Μηχανική Μάθηση

Μιχάλης Τίτσιας

Διάλεξη 9ή

Μέθοδοι μείωσης διάστασης

(*PCA*, *Probabilistic PCA*, *Factor Analysis*)

Περιεχόμενα

- Τι είναι μείωση διάστασης
- Λόγοι που χρησιμοποιούμε μείωση διάστασης
- Principal Component Analysis
- Παραδείγματα εφαρμογής του PCA
- Probabilistic PCA
- Factor Analysis
- Διαφορές μεταξύ Factor Analysis και Probabilistic PCA

Τι είναι μείωση διάστασης

Μείωση διάσταση αφορά τεχνικές όπου ένα σύνολο δεδομένων

$$(\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad \mathbf{x}_n \in \mathbb{R}^D$$

μετασχηματίζεται σε ένα σύνολο

$$(\mathbf{z}_1, \ldots, \mathbf{z}_N), \quad \mathbf{z}_n \in \mathbb{R}^M$$

μικρότερης διάστασης, δηλ. όπου $M < D$ και συνήθως $M \ll D$

Οι τεχνικές αυτές αποτελούν συστήματα μάθησης χωρίς επίβλεψη

- αν και υπάρχουν εξαιρέσεις (που δεν θα μας απασχολήσουν)

Λόγοι που χρησιμοποιούμε μείωση διάστασης

Συμπίεση δεδομένων $\Rightarrow$ πλεονέκτημα όσον αφόρα την αποθήκευση και ταχύτητα στην επεξεργασία

Οπτικοποίηση δεδομένων $\Rightarrow$ γραφική αναπαράσταση της δομής των δεδομένων στο δισδιάστατο ή τρισδιάστατο χώρο ($M = 2$ ή $M = 3$)

Ως ενδιάμεση προεπεξεργασία δεδομένων ενός συστήματος μάθησης με επίβλεψη (π.χ. κατηγοριοποίησης)

- Λιγότερες διαστάσεις $\Rightarrow$ ενδεχόμενη καλύτερη γενίκευση

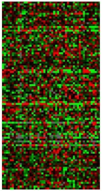
Λόγοι που χρησιμοποιούμε μείωση διάστασης

face images

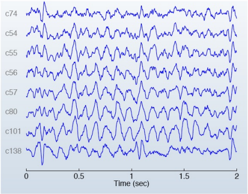
Zambian President Levy Mwanawasa has won a second term in office in an election his challenger Michael Sata accused him of rigging, official results showed on Monday.

According to media reports, a pair of hackers said on Saturday that the Firefox Web browser, commonly perceived as the safer and more customizable alternative to market leader Internet Explorer, is critically flawed. A presentation on the flaw was shown during the ToorCon hacker conference in San Diego.

documents



gene expression data

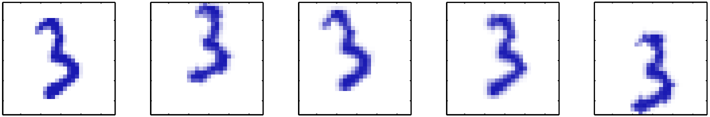


MEG readings

Υπάρχει πληθώρα δεδομένων στα οποία θα θέλαμε να εφαρμόσουμε τεχνικές μείωσης διάστασης

Λόγοι που χρησιμοποιούμε μείωση διάστασης

Παράδειγμα συμπίεσης



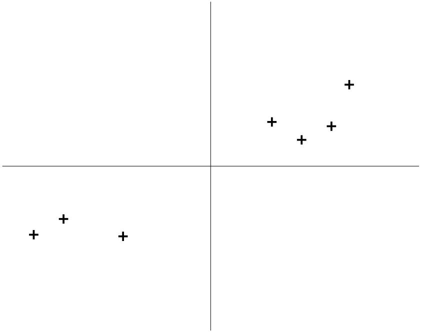
Δεδομένα: Η κάθε εικόνα έχει διάσταση $100 \times 100 = 10000$ αλλά ωστόσο έχει προέρθει από την ίδια εικόνα εφαρμόζοντας κάποια μετατόπιση και περιστροφή

Έχουμε μία οριζόντια και μία κάθετη μετατόπιση καθώς και την περιστροφή $\Rightarrow$ δηλ. 3 βαθμούς ελευθερίας

Οπότε όλο το σύνολο δεδομένων θα μπορούσε να περιγραφεί θεωρητικά

- με μόνο μία είκονα
- και κάθε $\mathbf{x}_n \in \mathbb{R}^{10000}$ με ένα διάνυσμα $\mathbf{z}_n \in \mathbb{R}^3$

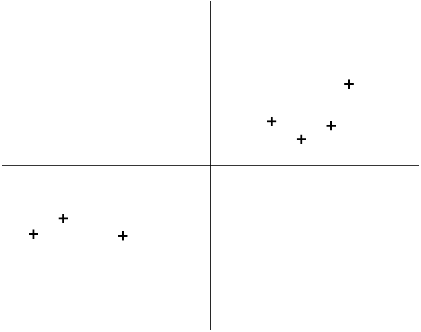
Principal Component Analysis



Έστω τα δεδομένα του σχήματος που έχουν μέση τιμή ίση με το μηδέν, διάσταση $D = 2$ και τιμές

$$X = \begin{bmatrix} -2.9714 & -1.4606 \\ -2.4714 & -1.1253 \\ -1.4714 & -1.4941 \\ 1.0286 & 0.9394 \\ 1.5286 & 0.5589 \\ 2.0286 & 0.8485 \\ 2.3286 & 1.7331 \end{bmatrix}$$

