CHAPTER 6 – Principal Components Analysis **(PCA)**

(The Karhunen – Loève transform):

- The goal: Given an original set of m measurements $\underline{x} \in \Re^m$ compute $\underline{y} \in \Re^\ell$

$$\underline{y} = A^T \underline{x}$$

for an orthogonal A, so that the elements of $\underline{y}$ are optimally mutually uncorrelated.

That is

$$E\left[y(i)y(j)\right] = 0, \ i \neq j$$

- Sketch of the proof:

$$R_y = E\left[\underline{y}\,\underline{y}^T\right] = E\left[A^T \underline{x}\,\underline{x}^T A\right] = A^T R_x A$$

- If A is chosen so that its columns $\underline{a}_i$ are the orthogonal eigenvectors of R_x, then

$$R_y = A^T R_x A = \Lambda$$

where Λ is diagonal with elements the respective eigenvalues λ_i.

- Observe that this is a sufficient condition but not necessary. It **imposes** a specific orthogonal structure on A.

➢ Properties of the solution

- Mean Square Error approximation.

Due to the orthogonality of A:

$$\underline{x} = \sum_{i=0}^{m} y(i)\underline{a}_i, \quad y(i) = \underline{a}_i^{\ T} \underline{x}$$

– Define

$$\underline{\hat{x}} = \sum_{i=0}^{\ell-1} y(i)\underline{a}_i$$

– The Karhunen – Loève transform minimizes the square error:

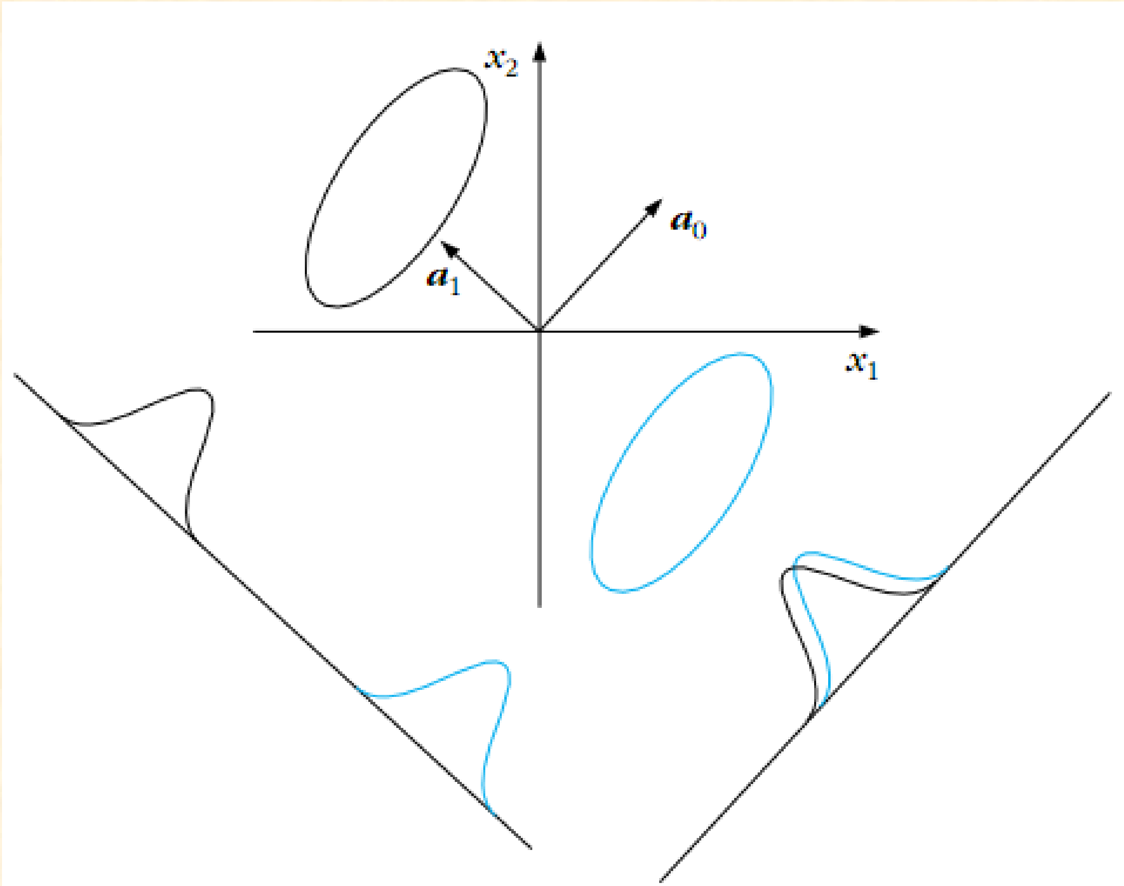
$$E\left[\left\|\underline{x} - \underline{\hat{x}}\right\|^2\right] = E\left[\left\|\sum_{i=\ell}^{m} y(i)\underline{a}_i\right\|^2\right]$$

