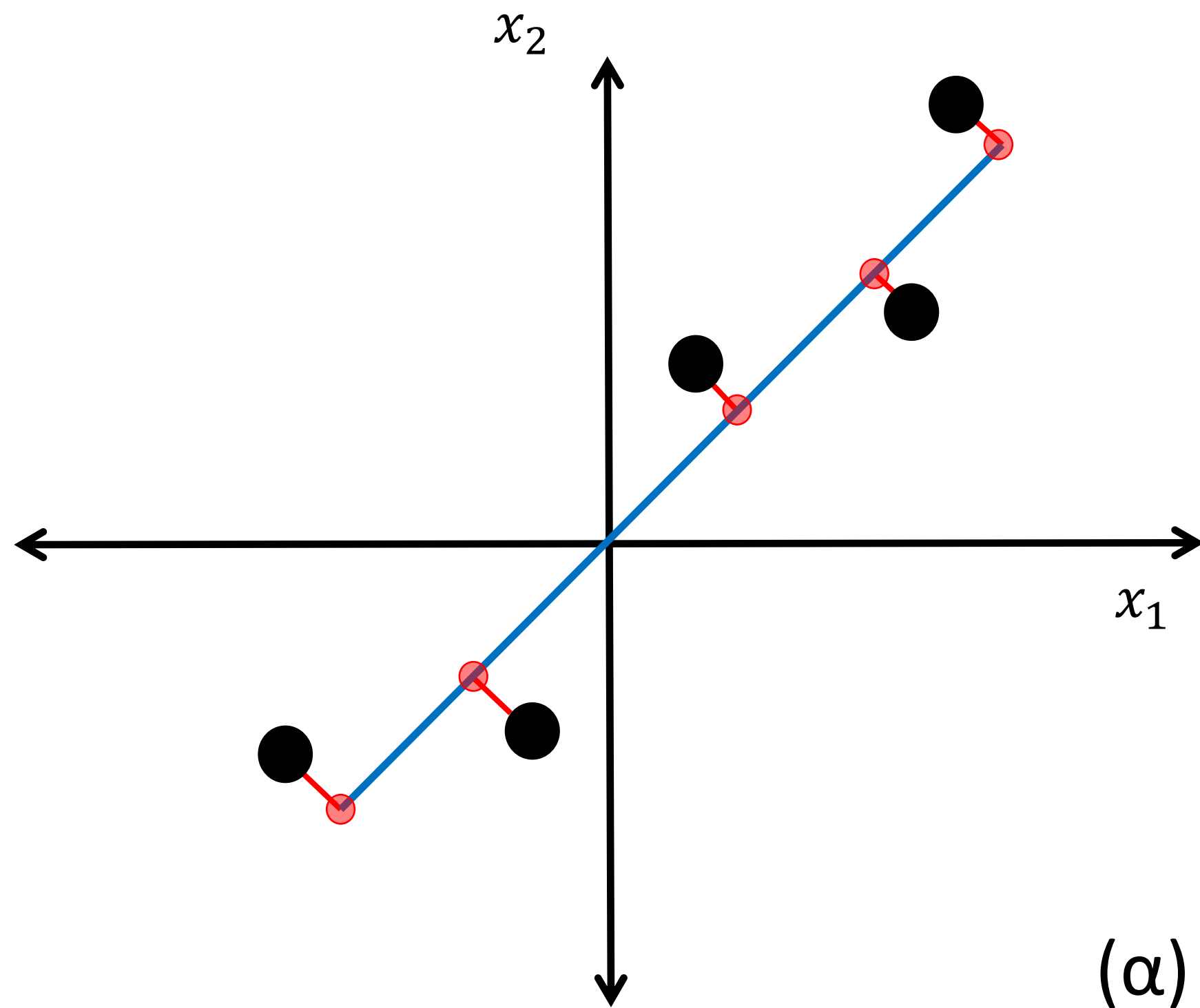


(β)

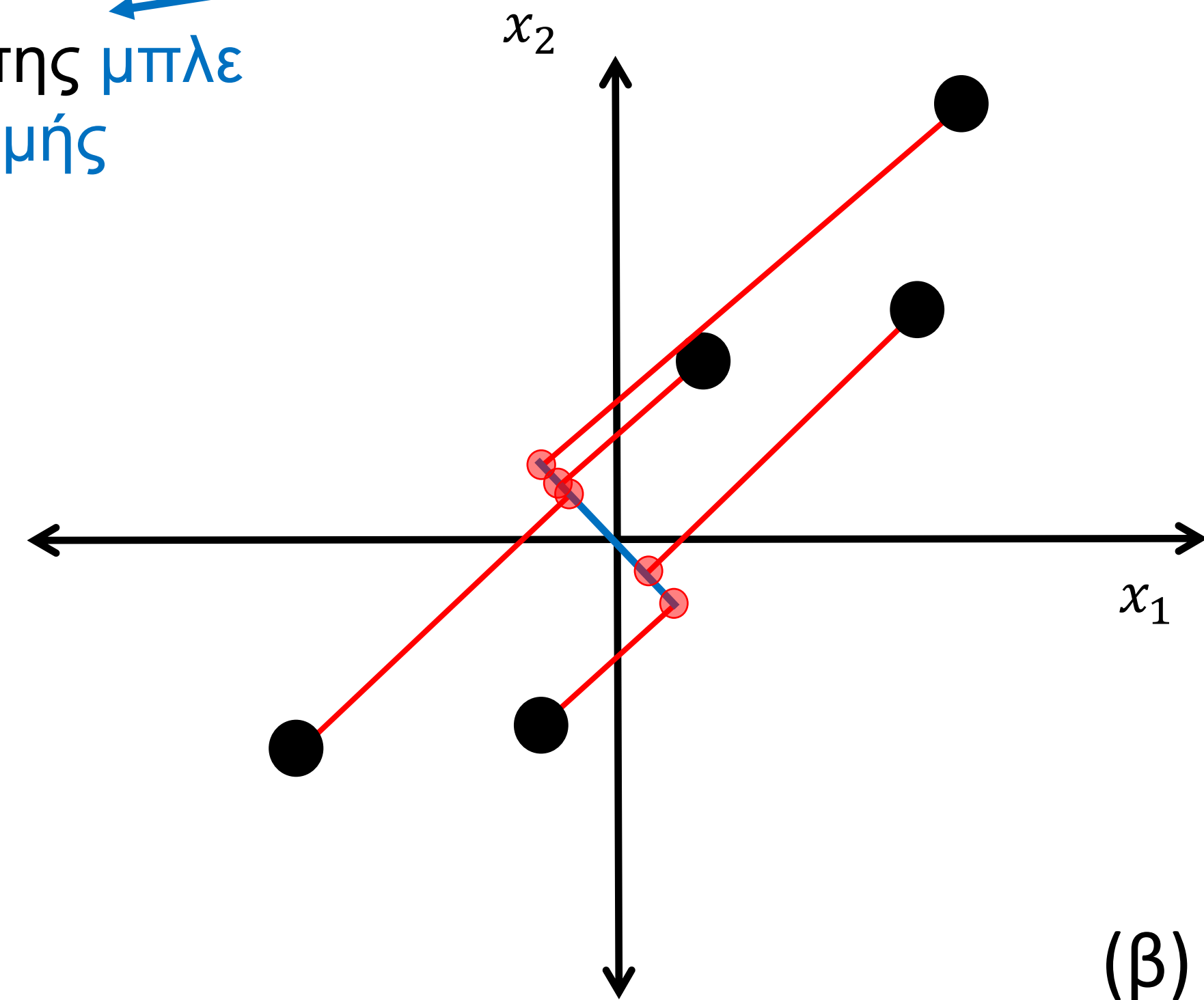




Σε ποια περίπτωση τα προβλεπόμενα σημεία έχουν **μεγαλύτερη διακύμανση;**

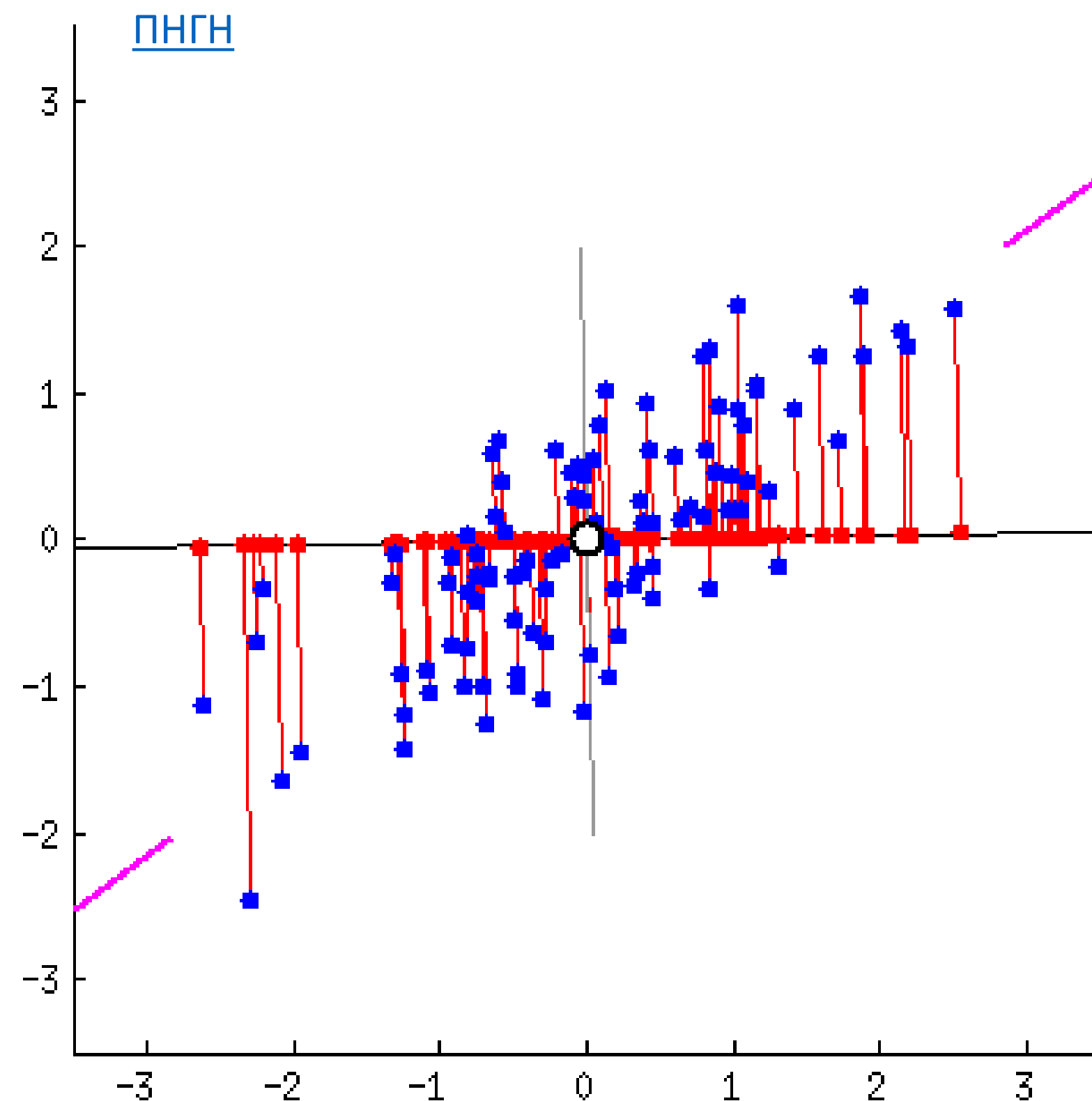


Μέγεθος της μπλε γραμμής





Principal Component Analysis

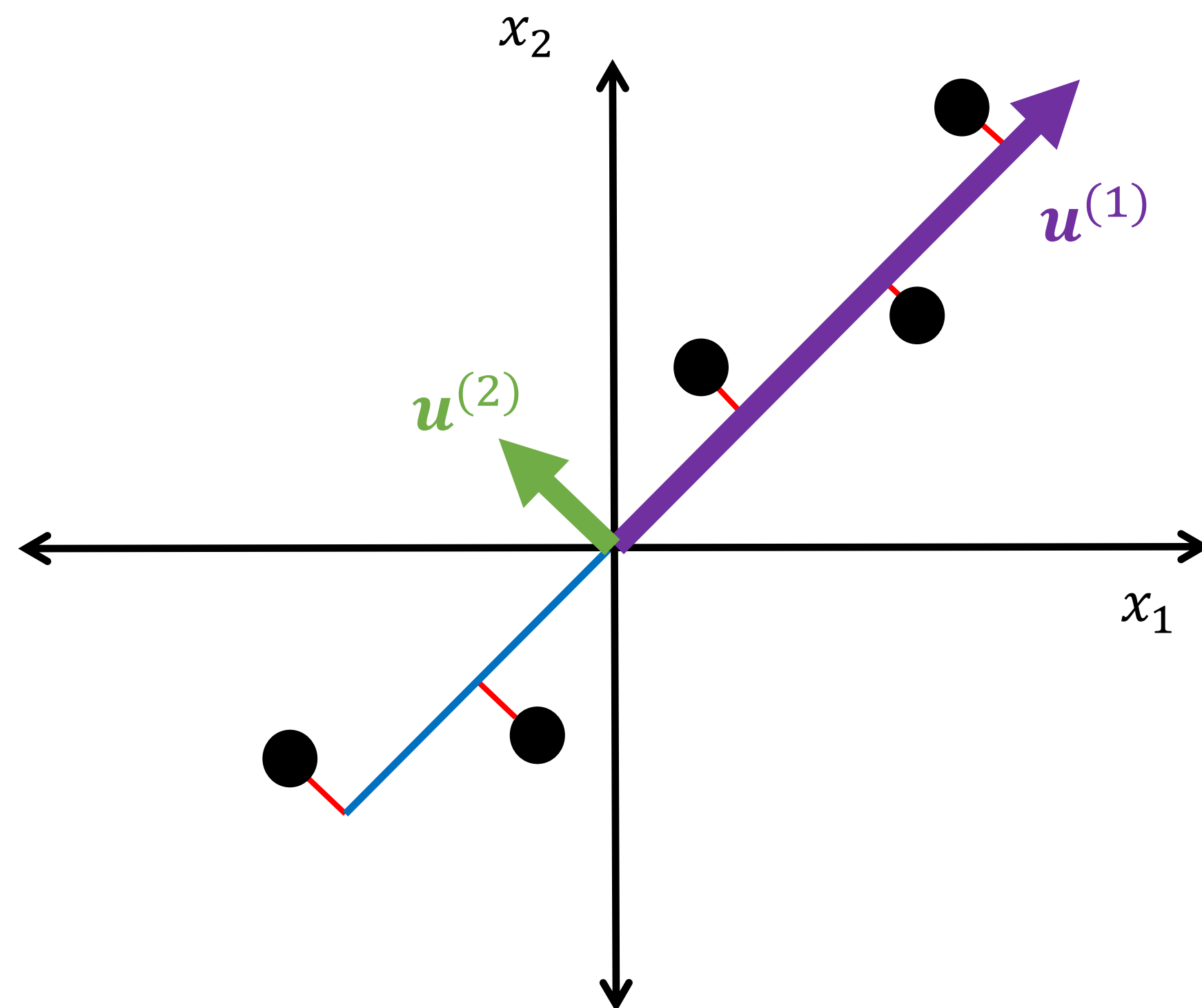


- Μέσος όρος των μπλε σημείων είναι ο λευκός κύκλος (το κέντρο)
- Διακύμανση των κόκκινων σημείων είναι η μέση τετραγωνική απόσταση από το κέντρο
- Από όλες τις πιθανές γραμμές που περιστρέφονται γύρω από το κέντρο, η PCA βρίσκει αυτή στην οποία τα προβαλλόμενα σημεία έχουν μέγιστη διακύμανση