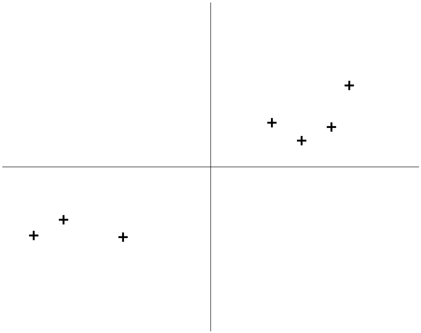
Principal Component Analysis



Θα θέλαμε να βρούμε μια αναπαράσταση στο χώρο $M = 1$

$$X = \begin{bmatrix} -2.9714 & -1.4606 \\ -2.4714 & -1.1253 \\ -1.4714 & -1.4941 \\ 1.0286 & 0.9394 \\ 1.5286 & 0.5589 \\ 2.0286 & 0.8485 \\ 2.3286 & 1.7331 \end{bmatrix} \quad \Rightarrow \quad Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \end{bmatrix}$$

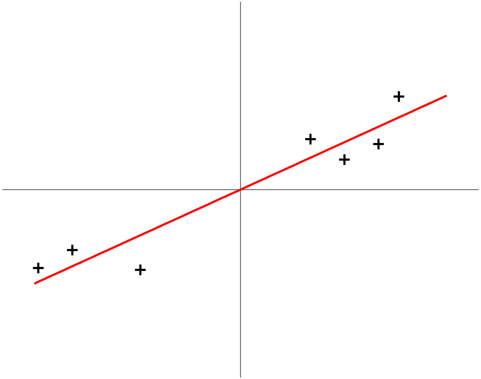
Principal Component Analysis



Πώς θα καθορίσουμε τα Z;

$$X = \begin{bmatrix} -2.9714 & -1.4606 \\ -2.4714 & -1.1253 \\ -1.4714 & -1.4941 \\ 1.0286 & 0.9394 \\ 1.5286 & 0.5589 \\ 2.0286 & 0.8485 \\ 2.3286 & 1.7331 \end{bmatrix} \Rightarrow Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \end{bmatrix}?$$

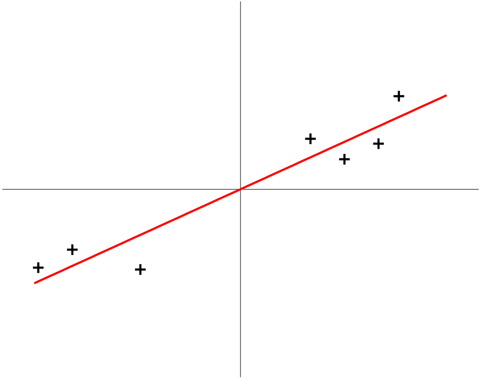
Principal Component Analysis



Η μείωση διάστασης θα βασιστεί σε μια γραμμή που παιρνάει από το μηδέν και επομένως καθορίζεται πλήρως από ένα δισδιάστατο διάνυσμα

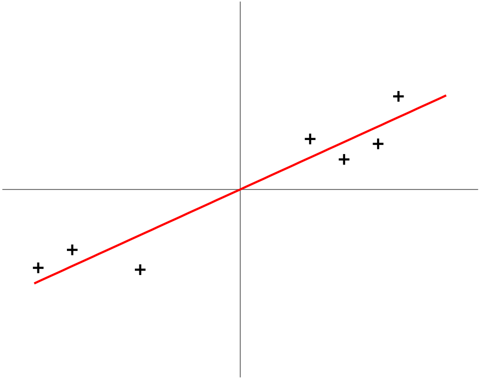
$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Principal Component Analysis



Διαισθητικά η γραμμή αυτή θα πρέπει να έχει μια τέτοια κατεύθυνση που να αντιστοιχεί στην μεγάλυτερη έκταση (διασπορά/διακύμανση) των δεδομένων

Principal Component Analysis

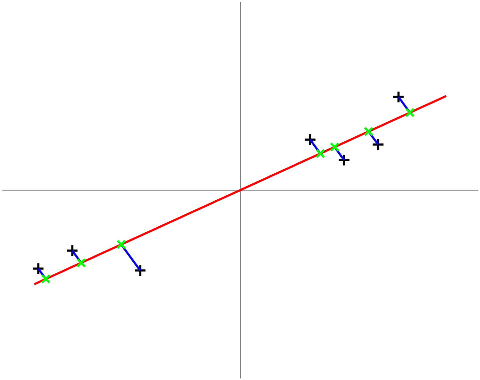


Κάθε σημείο της κόκκινης γραμμής γράφεται ως

$$z\mathbf{u} = z \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} z * u_1 \\ z * u_2 \end{bmatrix}$$

$z \in \mathbb{R}$ (δηλ. μονοδιάσταση τιμή)

