Appearance based methods (classification)

As opposed to model-based methods, appearance based methods use no a priori knowledge on the data present in the image. Instead, they try through statistical analysis of the available dataset (either an image or image characteristics database) to extract the different variation modes of the database and provide a set of subclasses which represents them best.
Usually 2 phases will be involved:
A training phase and a classification phase

- Training:
  - Acquire initial set of hand sign images (training set).
  - Calculate the eigenhands from the training set, keeping only the M eigenvectors corresponding to the highest eigenvalues.
  - Calculate representation of each hand signs in hand space.

- Testing:
  - Project input image into hand space.
  - Find most likely candidate by distance computation

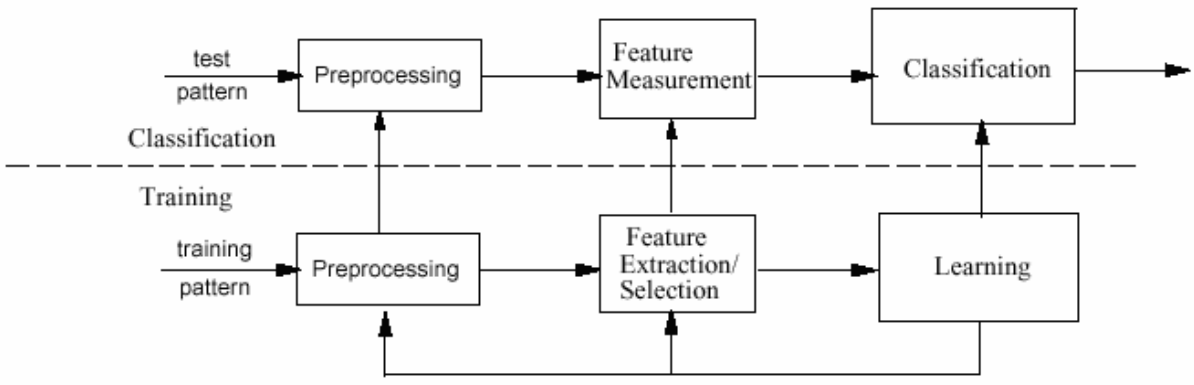


Figure 1: Model for statistical pattern recognition.

(From Jain et al 2000)

PCA

Principal components analysis is a method developed in statistics for regression, reduction of dimensionality of data and noise reduction. It has been introduced by:

- Pearson for biology (1901)
- Hotelling for psychometry (1933)
- Karhunen-Loeve (1947-1963) for infinite-dimensional and continuous cases in probability theory.

Reduction of dimensionality:
Select a minimum set of features such that the probability distribution of different classes given the values for those features is as close as possible to the original distribution given the values of all features.

Why?

- The curse of dimensionality (Bellman 1961)
- Measurement cost
- Classification accuracy
- Visualization in 2D or 3D of a dataset help to understand its underlying pattern structure
- pattern extraction (side effect)->reduction of number of attributes

Although a hand image defines a point in the high dimensional image space, different hand images share a number of similarities with each other.

- It seems plausible to find a lower dimensional subspace, which can represent them best.

Then?

- Project the face images into an appropriately chosen subspace and perform classification by similarity computation (distance).

Goal?

- Find the vectors that best account for the distribution of hand images within the entire image database space

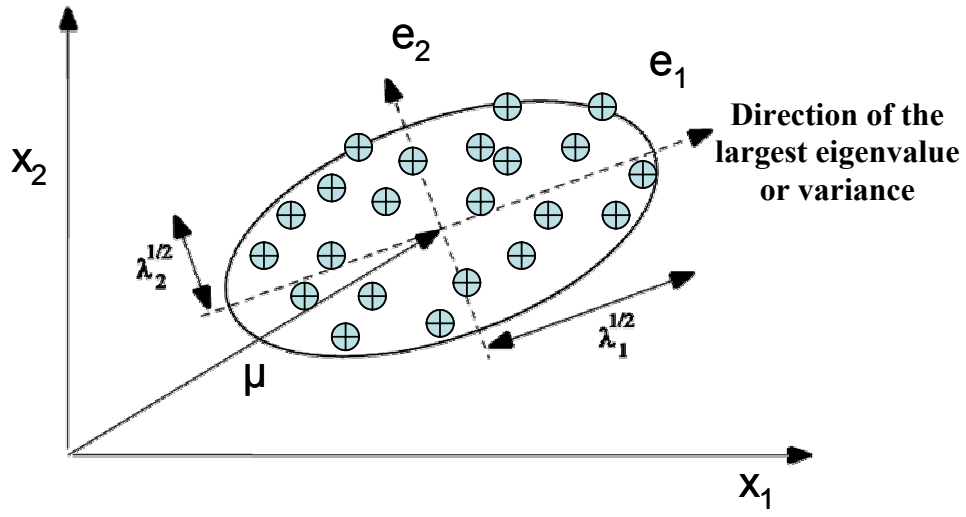
PCA (principle)

PCA removes correlations between variables or signals, while at the same time finding directions with maximal variance.

Let us suppose we have access to N samples of a vector x with n elements.
Elements of x can be measurements such pixel grey levels, or values (image features) of a signal at different time instants. These vectors will not be uniformly distributed in the n-dimensional space but more likely scattered.

Given *N* data vectors of dimension n from the dataset, you have to find *c* <= *N* orthogonal vectors that can be best used to represent data.

- Same thing as saying: you need to find an orthonormal basis which maximizes the variance of the projection of the dataset vectors along the new coordinate axis.
- The first axis corresponds to the maximal variance; the second axis corresponds to the maximal variance in the direction orthogonal to the first axis, etc...

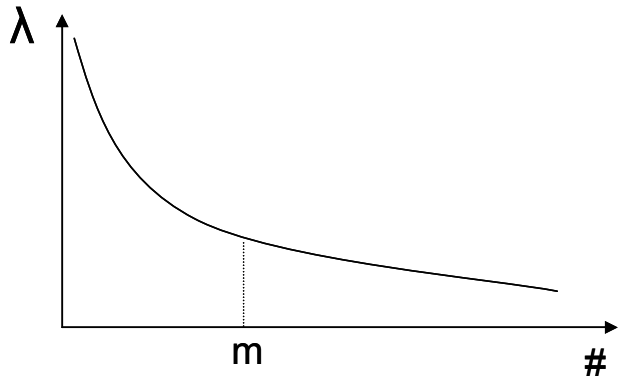


Principal axes of a density: e_1 and e_2. Eigenvalues λ_1 and λ_2 provide the variances in the Principal component directions.

- In the new basis (with ON vectors e_i) the dataset vectors (x) can be written as a linear combination:

$$x = \omega_1 e_1 + \omega_2 e_2 + \ldots + \omega_N e_N$$

- The smallest variances account for little data change in some directions. These directions can be dropped without significant changes.



- The original data set is reduced (projected) from one consisting of N data vectors onto c principal components (reduced dimensions)
- Each data vector is a linear combination of the *c* principal component vectors

$$x = \omega_1 e_1 + \omega_2 e_2 + \ldots + \omega_c e_c$$

- In the new coordinate system, axes are ordered according to variance values (i.e. size of the eigenvalue).

How to find the ON basis which maximizes the variance of the projection of the dataset vectors along the new coordinate axis?

- Compute the Covariance matrix of the centered dataset
- Find its eigenvalues and eigenvectors.
  - The unit eigenvectors of the covariance matrix are the principal directions of the data.
  - The projections of x to the principal directions are the principal components of x.
- Variance -> eigenvalue.
- Orthogonal axis ->direction of the eigenvectors
  - Take eigenvectors of unit norm for unicity of solution.