Principal Component Analysis



- Θέλουμε να μειώσουμε τα δεδομένα από 2D σε 1D

$$\mathbf{x}^{(i)} \in \mathbb{R}^2 \rightarrow \mathbf{z}^{(i)} \in \mathbb{R}$$
- Το PCA βρίσκει ένα χαμηλού διαστάσεων υποδιάστημα πάνω στο οποίο να προβάλει τα δεδομένα, έτσι ώστε να ελαχιστοποιηθεί το άθροισμα των σφαλμάτων τετραγωνικής προβολής
- Αυτό σημαίνει ότι βρίσκει ένα διάνυσμα $\mathbf{u}^{(1)}$ που καθορίζει αυτή την κατεύθυνση.
- Σε γενικές γραμμές, βρίσκει τα n **ορθογώνια διανύσματα**, που **ονομάζονται κύρια συστατικά (ή ιδιοδιανύσματα)**, όπου το πρώτο εξηγεί το μεγαλύτερο μέρος της διακύμανσης, το δεύτερο εξηγεί τη μεγαλύτερη διακύμανση σε ό,τι απομένει μόλις αφαιρεθεί η επίδραση του πρώτου συστατικού, και ούτω καθεξής μέχρι να εξηγηθεί όλη η διακύμανση.





Αλγόριθμος ανάλυσης κύριων συστατικών

Μειώνει τα δεδομένα από n τις διαστάσεις στις k διαστάσεις

Βήμα 0: Χρησιμοποιήστε τη δυνατότητα κλιμάκωσης για να εξασφαλίσετε ότι κάθε χαρακτηριστικό είναι κανονικοποιημένο (μηδενικό μέσο)

Βήμα 1: Υπολογίστε τον πίνακα συνδιακύμανσης:

$$C = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)}) (\mathbf{x}^{(i)})^T = (1/m) * X^T * X \quad \text{που } X \in \mathbb{R}^{m \times n}$$

$$C = \begin{bmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) \\ \text{cov}(x_1, x_2) & \text{var}(x_2) \end{bmatrix}$$

Βήμα 2: Υπολογίστε τα «ιδιοδιανύσματα» του πίνακα C :

$$[U, S, V] = \text{numpy.linalg.svd}(C) \quad U = [\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(n)}] \in \mathbb{R}^{n \times n} \quad \text{που } \mathbf{u}^{(i)} \in \mathbb{R}^n$$

Βήμα 3: Πάρτε τις k πρώτες στήλες του πίνακα U :

$$W = U(:, 1:k) \quad W = [\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(k)}] \in \mathbb{R}^{n \times k}$$

Βήμα 4: Μετατροπή δεδομένων

$$\mathbf{z}^{(i)} = W^T \mathbf{x}^{(i)} \quad (\mathbf{z}^{(i)} \in \mathbb{R}^k)$$

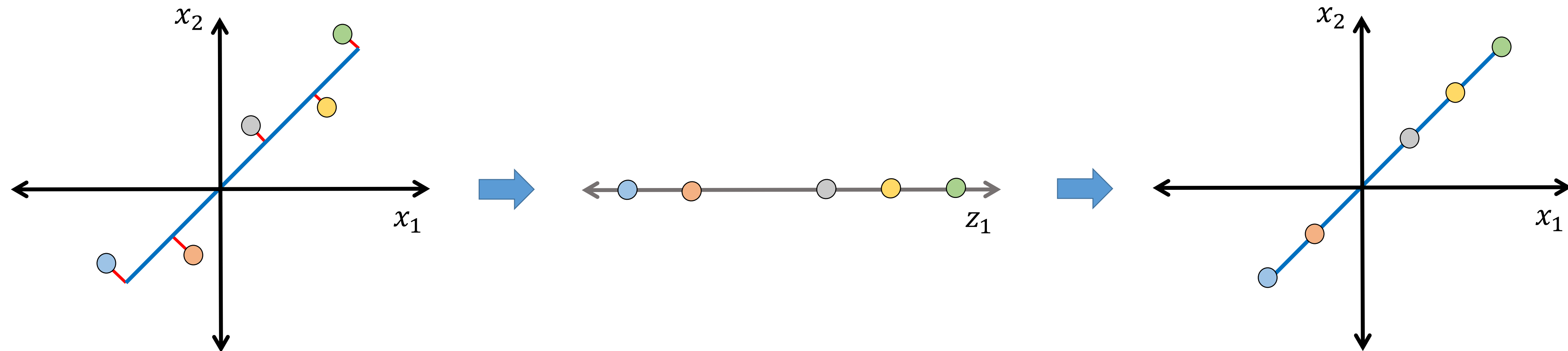




Ανακατασκευή από συμπιεσμένη αναπαράσταση

Προβολή: κωδικοποιεί x και δίνει $z = W^T x$

Ανακατασκευή: αποκωδικοποιεί z και δίνει $\hat{x} = Wz$





Επιλογή k (αριθμός principal components)

Μέσο τετραγωνικό σφάλμα προβολής: $\frac{1}{m} \sum_{i=1}^m \left\| \mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)} \right\|^2$

Συνολική διακύμανση των δεδομένων: $\frac{1}{m} \sum_{i=1}^m \left\| \mathbf{x}^{(i)} \right\|^2$

Συνήθως, επιλέξτε k να είναι η μικρότερη αξία, έτσι ώστε

$$\frac{\frac{1}{m} \sum_{i=1}^m \left\| \mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)} \right\|^2}{\frac{1}{m} \sum_{i=1}^m \left\| \mathbf{x}^{(i)} \right\|^2} < 0.01 \quad (1 \%)$$

«Το 99 % της διακύμανσης διατηρείται»

Διαφάνεια προσαρμοσμένη από Andrew Ng – Machine Learning Course (Coursera)





Επιλογή k (αριθμός principal components)

Προσέγγιση 1:

Θέσε $k=1$

1. Δοκιμάστε το PCA(k) και υπολογίστε $W, \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}, \hat{\mathbf{x}}^{(1)}, \dots, \hat{\mathbf{x}}^{(m)}$

2. Αν είναι $\frac{\frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \hat{x}^{(i)}\|^2}{\frac{1}{m} \sum_{i=1}^m \|x^{(i)}\|^2} < 0.01$;

επέστρεψε K

Διαφορετικά

$K=k+1$

Πήγαινε στο 1.





Επιλογή k (αριθμός principal components)

Προσέγγιση 2:

$[U, S, V] = \text{numpy.linalg.svd}(\text{Sigma})$

$$S = \begin{bmatrix} S_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & S_{nn} \end{bmatrix} \text{ διαγώνιος πίνακας}$$

Για $k=1$ έως n :

εάν $\frac{\sum_{i=1}^k S_{ii}}{\sum_{i=1}^n S_{ii}} \geq 0.99$:
επέστρεψε K

Λόγος της εξηγημένης διακύμανσης

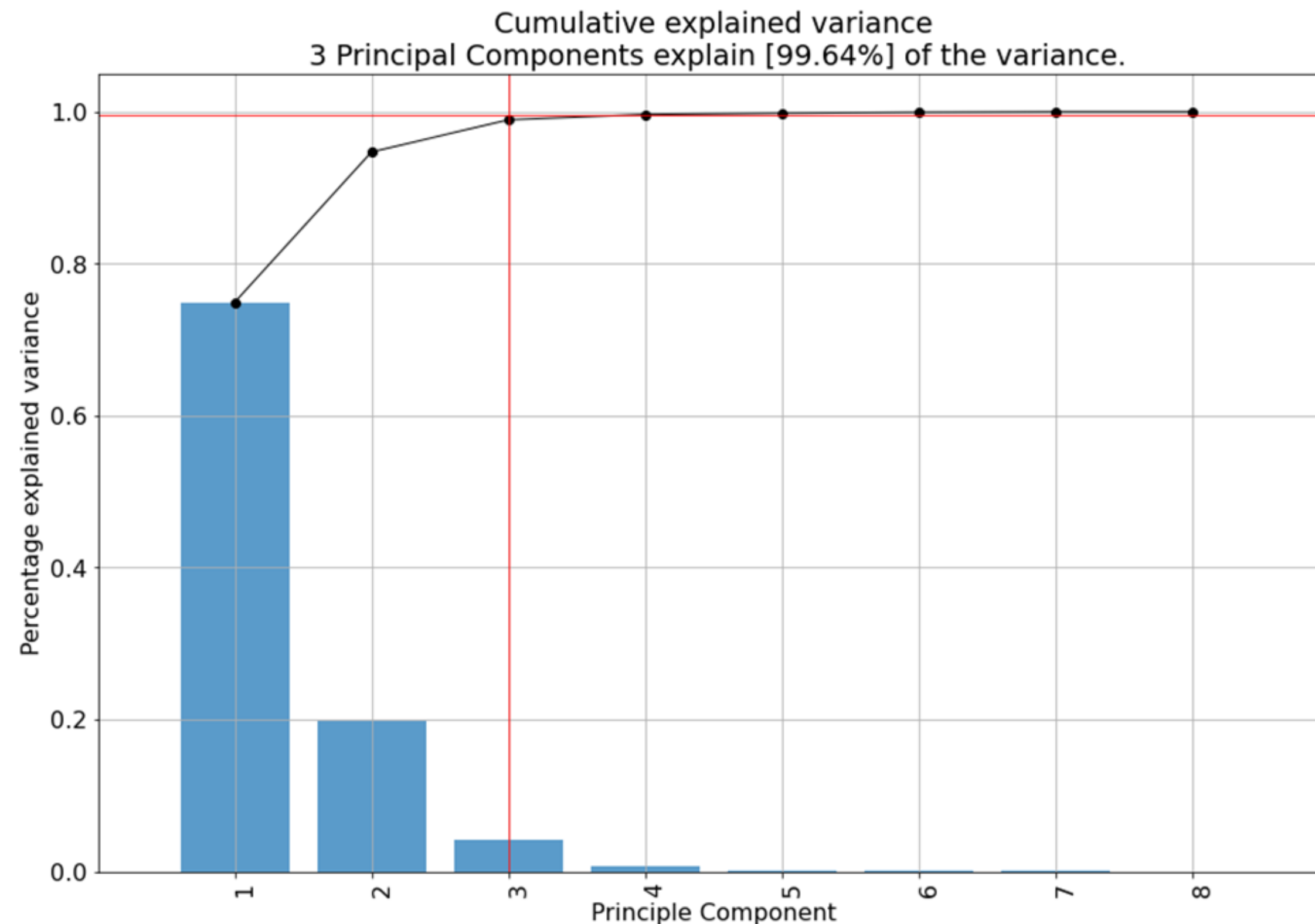
(99 % της διακύμανσης που διατηρείται)

Διαφάνεια προσαρμοσμένη από Andrew Ng – Machine Learning Course (Coursera)





Επιλογή k (αριθμός principal components)



[ΠΗΓΗ](#)

Μπορούμε να χρησιμοποιήσουμε μια πλοκή για να μας βοηθήσει να απεικονίσουμε πόση διακύμανση διατηρείται

2 PC καλύπτουν περισσότερο από το 95 % της διακύμανσης

3 PC καλύπτουν περισσότερο από το 99 % της διακύμανσης

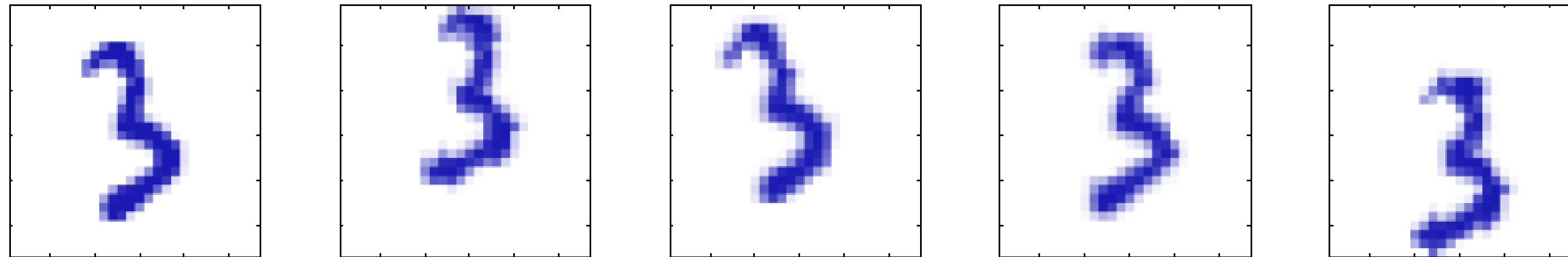
→ Μείωση του προβλήματος από 8D σε 3D



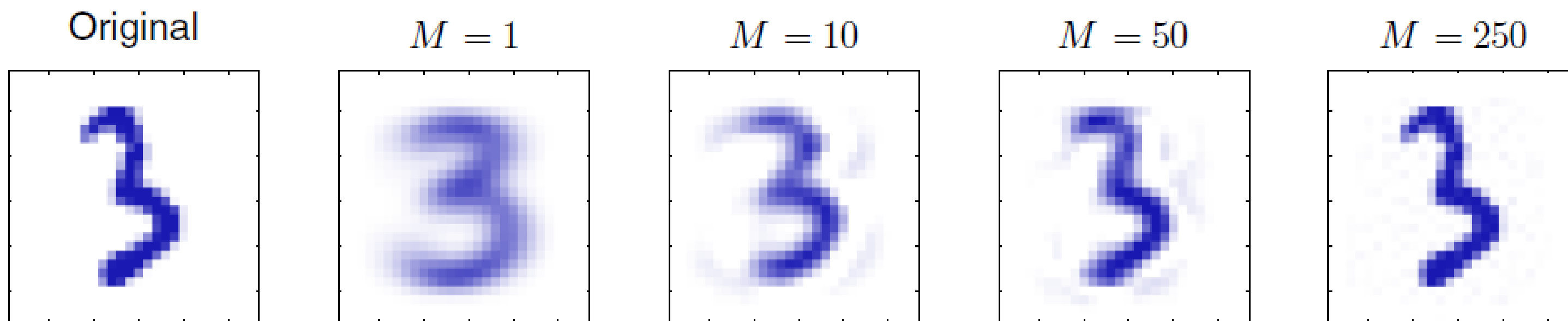


Παράδειγμα

Σύνολο δεδομένων των 3 μετατοπισμένων και περιστρεφόμενων ψηφίων (100x100 = 10.000 pixels)