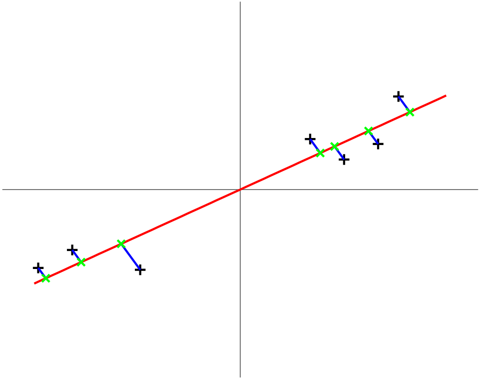
Principal Component Analysis



Θα μπορούσαμε κάθε δεδομένο να το προβάλλουμε πάνω στην κόκκινη γραμμή

- Κάθε μπλε γραμμή σχηματίζει ορθή γωνία με την κόκκινη
- $\Rightarrow$ ώστε η απόσταση του κάθε δεδομένου $+$ από το x είναι η ελάχιστη δυνατή (το x είναι προβολή του $+$)

Principal Component Analysis

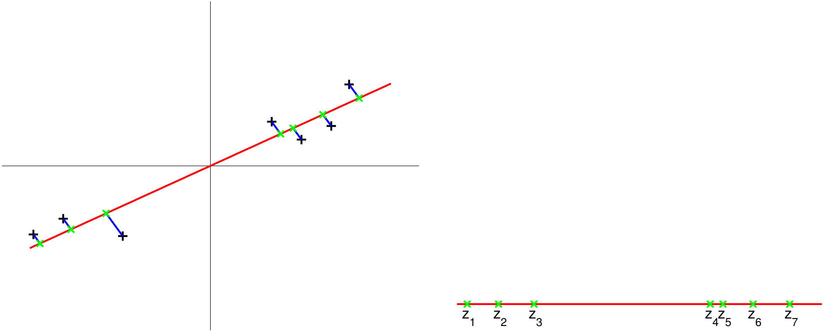


Έτσι τα αρχικά δεδομένα $\mathbf{x}_n = \begin{bmatrix} x_{n1} \\ x_{n2} \end{bmatrix}$, $n = 1, \ldots, 7$ τα ανακατασκευάσουμε με τις προβολές

$$\widetilde{\mathbf{x}}_n = z_n \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad n = 1, \ldots, 7$$

η τιμή του κάθε $z_n \in \mathbb{R}$ είναι η αναπαράσταση στην μειωμένη διάσταση

Principal Component Analysis



Αρχικά δεδομένα: $\mathbf{x}_n = \begin{bmatrix} x_{n1} \\ x_{n2} \end{bmatrix}, \; n = 1, \ldots, 7$

Ανακατασκευασμένα δεδομένα: $\widetilde{\mathbf{x}}_n = z_n \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad n = 1, \ldots, 7$

Αναπαραστάσεις στην μειωμένη διάσταση: $z_n, n = 1, \ldots, 7$

Principal Component Analysis

Αρχικά δεδομένα: $X = \begin{bmatrix} -2.9714 & -1.4606 \\ -2.4714 & -1.1253 \\ -1.4714 & -1.4941 \\ 1.0286 & 0.9394 \\ 1.5286 & 0.5589 \\ 2.0286 & 0.8485 \\ 2.3286 & 1.7331 \end{bmatrix}$

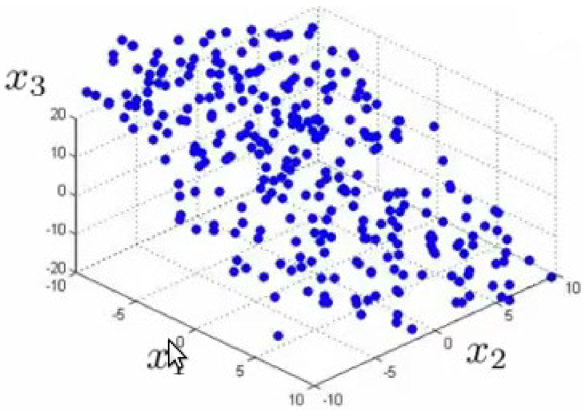
Ανακατασκευασμένα: $\widetilde{X} = \begin{bmatrix} -2.8604 & -1.6526 \\ -2.3404 & -1.3522 \\ -1.7504 & -1.0113 \\ 1.1781 & 0.6806 \\ 1.3881 & 0.8020 \\ 1.8884 & 1.0911 \\ 2.4965 & 1.4424 \end{bmatrix}$

Αναπαραστάσεις στην μειωμένη διάσταση:

$Z = [-3.3035 \;\; -2.7029 \;\; -2.0215 \;\; 1.3606 \;\; 1.6032 \;\; 2.1810 \;\; 2.8833]^T$

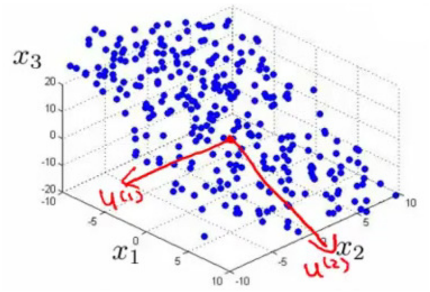
Διάνυσμα προβολής: $\mathbf{u} = [0.8659 \;\; 0.5003]^T$

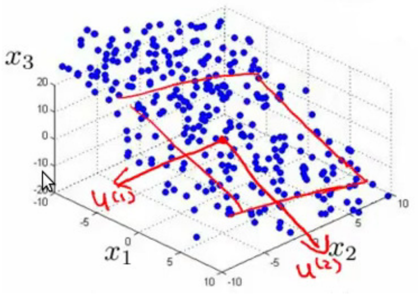
Principal Component Analysis



Έστω ένα παράδειγμα με 3-διάστατα δεδομένα ($D = 2$) για τα οποία θα θέλαμε να μειώσουμε την διάσταση τους και να οδηγηθούμε σε αναπαραστάσεις στις 2 διαστάσεις ($M = 2$)

