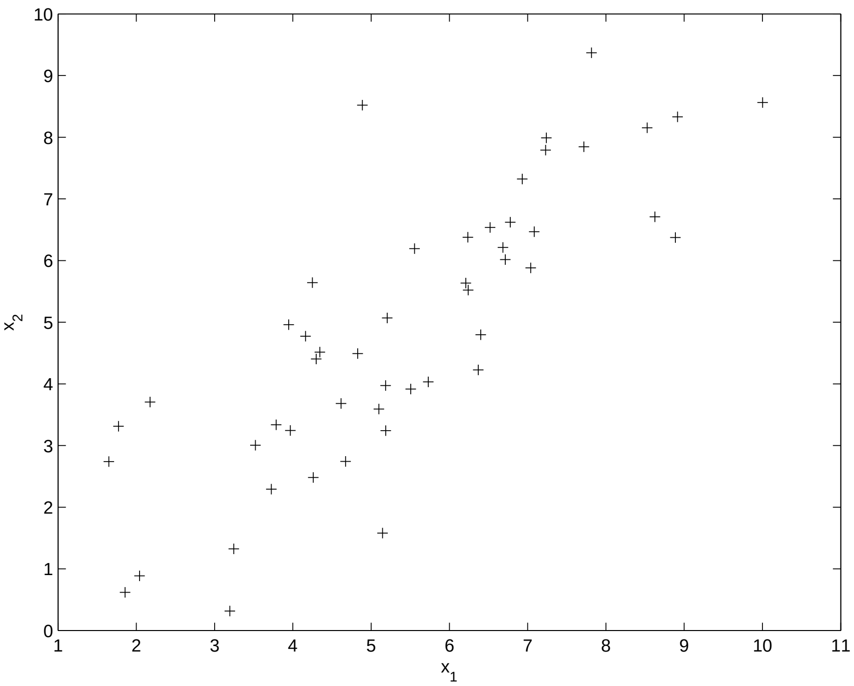
Principal Components Analysis

Debrup Chakraborty

CINVESTAV
email: debrup@cs.cinvestav.mx

PCA (Motivation)



- This data is in two dimensions but the two attributes x_1 and x_2 are strongly correlated
- We can view that the data originally lies along a diagonal axis, with some noise

PCA (Motivation)

- We shall try to device a method to try to find the best direction in which the data can be projected so as to maximize the variance in the data.
- By PCA we can find a lower dimensional representation of a given data. Thus, this is a dimensionality reduction technique.

Normalization prior to PCA

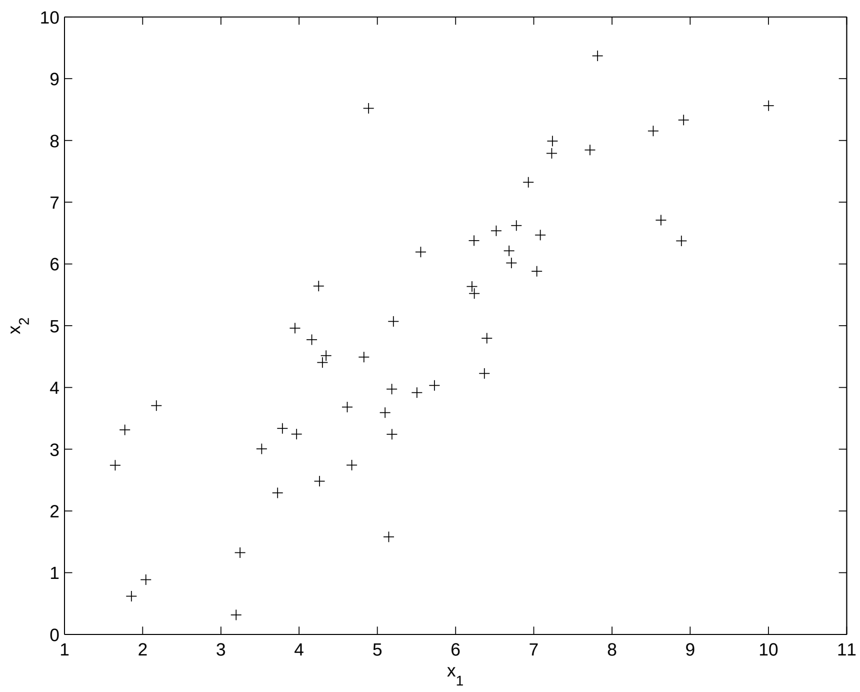
- Assume the data set to be $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$

1. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{x}^{(i)}$
2. Replace each $\boldsymbol{x}^{(i)}$ with $\boldsymbol{x}^{(i)} - \mu$.
3. Let $\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m} (\boldsymbol{x}_j^{(i)})^2$
4. Replace each $\boldsymbol{x}_j^{(i)}$ with $\boldsymbol{x}_j^{(i)}/\sigma_j$