Ανακατασκευές PCA που λαμβάνονται με διατήρηση των κύριων συστατικών στοιχείων M

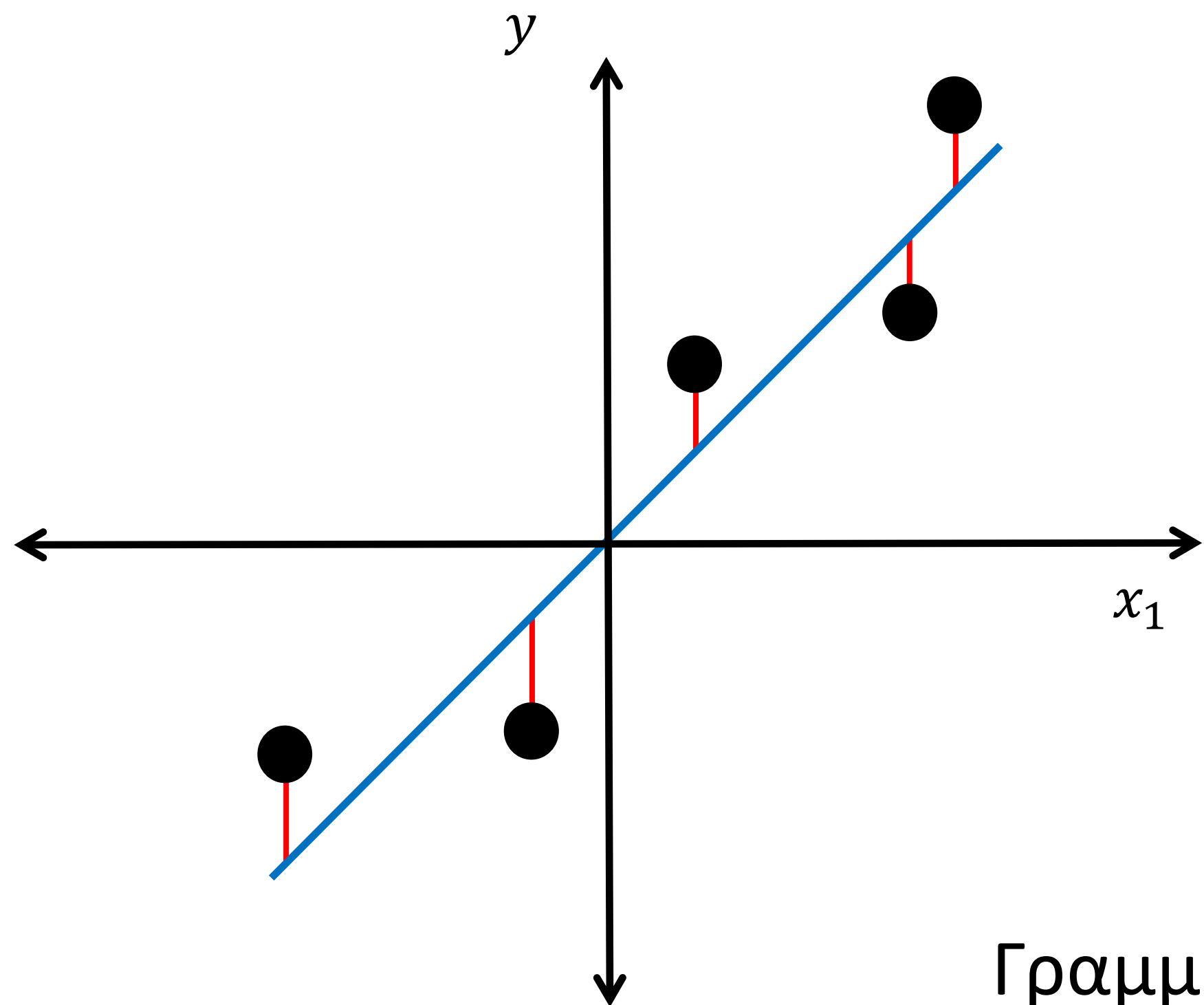


Προσαρμοσμένο από: Επίσκοπος (2006). Αναγνώριση προτύπων και μηχανική μάθηση.

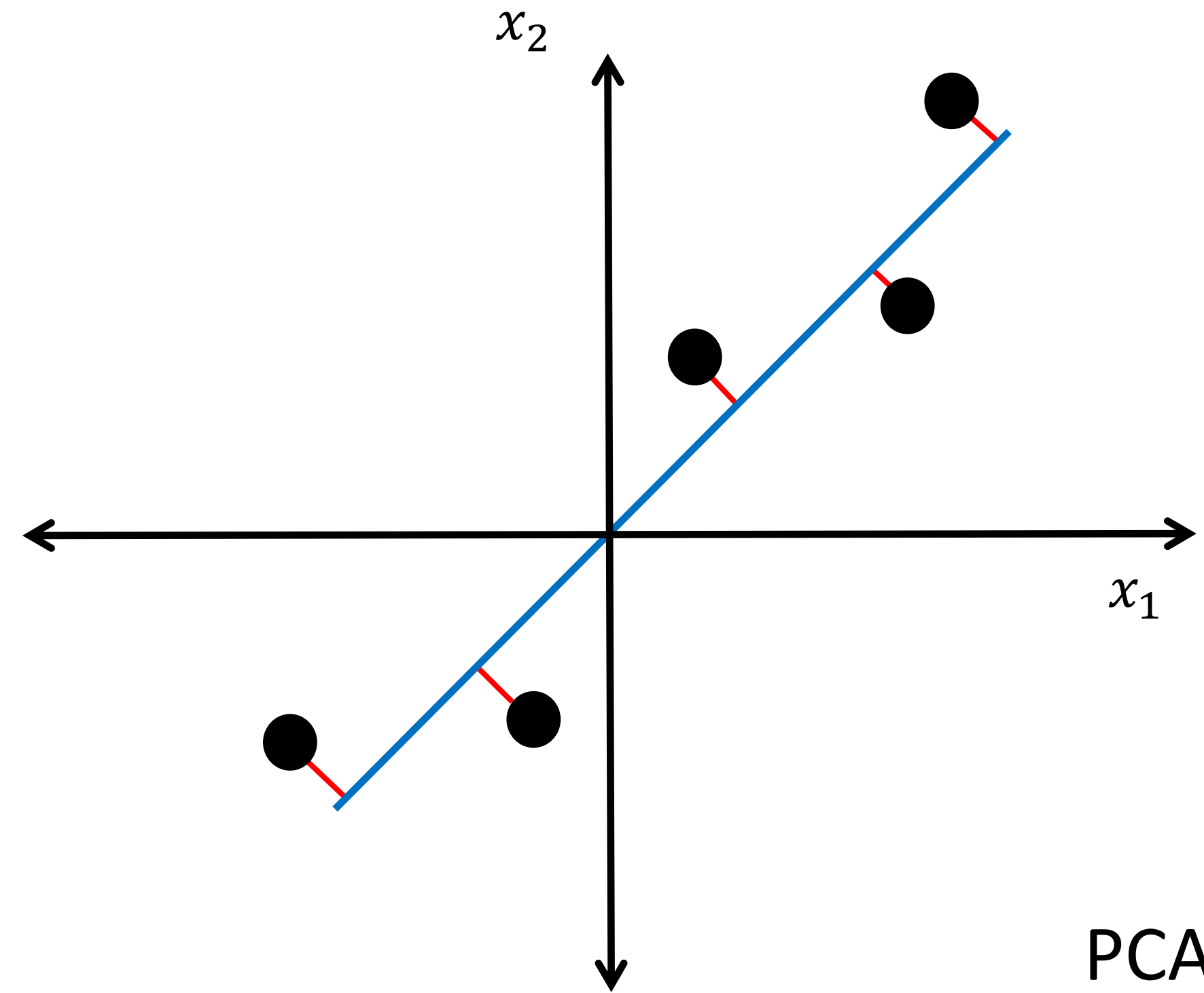




Η PCA δεν είναι γραμμική παλινδρόμηση



Γραμμική
παλινδρόμηση



PCA





PCA στο scikit-learn

```
>>> import numpy as np
>>> from sklearn.decomposition import PCA
>>> X = np.array([[ -1, -1], [-2, -1], [-3, -2], [ 1,  1], [ 2,  1], [ 3,  2]])
>>> pca = PCA(n_components=2)
>>> pca.fit(X)
PCA(n_components=2)
>>> print(pca.explained_variance_ratio_)
[0.9924... 0.0075...]
```





Η μείωση της διάστασης μπορεί να βοηθήσει την εποπτευόμενη μάθηση

$D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ όπου $x^{(i)}$ είναι υψηλής διάστασης, π.χ., $x^{(i)} \in \mathbb{R}^{10000}$

1. Εξαγάγετε εισροές από τα cluster και δημιουργήστε ένα μη επισημασμένο σύνολο δεδομένων: $x^{(1)}, \dots, x^{(m)} \in \mathbb{R}^{10000}$
2. Εφαρμόστε το PCA στο μη επισημασμένο σύνολο δεδομένων και λάβετε μειωμένη διαστατική αναπαράσταση: $z^{(1)}, \dots, z^{(m)} \in \mathbb{R}^{500}$

Η χαρτογράφηση $x^{(i)} \rightarrow z^{(i)}$ θα πρέπει να υπολογιστεί με την εκτέλεση PCA μόνο στο σύνολο εκπαίδευσης!

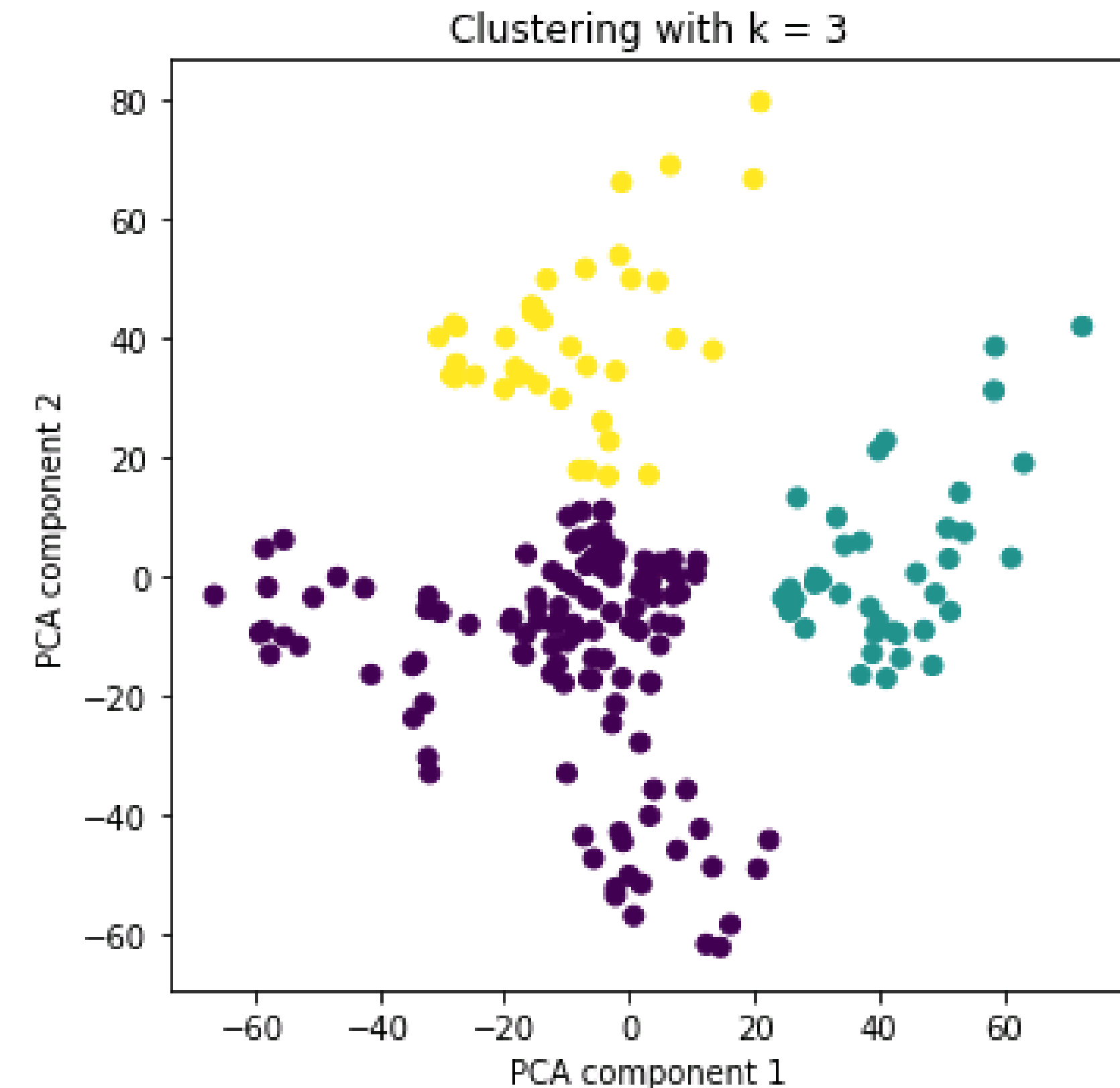
3. Δημιουργήστε ένα νέο σετ εκπαίδευσης με ετικέτα: $D_{lowdim} = \{(z^{(1)}, y^{(1)}), \dots, (z^{(m)}, y^{(m)})\}$
4. Χρήση του νέου συνόλου δεδομένων με ετικέτα για την κατάρτιση εποπτευόμενου μοντέλου μάθησης (π.χ. logistic regression): $f_{\theta}(z)$
5. Για μια εισαγωγή επικύρωσης/δοκιμή, μετατρέπουμε την $x^{(new)}$ σε $z^{(new)}$, και την τροφοδοτούμε με το μοντέλο