Principal Component Analysis

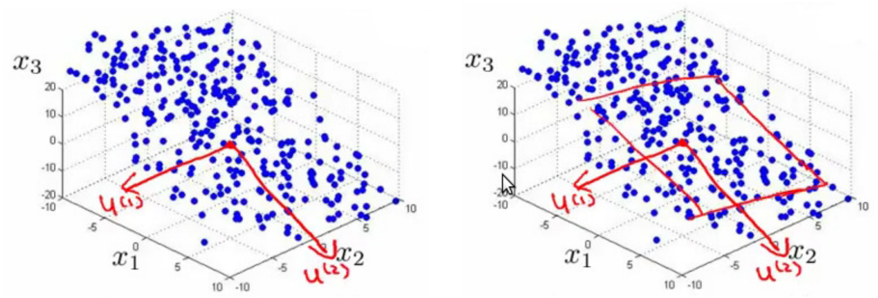




Τώρα θα βασιστούμε σε μια **επιφάνεια προβολής** η οποία καθορίζεται πλήρως από δύο ανεξάρτητα διανύσματα

$$\mathbf{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} u_{21} \\ u_{22} \\ u_{23} \end{bmatrix}$$

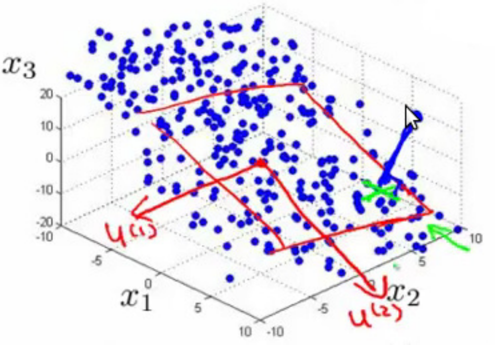
Principal Component Analysis



Αυτά τα διανύσματα αποτελούν βάση για το χώρο της επιφάνειας προβολής και μπορούμε πάντα να τα ορίζουμε ως ορθοκανονικά, δηλ.

$$\mathbf{u}_1^T \mathbf{u}_2 = 0, \quad \mathbf{u}_1^T \mathbf{u}_1 = 1, \quad \mathbf{u}_2^T \mathbf{u}_2 = 1$$

Principal Component Analysis



Αρχικό δεδομένο: $\mathbf{x}_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ x_{n3} \end{bmatrix}$

Ανακατασκευασμένο δεδομένο: $\widetilde{\mathbf{x}}_n = z_{n1}\mathbf{u}_1 + z_{n2}\mathbf{u}_2$

Αναπαράσταση στην μειωμένη διάσταση: $\mathbf{z}_n = \begin{bmatrix} z_{n1} \\ z_{n2} \end{bmatrix}$

Principal components analysis

Στη γενική μορφή

- Έχουμε δεδομένα $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, με $\mathbf{x}_n \in \mathbb{R}^D$, που έχουν κανονικοποιηθεί ώστε να έχουν μέση τιμή μηδέν
- Έχουμε μια ορθακανονική βάση $M < D$ διανυσμάτων $(\mathbf{u}_1, \ldots, \mathbf{u}_M)$, με $\mathbf{u}_n \in \mathbb{R}^D$, χρησιμοποιώντας την οποία μπορούμε να ανακατασκευάσουμε κάθε δεδομένο $\mathbf{x}_n$ ως

$$\widetilde{\mathbf{x}}_n = \sum_{j=1}^{M} z_{nj}\mathbf{u}_j = U\mathbf{z}_n$$

- Το $\mathbf{z}_n \in \mathbb{R}^M$ είναι η αναπαράσταση του $\mathbf{x}_n$ στην μειωμένη διάσταση

PCA: Υπολογίζει τα $(\mathbf{u}_1, \ldots, \mathbf{u}_M)$ και $(\mathbf{z}_1, \ldots, \mathbf{z}_N)$ μέσω μιας ορισμένης συνάρτηση κόστους

Ελαχιστοποίηση του σφάλματος ανακατασκευής

- Συνάρτηση κόστους

$$J = \frac{1}{N}\sum_{n=1}^{N} ||\mathbf{x}_n - \widetilde{\mathbf{x}}_n||^2$$

  ή

$$J = \frac{1}{N}\sum_{n=1}^{N} ||\mathbf{x}_n - U\mathbf{z}_n||^2$$

- Ελαχιστοποιώντας την J ως προς τα z_{nj} και τα διανύσματα $\mathbf{u}_j$ (υπό τον περιορισμό ότι είναι ορθοκανονικά) καταλήγουμε στην λύση του PCA

Λύση
Το $U \in R^{D \times M}$ είναι τα M μεγαλύτερα/πρωτεύοντα, δηλ. που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές, ιδιοδιανύσματα (ονομάζονται principal components) του πίνακα συμμεταβλητότητας των δεδομένων

$$S = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T = \frac{1}{N} X^T X$$

