Linear Dimension Reduction

Machine learning 2021

Mansoor Rezghi

Ref

Zaki: chap 7

Bishop 12.1(PCA)

4.1.4(LDA)

Morphy 12.2.1-12.2.3

Introduction:

SVD: Singular Value Decomposition.

$\forall A \in \mathbb{R}^{m \times n}$ $A = U \Sigma V^T$ $U \in \mathbb{R}^{m \times m}$

$V \in \mathbb{R}^{n \times n}$

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$\Sigma_r = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$

Complete.

$A = U_r \Sigma_r V_r^T$ $\longrightarrow$ Reduced Form.

$U = [U_r, U_{m-r}]$ $V = [V_r, V_{n-r}]$

$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$$= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

$$A_k = \underset{\text{Rank}(B) \leq k}{\arg\min} \| A - B \|_F^2$$

$$A_k = U_k \Sigma_k V_k^T$$

$$U_k = [u_1, \dots, u_k]$$

$$\Sigma_k = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}$$

$$V_k = [v_1, \dots, v_k]$$

$\sigma_k \quad \sigma_1 > \sigma_2 > \sigma_3$ ——

$U_i \longrightarrow$

با افزایش اندیس i، U_i ها نوسانی تر خواهند شد.

$U = [U_1, \dots, U_n]$ Base R^n

if $x \in R^n \quad \exists\, y$ s.t $\underline{x} = Uy = \sum y_i U_i = \underline{\sum_{i=1}^{k} y_i U_i} + \sum_{i=k+1}^{n} \cancel{y_i U_i}$

ضریب U_i در ساختن بردار x است.

$U = [\underbrace{U_1, \dots, U_k}, U_{k+1}, \dots, U_n]$

project span$\{U_1, \dots, U_k\}$

$x \simeq \sum_{i=1}^{k} y_i U_i = [U_1, \dots, U_k] \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$

project $x \longrightarrow$

$\underline{x \simeq U_k y^k}$

$X \longrightarrow$ project $[U_1, \dots, U_k]$

$\min \| x - \overset{y}{\overline{U_k \alpha}} \|_F^2 \rightarrow \boxed{\alpha = U^T x}$

$y = \underline{U}^T x$

eigenvalue.

$A \in \mathbb{R}^{n \times n} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$\forall \lambda$ if $\exists\, x \neq 0$ s.t. $\underbrace{Ax}_{\text{corresponding eigenvalue } \lambda} = \underbrace{\lambda}_{\text{eigenvalue}} x$

$$Ax = \lambda x \Rightarrow \underbrace{(A - \lambda I)}_{\text{singular}} x = 0 \qquad \det(A - \lambda I) = 0$$

A SPD semi positiv definit $\quad \forall x \neq 0 \;\; x^T A x \geq 0$ ($>$ $\rightarrow$ PD)

$$A \text{ SPD} \Rightarrow Ax = \lambda x \rightarrow x^T A x = \lambda \|x\|_2^2$$

$$A \text{ SPD} \Leftrightarrow \lambda_i \geq 0$$

$\{x_1 \dots x_N\}$ $\quad X = [x_1, \dots, x_N] \in \mathbb{R}^{D\times N}$

$\mathbb{R}^D \ni x_i \to U y_i$, $\quad y_i \in \mathbb{R}^d$

$U \in \mathbb{R}^{D\times d}$

$$\min \sum_{i=1}^{N} \| x_i - U_d y_i \| = \| X - U_d Y_d \|_F^2$$

$Y_d = [y_1, \dots, y_N] \in \mathbb{R}^{d\times N}$

$$\min_{U, Y} \| X - \bar{U}_d Y_d \|_F^2$$

$\mathbb{R}^{D\times d}$ $\quad \mathbb{R}^{d\times N}$

$\mathrm{Rak}(U) = d \leq d$

$\mathrm{Rak}(Y_d) \leq d$ $\implies$ if $B = U_d Y_d \implies \mathrm{Rak}(B) \leq d$

solution $\quad \bar{U}_d Y_d = U_d \Sigma_d V_d^T \implies \bar{U}_d = U_d$

$Y_d = \Sigma_d V_d^T$

$$Y_d = \Sigma_d V_d^T = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_d^T \end{pmatrix} =$$

~~for~~ $x \sim \bar{U} y$

By SVD $\longrightarrow$ PR $\quad X = U \Sigma V^T \quad$ if $\bar{U} = [u_1 .., u_d]$

Left $\overline{\text{singh vect}}$

$$\min \| x_i - U_d y \| \Rightarrow y = U_d^T x \qquad y = \bar{U}^T x$$

$$Y_d = U_d^T X$$

$$\| X - U_d Y_d \|_F^2 = \sum_{i=d+1}^{r} \sigma_i^2$$

$$X = [U_d \;\; U_{D-d}] \begin{bmatrix} \Sigma_d & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} V_d^T \\ V_{r-d}^T \end{bmatrix}$$

$$U_d^T X = [I \;\; 0] \qquad = \Sigma_d V_d^T$$

$X = [x_1, \dots, x_N] \in \mathbb{R}^{D \times N}$

if $\underbrace{\mathrm{Rank}(X)}_{r} < D$

$\mathrm{Rank}(X) \leq \{D, N\}$

$$X = [u_1 \dots u_D] \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & \ddots \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_N^T \end{bmatrix}$$

$\cancel{X} = U \Sigma \underline{V^T}$

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^T + \boxed{0 \cdot u_{r+1} v_{r+1}^T + \dots + 0 \cdot u_D v_D^T}$$

$$= \underbrace{[u_1, \dots, u_r]}_{\in \mathbb{R}^{D \times r}} \underbrace{\Sigma_r V_r^T}_{Y \in \mathbb{R}^{r \times N}}$$

$r < D$

$$\cancel{A} \quad \cancel{X} = U \underbrace{\Sigma V^T}_{y} \Rightarrow \quad x_i = \sum_{j=1}^{r} y_i^j u_j = \sum_{i=1}^{k} y_i^j u_j + \boxed{\quad\quad}$$

u_j

PCA: principle Compond Analysis.

$$\min \|x - Uy\| \to y = U^T x$$

$$\min \left\| \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} y \right\|_2^2$$

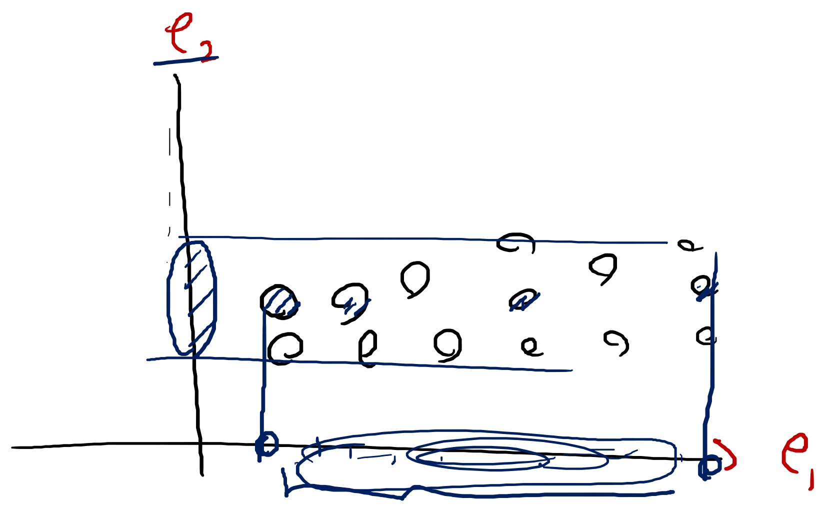
$$\to y = 3$$

$$\begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 3 \\ 4 \end{pmatrix} \to 3$$

$$R^2 \quad \begin{pmatrix} 3 \\ 4 \end{pmatrix} \to 3 \to \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

span $\{u_1\}$



$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$x \to y$$

$$x_i \xrightarrow{U_d^T} y_i \qquad y_i = U_d^T x_i \qquad U_d = [U_1, \dots, U_d]$$