Η μείωση της διάστασης μπορεί να βοηθήσει στην ομαδοποίηση

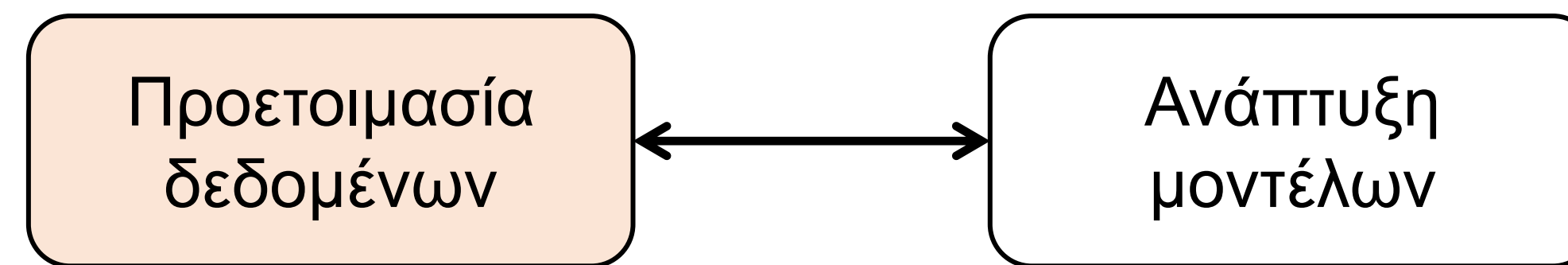
- Αλγόριθμοι ομαδοποίησης βασίζονται σε υπολογισμούς απόστασης
- Οι υπολογισμοί απόστασης σε υψηλές διαστάσεις είναι προβληματικοί
- Όταν θέλουμε να συγκεντρώσουμε δεδομένα υψηλής διάστασης, συχνά κάνουμε μείωση της διάστασης για να αποκτήσουμε τη χαμηλότερη διάσταση αναπαράστασης, $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}$ και να εκτελέσουμε ομαδοποίηση χρησιμοποιώντας αυτά ως είσοδο.



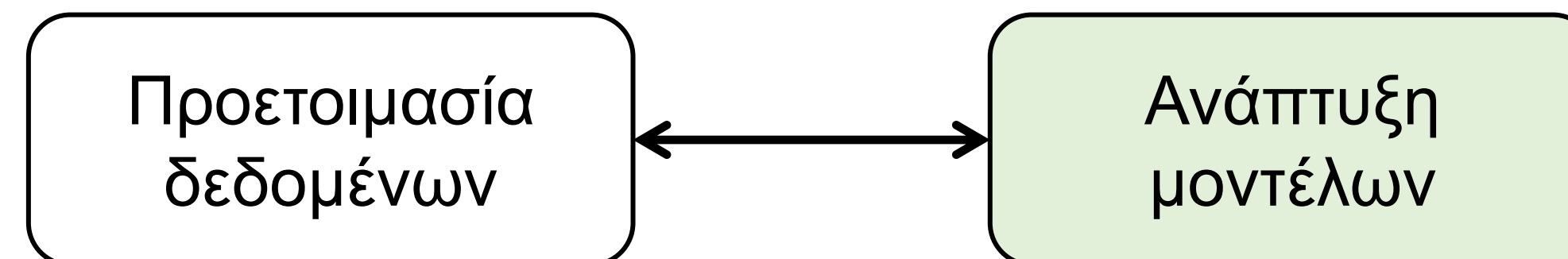


Που η βρίσκεται μείωση της διάστασης στον κύκλο ενός project MM;

- Εξαρτάται από τη χρήση:
 - Τυπική χρήση είναι η απεικόνιση ή η συμπίεση για τη βοήθεια άλλων αλγορίθμων μάθησης



- Ωστόσο, μια εργασία μπορεί να αφορά τη δημιουργία ενός μοντέλου που μειώνει τη μνήμη/το δίσκο που απαιτείται για την αποθήκευση ορισμένων δεδομένων





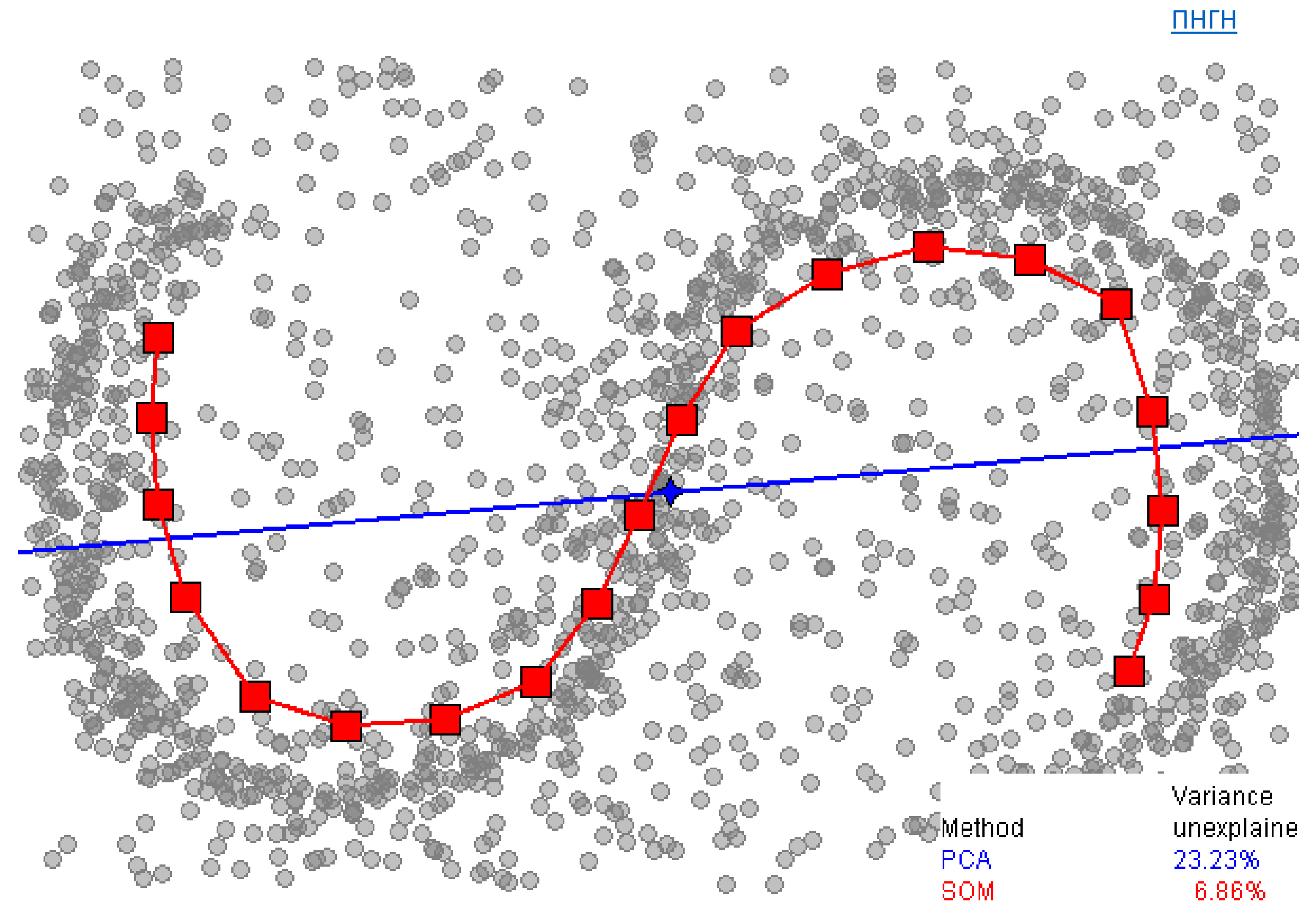
Μη-γραμμική μείωση διάστασης





Nonlinear Dimensionality Reduction

- Το PCA προβάλλει τα δεδομένα σε ένα **γραμμικό** υποδιάστημα χαμηλότερης διάστασης
- Τι γίνεται αν τα δεδομένα βρίσκονται σε μια χαμηλής διάστασης μη **γραμμική** επιφάνεια;





Πυρήνας PCA

Το **PCA** υπολογίζει τον πίνακα συνδιακύμανσης $X \in \mathbb{R}^{m \times n}$ ως εξής:

$$C = \frac{1}{m} \sum_{i=1}^m (x^{(i)}) (x^{(i)})^T$$

Στη συνέχεια προβάλλει τα δεδομένα στα πρώτα k ιδιοδιανύσματα αυτού του πίνακα

Ο **πυρήνας PCA** υπολογίζει τον πίνακα συνδιακύμανσης των δεδομένων μετά τη μετατροπή του σε ένα χώρο υψηλότερης διάστασης χρησιμοποιώντας το τέχνασμα του πυρήνα:

$$C = \frac{1}{m} \sum_{i=1}^m \varphi(x^{(i)}) \varphi(x^{(i)})^T$$

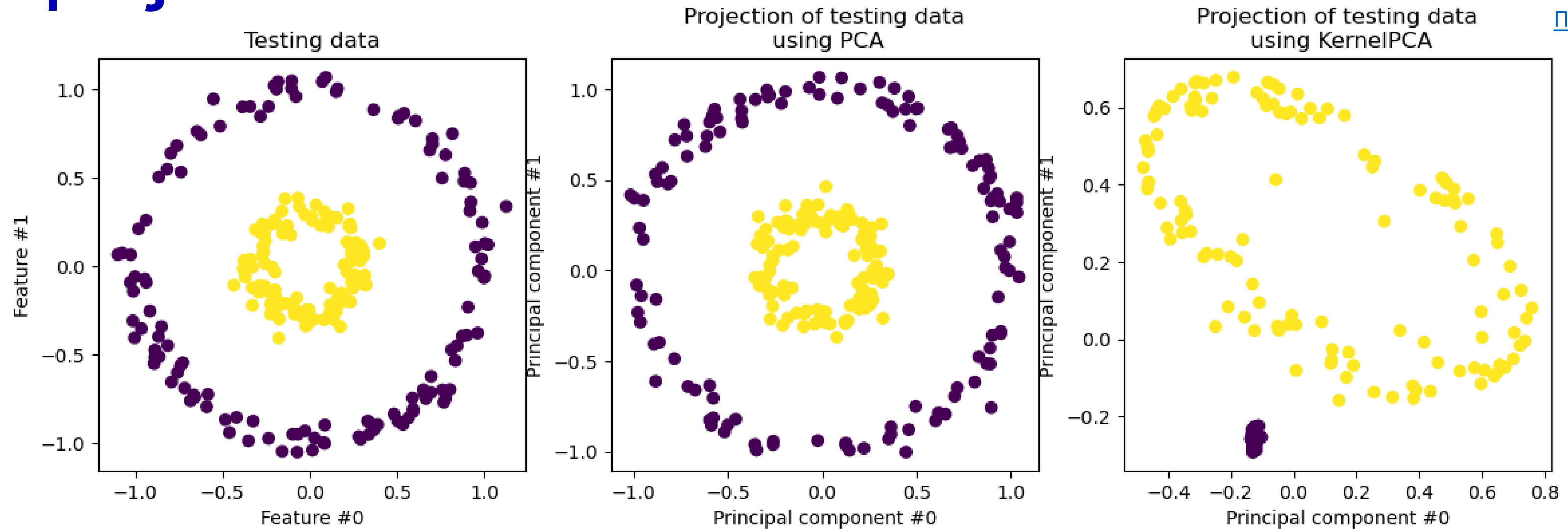
Στη συνέχεια προβάλλει τα δεδομένα στα πρώτα k ιδιοδιανύσματα αυτού του πίνακα ακριβώς όπως το PCA

Υπενθύμιση κόλπου πυρήνα: το $\varphi(x^{(i)})$ δεν έχει υπολογιστεί ρητά





Πυρήνας PCA



[ΠΗΓΗ](#)

