PCA: Find Projections to Minimize Reconstruction Error

Assume data is set of d-dimensional vectors, where nth vector is

$$\mathbf{x}^n = \langle x_1^n \ldots x_d^n \rangle$$

We can represent these in terms of any d orthogonal vectors $u_1 \ldots u_d$

$$\mathbf{x}^n = \sum_{i=1}^{d} z_i^n \mathbf{u}_i; \quad \mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$$

PCA: given M<d. Find $\langle \mathbf{u}_1 \ldots \mathbf{u}_M \rangle$

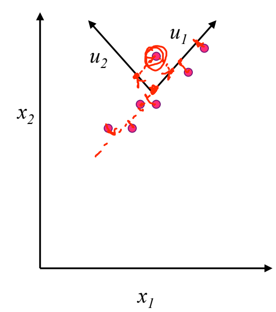
that minimizes $E_M \equiv \sum_{n=1}^{N} ||\mathbf{x}^n - \hat{\mathbf{x}}^n||^2$

where $\hat{\mathbf{x}}^n = \bar{\mathbf{x}} + \sum_{i=1}^{M} z_i^n \mathbf{u}_i$

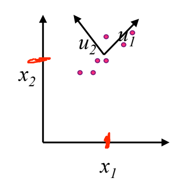
original reconstructed

Mean

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}^n$$



PCA



PCA: given M<d. Find $\langle \mathbf{u}_1 \ldots \mathbf{u}_M \rangle$

that minimizes $E_M \equiv \sum_{n=1}^{N} ||\mathbf{x}^n - \hat{\mathbf{x}}^n||^2$

where $\hat{\mathbf{x}}^n = \bar{\mathbf{x}} + \sum_{i=1}^{M} z_i^n \mathbf{u}_i$

dimensions to encode x's

dimension of X

Note we get zero error if M=d, so all error is due to missing components.

Therefore,

$$E_M = \sum_{i=M+1}^{d} \sum_{n=1}^{N} [\mathbf{u}_i^T (\mathbf{x}^n - \bar{\mathbf{x}})]^2$$

$$= \sum_{i=M+1}^{d} \mathbf{u}_i^T \Sigma \, \mathbf{u}_i$$

This minimized when $\boldsymbol{u}_i$ is eigenvector of Σ, the covariance matrix of X. i.e., minimized when:

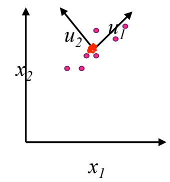
$$\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

Covariance matrix: $\Sigma = \sum (\mathbf{x}^n - \bar{\mathbf{x}})(\mathbf{x}^n - \bar{\mathbf{x}})^T$

$$\Sigma_{ij} = \sum_{n=1}^{N} (x_i^n - \bar{x}_i)(x_j^n - \bar{x}_j)$$

PCA

Minimize $E_M = \sum_{i=M+1}^{d} \mathbf{u}_i^T \Sigma \; \mathbf{u}_i$



$\rightarrow \quad \Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i$

Eigenvector of Σ

Eigenvalue (scalar)

$\rightarrow \quad E_M = \sum_{i=M+1}^{d} \lambda_i$

PCA algorithm 1:

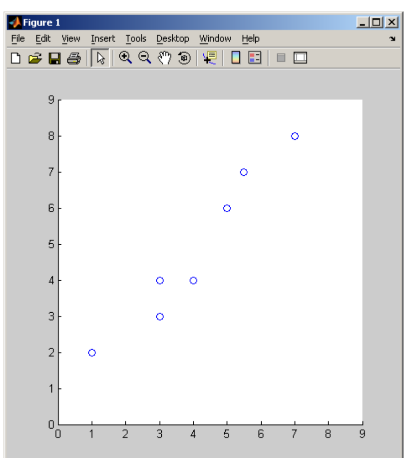
# of pts

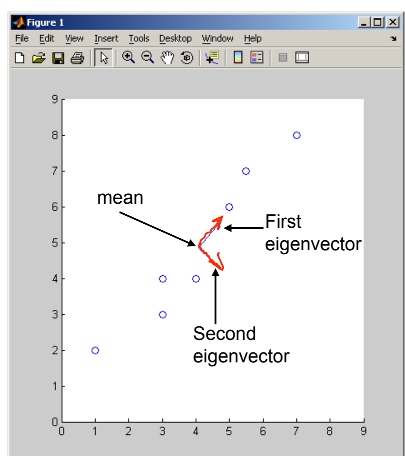
dimension

1. X ← Create N x d data matrix, with one row vector x^n per data point
2. X ← subtract mean $\bar{x}$ from each row vector x^n in X
3. Σ ← covariance matrix of X ✓
4. Find eigenvectors and eigenvalues of Σ
5. PC's ← the M eigenvectors with largest eigenvalues