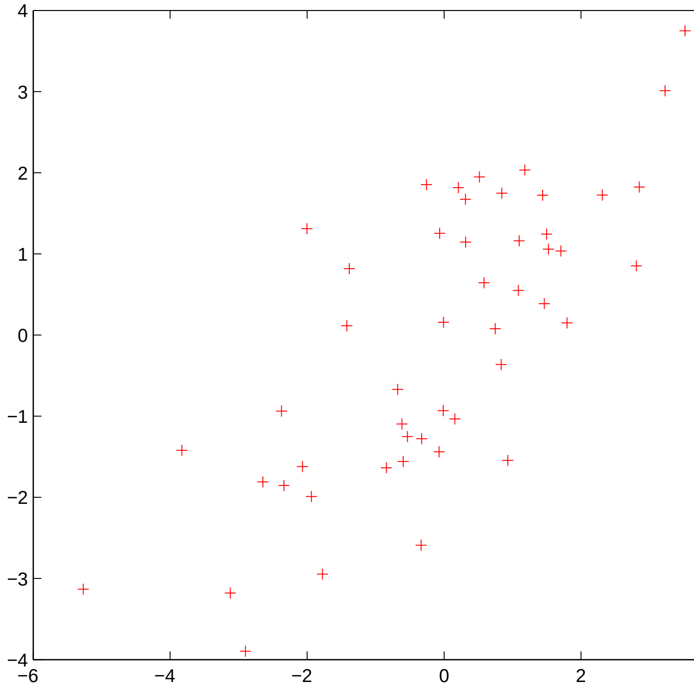
Normalization prior to PCA

- Steps 1-2 zero out the mean
- Steps 3-4 rescale each co-ordinate to have unit variance
- These steps can be omitted for data which are known to have zero mean and unit variance in all attributes.
- This rescaling keeps the data in same scale for each attribute, and makes the individual attributes comparable.
- This does not hamper the structure of the data.