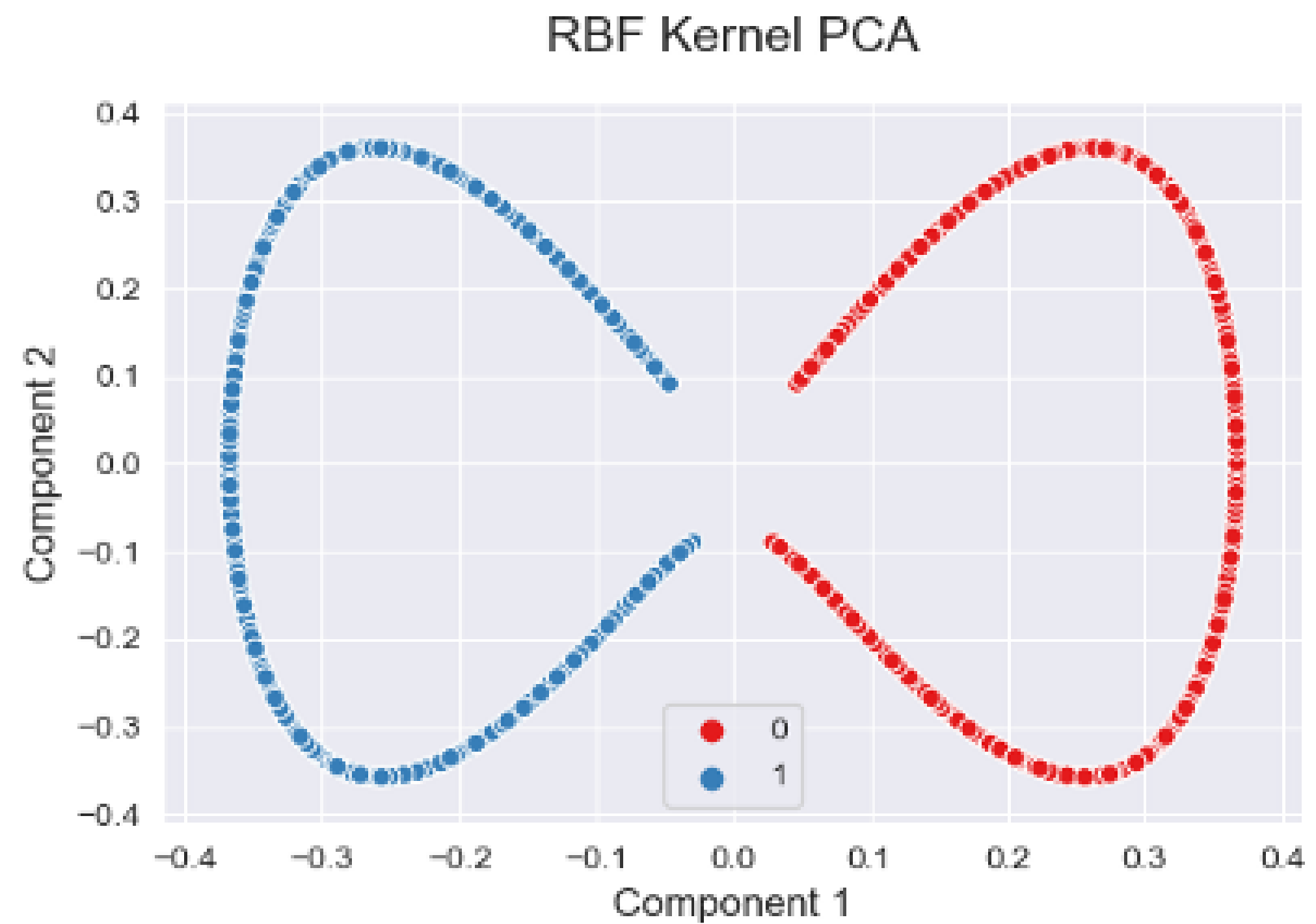
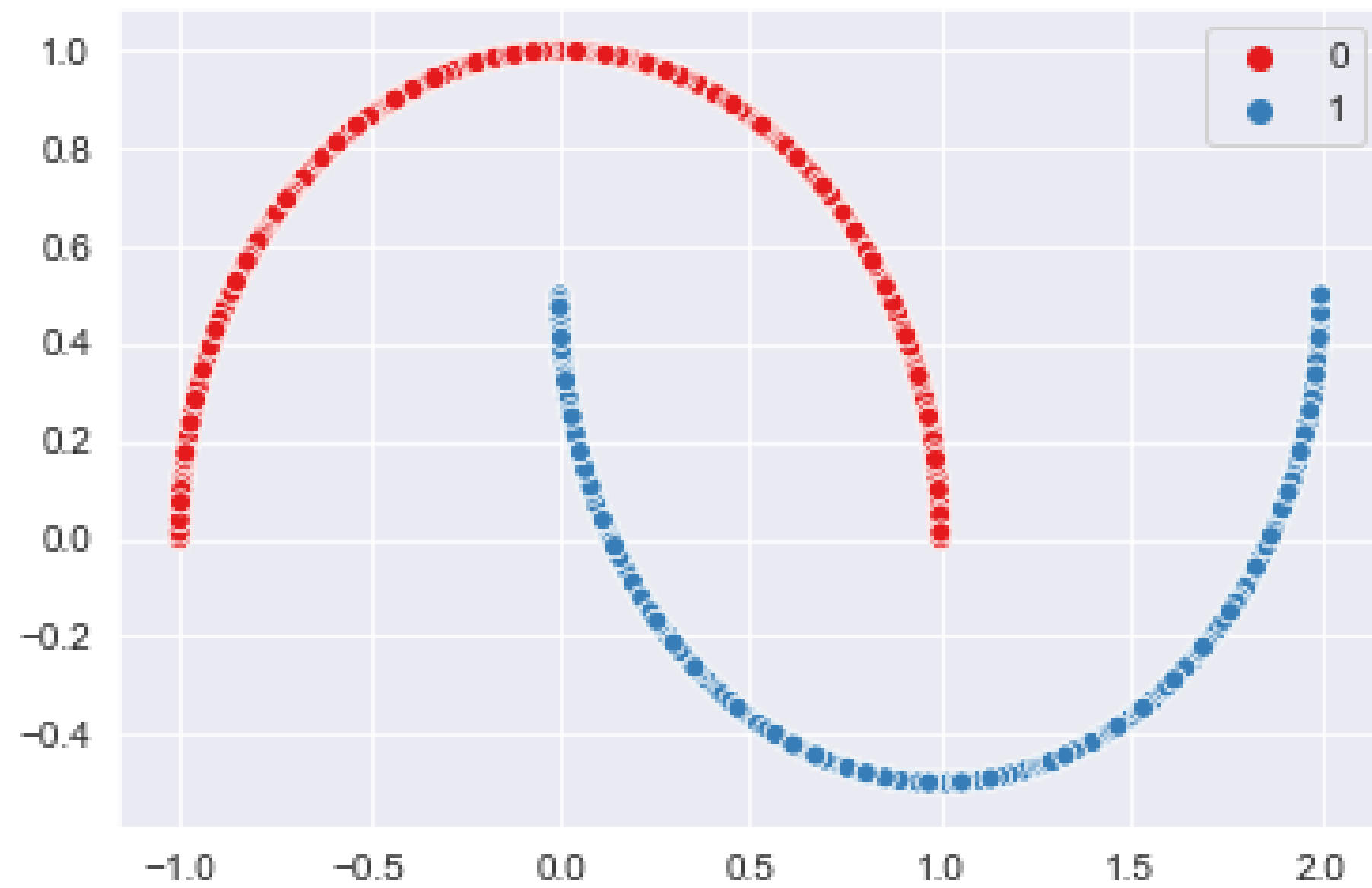
Χρησιμοποιώντας έναν πυρήνα RBF επιτρέπει να κάνει μια μη γραμμική προβολή, η οποία ξεδιπλώνει τα δεδομένα, διατηρώντας παράλληλα τις σχετικές αποστάσεις των ζευγών των σημείων δεδομένων που είναι κοντά το ένα στο άλλο στον αρχικό χώρο





Πυρήνας PCA

[ΠΗΓΗ](#)





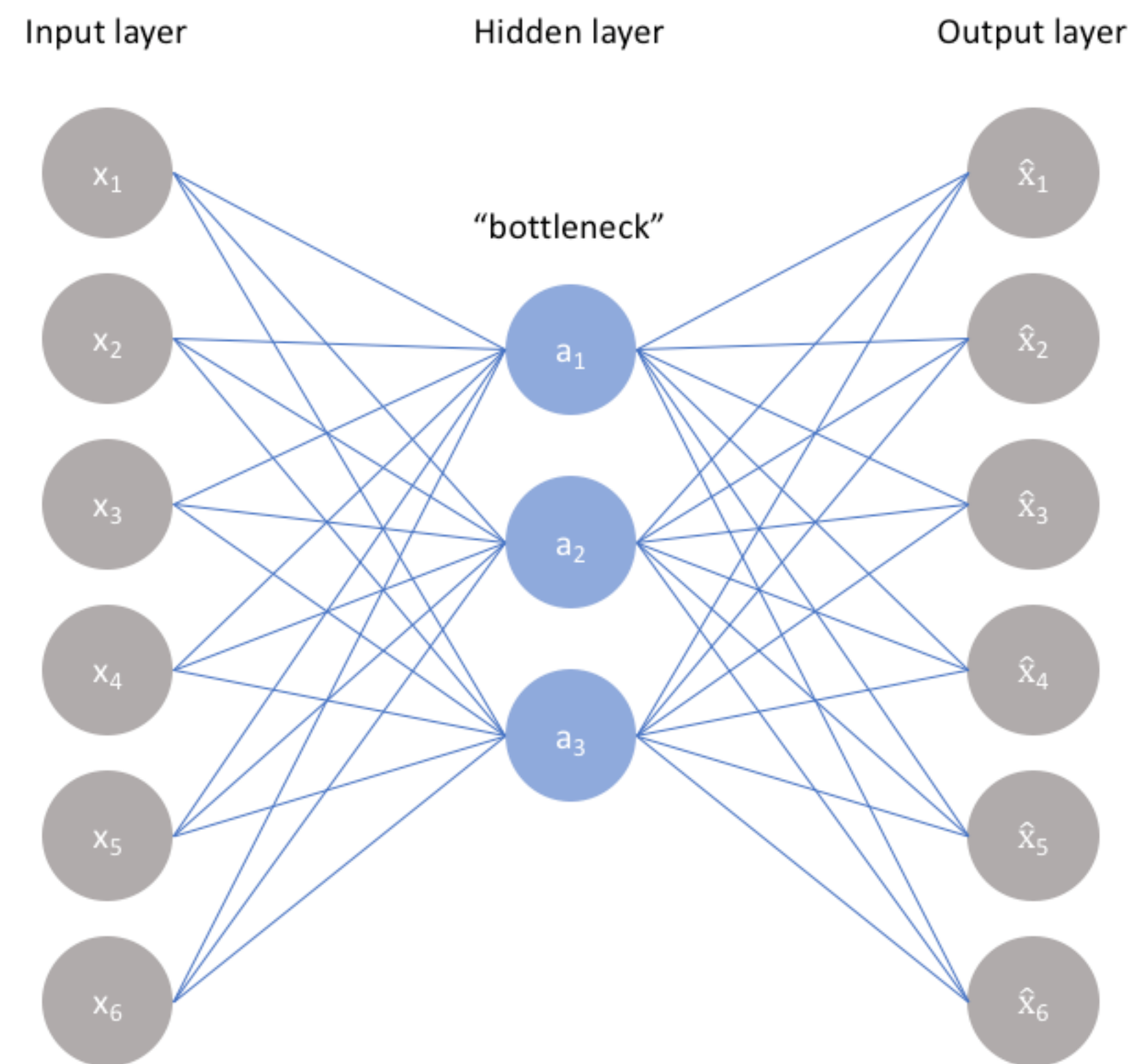
Αυτόματοι κωδικοποιητές

- Μπορούμε να σκεφτούμε το PCA ως:
 - Την εκμάθηση μιας γραμμικής χαρτογράφησης $x \rightarrow z$, που ονομάζεται **κωδικοποιητής** f_e , και την εκμάθηση μιας άλλης γραμμικής χαρτογράφησης $z \rightarrow x$, που ονομάζεται **αποκωδικοποιητής**, f_d .
 - Η λειτουργία ανασυγκρότησης έχει τη μορφή $r(x) = f_d(f_e(x))$
 - Το μοντέλο είναι εκπαιδευμένο για να ελαχιστοποιεί: $L(\theta) = \|r(x) - x\|^2$
- Οι αυτόματοι κωδικοποιητές είναι **νευρωνικά δίκτυα** που χρησιμοποιούν μη γραμμικές αντιστοιχίσεις για τον κωδικοποιητή και τον αποκωδικοποιητή
 - Εκπαιδεύονται χρησιμοποιώντας οπισθοδιάδοση και gradient descent





Autoencoders



Πηγή [εικόνας](#)

Κρυφό στρώμα του μεγέθους k

- Σε $k < n$: **περίπτωση ελλιπούς εκπροσώπησης**:
 - χαμηλός βαθμός συμφόρησης που αναγκάζει να κωδικοποιήσει (συμπιέζει) τις πληροφορίες χρησιμοποιώντας λιγότερα bits
- Σε $k \geq n$: **περίπτωση υπερπλήρους εκπροσώπησης**:
 - το μοντέλο μπορεί να μάθει τη λειτουργία αναγνώρισης!
 - ανάγκη επιβολής κάποιου είδους τακτοποίησης, όπως η προσθήκη θορύβου στις εισροές, ο εξαναγκασμός της ενεργοποίησης των κρυφών μονάδων να είναι αραιή, κ.λπ.

Τι γίνεται αν έχουμε 1 κρυφό στρώμα γραμμικών λειτουργιών ενεργοποίησης;

➤ αυτό είναι ισοδύναμο με το PCA





Παραλλαγές Autoencoder

- **Bottleneck (undercomplete) autoencoder** : χαμηλότερο διαστατικό κρυφό στρώμα από το στρώμα εισαγωγής
- **Denoising autoencoder**: ανακατασκευή ενός καθαρού σήματος από ένα θορυβώδες
- **Sparse autoencoder**: περισσότεροι κρυμμένοι κόμβοι από εισόδους, αλλά αραιά περιορισμένοι στις κρυφές μονάδες ενεργοποίησης
- **Contractive autoencoder**: ποινή για τα παράγωγα των κρυφών μονάδων
- **Convolutional autoencoder**: χρησιμοποιείται για δεδομένα εικόνας
- **Variational autoencoder**: γεννητικό μοντέλο, μπορεί να δημιουργήσει νέα δείγματα και να παρεμβληθεί μεταξύ τους
- ...