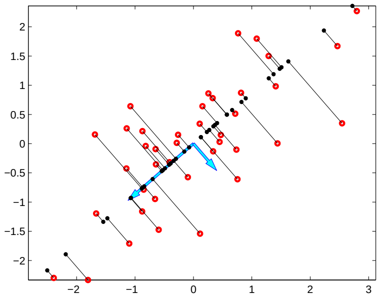
Η αναπαράσταση στην μειωμένη διάσταση για το n δεδομένο παίρνει την μορφή

$$\mathbf{z}_n = U^T \mathbf{x}_n$$

Το ανακασκευασμένο δεδομένο $\widetilde{\mathbf{x}}_n$ γράφεται ως

$$\widetilde{\mathbf{x}}_n = U \mathbf{z}_n = U \left(U^T \mathbf{x}_n \right)$$

Principal components analysis



- Οι κόκκινες κουκίδες δείχνουν τα δεδομένα $\{\mathbf{x}_n\}$ (με μέση τιμή 0)
- Οι **μαύρες κουκίδες** είναι τα ανακατασκευασμένα δεδομένα $\{\widetilde{\mathbf{x}}_n\}$
- Το μέγαλο γαλάζιο βελάκι δείχνει το πρώτο principal component, ενώ το μικρό βελάκι είναι το δεύτερο principal component
- Η απόσταση (με πρόσιμο, δηλ. αρνητική ή θετική) των μαύρων κουκκίδων από την αρχή των αξόνων (το κέντρο των δεδομένων) είναι η αναπαράσταση των $z_n = \mathbf{u}_1^T \mathbf{x}_n$ (**πρόσεξε ότι έχουν μέγιστη διακύμανση κατά μήκος του μεγάλου ιδιοδιανύσματος**)

Principal components analysis

Μπορούμε να οδηγηθούμε ακριβώς στην ίδια λύση χρησιμοποιώντας singular value decomposition (SVD) του $N \times D$ πίνακα των δεδομένων X

$$
\begin{aligned}
[V, S, U] &= svd(X) \\
U &= U(:, 1 : M)
\end{aligned}
$$

Ο τρόπος αυτός αποφεύγει τον υπολογισμό του $D \times D$ πίνακα συμμεταβλητότητας S, οπότε μπορεί να εφαρμοστεί σε περιπτώσεις όπου το D είναι πολύ μεγάλο

Principal components analysis

Αφού ελαχιστοποιηθεί η συνάρτηση κόστους

$$J = \frac{1}{N} \sum_{n=1}^{N} ||\mathbf{x}_n - U\mathbf{z}_n||^2$$

μπορεί να δειχθεί ότι η τελική τιμή της γράφεται ως

$$J = \sum_{i=M+1}^{D} \lambda_i$$

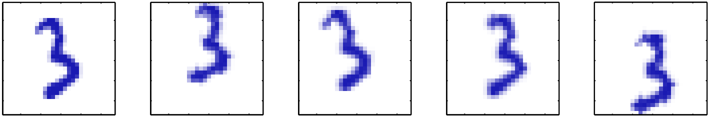
το οποίο είναι το άθροισμα των μικρότερων ιδιοτιμών, δηλ. αυτών που αντιστοιχούν στα εναπομείναντα ιδιοδιανύσματα τα αποία δεν συμπεριλαμβάνονται στα M πρωτεύοντα ιδιοδιανύσματα

Ο λόγος

$$\frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{D} \lambda_i}$$

ουσιαστικά υποδεικνύει το ποσοστό της διακύμανσης/δομής των αρχικών δεδομένων που ανακατασκευάζεται μέσω του PCA

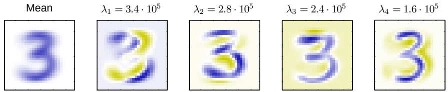
Παραδείγματα εφαρμογής του PCA



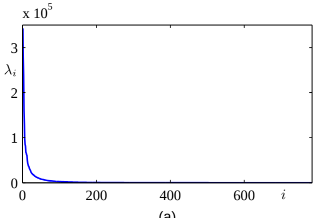
- Εφαρμογή PCA στο σύνολο δεδομένων που περιέχει μετατοπισμένες και περιστράμμες εικόνες της ίδιας αρχικής είκονας του χειρόγραφου αριθμού 3
- Είχαμε πει ότι οι βαθμοί ελευθερίας αυτών των δεδομένων είναι 3
    - ωστόσο το αντίστοιχο manifold, δηλαδή ο χώρος χαμηλής διάστασης είναι υπερβολικά μη γραμμικός
- Το PCA είναι γραμμική μέθοδος $\Rightarrow$ θα χρειαστεί πολύ μεγαλύτερη διάσταση από 3 προκειμένου να ανακατασκευάσει τα δεδομένα ικανοποιητικά

Παραδείγματα εφαρμογής του PCA

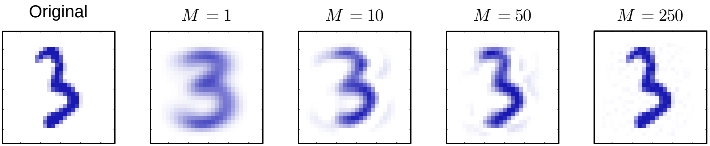
Σχήμα: Μέση εικόνα καθώς και τα 4 πρώτα ιδιοδιανύσματα (δηλ. principal components) με τις αντίστοιχες ιδιοτιμές



Σχήμα: Φάσμα ιδιοτιμών



Παραδείγματα εφαρμογής του PCA



Ανακατασκευή μιας εικόνας του 3 με PCA χρησιμοποιώντας διαφορετικές τιμές της διαστασης M

Παραδείγματα εφαρμογής του PCA

Παραπάνω εμφανίζονται εικόνες από τη βάση Olivetti faces.
Κάθε εικόνα αποτελείται από $92 \times 112 = 10304$ pixels.

