7.5 Applications to Image Processing and Statistics

A technique called *principal component analysis* is used to analyze multivariate data, using orthogonal diagonalization and singular value decomposition.

Example 1. *Suppose we measure the heights and the weights of students. We put these observations in a matrix called the* matrix of observations*. Each column of this matrix is an* observation vector *in* $\mathbb{R}^2$*. The observations matrix would look like the following:*

$$\begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix}.$$

We may plot these observations on a two-dimensional scatterplot.

Suppose $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is a matrix of observations. Then the *sample mean* of the observations is

$$\mathbf{M} = \frac{1}{n}(\mathbf{x}_1 + \cdots + \mathbf{x}_n).$$

For $k = 1, \ldots, n$, let $\hat{\mathbf{x}}_k = \mathbf{x}_k - \mathbf{M}$. Let B be the $p \times n$ matrix

$$B = [\hat{\mathbf{x}}_1 \quad \cdots \quad \hat{\mathbf{x}}_n].$$

The *sample covariance matrix* is the $p \times p$ matrix S defined by

$$S = \frac{1}{n-1} BB^T.$$

Suppose $\mathbf{x} = (x_1, \ldots, x_p)$ be a vector that varies over the set of observation vectors. For $j = 1, \ldots, p$, the diagonal entry s_{jj} in S is called the *variance* of x_j. The *total variance* of the data is the sum of the variances on the diagonal of S. The sum of the diagonal entries of a square matrix S is called the *trace* of S, denoted by $\mathrm{tr}(S)$. Thus

$$\text{total variance} = \mathrm{tr}(S).$$

The entry s_{ij} for $i \neq j$ is called the *covariance* of x_i and x_j. When the covariance between x_i and x_j is zero, statisticians say that x_i and x_j are *uncorrelated*. As usual, the more zeros a matrix has, the easier it is to work with it. Thus analysis of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is greatly simplified when most of the variables $x_1, \ldots, x_n$ are uncorrelated.

The goal of principal component analysis is to find an orthogonal $p \times p$ matrix $P = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p]$ that determines a change of variable, $\mathbf{x} = P\mathbf{y}$, or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

with the property that the new variables $y_1, \ldots, y_p$ are uncorrelated and are arranged in order of decreasing variance. Let $c_1, \ldots, c_p$ be the entries in $\mathbf{u}_1$. Since $\mathbf{u}_1^T$ is the first row of P^T, the equation $\mathbf{y} = P^T\mathbf{x}$ shows that

$$y_1 = \mathbf{u}_1^T\mathbf{x} = c_1x_1 + \cdots + c_px_p.$$

In a similar fashion, $\mathbf{u}_2$ determines the variable y_2, and so on.

It can be shown that an orthogonal change of variables, $\mathbf{x} = P\mathbf{y}$, does not change the total variance of the data, because left-multiplication by P does not change the lengths of vectors or the angles between them. Thus, if $S = PDP^T$, then

$$\text{total variance of } x_1, \ldots, x_p = \text{total variance of } y_1, \ldots, y_p = \mathrm{tr}(D) = \lambda_1 + \cdots + \lambda_p.$$