PCA (aka Karhunen-Loeve transform aka Hotelling)

We start with a database of N images $\mathbf{I_i}$ represented by their feature vectors x_i

$$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N\}$$

(whatever they might be) of dimension M by 1.
The feature vector might be a concatenation of the image pixels, curves describing the hand contours, moments, Fourier descriptors, Corners or other points of interests (Curvature related)

We need to define the average feature vector of the database and the corresponding centered feature values:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_N \qquad\qquad \mathbf{y}_i = \mathbf{x}_i - \overline{\mathbf{x}}$$

The covariance matrix **C** of the database is defined by**:**

$$C = \sum_{i=1}^{N}\mathbf{y_i}\mathbf{y_i}^T \qquad C = \sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})\ \ (\mathbf{x}_i - \overline{\mathbf{x}})^T = \sum_{i=1}^{N}\mathbf{y}_i\ \ \mathbf{y}_i^T$$

$\mathbf{e_i}$ and λ_i the eigenvectors and eigenvalues of the covariance matrix C verify:

$$C\,\mathbf{e}_i = \lambda_i\,\mathbf{e}_i$$

The eigenvectors of C are also called the **principal component** of C.
The matrix is symmetric, positive definite.

- its eigenvectors are orthonormal
- its eigenvalues are all non-negative

PCA applied directly to the covariance matrix of the images leads to extensive computation, as the dimension of the covariance matrix is equal to the square of the dimension of a feature vector. For a 256x256 image, a feature vector will be of dimension 65536x1. The covariance matrix will be of size 65536x65536 and you will have to search for 65536 eigenvectors each of size 65536x1.

Turk and Pentland introduced a method to obtain the eigenvalues and eigenvectors of C by studying (finding the eigenvalues and eigenvectors) a smaller matrix.
Instead of working directly on C, the eigenvectors and eigenvalues of an NxN matrix (N size of your database) can be used to deduce the eigenvectors and eigenvalues of

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i\mathbf{y}_i^T\mathbf{e} = \lambda\,\mathbf{e}$$

$$\Rightarrow \quad \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_i^T\mathbf{e})\mathbf{y}_i = \lambda\,\mathbf{e}$$

$$\Rightarrow \quad \frac{1}{N}\sum_{i=1}^{N}a_i\mathbf{y}_i = \lambda\,\mathbf{e}$$

C. Let's write the conditions leading to **e** being an eigenvector of **C** with **λ** as associated eigenvalues:
as:
Therefore, if **e** is an eigenvector of **C**, **e** can be expressed as a linear combination of the vectors **{y_i}**

$$\mathbf{e} = \sum_{i=1}^{N} b_i \mathbf{y}_i$$

with **b_i** defined as:

$$b_i = \frac{a_i}{\lambda N}$$

$$\mathbf{y}_i \mathbf{y}_i^T \mathbf{e} = (\mathbf{y}_i^T \mathbf{e}) \mathbf{y}_i$$

Furthermore:

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^T \mathbf{e} = \lambda \sum_{i=1}^{N} b_i \mathbf{y}_i$$

$$\Rightarrow \quad \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^T (\sum_{j=1}^{N} b_j \mathbf{y}_j) = \lambda \sum_{i=1}^{N} b_i \mathbf{y}_i$$

$$\Rightarrow \quad \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \sum_{j=1}^{N} b_j \, \mathbf{y}_i^T \mathbf{y}_j = \lambda \sum_{i=1}^{N} b_i \mathbf{y}_i$$

$$\Rightarrow \quad \frac{1}{N} \sum_{j=1}^{N} \mathbf{y}_i^T \mathbf{y}_j \, b_j = \lambda b_i$$

The vector **b (b_1, b_2,…,b_N)^T** is the eigenvector of the matrix with (i,j) elements
The associated eigenvalue is **λ**
Eigenvectors (**e**) of matrix **C** can be computed as linear combination of the previous matrix with eigenvectors **b**:

$$\mathbf{e} = \sum_{i=1}^{N} b_i \mathbf{y}_i$$
$$= \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{bmatrix} \mathbf{b}$$

Conclusion: we replace the task of finding eigenvectors of matrix M by M by finding the eigenvectors of matrix N by N.

