Principal Component Analysis

Hernán Felipe García Arias PhD

Especialización en Analítica de Datos
Facultad de Ingeniería
Universidad de Antioquia

Febrero, 2025

Contenido

Feature Selection/Extraction I

- Solution to a number of problems in Pattern Recognition can be achieved by choosing a better feature space.
- **Curse of Dimensionality**: $\#$examples needed to train classifier function grows exponentially with $\#$dimensions.
- **What features best characterize class**: What words best characterize a document class.

Feature Selection/Extraction II

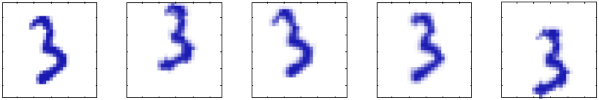


Figure: A synthetic data set obtained by taking one of the off-line digit images and creating multiple copies in each of which the digit has undergone a random displacement and rotation within some larger image field. The resulting images each have $100 \times 100 = 10000$ pixels.

The model I

- PCA is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction, and data visualization (Jolliffe, 2002). It is also known as the Karhunen-Lo‘eve transform.

- Consider a data set of observations $\{\mathbf{x}_n\}$ where $n = 1, \ldots, N$, and $\mathbf{x}_n$ is a Euclidean variable with dimensionality D.

- Our goal is to project the data onto a space having dimensionality $M < D$ while maximizing the variance of the projected data.

The model II

- The mean of the projected data is,

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

and the variance of the projected data is given by

$$\frac{1}{N}\sum_{n=1}^{N}\{\mathbf{u}_1^\top\mathbf{x}_n - \mathbf{u}_1^\top\bar{\mathbf{x}}\} = \mathbf{u}_1^\top\mathbf{S}\mathbf{u}_1, \tag{2}$$

where **S** is the data covariance matrix defined by

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \tag{3}$$

The model III

- We introduce a Lagrange multiplier that we shall denote by λ_1, and then make an unconstrained maximization of

$$\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^\top \mathbf{u}_1) \tag{4}$$

- By setting the derivative with respect to u1 equal to zero, we see that this quantity will have a stationary point when

$$\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \tag{5}$$

which says that $\mathbf{u}_1$ must be an eigenvector of $\mathbf{S}$.

The model IV

Tip

To summarize, principal component analysis involves evaluating the mean $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{S}$ of the data set and then finding the M eigenvectors of $\mathbf{S}$ corresponding to the M largest eigenvalues λ_i [1].

Applications I

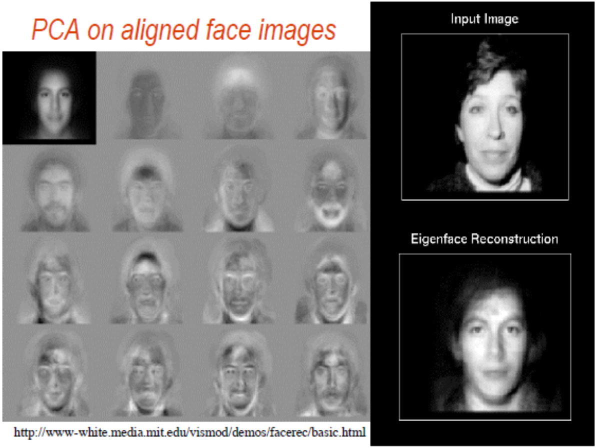
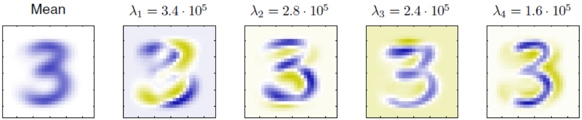


Figure: Face Recognition

Applications II



The mean vector $\overline{\mathbf{x}}$ along with the first four PCA eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_4$ for the off-line digits data set, together with the corresponding eigenvalues.

Figure: Handwrite Digit Recognition

Algorithm I

PCA algorithm $(\mathbf{X}, k)$: top k eigenvalues/eigenvectors
DATA: $\mathbf{X}$ $m \times N$ data matrix
Each point $\mathbf{x}_n$ is a column vector of $\mathbf{X}$

1. Substract the mean from each column of $\mathbf{X}$ (center the data)
2. Compute de covariance matrix of $\mathbf{X}$
3. Perform SVD decomposition of $\mathbf{S}$ so that $\{\lambda_i, \mathbf{u}_i\}$ are the eigenvalues and eigenvectors of $\mathbf{S}$
4. Return the k principle components so that $k <= m$ (retain a given percentage of the data variance)

Bibliography

Christopher M Bishop et al.
Pattern recognition and machine learning.
Springer New York, 2006.