
Appendix II: Derivation of Principal Component Analysis

2.1 Eigenvectors of the Covariance Matrix Derivation

2.1.1 Principal Component Feature Extraction

Consider a random vector $\mathbf{x} \in \mathfrak{R}^n$ for which a basis \mathbf{u} is to be determined such that an approximation of \mathbf{x} can be obtained by a linear combination of m orthogonal basis vectors:

$$\hat{\mathbf{x}} = y_1 \mathbf{u}_1 + y_2 \mathbf{u}_2 + \dots + y_m \mathbf{u}_m \quad [154]$$

where $m < n$ and y_j is the weighting coefficient of basis vector \mathbf{u}_j formed by taking the inner product of \mathbf{x} with \mathbf{u}_j :

$$y_j = \mathbf{x}^T \mathbf{u}_j . \quad [155]$$

By forming a $k \times n$ observation matrix \mathbf{X} , the rows of which are the observations \mathbf{x}_k^T , we can express Equation 154 and Equation 155 respectively in the following form:

$$\hat{\mathbf{X}} = \mathbf{Y} \mathbf{U}_m^T \quad [156]$$

$$\mathbf{Y}_m = \mathbf{X} \mathbf{U}_m \quad [157]$$

where \mathbf{U}_m is an $n \times m$ matrix whose columns are an uncorrelated basis for \mathbf{X} , and \mathbf{Y}_m is a $k \times m$ matrix whose columns are the coefficients for each column vector in \mathbf{U}_m . By the orthog-

onality of \mathbf{U}_m and by the imposition of an additional constraint that the columns of \mathbf{U}_m have unit norm, i.e. $\|\mathbf{U}_j\| = 1$, it follows that:

$$\mathbf{U}_m^T \mathbf{U}_m = \mathbf{I}_m \quad [158]$$

where \mathbf{I}_m is a $m \times m$ diagonal matrix with unit entries. We refer to Equation 156 as the PCA-feature re-synthesis equation and Equation 157 as the PCA-feature projection equation where the columns of \mathbf{Y}_m correspond to the estimated features and the columns of \mathbf{U}_m are the projection vectors which perform the linear transformation of the input to the new uncorrelated basis.

The problem, then, for deriving a PCA is to obtain the matrix \mathbf{U}_m such that the residual error in approximating \mathbf{x} with $\hat{\mathbf{x}}$ is minimized:

$$\boldsymbol{\varepsilon} = \mathbf{x} - \hat{\mathbf{x}} = \sum_{j=m+1}^n y_j \mathbf{u}_j \quad [159]$$

that is, the expansion of all the unused features results in a minimal signal which is the residual error $\boldsymbol{\varepsilon}$. A suitable criteria is the minimization of the expectation of the mean-square residual error:

$$\xi = E[|\boldsymbol{\varepsilon}|^2] = E[|\mathbf{x} - \hat{\mathbf{x}}|^2] \quad [160]$$

where the expectation of an arbitrary function of \mathbf{x} , say $\psi(\mathbf{x})$, is defined as the element-wise operation:

$$E[\psi(\mathbf{x})] = \int_{-\infty}^{\infty} \psi(\mathbf{x}) p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \quad [161]$$

where $p_{\mathbf{x}}(\mathbf{x})$ is the probability density function of the random variable \mathbf{x} . Since the expectation operator is linear and due to the condition that the columns of \mathbf{U}_m are orthonormal it follows that:

$$\zeta = E[\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}] = E\left[\left(\sum_{i=m+1}^n y_i \mathbf{u}_i^T\right)\left(\sum_{j=m+1}^n y_j \mathbf{u}_j\right)\right] = \sum_{j=m+1}^n E[y_j^2] \quad [162]$$

from which it follows that:

$$E[y_j^2] = E[(\mathbf{u}_j^T \mathbf{x})(\mathbf{x}^T \mathbf{u}_j)] = \mathbf{u}_j^T E[\mathbf{x}\mathbf{x}^T] \mathbf{u}_j = \mathbf{u}_j^T \mathbf{R} \mathbf{u}_j \quad [163]$$

where \mathbf{R} is the correlation matrix for \mathbf{x} . Now by substitution of Equation 163 into Equation 162 we arrive at the quantity to be minimized:

$$\zeta = \sum_{j=m+1}^n \mathbf{u}_j^T \mathbf{R} \mathbf{u}_j \quad [164]$$

We can express the minimization as a multi-variate differential equation by using a set of Lagrangian multipliers λ_j and setting the derivative of the expectation of the mean-square error with respect to the basis components \mathbf{u}_j to zero. This is a necessary condition for the minimum and gives the characteristic equation:

$$\frac{\partial}{\partial \mathbf{u}_j} \zeta = 2(\mathbf{R} \mathbf{u}_j - \lambda_j \mathbf{u}_j) = 0, \quad j = m+1, \dots, n \quad [165]$$

It is well known that the solutions to this equation constitute the eigenvectors of the correlation matrix \mathbf{R} . It is also worth noting that the correlation matrix is related to the covariance matrix by the following expression:

$$\mathbf{Q}_x = \mathbf{R}_x - \mathbf{m}\mathbf{m}^T \quad [166]$$

Thus, for zero-mean or *centered* data, the problem is equivalent to finding the eigenvectors of the covariance matrix \mathbf{Q}_x . Since the columns of \mathbf{U} are now determined to be the eigenvectors, we can re-express the residual error as the sum of the eigenvalues of the unused portion of the basis:

$$\zeta = \sum_{j=m+1}^n \lambda_j \quad [167]$$

Eigenvectors of the Covariance Matrix Derivation

and the solution to the minimization reduces to ordering the basis vectors \mathbf{u}_j such that the columns with the smallest eigenvalues occur in the unused portion of the basis which also implies that the m columns of \mathbf{U}_m which *are* used for reconstruction should comprise the eigenvectors with the m largest eigenvalues.

Now that we have arrived at the form of the solution for optimal orthonormal basis reconstruction (in the square-error sense) we must find a general form for representing the solution. Since the eigenvalues form the diagonal elements of the covariance of the transformed data with all other elements equal to zero we can express the solution to the eigenvalue decomposition as a diagonalization of the input covariance matrix \mathbf{Q}_x .