$$\left\{ \frac{1}{N} \mathbf{y}_i^T \mathbf{y}_j \right\}$$

Summary

1. Take the mean of the feature vectors (x_i) of your database:

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

2. Compute their centered difference

$$\mathbf{y}_i = \mathbf{x}_i - \bar{\mathbf{x}}$$

3. Form the matrix M

$$M = \begin{bmatrix} \mathbf{y}_1^T\mathbf{y}_1 & \mathbf{y}_1^T\mathbf{y}_2 & \cdots & \mathbf{y}_1^T\mathbf{y}_N \\ \mathbf{y}_2^T\mathbf{y}_1 & \mathbf{y}_2^T\mathbf{y}_2 & \cdots & \mathbf{y}_2^T\mathbf{y}_N \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_N^T\mathbf{y}_1 & \mathbf{y}_N^T\mathbf{y}_2 & \cdots & \mathbf{y}_N^T\mathbf{y}_N \end{bmatrix}$$

4. Find the eigenvectors and eigenvalues of M

$$\mathbf{M}\,\mathbf{b} = \alpha\ \mathbf{b}$$

5. Compute eigenvalues λ and eigenvectors e_i of the covariance matrix $\mathbf{C}$ of your database

$$\mathbf{e}_i = \begin{bmatrix}\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3 \cdots \mathbf{y}_N\end{bmatrix}\mathbf{b}_i \qquad\qquad \lambda_i = \alpha_i$$

6. Find the most meaningful eigenvectors.
   - Keep the first *k* dimensions, which account for a given fraction of all the variance.
   - Usually above 70% is sufficient for rough description of the dataset modes.

$$R = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{N}\lambda_i}$$

7. 2 approaches
   - Find the most significant classes of your database via distance computation.
   - Express feature vectors of the database as linear combination of eigenvectors
   -

8. Classify the images not belonging to the database.

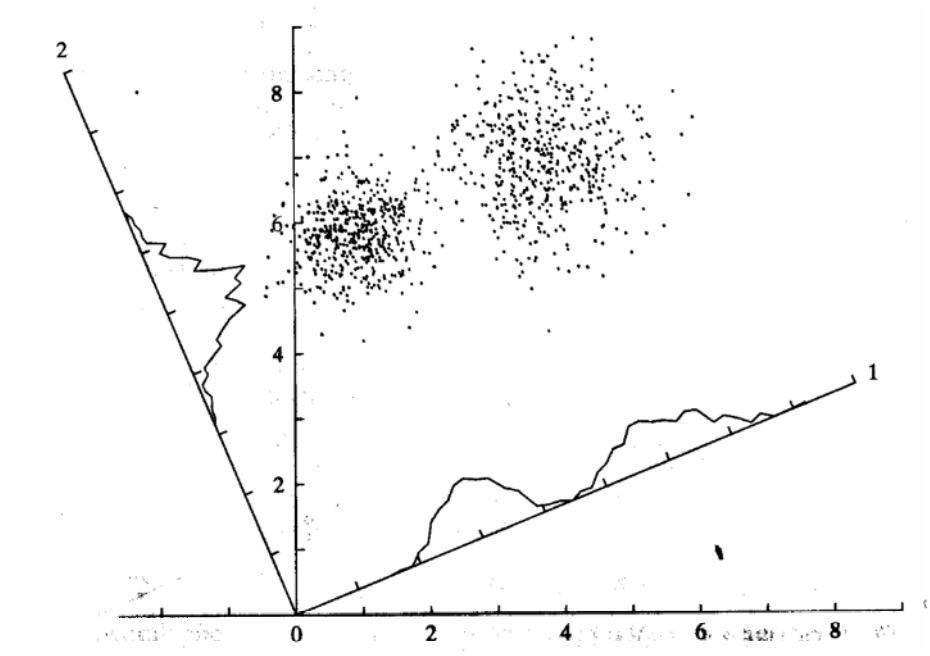
- The distance between the feature vectors of the image and the closest point of each class in the training set is determined. The image is classified into the class whom minimal distance sample point belongs.
- The distance between the feature vectors of the image and the representative feature vector of each class. The Image is similarly classified into the class of the representative image with minimal distance.