PCA Example

$$\hat{\mathbf{x}}^n = \bar{\mathbf{x}} + \sum_{i=1}^{M} z_i^n \mathbf{u}_i$$

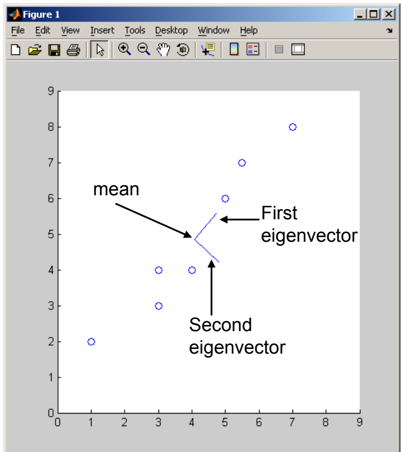


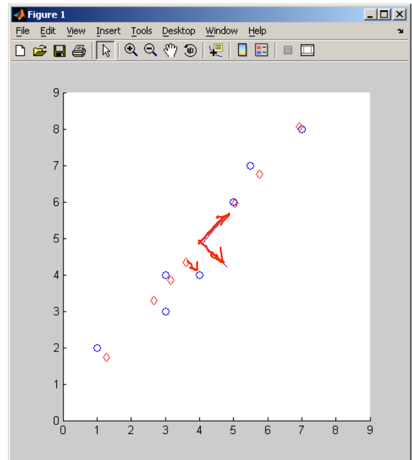


PCA Example

$$\hat{\mathbf{x}}^n = \bar{\mathbf{x}} + \sum_{i=1}^{M} z_i^n \mathbf{u}_i$$

Reconstructed data using only first eigenvector (M=1)





Very Nice When Initial Dimension Not Too Big

What if very large dimensional data?

• e.g., Images (d ¸ 10^4)

Problem:

• Covariance matrix Σ is size (d x d)

• $d=10^4 \rightarrow |\Sigma| = \boxed{10^8}$

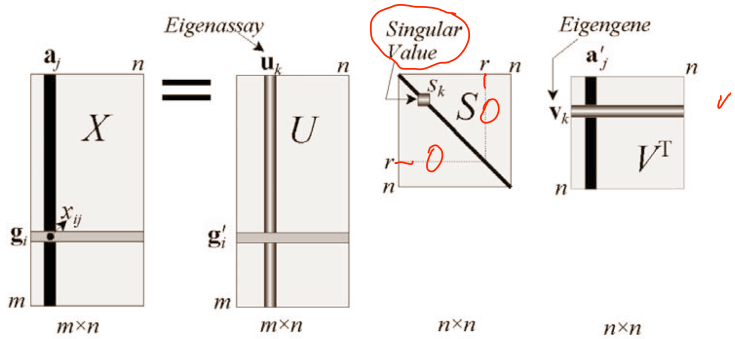
Singular Value Decomposition (SVD) to the rescue!

• pretty efficient algs available, including Matlab SVD

• some implementations find just top N eigenvectors

SVD

$$X = USV^{T}$$



Data X, one row per data point

US gives coordinates of rows of X in the space of principle components

S is diagonal, $S_k > S_{k+1}$, S_k^2 is kth largest eigenvalue

Rows of V^T are unit length eigenvectors of X^TX

If cols of X have zero mean, then $X^TX = c\,\Sigma$ and eigenvects are the Principle Components

[from Wall et al., 2003]

Singular Value Decomposition

To generate principle components:

- Subtract mean $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x^n$ from each data point, to create zero-centered data
- Create matrix X with one row vector per (zero centered) data point
- Solve SVD: $X = USV^T$
- Output Principle components: columns of V (= rows of V^T)
  - Eigenvectors in V are sorted from largest to smallest eigenvalues
  - S is diagonal, with s_k^2 giving eigenvalue for kth eigenvector

Singular Value Decomposition

To project a point (column vector x) into PC coordinates:

$V^T x$

$X = U S V^T$

If x_i is i^{th} row of data matrix X, then

- (i^{th} row of US) $= V^T x_i^T$
- $(US)^T = V^T X^T$

To project a column vector x to M dim Principle Components subspace, take just the first M coordinates of $V^T x$

Independent Components Analysis (ICA)

- PCA seeks orthogonal directions $<Y_1 \ldots Y_M>$ in feature space X that minimize reconstruction error

- ICA seeks directions $<Y_1 \ldots Y_M>$ that are most *statistically independent*. I.e., that minimize $I(Y)$, the mutual information between the Y_j :

$$I(Y) = \left[\sum_{j=1}^{J} H(Y_j) \right] - H(Y)$$