بیشینه

پراکندگی y_i ها ماکزیمم شود

$$\min \sum_{i=1}^{N} \| y_i - \mu_y \|_2^2$$

$$y_i = U^T x_i \quad ,$$

$$\mu_y = \frac{\sum y_i}{N} = \frac{\sum U^T x_i}{N} = U^T \left(\overline{\frac{\sum x_i}{N}}^{\mu_x} \right) = U^T \mu_x$$

$$\sum \| y_i - \mu_y \|_2^2 = \sum_{i=1}^{N} \| U^T (x_i - \mu_x) \|_2^2$$

$$\max_{U} \sum_{i=1}^{N} \|U^T(x_i-\mu_x)\|_2^2 = \sum_{j=1}^{N} \|z_i\|_2^2 = tr(ZZ^T) = tr(U^T \bar{X} \bar{X}^T U)$$

Let $\bar{X} = [\bar{x}_1, \ldots, \bar{x}_N] \to \underbrace{U^T \bar{X}}_{Z} = [U^T \bar{x}_1, \ldots, U^T \bar{x}_N] = [U^T(x_1-\mu_x), \ldots, U^T(x_N-\mu_x)]$

$\bar{x}_i = x_i - \mu_x$

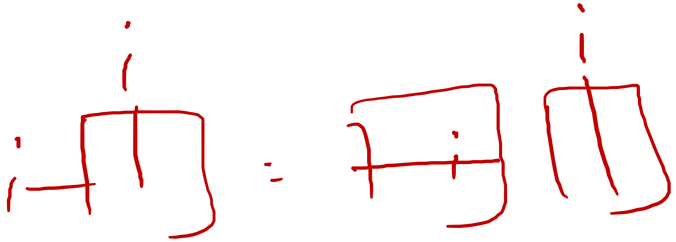
But. $\boxed{\sum \|z_i\|_2^2 = tr(Z^T Z) = tr(ZZ^T)}$

$$\max_{U} tr(U^T \bar{X} \bar{X}^T U)$$

$$= tr(U^T \underbrace{\frac{1}{N-1} \bar{X} \bar{X}^T}_{\text{Cov.}} U)$$

$$C = \frac{1}{N-1} \sum (x_i-\mu)(x_i-\mu)^T = \frac{1}{N-1} \bar{X} \bar{X}^T$$

$$Z^T Z = \begin{bmatrix} z_1^T \\ \vdots \\ z_n^T \end{bmatrix} [z_1 \quad z_n] = \begin{pmatrix} \|z_1\|_2^2 & \\ & \|z_n\|_2^2 \end{pmatrix}$$



$$\max_{U} \operatorname{tr}\left(U^T \bar{X} \bar{X}^T U \right) = \operatorname{tr}_{U}\left(U^T C U \right)$$

کوواریانس

$y = U^T x$

A semi positive definite if $U \in R^{n \times k}$

$\quad$ orthonormal.

$A \in R^{n \times n}$

$U = \arg\max \operatorname{trace}\left(\bar{U}^T A \bar{U} \right) \Rightarrow U = [u_1, \dots, u_k]$

$\bar{U}$ orthonormal

u_i : بردار ویژه (singular vector) متناظر با k امین بردار ویژه بزرگ A است

$$A u_i = \lambda_i u_i \qquad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq \lambda_{k+1} \dots \lambda_n$$

$$\begin{array}{|ccc|} \hline u_1 & u_2 & u_k \\ \hline \end{array} \longrightarrow U = [u_1, \dots, u_k]$$