Conclusion

PCA is useful for:

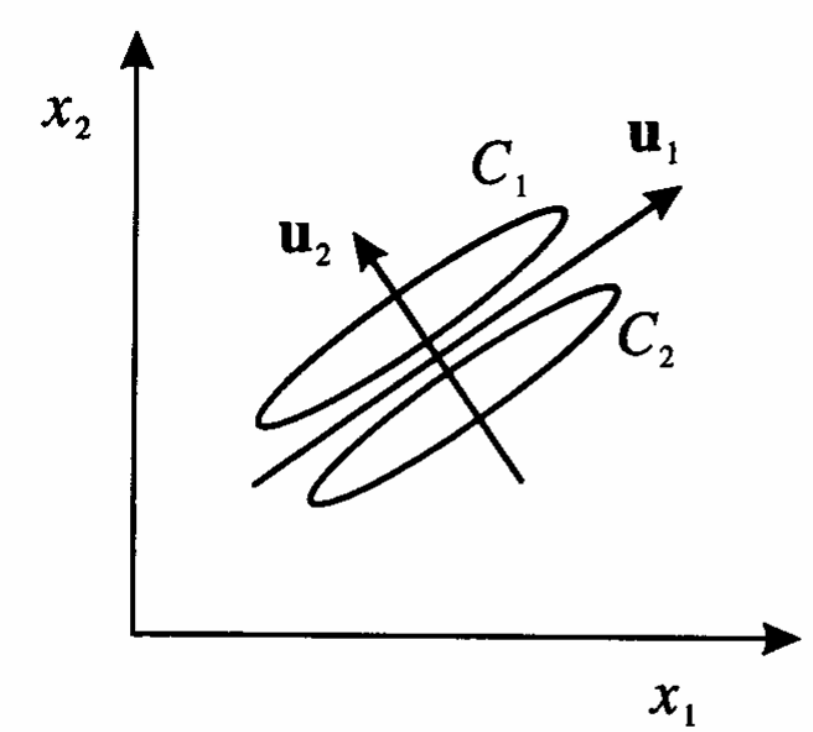
- Finding new, more informative, uncorrelated features
- Reducing dimensionality by rejecting features associated to low variances

PCA maximises while minimizing the sums of squares of distances from original points to projections.



Projection on the axis of the largest eigenvalue

Drawbacks:
The largest variance determines which components are used, but does not guarantee that this will provide an interesting viewpoint for clustering data.



First component will be chosen along the largest variance direction; both clusters will strongly overlap, no interesting structure will be visible.

- A Projection onto the orthogonal axis to the first PCA component has much more discriminating power.
- Find the basis which maximizes the distance of projected mean values while minimizing the within class scatter (variance without the normalization constant)

- **Bibliography**

A. K. Jain, R. P. W. Duin, and J. Mao, "*Statistical pattern recognition: A review*". IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 4--37, Jan. 2000.

Murakami, H. and Vijaya Kumar, B.V.K., "*Efficient Calculation of Primary Images from a Set of Images*". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-4, No. 5, Sep 1982

N. Har'El. "*Finding the largest eigenvalues of a real symmetric matrix, and corresponding eigenvectors*". Research Report, Department of Mathematics, Israel Institute of Technology, 1992.

M. Turk and A. Pentland, "*Eigenfaces for recognition*". Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, Winter 1991.

A.L.Yuille, P.W.Hallinan and D.S.Cohen, "*Feature extraction from faces using deformable templates*". International Journal of Computer Vision,vol.8,no.2,pp.99 –111,1992.

P.S.Penev and L.Sirovich. "The global dimensionality of face space". Proc. 4th International Conference on Automatic Face and Gesture Recognition, pp.264 –270, Grenoble, France, 2000.