and

$$VD = \begin{pmatrix} 51.4683 & 3.3013 & -3.8569 \\ -9.9623 & -6.6467 & -2.7082 \\ 76.6327 & 3.1845 & 0.2348 \\ 2.2393 & -8.6943 & 5.2872 \\ -33.6038 & 4.1334 & -3.6415 \\ -25.5941 & 1.3833 & -0.4350 \\ -53.4333 & 7.2258 & -1.3547 \\ -13.0100 & 6.8594 & 4.2010 \\ -6.2843 & 4.6254 & 4.3212 \\ -15.2173 & -14.3266 & -1.1581 \end{pmatrix}, \quad X - \mu = \begin{pmatrix} -1.2000 & -51.4000 & -5.6000 \\ -4.2000 & 9.6000 & 6.4000 \\ 3.8000 & -76.4000 & -5.6000 \\ 3.8000 & -2.4000 & 9.4000 \\ -4.2000 & 33.6000 & -3.6000 \\ -1.2000 & 25.6000 & -0.6000 \\ -2.2000 & 53.6000 & -5.6000 \\ 4.8000 & 13.4000 & -5.6000 \\ 4.8000 & 6.6000 & -3.6000 \\ -4.2000 & 14.6000 & 14.4000 \end{pmatrix}.$$

The first principal direction $u_1 = (0.0394, -0.9987, -0.0327)$ is basically the opposite of the y-axis, and the most significant feature is the year of birth. The second principal direction $u_2 = (0.1717, 0.0390, -0.9844)$ is close to the opposite of the z-axis, and the second most significant feature is the lenght of beards. A best affine plane is spanned by the vectors u_1 and u_2.

There are many applications of PCA to data compression, dimension reduction, and pattern analysis. The basic idea is that in many cases, given a data set $X_1, \ldots, X_n$, with $X_i \in \mathbb{R}^d$, only a "small" subset of $m < d$ of the features is needed to describe the data set accurately.

If $u_1, \ldots, u_d$ are the principal directions of $X - \mu$, then the first m projections of the data (the first m principal components, i.e., the first m columns of VD) onto the first m principal directions represent the data without much loss of information. Thus, instead of using the original data points $X_1, \ldots, X_n$, with $X_i \in \mathbb{R}^d$, we can use their projections onto the first m principal directions $Y_1, \ldots, Y_m$, where $Y_i \in \mathbb{R}^m$ and $m < d$, obtaining a compressed version of the original data set.

For example, PCA is used in computer vision for *face recognition*. Sirovitch and Kirby (1987) seem to be the first to have had the idea of using PCA to compress facial images. They introduced the term *eigenpicture* to refer to the principal directions, u_i. However, an explicit face recognition algorithm was given only later by Turk and Pentland (1991). They renamed eigenpictures as *eigenfaces*.

For details on the topic of eigenfaces, see Forsyth and Ponce [64] (Chapter 22, Section 22.3.2), where you will also find exact references to Turk and Pentland's papers.

Another interesting application of PCA is to the *recognition of handwritten digits*. Such an application is described in Hastie, Tibshirani, and Friedman, [88] (Chapter 14, Section 14.5.1).

23.6 Summary

The main concepts and results of this chapter are listed below: