In general, since an SVD of X is not unique, *the principal directions* $u_1, \ldots, u_d$ *are not unique*. This can happen when a data set has some *rotational symmetries*, and in such a case, PCA is not a very good method for analyzing the data set.

23.5 Best Affine Approximation

A problem very close to PCA (and based on least squares) is to *best approximate a data set of* n *points* $X_1, \ldots, X_n$*, with* $X_i \in \mathbb{R}^d$*, by a* p*-dimensional affine subspace* A *of* $\mathbb{R}^d$*, with* $1 \le p \le d-1$ (the terminology rank $d-p$ is also used).

First consider $p = d-1$. Then $A = A_1$ is an affine hyperplane (in $\mathbb{R}^d$), and it is given by an equation of the form

$$a_1x_1 + \cdots + a_dx_d + c = 0.$$

By *best approximation*, we mean that $(a_1, \ldots, a_d, c)$ solves the homogeneous linear system

$$\begin{pmatrix} x_{1\,1} & \cdots & x_{1\,d} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n\,1} & \cdots & x_{n\,d} & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_d \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

in the *least squares sense, subject to the condition that* $a = (a_1, \ldots, a_d)$ *is a unit vector*, that is, $a^\top a = 1$, where $X_i = (x_{i\,1}, \cdots, x_{i\,d})$.

If we form the symmetric matrix

$$\begin{pmatrix} x_{1\,1} & \cdots & x_{1\,d} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n\,1} & \cdots & x_{n\,d} & 1 \end{pmatrix}^\top \begin{pmatrix} x_{1\,1} & \cdots & x_{1\,d} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n\,1} & \cdots & x_{n\,d} & 1 \end{pmatrix}$$

involved in the normal equations, we see that the bottom row (and last column) of that matrix is

$$n\mu_1 \quad \cdots \quad n\mu_d \quad n,$$

where $n\mu_j = \sum_{i=1}^n x_{i\,j}$ is n times the mean of the column C_j of X.

Therefore, if $(a_1, \ldots, a_d, c)$ is a least squares solution, that is, a solution of the normal equations, we must have

$$n\mu_1a_1 + \cdots + n\mu_da_d + nc = 0,$$

that is,

$$a_1\mu_1 + \cdots + a_d\mu_d + c = 0,$$

which means that the *hyperplane* A_1 *must pass through the centroid* μ *of the data points* $X_1, \ldots, X_n$. Then we can rewrite the original system with respect to the centered data $X_i - \mu$, find that the variable c drops out, get the system

$$(X - \mu)a = 0,$$