Παραδείγματα εφαρμογής του PCA

Δεξιά: Ανακατασκευασμένες εικόνες με $M = 49$. **Αριστερά**: Τα 49 principal components/eigenfaces

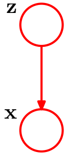
Probabilistic PCA

Υπάρχει μια ερμηνεία της PCA ως εκπαίδευση ενός πιθανοτικού μοντέλου

Η ερμηνεία αυτή έχει πολλά πλεονεκτήματα

- Χρήση του EM αλγορίθμου για εκπαίδευση
- Μάθηση του μοντέλου ακόμα και όταν τα δεδομένα έχουν ελλειπής (missing) τιμές
  - Ένα σημαντικό παράδειγμα τέτοια εφαρμογής είναι τα recommendation systems
- Γενικεύσεις της μεθόδου όπως μίξεις από PCA
- και άλλα (δες Bishop σελίδες 570-571)

Probabilistic PCA



- Η ιδέα είναι να διατυπώσουμε γενικά το πρόβλημα μείωσης διάστασης ως ένα περιγραφικό (generative) μοντέλο κρυμμένων μεταβλητών
  - Για να παραχθεί ένα δεδομένο $\mathbf{x} \in \mathbb{R}^D$, πρώτα υποθέτουμε ότι παράγεται η κρυμμένη (αιτία) μεταβλητή χαμηλής διάστασης $\mathbf{z} \in \mathbb{R}^M$

$$p(\mathbf{z}) = \prod_{j=1}^{M} \mathcal{N}(z_j|0,1) = \mathcal{N}(\mathbf{z}|\mathbf{0}, I)$$

  - Έπειτα παράγεται το δεδομένο από

$$p(\mathbf{x}|\mathbf{z}) = \prod_{d=1}^{D} \mathcal{N}(x_d|\mathbf{w}_d^T\mathbf{z} + \mu_d, \sigma^2) = \mathcal{N}(\mathbf{x}|W\mathbf{z} + \boldsymbol{\mu}, \sigma^2 I)$$

Probabilistic PCA

$$p(\mathbf{z}) = \prod_{j=1}^{M} \mathcal{N}(z_j|0,1) = \mathcal{N}(\mathbf{z}|\mathbf{0}, I)$$

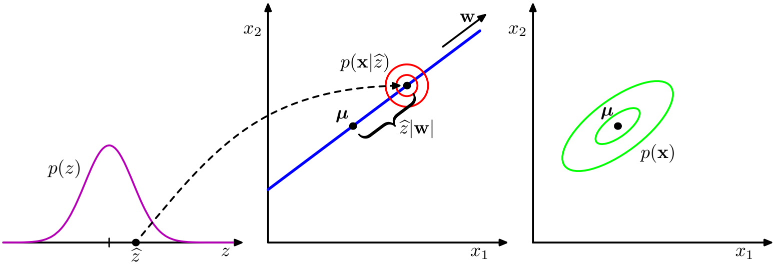
$$p(\mathbf{x}|\mathbf{z}) = \prod_{d=1}^{D} \mathcal{N}(x_d|\mathbf{w}_d^T\mathbf{z} + \mu_d, \sigma^2) = \mathcal{N}(\mathbf{x}|W\mathbf{z} + \boldsymbol{\mu}, \sigma^2 I)$$

όπου $W \in \mathbb{R}^{D\times M}$ είναι ο πίνακας απεικόνισης που μας μεταφέρει από το χώρο της μειωμένης διάστασης στο χώρο των παρατηρούμενων δεδομένων, ενώ $\boldsymbol{\mu}$ είναι ένα διάνυσμα μέσης τιμής (μπορεί να παραληφθεί αν τα δεδομένα έχουν μέση τιμή ίση με το 0)

- Αφού το $\mathbf{z}$ είναι κρυμμένη (δηλ. μη παρατηρήσιμη) μεταβλητή, η κατανομή του δεδομένου $\mathbf{x}$ προκύπτει από το κανόνα αθροίσματος

$$\begin{aligned} p(\mathbf{x}) &= \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \\ &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, WW^T + \sigma^2 I) \end{aligned}$$

Probabilistic PCA



Τα βήματα παραγωγής ενός δισδιάστατου δεδομένου από μία μονοδιάστατη κρυμμένη μεταβλητή

Probabilistic PCA

Αν δοθέντος ενός **x** θέλουμε να βρούμε μια συγκεκριμένη αναπαράσταση του στην μειωμένη διάσταση, τότε θα χρησιμοποιήσουμε τον θεώρημα του Bayes και θα υπολογίσουμε την εκ των υστέρων κατανομή του **z**

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$$

Η κατανομή αυτή θα είναι Gaussian (δες Bishop) και η μέση τιμή της θα μπορούσε να χρησιμοποιηθεί ως μια συγκεκριμένη τιμή για το διάνυσμα μειωμένης αναπαράστασης του δεδομένου **x**