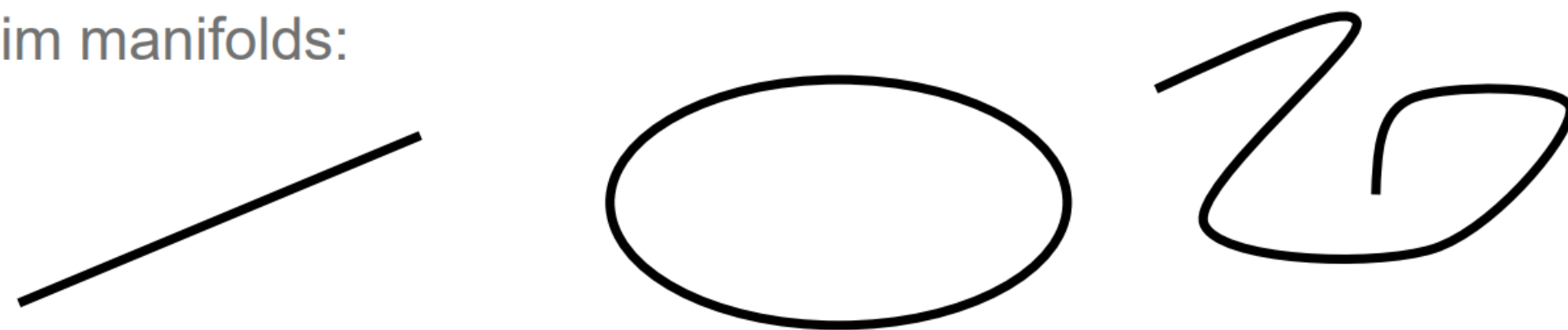


Μάθηση manifold

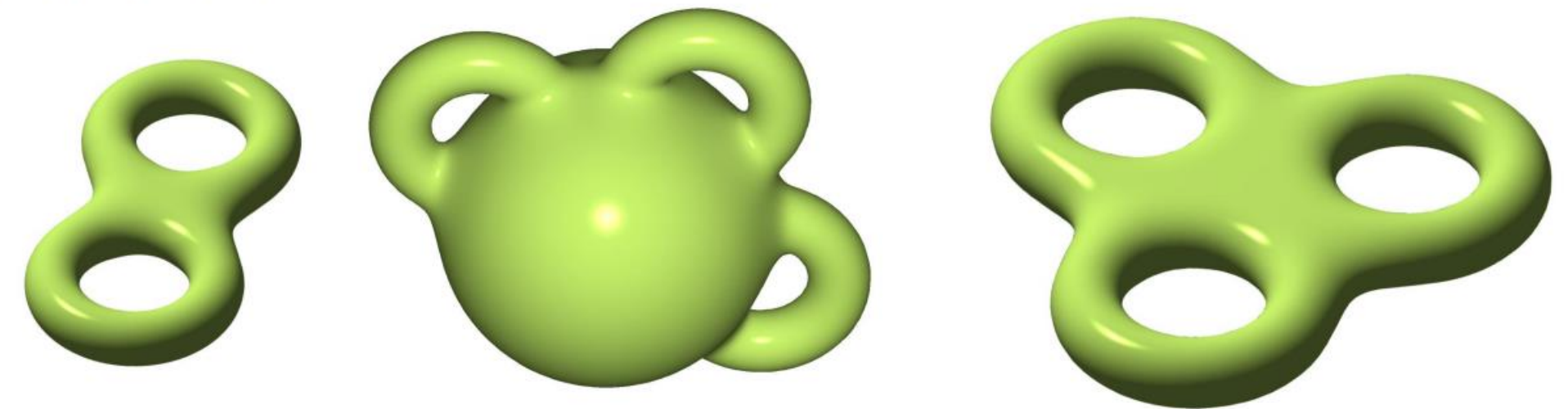
Υπόθεση manifold: Πολλά υψηλής διάστασης σύνολα δεδομένων που συμβαίνουν στον πραγματικό κόσμο βρίσκονται στην πραγματικότητα κατά μήκος των χαμηλών διαστάσεων πολλαπλοτήτων **μέσα** σε αυτό το χώρο υψηλής διάστασης.

- **Manifold:** κάθε αντικείμενο που είναι σχεδόν «επίπεδο» σε μικρές κλίμακες

1dim manifolds:



2dim manifolds:



- Πολλά σύνολα δεδομένων που φαίνεται να απαιτούν πολλές μεταβλητές για να περιγράψουν, μπορούν στην πραγματικότητα να περιγραφούν από μικρότερο αριθμό μεταβλητών.
- Οι κατώτερες διαστάσεις αναπαραστάσεις των δεδομένων αναφέρονται συχνά ως «**εσωτερικές μεταβλητές**»: τιμές από τις οποίες παρήχθησαν τα δεδομένα

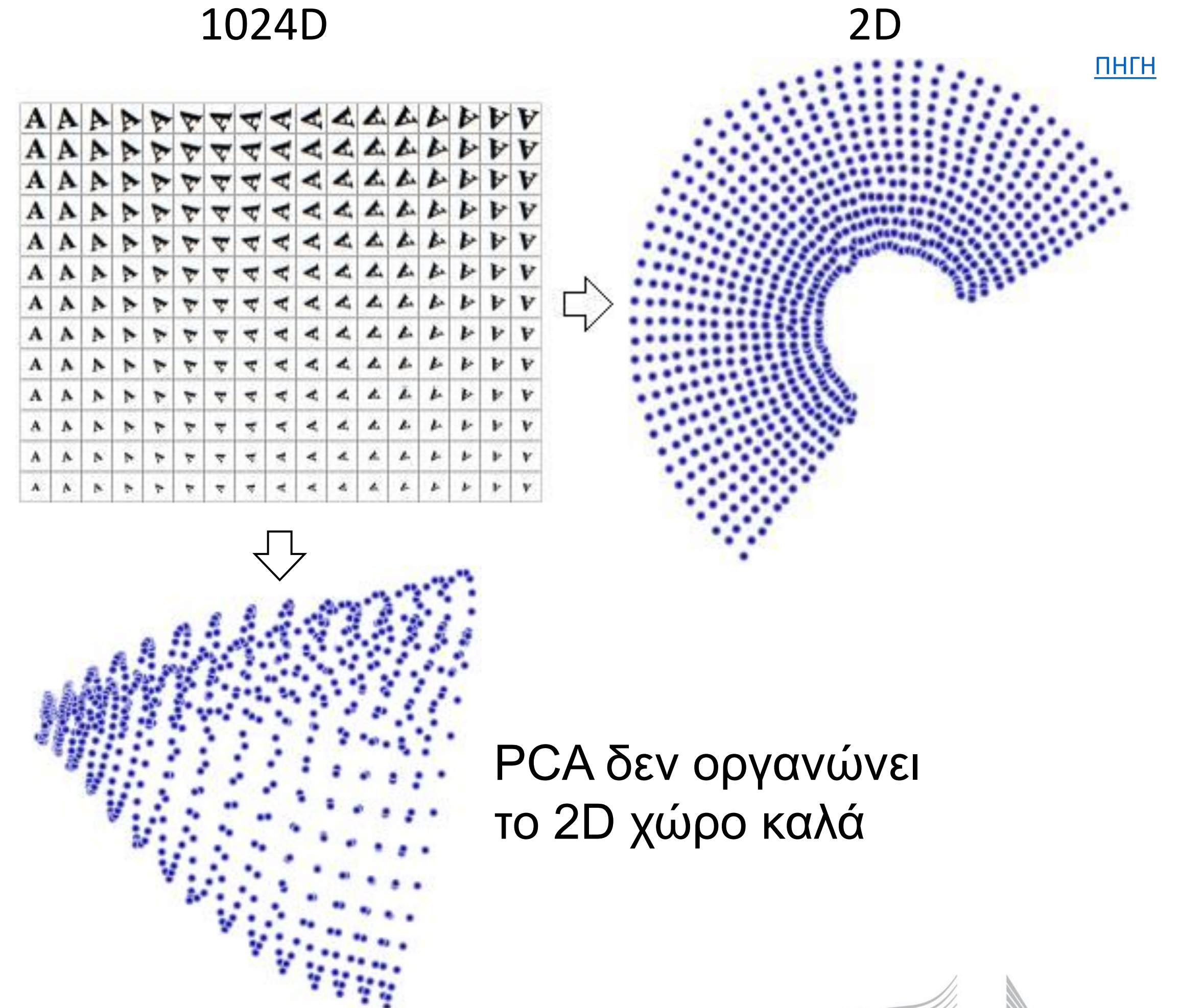




Παράδειγμα

Σύνολο δεδομένων που περιέχει εικόνες του γράμματος «Α» το οποίο έχει **κλιμακωθεί** και **περιστραφεί** κατά διάφορα ποσά

- Η εικόνα: 32x32 pixels = 1024 τιμές εικονοστοιχείων (διαστάσεις)
- Εγγενής διάσταση = 2 (περιστροφή και κλίμακα)
- Οι πληροφορίες σχετικά με το σχήμα ή την εμφάνιση του γράμματος «Α» δεν αποτελούν μέρος των εγγενών μεταβλητών
- Ο μη γραμμικός αλγόριθμος μείωσης διάστασης μπορεί να μειώσει το σύνολο δεδομένων σε 2 διαστάσεις και να οργανώσει τα σημεία καλά



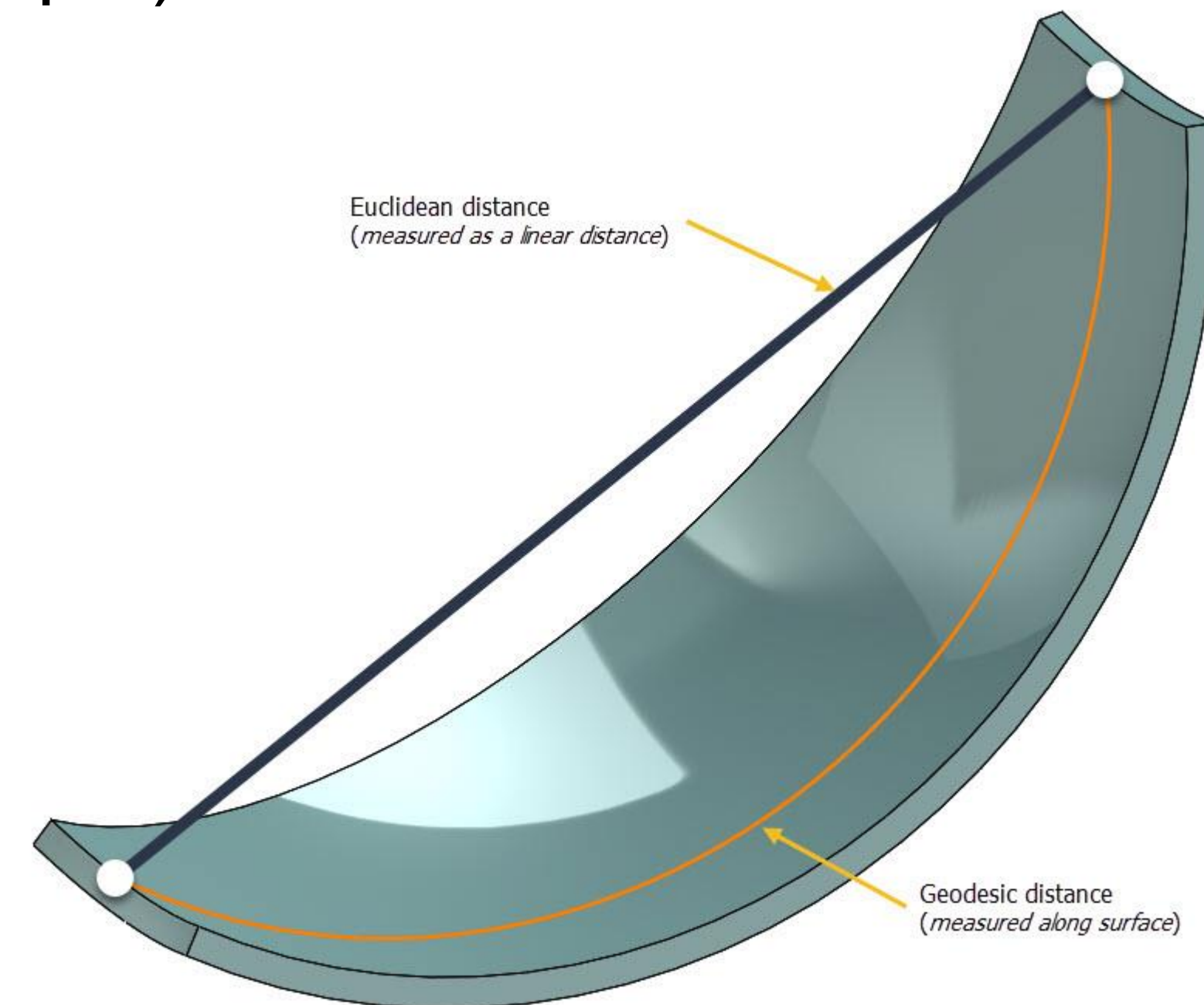
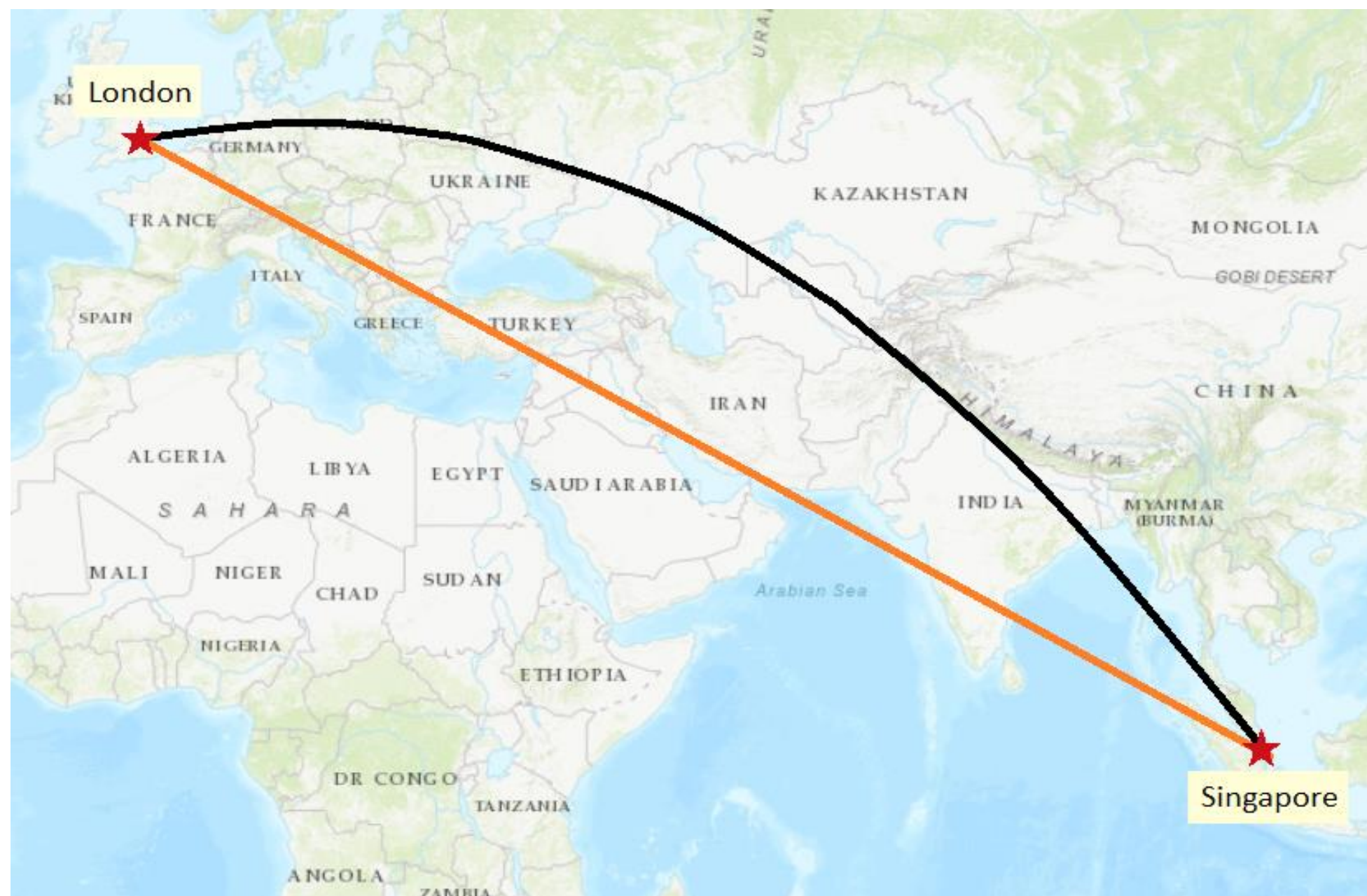