– The error is:

$$E\left[\left\|\underline{x} - \underline{\hat{x}}\right\|^2\right] = \sum_{i=\ell}^{m} \lambda_i$$

It can be also shown that this is the minimum mean square error compared to **any** other representation of $\underline{x}$ by an ℓ-dimensional vector.

– In other words, $\underline{\hat{x}}$ is the projection of $\underline{x}$ into the subspace spanned by the principal ℓ eigenvectors. However, for Pattern Recognition this is not the always the best solution.



- Total variance: It is easily seen that

$$\sigma^2_{y(i)} = E\left[y^2(i)\right] = \lambda_i$$

Thus Karhunen – Loève transform makes the total variance maximum.

- Assuming $\underline{y}$ to be a zero mean multivariate Gaussian, then the K-L transform maximizes the entropy:

$$H_y = -E\left[\ln P_{\underline{y}}(\underline{y})\right]$$

of the resulting $\underline{y}$ process.

- Summary of the PCA algorithm
  - Estimate the autocorrelation matrix

$$R \cong \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T = \frac{1}{n} X^T X$$

where $$X = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix}$$

  - Perform eigenanalysis of R.
  - Choose the l largest eigenvalues / eigenvectors

$$\lambda_i , \underline{a}_i , i = 0,1,\ldots, l-1$$

  - Project a given $\underline{x}$ onto the low l-dimensional space, spanned by $\underline{a}_0, \underline{a}_1, \ldots, \underline{a}_{l-1}$.

$$y_i = \underline{a}_{\mathrm{i}}^{\mathrm{T}} \underline{x} , i = 0,1,\ldots, l-1$$

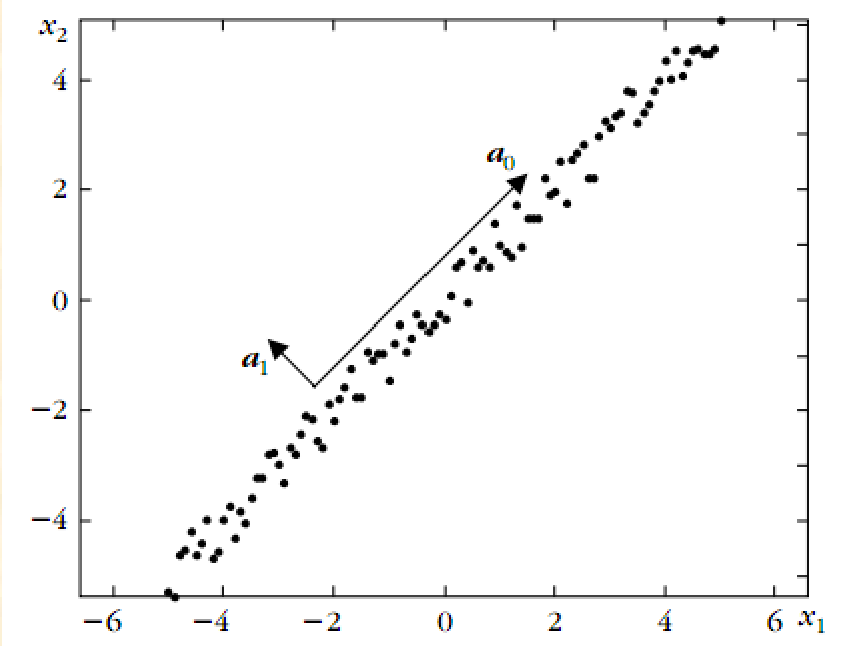
➢ Remark

An equivalent space results, if the eigenanalysis is performed on

$$XX^T$$

This is because XX^T and $X^T X$ have the same eigenvalues and related eigenvectors. This is usually referred as the Metric Multidimensional Scaling (MDS) method.

- PCA, being a linear projection method, is appropriate for dimensionality reduction if the data are spread around a hyperplane.



The eigenvalue – eigendecomposition of the correlation matrix reveals the dimensionality of the hyperplane, across which data are spread. In other words, dimensionality is a measure of the number of the underlying modes of variability.