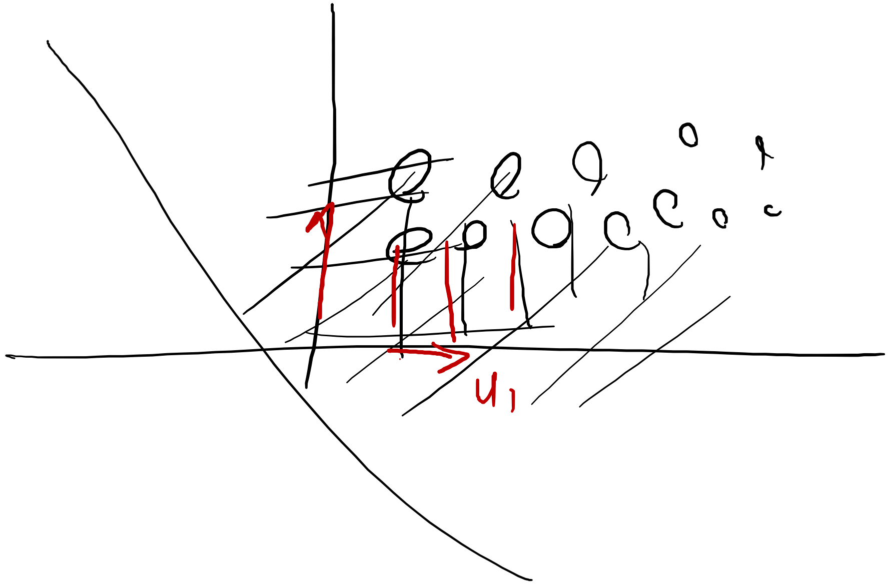
$C = \bar{X} \bar{X}^T$

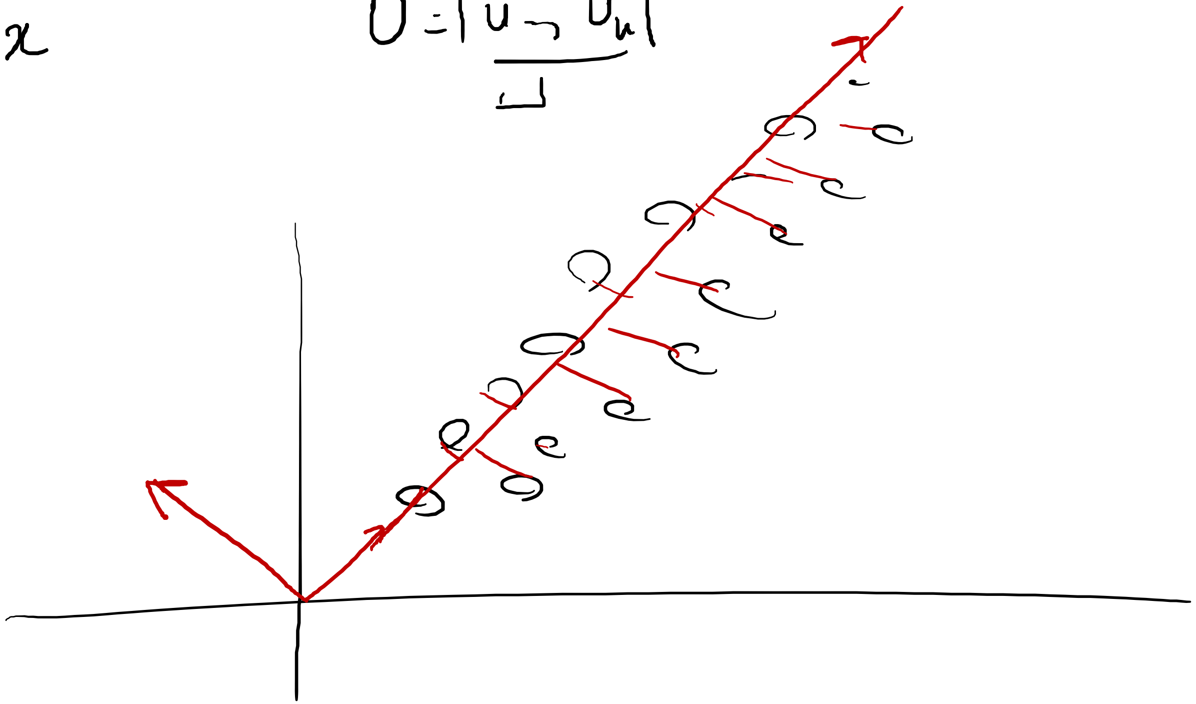
$\forall p \neq 0. \quad p^T C p = \underbrace{p^T \bar{X}}_{q^T} \underbrace{\bar{X}^T p}_{q} = q^T q = \|q\|_2^2$

if U solution $\quad Tr(U^T A U) = \sum_{i=1}^{k} \lambda_i$

$X = [x_1 \ldots x_N]$ $\quad \bar{x}_i = x_i - \mu$ $\quad C = \bar{X}\bar{X}^T \longrightarrow$ eigenvectors

$y = U^T x$ $\quad U = [u_1 \ldots u_k]$





$C = \bar{X}\bar{X}^T$

Let $\bar{X} = U \Sigma V^T$

$\Downarrow$

$$C = \bar{X}\bar{X}^T = U \Sigma \underbrace{V^T V}_{I} \Sigma U^T$$

PCA on $\bar{X}$

$\Downarrow$ dual

SVD $\bar{X}$

$\bar{X} = [x_1 - \mu_n, \ldots, x_N - \mu_n]$

$$C = U \Sigma^2 U^T$$

$$C U = U \Sigma^2$$

$$\underbrace{C[u_1, \ldots, u_n]}= [u_1 \ldots u_n] \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$

if $\mu_n = 0$

Centerlaized

$\Downarrow$

$$C u_i = \sigma_i^2 u_i \qquad C u_i = \lambda_i u_i$$

λ_i

$\bar{X} = X$

$\boxed{\text{SVD} = \text{PCA}}$

$X \xrightarrow{PCA} \quad C = \bar{X}\bar{X}^T \Longrightarrow \bar{X}$ SVD

$\boxed{\text{AutoEncoder}}$

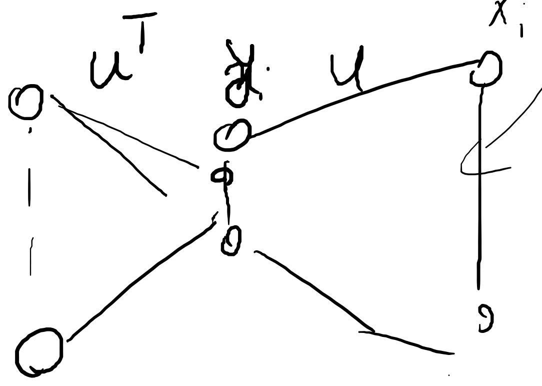
SVD $\min \|X - UY\| = \sum \|x_i - \underbrace{U y_i}_{\bar{x}_i}\|$

$y_i = U^T x_i$

PCA

$\Downarrow$

$\boxed{\text{PPCA}}$



$\min \|x_i - \bar{x}_i\|$

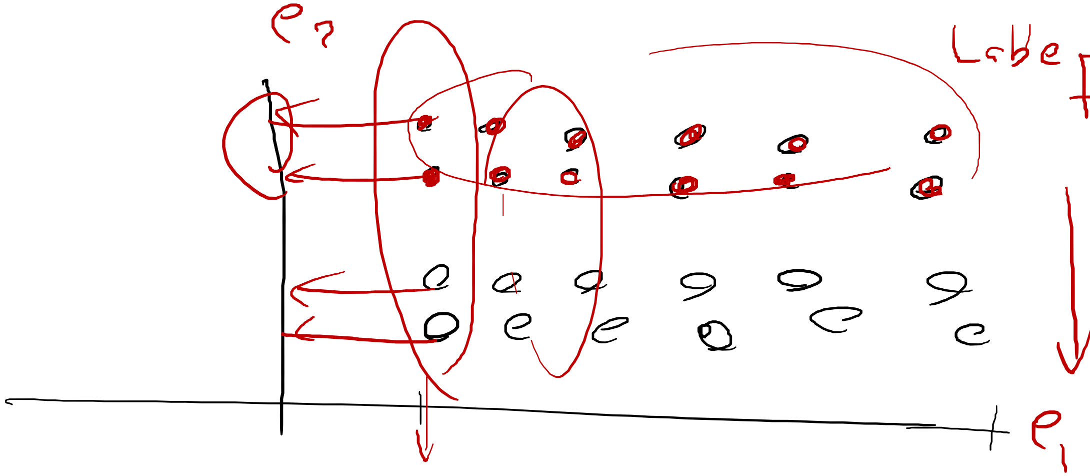
SVD $\|\bar{X} - UY\|$

$\boxed{\text{SVD } X}$

PCA X

SVD $\bar{X}$

PCA X



LDA

Fisher ℓ_n

$y_i = U_p^T x_i$

output.

داده‌های درون کلاسی به هم نزدیک (اسپرندی درون کلاسی کم)

فاصله بین کلاسی زیاد (اسپرندگی بیرون کلاسی زیاد)

min f_1

max f_2

multi objective. $\max \dfrac{f_2}{f_1}$

Bishop
4

$$S_i \quad \sum_{x_j \in C_i} \|y - \mu_i^y\|_2^2$$

$$S \quad \sum \|\mu_i^y - \mu^y\|_2^2$$

$$\max \frac{\sum \|y_i^y - \mu^y\|_2^2}{\sum_{i=1}^{k} \sum_{x \in C_i} \|y - \mu_i^y\|_2^2} = \frac{\mathrm{tr}(U^T S_B U)}{\mathrm{tr}(U^T S_W U)}$$

$$S_B = \sum (\cancel{x}\mu_i - \mu)(\cancel{x}\mu_i - \mu)^T \; ?$$

مین کلاس ام

استفاده

$$S_W = \sum S_i = \cancel{\mathrm{tr}}($$

$$S_i = \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)$$

Bishop & Duda

Linear Vs Nonlinear FE

Linear:

Input Data	$\{x_1, \dots, x_N\} \in R^D$
Projection Matrix	$U \in R^{d*D}$
Reduced Data	$y_i = Ux_i$

Data depended Vs Data Independent

- Predefined Operators like DCT,DST, Fourier, Wavelet,….
- Data depended: PCA , LDA, …

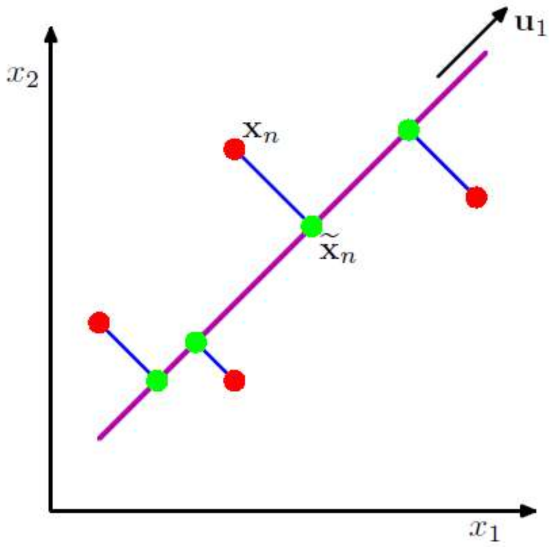
Supervised Vs Unsupervised

Unsupervised: PCA, SVD,...