Ευκλείδεια vs γεωδесική απόσταση

«Η μικρότερη απόσταση μεταξύ δύο σημείων είναι μια ευθεία γραμμή»

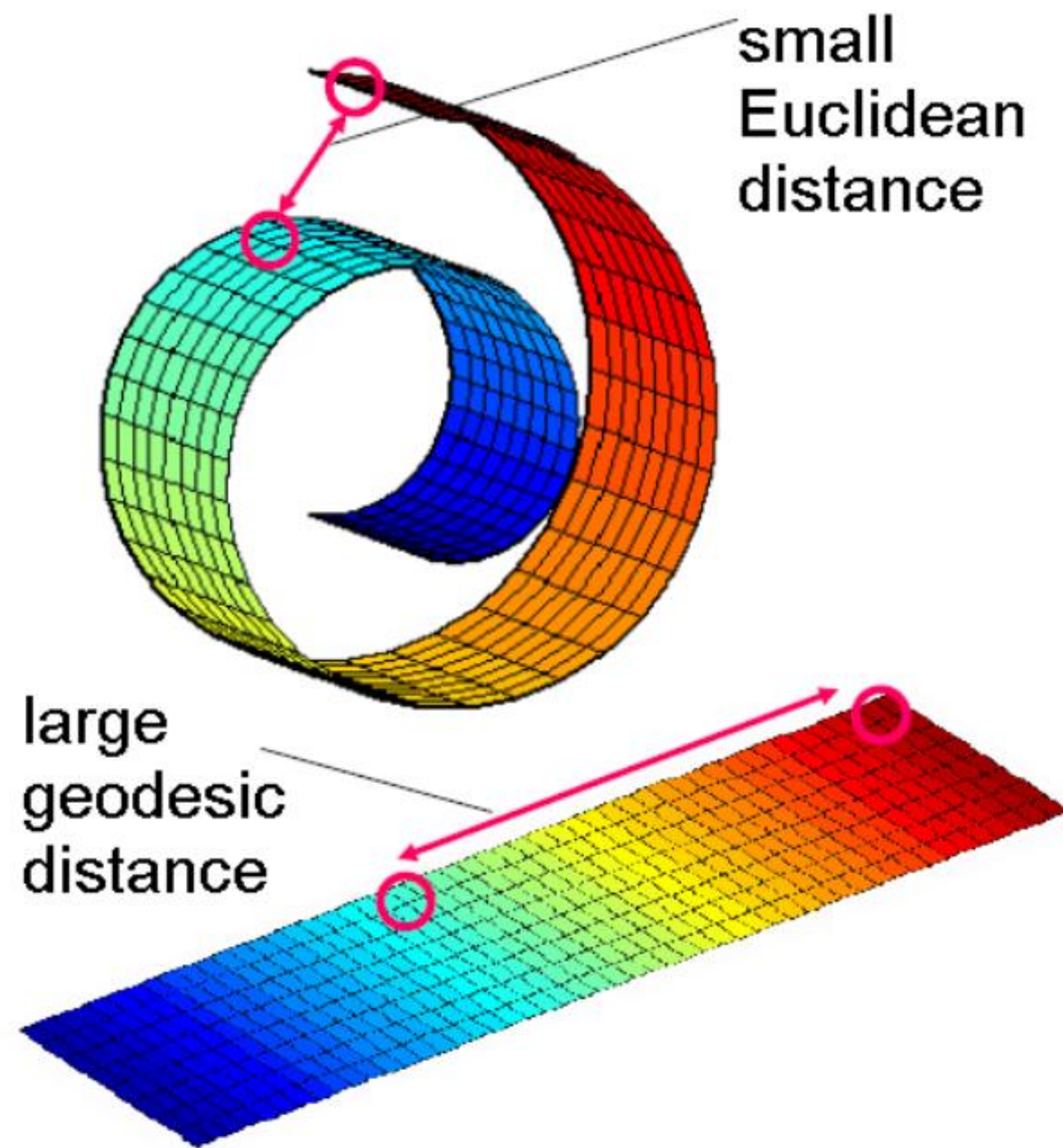
... στην **Ευκλείδεια Γεωμετρία**

Γεωδесική απόσταση: μετρείται κατά μήκος της επιφάνειας (λαμβάνοντας υπόψη τη γεωμετρία)

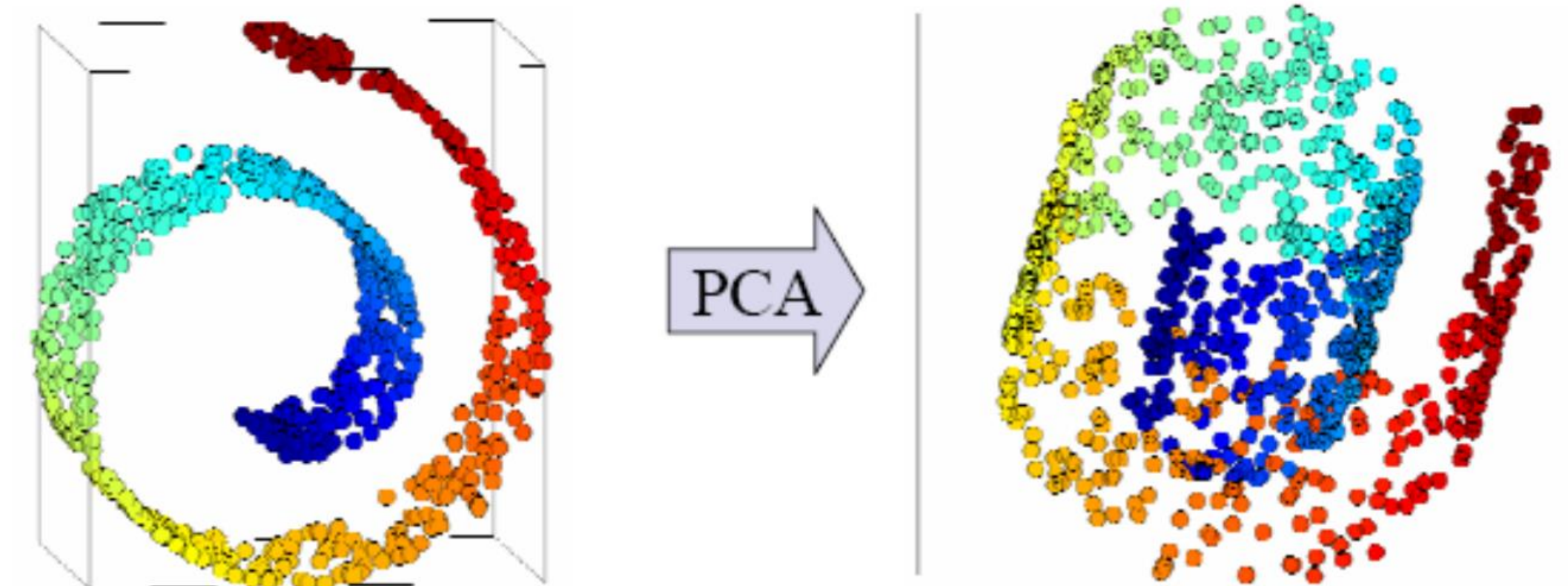




Παράδειγμα «Swiss roll»: 2D πολλαπλή ενσωματωμένη σε 3D χώρο



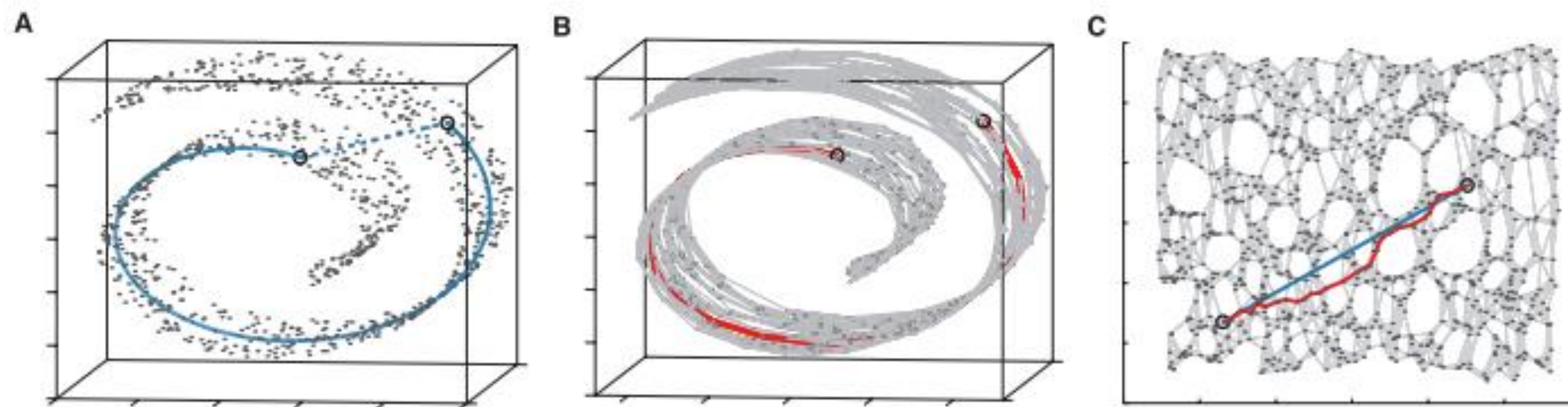
Η PCA είναι γραμμική μέθοδος: αποτυγχάνει να βρει τη μη γραμμική δομή στα δεδομένα επειδή χρησιμοποιεί την Ευκλείδεια απόσταση



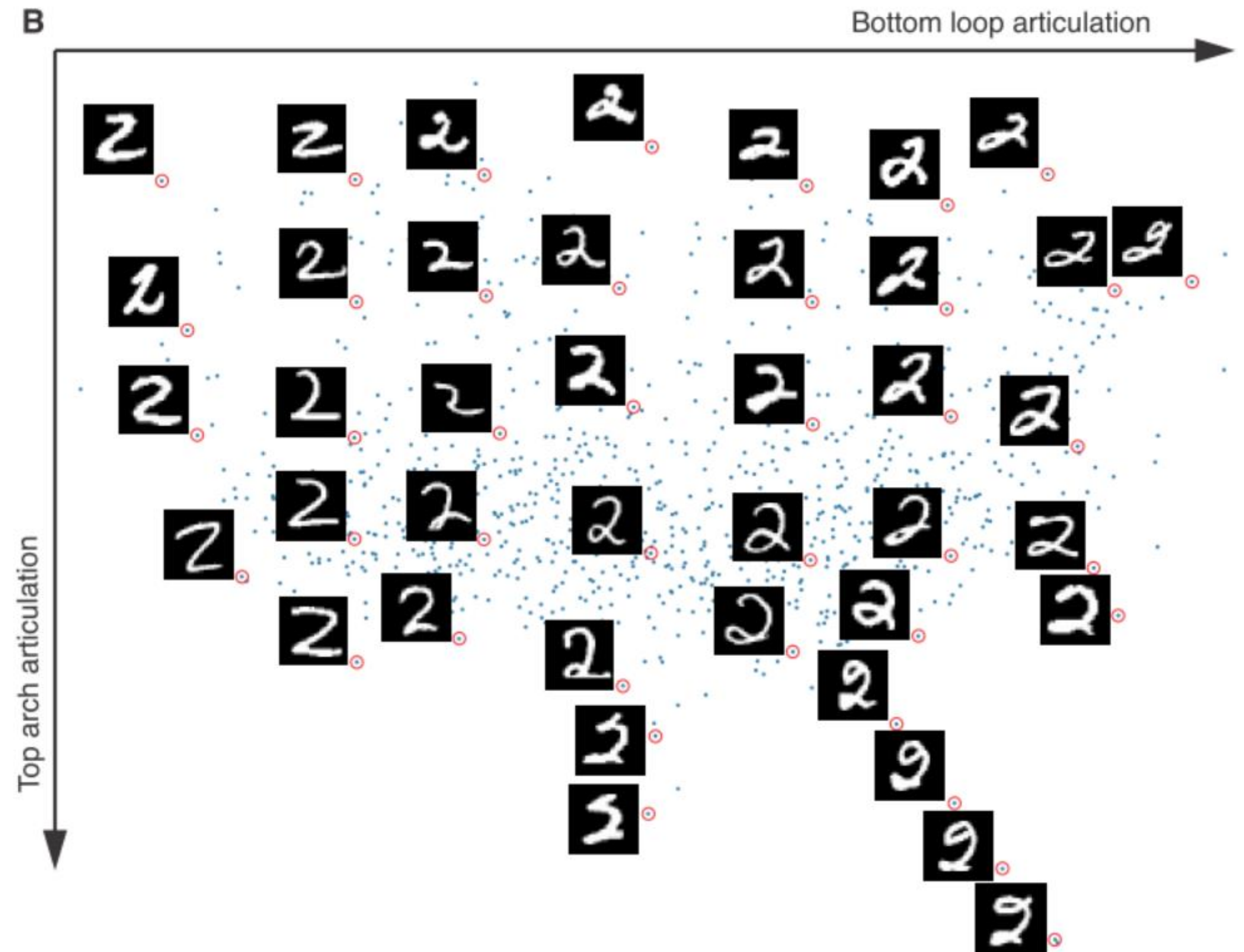


Isometric mapping (Isomap)

Το Isomap επιδιώκει μια χαμηλότερη διάσταση ενσωμάτωσης που διατηρεί γεωδαιτικές αποστάσεις μεταξύ όλων των σημείων.



