FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

Lecture Pattern Analysis

Part 12: Principal Component Analysis

Christian Riess
IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
May 30, 2021

Introduction

- Principal Component Analysis (PCA), a.k.a. “Karhunen-Loeve Transform” or “KL-Transform” is a workhorse all across science and engineering
- PCA provides a more compact representation in a lower-dim. space
- Brief overview[1]:
  - PCA is a linear projection onto an orthogonal basis **U**, i.e.,

$$\mathbf{u}_i^T \cdot \mathbf{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

  - This basis are the eigenvectors of the (mean-free) data covariance
  - Thus, the calculation of PCA is essentially to normalize the data and to perform an eigenvalue decomposition
  - The magnitude of the eigenvalues indicates the contribution of a dimension to the covariance of the data

[1] The literature source for this lecture is Bishop Sec. 12.1.1

Objective Function and Normalization

- Core idea: find a linear mapping $\Phi : \mathbb{R}^d \to \mathbb{R}^{d'}$, $d' \ll d$, that maximizes the variance (spread) of the data along each dimension
- Objective function:

$$J = \sum_{i,j=1}^{N} (\Phi \mathbf{x}_i - \Phi \mathbf{x}_j)^{\mathsf{T}} (\Phi \mathbf{x}_i - \Phi \mathbf{x}_j) + \lambda(\Phi^{\mathsf{T}}\Phi - 1) \tag{2}$$

  where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d$ are the data points
- Assume zero-mean samples. Hence, in practice, subtract the mean of the samples to obtain

$$\sum_{i=1}^{N} \mathbf{x}_i = 0 \tag{3}$$

Derivation of the Principal Components

- We seek a projection $\mathbf{u}$ onto the 1-D subspace that maximizes the variance, and show that $\mathbf{u}$ is the largest eigenvector of the covariance matrix
- $\mathbf{u}$ is the first column of Φ, further vectors are obtained by induction:
  - Project the data on the $d-1$-dim. subspace orthogonal to $\mathbf{u}$
  - Repeat the reasoning $d'-1$ times
- To begin, let $\mathbf{u} \in \mathbb{R}^d$ be an arbitrary direction of unit length, i.e., $\mathbf{u}^\mathsf{T}\mathbf{u} = 1$
- The inner product $\mathbf{u}^\mathsf{T}\mathbf{x}$ projects $\mathbf{x}$ onto a 1-D space
- The variance of the projected data is

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{u}^\mathsf{T}\mathbf{x}_i - \mathbf{u}\bar{\mathbf{x}})^2 = \mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u} \tag{4}$$

where $\bar{\mathbf{x}}$ is the (component-wise) mean of all $\mathbf{x}_i$ and $\mathbf{S}$ is the covariance matrix

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\mathsf{T} \tag{5}$$

Maximizing the Variance

- We seek a unit-length direction $\mathbf{u}$ that maximizes the variance:

$$\mathbf{u}^{\mathsf{T}}\mathbf{S}\mathbf{u} + \lambda(1 - \mathbf{u}^{\mathsf{T}}\mathbf{u}) \rightarrow \max \tag{6}$$

  where λ is a Lagrange multiplier to include the constraint $\mathbf{u}^{\mathsf{T}}\mathbf{u} = 1$
- The maximum is found by calculating the derivative w.r.t. $\mathbf{u}$, and to set the equation equal to 0:

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{u}^{\mathsf{T}}\mathbf{S}\mathbf{u} + \lambda(1 - \mathbf{u}^{\mathsf{T}}\mathbf{u}) \stackrel{!}{=} 0 \tag{7}$$

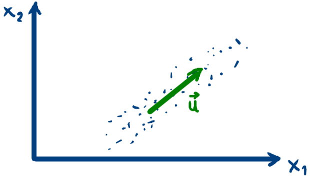
$$\Leftrightarrow \quad 2\mathbf{S}\mathbf{u} = 2\lambda\mathbf{u} \tag{8}$$

$$\Leftrightarrow \quad \mathbf{S}\mathbf{u} = \lambda\mathbf{u} \tag{9}$$

- This is just the eigenvector decomposition of $\mathbf{S}$. Hence, the eigenvector associated with the largest eigenvalue provides maximum covariance
- This vector is called a “principal component”

Remarks

- You probably know sketches of the direction of maximum covariance:



- The relative magnitude of an eigenvalue indicates the percentage of variance that is represented. For example, if

$$\left(\sum_{i=1}^{d'} \lambda_i\right) \Big/ \left(\sum_{i=1}^{d} \lambda_i\right) = 0.98 \ , \tag{10}$$

then a d'-dim. subspace preserves 98% of the variance of the data

- This argument is used, e.g., in remote sensing to compress 100s of (correlated) color bands to less than 10 dimensions