APPENDIX C

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) maps the input variables — in our case usually a vector of features $\boldsymbol{v}$ — to a new coordinate system by a linear combination of the individual features:

$$\boldsymbol{u}(n) = \boldsymbol{T}^{\mathrm{T}} \cdot \boldsymbol{v}(n). \tag{C.1}$$

The resulting vector $\boldsymbol{u}(n)$ is the data in the new coordinate system for observation n and transformation matrix $\boldsymbol{T}^{\mathrm{T}}$ contains different linear combinations for the input feature vector $\boldsymbol{v}(n)$. The number of features in the vector will be referred to as $\mathcal{F}$. Formulating Eq. (C.1) not only for one observation but for a series of feature vectors $\boldsymbol{V}$ leads to

$$\boldsymbol{U} = \boldsymbol{T}^{\mathrm{T}} \cdot \boldsymbol{V}. \tag{C.2}$$

The transformation matrix is a square matrix with the dimensions $\mathcal{F} \times \mathcal{F}$. It is composed of vectors defining the linear combinations of the input features:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{c}_0 & \boldsymbol{c}_1 & \dots & \boldsymbol{c}_{\mathcal{F}-1} \end{bmatrix}. \tag{C.3}$$

The transformation matrix has the following main properties:

- the vectors $\boldsymbol{c}_i$ are in the direction of the highest variance in the data and the variance is concentrated in as few output components as possible,
- the vectors $\boldsymbol{c}_i$ are orthogonal to each other

$$\boldsymbol{c}_i^{\mathrm{T}} \cdot \boldsymbol{c}_j = 0 \quad \forall\ i \neq j \tag{C.4}$$

An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics, First Edition. Alexander Lerch.

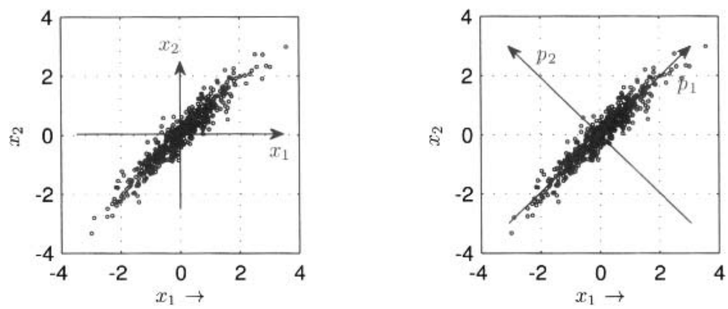

Figure C.1 Scatter plot of a two-dimensional data set with variables x_1, x_2, and the rotated coordinate system after PCA with the component axes p_1, p_2

- and the transformation is invertible:

$$v(n) = T \cdot u(n). \tag{C.5}$$

Figure C.1 shows the scatter plots of two example variables x_1 and x_2 with the original axes on the left and rotated axes on the right.

C.1 Computation of the Transformation Matrix

The first step in computing the matrix $\boldsymbol{T}$ is the calculation of the feature covariance matrix:

$$\boldsymbol{R} = \frac{1}{\mathcal{F} - 1} \cdot (\boldsymbol{v} - \boldsymbol{\mu}_v)\left(\boldsymbol{v}^{\mathrm{T}} - \boldsymbol{\mu}_v^{\mathrm{T}}\right) \tag{C.6}$$

with the vector $\boldsymbol{\mu}_v$ containing the arithmetic mean of each feature. The covariance matrix is square, symmetric, and has only positive entries.

The eigenvectors of the covariance matrix represent the axes of the new coordinate system and thus comprise the transformation matrix. Usually, they are ordered with decreasing eigenvalues; the vector in the first column c_0 is the vector with the highest eigenvalue and the vector in the last column has the lowest eigenvalue.

C.2 Interpretation of the Transformation Matrix

The transformation matrix $\boldsymbol{T}$ contains useful information on the input data set. Each column is a different linear combination of the input features, and they are sorted according to their eigenvalues or, in other words, according to the variance this component contributes to the overall variance. Features with high influence on the first components can be assumed to be of higher importance than features with high influence on the last components. Thus, PCA can be useful for both feature subset selection and feature space transformation. As the last components have only limited impact on the result, they might be discarded. The matrix $\boldsymbol{T}$ can then be truncated from the dimensions $\mathcal{F} \times \mathcal{F}$ to $\mathcal{F} \times \mathcal{L}$ with $\mathcal{L}$ being the required number of components in the space transformation process.