Probabilistic PCA

- Όπως και σε κάθε άλλο πιθανοτικό μοντέλο, δοσμένων των δεδομένων $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ θέλουμε να μεγιστοποιήσουμε την πιθανοφάνεια

$$p(X|W, \boldsymbol{\mu}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}, WW^T + \sigma^2 I)$$

ως προς τις παραμέτρους $(W, \boldsymbol{\mu}, \sigma^2)$

- Η λογαριθμική πιθανοφάνεια γράφεται στην μορφή

$$\mathcal{L}(W, \boldsymbol{\mu}, \sigma^2) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log|C| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^T C^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

όπου

$$C = WW^T + \sigma^2 I$$

Probabilistic PCA

$$\mathcal{L}(W, \boldsymbol{\mu}, \sigma^2) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log|C| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^T C^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

όπου

$$C = WW^T + \sigma^2 I$$

Παίρνωντας παραγώγους και έπειτα από αρκετές πράξεις μπορεί να δειχθεί ότι η λογαριθμική πιθανοφάνεια μεγιστοποιείται όταν ο πίνακας απεικόνισης αναπαριστά τα principal components (δες Bishop για τις λεπτομέρειες)

Όποτε το PCA μπορεί να εξηγηθεί ως η εκπαίδευση ενός πιθανοτικού μοντέλου

Probabilistic PCA

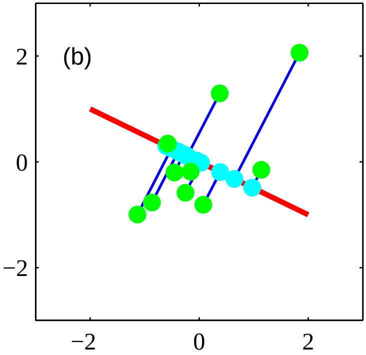
Ένα σημαντικό πλεονέκτημα της πιθανοτικής ερμηνείας είναι ότι μας επιτρέπει την χρήση του EM

1. Ε βήμα: $p(\mathbf{z}_n | \mathbf{x}_n, \boldsymbol{\theta}^{(t)})$
2. Μ βήμα: $\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ όπου
   $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = \sum_{n=1}^{N} \int p(\mathbf{z}_n | \mathbf{x}_n, \boldsymbol{\theta}^{(t)}) \log p(\mathbf{x}_n, \mathbf{z}_n | \boldsymbol{\theta}) d\mathbf{z}_n$

Ο EM αποφεύγει την εφαρμογή ανάλυσης ιδιοτιμών/ιδιοδιανυσμάτων που έχει πολυπλοκότητα $O(D^3)$ και για πολύ μεγαλο D γίνεται υπερβολικά δαπανηρή

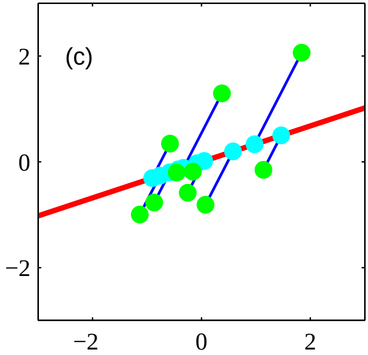
Πιο συγκεκριμένα ο EM βρίσκει τα principal components με επαναληπτικό τρόπο όπου η κάθε επανάληψη έχει μικρή πολυπλοκότητα

Probabilistic PCA



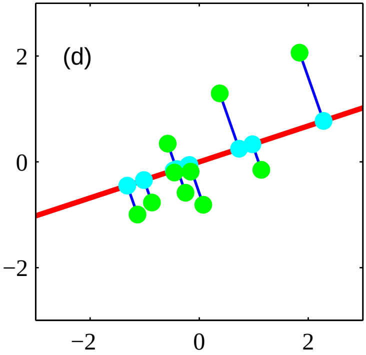
Επάληψη 1, Ε βημα

Probabilistic PCA



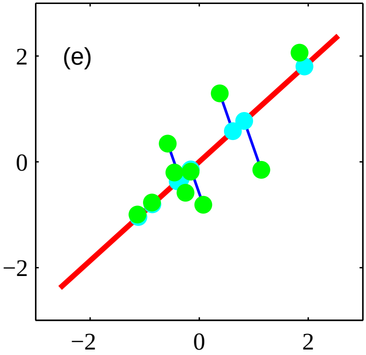
Επάληψη 1, M βημα

Probabilistic PCA



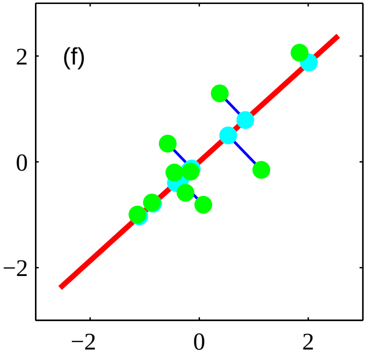
Επάληψη 2, Ε βημα

Probabilistic PCA



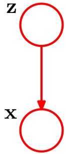
Επάληψη 2, Μ βημα

Probabilistic PCA



Σύγκληση

Factor Analysis



- Factor analysis είναι ένα μοντέλο όμοιο του πιθανοτικού PCA που δίνεται από

$$p(\mathbf{z}) = \prod_{j=1}^{M} \mathcal{N}(z_j|0,1) = \mathcal{N}(\mathbf{z}|\mathbf{0}, I)$$

$$p(\mathbf{x}|\mathbf{z}) = \prod_{d=1}^{D} \mathcal{N}(x_d|\mathbf{w}_d^T\mathbf{z} + \mu_d, \psi_d) = \mathcal{N}(\mathbf{x}|W\mathbf{z} + \mu, \Psi)$$

όπου η διαφορά είναι ότι τώρα υπάρχει μια διαφορετική διακύμανση ψ_d για κάθε συνιστώσα των δεδομένων

  - Αυτό επιτρέπει την μοντελοποίηση συμμεταβλητότητας σε αντίθεση με το PCA που μοντελοποιεί διακύμανση

Η εκπαίδευση του factor analysis γίνεται μόνο με τον αλγόριθμο EM $\Rightarrow$ δεν είναι δυνατή η λύση με ανάλυση ιδιοτιμών και ιδιοδιανυσμάτων!