Η εκμάθηση της πολλαπλής επιτρέπει την παρεμβολή μεταξύ των σημείων (δημιουργία σημείων όχι στο σύνολο κατάρτισης)





t-distributed stochastic neighbor embedding (t-SNE)

- Ευρέως χρησιμοποιούμενη μέθοδος για **την απεικόνιση** των υψηλών διαστάσεων συνόλων δεδομένων ($k = 2$ ή $k = 3$)
- Υπολογίζει την πιθανότητα ότι τα ζεύγη των σημείων δεδομένων στον χώρο υψηλής διάστασης σχετίζονται και επιλέγει χαμηλής διάστασης ενσωματώσεις που παράγουν μια παρόμοια κατανομή
- Διατηρεί μόνο μικρές κατά ζεύγη αποστάσεις (τοπικές ομοιότητες), ενώ η PCA διατηρεί μεγάλες αποστάσεις κατά ζεύγη για να μεγιστοποιήσει τη διακύμανση
- Βελτιστοποιεί μια μη κυρτή συνάρτηση κόστους
 - Διαφορετικές αρχικοποιήσεις έχουν διαφορετικά αποτελέσματα (στοχαστική)
 - Μπορεί να είναι αργή
- Χρήσιμο για οπτικοποίηση, αλλά δεν συνιστάται να χρησιμοποιηθεί για ομαδοποίηση ή ανίχνευση ακραίων τιμών
 - Μπορεί να δείξει συστάδες που δεν υπάρχουν
- Σύσταση: χρησιμοποιήστε το PCA ως ένα βήμα εκκίνησης για τη μείωση των διαστάσεων σε λογική ποσότητα (π.χ. 50) για την μείωση του θορύβου και την επιτάχυνση του υπολογισμού





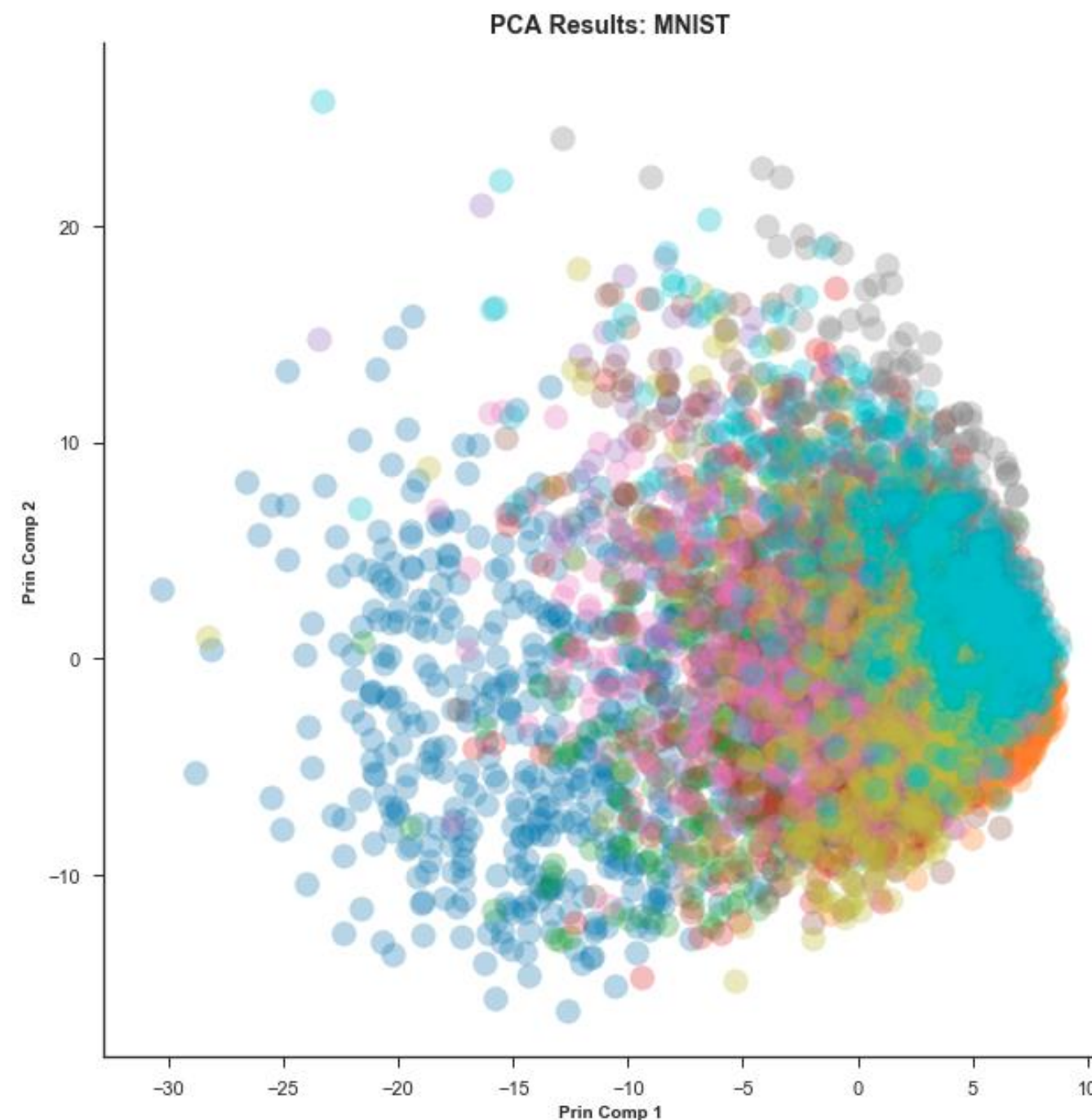
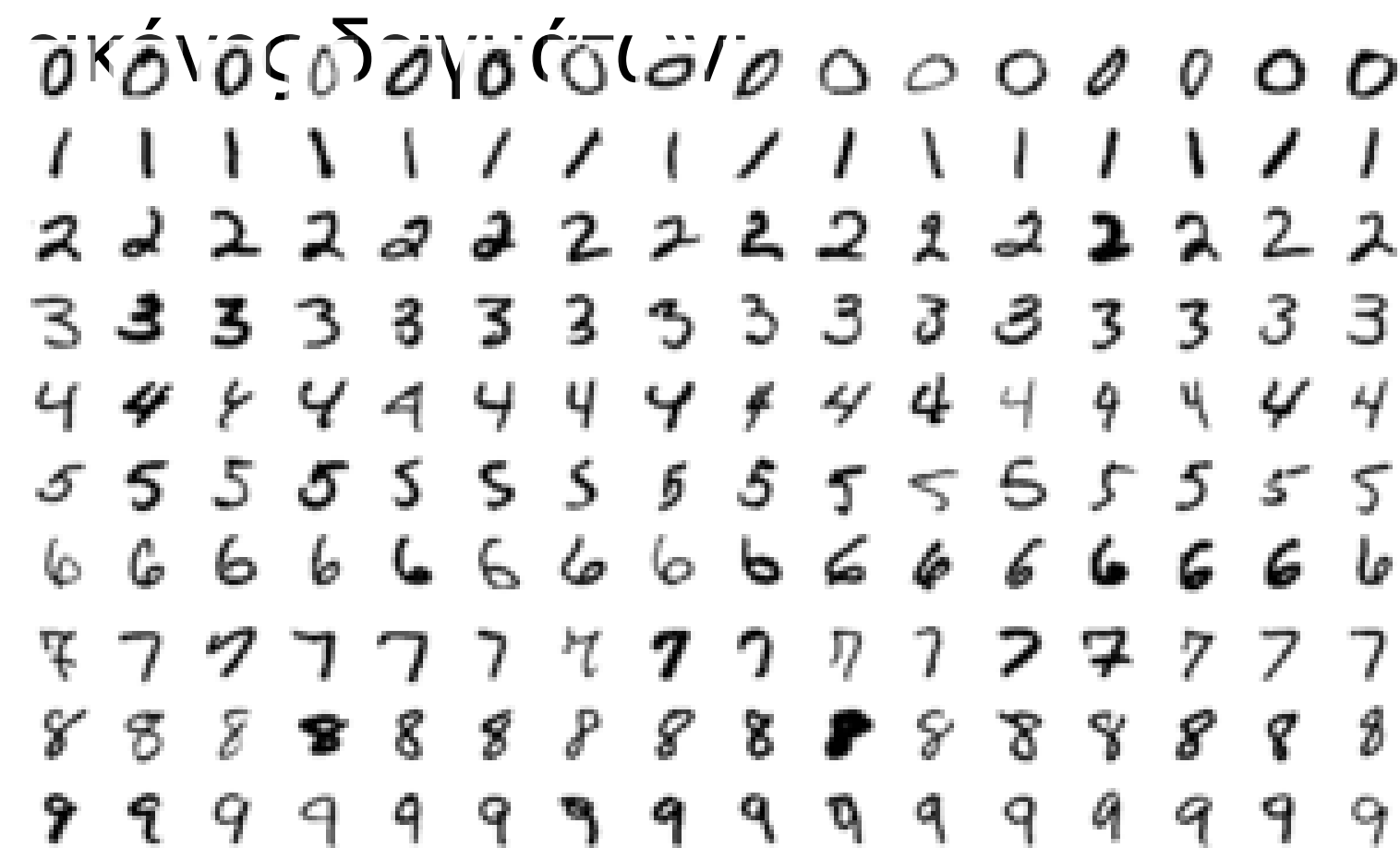
PCA vs t-SNE στα ψηφία MNIST

Πηγή [εικόνων](#)

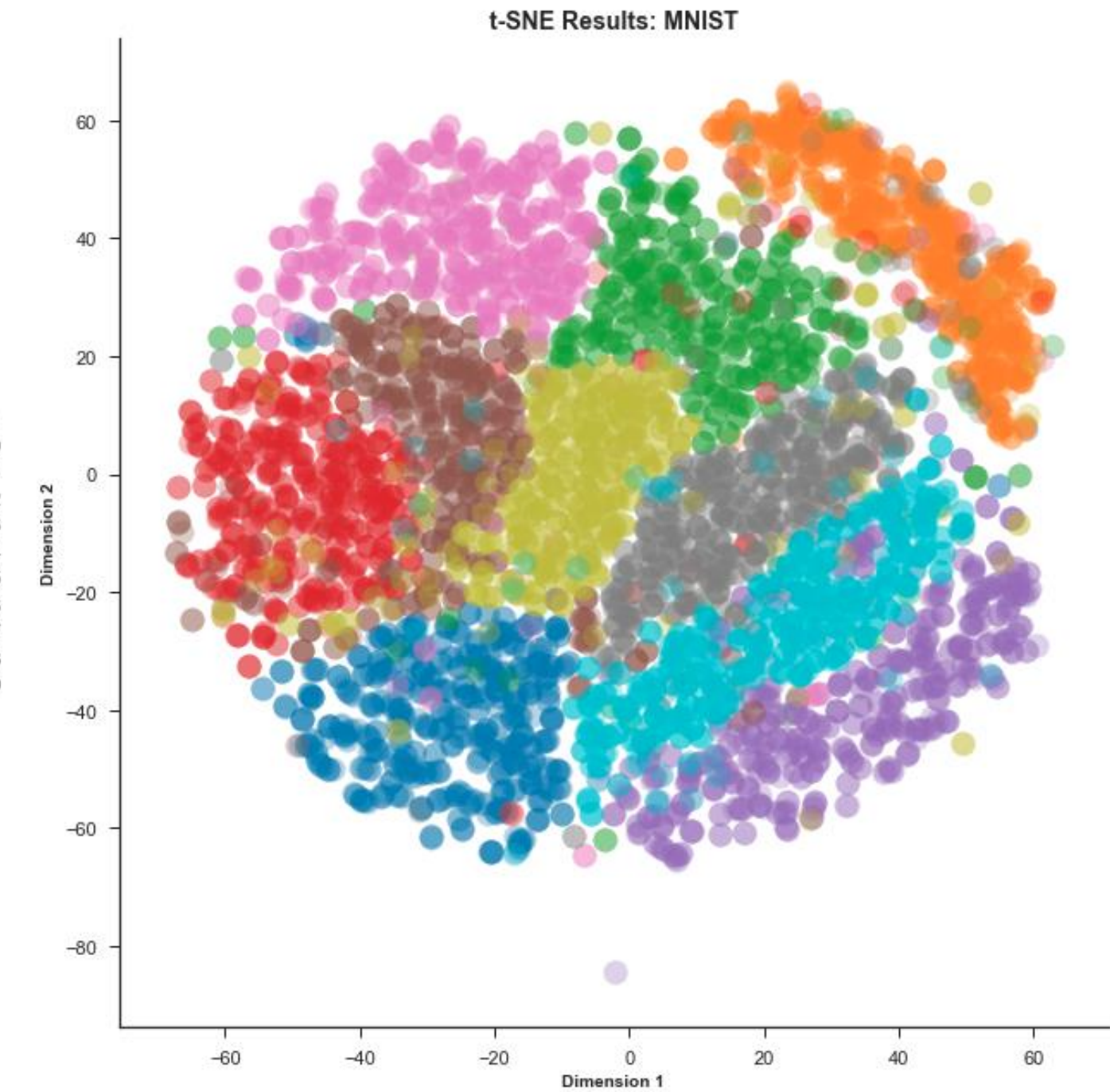
ΤΟ MNIST: σύνολο δεδομένων χειρόγραφων ψηφίων

60000 εκπαιδευτικές εικόνες

10000 εικόνες δοκιμής



PCA



T-SNE





Μέθοδοι μείωσης της διάστασης

- PCA
- Probabilistic PCA
- Kernel PCA
- Factor analysis
- Kohonen Self-organizing Maps
- Independent Component Analysis
- Autoencoders
- Multi-dimensional scaling
- Maximum variance unfolding
- Locally-Linear Embedding
- Isomap
- Hessian Eigenmaps
- Manifold sculpting
- ...





Επόμενες Διαλέξεις

- Ανίχνευση Ανωμαλιών
- Recommender Systems



MAI4CAREU

Master programmes in Artificial
Intelligence 4 Careers in Europe



Σας ευχαριστούμε



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