Supervised: LDA

PCA

Figure 12.2 Principal component analysis seeks a space of lower dimensionality, known as the principal subspace and denoted by the magenta line, such that the orthogonal projection of the data points (red dots) onto this subspace maximizes the variance of the projected points (green dots). An alternative definition of PCA is based on minimizing the sum-of-squares of the projection errors, indicated by the blue lines.



Maximize the Variance in the embedded Space

Dimension Reduction by orthogonal matrix

Data $X = [x_1, \ldots, x_N] \in \mathbb{R}^{D \times N}$ $\Rightarrow$

$$y = U^T x \in \mathbb{R}^d, \quad U \in \mathbb{R}^{D \times d}$$

$$\max \sum_{i=1}^{N} \|y_i - \mu_y\|_2^2$$

$$\max_U \text{Trace}(U^T X X^T U)$$

$$U = [u_1, \ldots, u_d],$$

$$X X^T u_i = \lambda_i u_i, \quad \lambda_1 \geq \ldots \geq \lambda_D$$

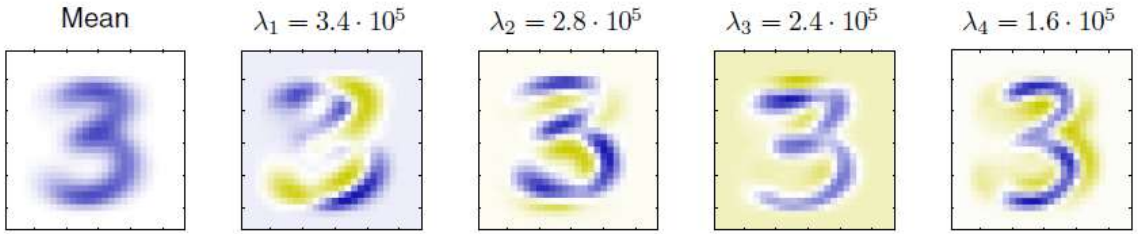


Figure 12.3 The mean vector $\overline{\mathbf{x}}$ along with the first four PCA eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_4$ for the off-line digits data set, together with the corresponding eigenvalues.

$$s_W = \sum_{k=1}^{K} \sum_{n \in C_k} (\mathbf{y}_n - \boldsymbol{\mu}_k)(\mathbf{y}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

$$s_B = \sum_{k=1}^{K} N_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^{\mathrm{T}}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n \in C_k} \mathbf{y}_n, \qquad \boldsymbol{\mu} = \frac{1}{N} \sum_{k=1}^{K} N_k \boldsymbol{\mu}_k.$$

> 2 classes,
ter than the
$\mathbf{v}_k^{\mathrm{T}}\mathbf{x}$, where
her to form
he columns

(4.39)

n of $\mathbf{y}$. The
ses follows

(4.40)

(4.41)

(4.42)

and N_k is the number of patterns in class $\mathcal{C}_k$. In order to find a generalization of the between-class covariance matrix, we follow Duda and Hart (1973) and consider first the total covariance matrix

$$\mathbf{S}_{\mathrm{T}} = \sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{m})(\mathbf{x}_n - \mathbf{m})^{\mathrm{T}} \tag{4.43}$$

where $\mathbf{m}$ is the mean of the total data set

$$\mathbf{m} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n = \frac{1}{N} \sum_{k=1}^{K} N_k \mathbf{m}_k \tag{4.44}$$

and $N = \sum_k N_k$ is the total number of data points. The total covariance matrix can be decomposed into the sum of the within-class covariance matrix, given by (4.40) and (4.41), plus an additional matrix $\mathbf{S}_{\mathrm{B}}$, which we identify as a measure of the between-class covariance

$$\mathbf{S}_{\mathrm{T}} = \mathbf{S}_{\mathrm{W}} + \mathbf{S}_{\mathrm{B}} \tag{4.45}$$

where

$$\mathbf{S}_{\mathrm{B}} = \sum_{k=1}^{K} N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^{\mathrm{T}}. \tag{4.46}$$

Ref

Zaki: chap 5

Tan-chapter 8