Normalization prior to PCA

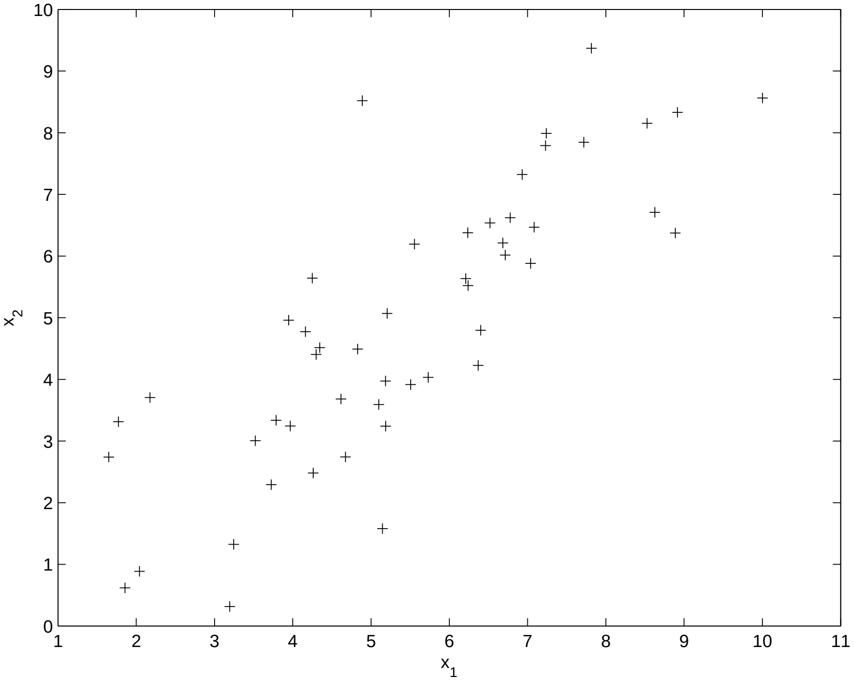


Original

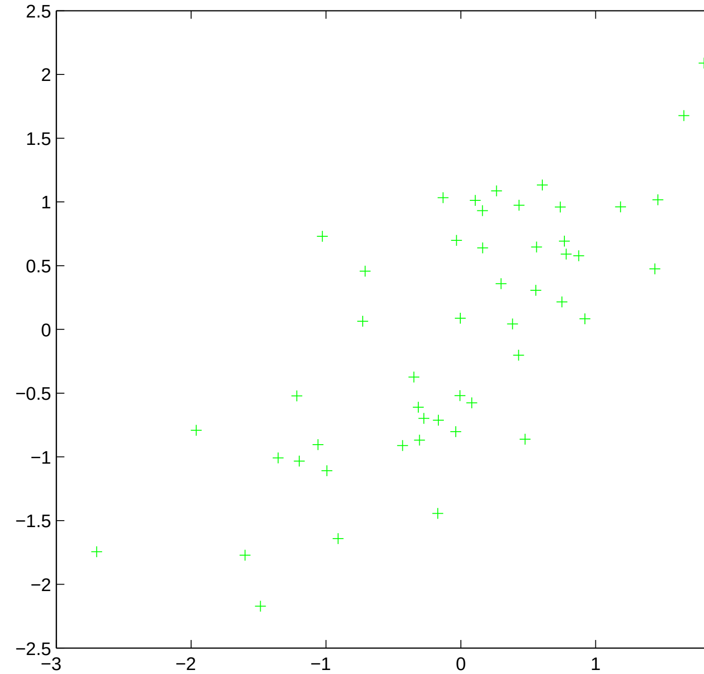


Step 2

Normalization prior to PCA



Original



Step 3

PCA

- Now after this normalization, our task would be to compute the major axis of variation of the data.
- We can pose this problem in the following manner:

> Find the unit vector $\boldsymbol{u}$ so that when the data is projected in the direction corresponding to $\boldsymbol{u}$, the variance of the projected data is maximized

PCA

- Given an unit vector $\boldsymbol{u}$ and a point $\boldsymbol{x}$, the length of the projection $\boldsymbol{x}$ onto $\boldsymbol{u}$ is given by $\boldsymbol{x}^T\boldsymbol{u}$.
- Hence to maximize the variance of the projections, we can propose the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{u}} \quad & \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{x}^{(i)^T}\boldsymbol{u})^2 \\ \text{such that} \quad & ||\boldsymbol{u}|| = 1 \end{aligned}$$

PCA

- Now,

$$\frac{1}{m}\sum_{i=1}^{m}({\boldsymbol{x}^{(i)}}^T\boldsymbol{u})^2 = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{u}^T\boldsymbol{x}^{(i)}{\boldsymbol{x}^{(i)}}^T\boldsymbol{u}$$

$$= \boldsymbol{u}^T\left(\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}^{(i)}{\boldsymbol{x}^{(i)}}^T\right)\boldsymbol{u}$$

- Let

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

- This is the empirical covariance matrix of the data (assuming, the data is zero mean)

PCA

- The optimization problem now can be posed as

$$\begin{aligned} \max_{\boldsymbol{u}} \quad & \boldsymbol{u}^T (\Sigma) \boldsymbol{u} \\ \text{such that} \quad & ||\boldsymbol{u}|| = 1 \end{aligned}$$

Result : The $\boldsymbol{u}$ which is a solution to this problem is the principal eigenvector of Σ.

Eigenvalues and eigenvectors

- Given a $d \times d$ matrix M, a very important class of linear equations is of the form

$$M\boldsymbol{x} = \lambda\boldsymbol{x}$$

  where λ is a scalar
- The above eq. can be rewritten as

$$(M - \lambda I)\boldsymbol{x} = \mathbf{0},$$

  where I is the identity matrix and $\mathbf{0}$ is the zero vector.
- The solution vector $\boldsymbol{x} = \boldsymbol{e}_i$ and the corresponding scalar λ_i is called the eigenvector and associated eigenvalue respectively.

Eigenvalues and eigenvectors

- If M is real symmetric, there are d (possibly nondistinct) solution vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$ each with an associated eigenvalue $\{\lambda_1, \lambda_2, \ldots, \lambda_d\}$.
- Under multiplication by M the eigenvectors are changed only in magnitude, not in direction:

$$M\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j$$

- If M is diagonal, then he eigenvectors are parallel to the coordinate axes.

Eigenvalues and eigenvectors

- One method of finding the eigenvectors and eigenvalues is to solve the *characteristic equation*

$$det(M - \lambda I) = 0$$

- The above equation in λ has d roots (possibly nondistinct).
- For each such root, we then solve a set of linear equations to find its associated eigenvector.

PCA

- In the general case if we wish to project our data into a k dimensional subspace where $k < n$, we should choose $u_1, u_2, ..., u_k$ to be the eigenvectors corresponding to the top k eigenvalues of the matrix Σ.
- To represent $\boldsymbol{x}^{(i)}$ in the new basis, we need only to compute the corresponding vector

$$\xi^{(i)} = \begin{bmatrix} u_1^T \boldsymbol{x}^{(i)} \\ u_2^T \boldsymbol{x}^{(i)} \\ \vdots \\ u_k^T \boldsymbol{x}^{(i)} \end{bmatrix} \in \Re^k$$

More references

- Andrew Ng's notes, (has a link in the website)
- Duda, Hart, Stork (Chap. 3, Page 115) (It has a bit different treatment)