Διαφορές μεταξύ Factor Analysis και Probabilistic PCA

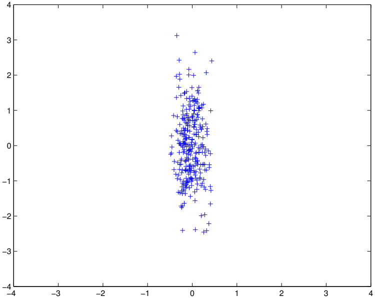
Το μοντέλο του Factor Analysis προσπαθεί να μοντελοποιήσει συμμεταβλητότητα ενώ το Probabilistic PCA διακύμανση

- ως εκ τούτου το Factor Analysis μπορεί να κάνει κάποιου είδους επιλογής χαρακτηριστικών όπου κάποιες συνιστώσες που αποτελούν απλώς ανεξάρτητο θόρυβο μπορούν να αγνοηθούν

Αν περιστρέψουμε τα δεδομένα η λύση του Probabilistic PCA ουσιαστικά δεν αλλάζει (απλά περιστρέφονται τα ιδιοδιανύσματα ενώ κατά τα αλλά το ταίριασμα στα δεδομένα παραμένει ίδιο). Αντιθέτως η λύση του Factor Analysis μπορεί να αλλάξει δραματικά

Η μεγιστοποίηση της πιθανοφάνειας για το μοντέλο του Factor Analysis απαιτεί τη χρήση EM και δεν έχει μοναδική λύση (υπάρχουν πολλά τοπικά μέγιστα στην συνάρτηση κόστους). Αντιθέτως το Probabilistic PCA δίνει μοναδική λύση

Διαφορές μεταξύ Factor Analysis και Probabilistic PCA



Στα δεδομένα του Σχήματος εφαρμόζουμε Probabilistic PCA όπου το κάθε $\mathbf{x} \in \mathbb{R}^2$ παράγεται βάσει

$$\mathcal{N}(\mathbf{x}|\mathbf{0}, \mathbf{w}z, \sigma^2 I)\mathcal{N}(z|0, 1),$$

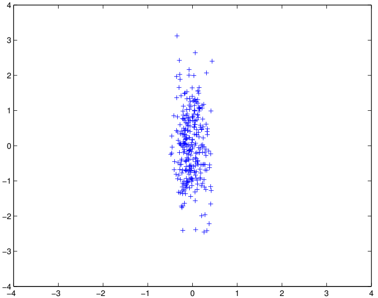
όπου $\mathbf{w} \in \mathbb{R}^2$ και η κρυμμμένη μεταβλητή z είναι μονοδιάστατη, Εφαρμόζουμε επίσης Factor Analysis όπου

$$\mathcal{N}(\mathbf{x}|\mathbf{0}, \mathbf{w}z, \Psi)\mathcal{N}(z|0, 1),$$

όπου ο Ψ είναι δισδιάστατος διαγώνιος πίνακας συμμεταβλητότητας

Πώς νομίζεις ότι το probabilistic PCA θα ταιριάξει στα δεδομένα όσον αφορά την τιμή του $\mathbf{w}$ και πώς το Factor Analysis;

Διαφορές μεταξύ Factor Analysis και Probabilistic PCA



Το Probabilistic PCA θα δώσει προσεγγιστικά ως $\mathbf{w}$ το δίανυσμα $[0\ 1]^T$ που περιγράφει τον κάθετο άξονα λόγω του ότι προς αυτή την κατεύθυνση τα δεδομένα έχουν την μέγιστη διακύμανση

Αντιθέτως το μοντέλο του Factor Analysis μπορεί να εξηγήσει πλήρως τα δεδομένα αυτά χρησιμοποιώντας μόνο τον διαγώνιο πίνακα Ψ. Επομένως θα θέσει προσεγγιστικά το $\mathbf{w}$ ίσο με το μηδέν και ουσιαστικά δεν θα το χρησιμοποιήσει προκειμένου να ερμηνεύσει τα δεδομένα

Γενικά στο Factor Analysis το W στοχεύει να ερμηνεύσει πιθανή συμμεταβλητότητα στις διαστάσεις του διανύσματος δεδομένων. Όταν τέτοια συμμεταβλητότητα δεν υφίσταται (όπως συμβαίνει στο σχήμα) τότε το μοντέλο μπορεί να εξηγήσει τα δεδομένα μόνο μέσω του Ψ

Επίλογος

- Διάβασμα για το σπίτι: Bishop: κεφάλαιο 12 (ως σελίδα 580)
- Επόμενο μάθημα: Μοντελοποίηση ακολουθιακών δεδομένων (Markov Models και Hidden Markov Models)