Principal Component Analysis

Wenjing Liao

School of Mathematics
Georgia Institute of Technology

Math 4803
Fall 2019

Approximate training data by a hyperplane

Training data: $x_1, x_2, \ldots, x_N \in \mathbb{R}^p$
Hyperplane:

$$f(\lambda) = \mu + \mathbf{V}_q \lambda,$$

where $\nu \in \mathbb{R}^p$, and $V_q \in \mathbb{R}^{p \times q}$ with orthonormal columns.

Least squares:

$$\min_{\mu, \{\lambda_i\}, \mathbf{V}_q} \sum_{i=1}^{N} \|x_i - \mu - \mathbf{V}_q \lambda_i\|^2.$$

How to solve the least squares?

Step 1: Find μ

$$\begin{aligned} \hat{\mu} &= \bar{x}, \\ \hat{\lambda}_i &= \mathbf{V}_q^T(x_i - \bar{x}). \end{aligned}$$

Step 1: Find the subspace V_q

$$\min_{\mathbf{V}_q} \sum_{i=1}^{N} ||(x_i - \bar{x}) - \mathbf{V}_q \mathbf{V}_q^T (x_i - \bar{x})||^2.$$

Use SVD

For convenience we assume that $\bar{x} = 0$ (otherwise we simply replace the observations by their centered versions $\tilde{x}_i = x_i - \bar{x}$). The $p \times p$ matrix $\mathbf{H}_q = \mathbf{V}_q\mathbf{V}_q^T$ is a *projection matrix*, and maps each point x_i onto its rank-q reconstruction $\mathbf{H}_q x_i$, the orthogonal projection of x_i onto the subspace spanned by the columns of $\mathbf{V}_q$. The solution can be expressed as follows. Stack the (centered) observations into the rows of an $N \times p$ matrix $\mathbf{X}$. We construct the *singular value decomposition* of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{14.54}$$

This is a standard decomposition in numerical analysis, and many algorithms exist for its computation (Golub and Van Loan, 1983, for example). Here $\mathbf{U}$ is an $N \times p$ orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$) whose columns $\mathbf{u}_j$ are called the *left singular vectors*; $\mathbf{V}$ is a $p \times p$ orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}_p$) with columns v_j called the *right singular vectors*, and $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ known as the *singular values*. For each rank q, the solution $\mathbf{V}_q$ to (14.53) consists of the first q columns of $\mathbf{V}$. The columns of $\mathbf{UD}$ are called the principal components of $\mathbf{X}$ (see Section 3.5.1). The N optimal $\hat{\lambda}_i$ in (14.52) are given by the first q principal components (the N rows of the $N \times q$ matrix $\mathbf{U}_q\mathbf{D}_q$).

Example

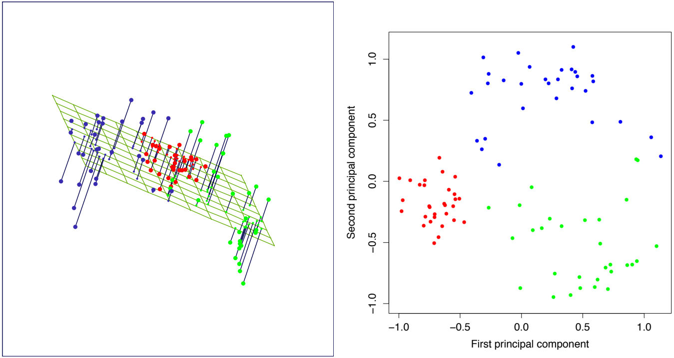


FIGURE 14.21. *The best rank-two linear approximation to the half-sphere data. The right panel shows the projected points with coordinates given by* $\mathbf{U}_2\mathbf{D}_2$*, the first two principal components of the data.*

Handwritten Digits

Data: Grayscale 16×16, 658 of 3's

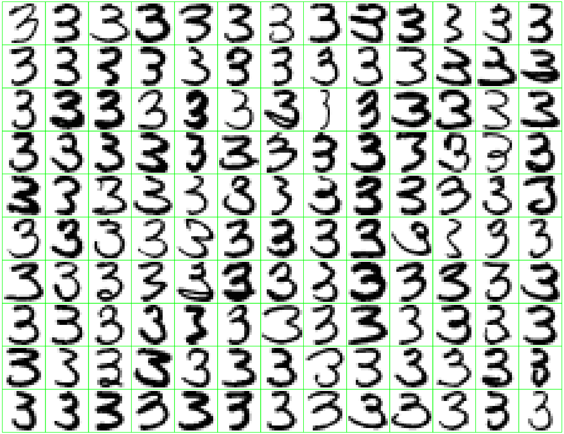


FIGURE 14.22. *A sample of 130 handwritten 3's shows a variety of writing styles.*

Singular values of X

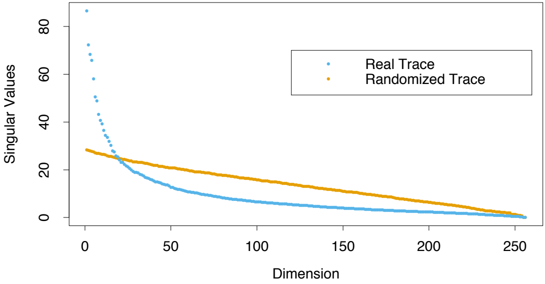
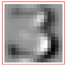
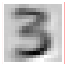


FIGURE 14.24. *The* 256 *singular values for the digitized threes, compared to those for a randomized version of the data (each column of* **X** *was scrambled).*

Representation using two principal components

$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$ .

Reference

Section 14.5.1: Trevor Hastie, Robert Tibshirani, The Elements of Statistical Learning, Second Edition.