

# 机器学习

## Machine learning

# 第七章 降维与特征选择

## Feature Reduction & Selection

授课人：周晓飞  
zhouxiaofei@iie.ac.cn  
2020-11-28

# 第七章 降维与特征选择

## 7.1 概述

## 7.2 特征选择

## 7.3 特征提取

# 特征提取

## 特征提取问题

**特征变换：**

$$y=W(x)$$

$$x: D\text{维} \rightarrow y: d\text{维}; \quad D \gg d$$

**优点：**

- 数据更紧致的压缩
- 优化预测性能
- 加快学习速度

**应用领域广：**模式识别、图像处理、信号处理、数据压缩、微波、雷达、...

# 特征提取

## 特征提取问题

不同的应用问题会有不同的特征提取研究问题：

例如：

图像：SIFT 特征、纹理特征

微波：频段、多普勒

文本：情感极性特征、依存关系

语音：音色、饱和度、爆破音

...

本讲从数据压缩表示和机器学习数据处理的角度

# 特征提取

## 特征提取问题

### 本章讲授的特征提取方法

- 线性变换

  - 方法 1: 主成分分析 (PCA)

  - 方法 2: 线性鉴别分析 (LDA)

- 非线性变换

  - 方法 3: 核方法的特征提取 KPCA、KFDA

  - 方法 4: 流行学习

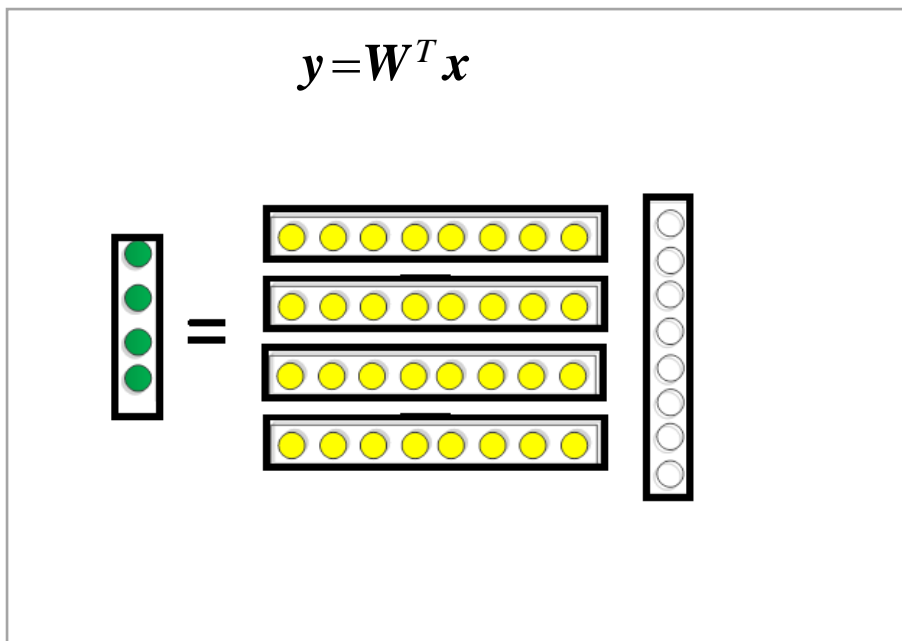
  - 方法 5: 非负矩阵分解

# 特征提取

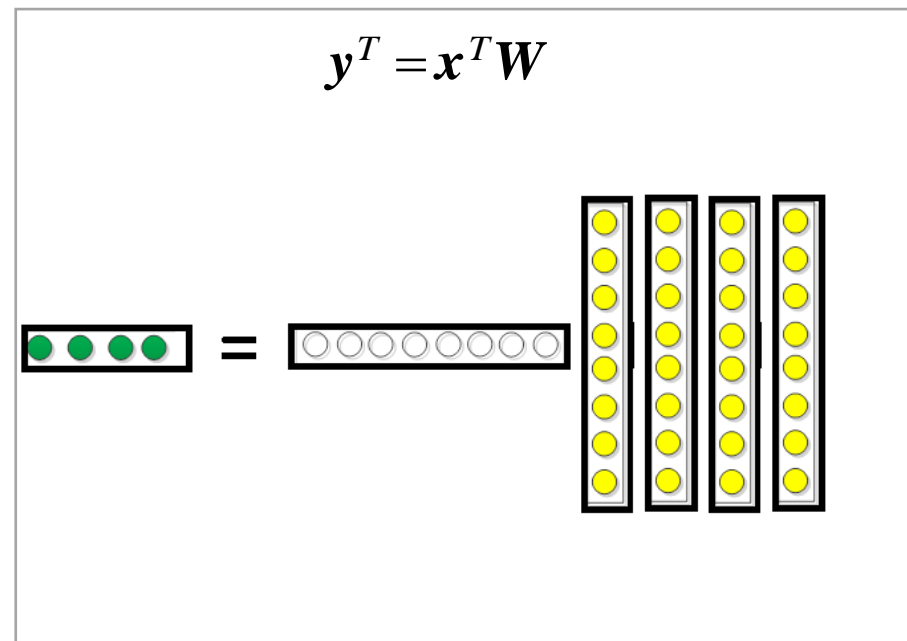
## 线性变换

线性变换的直观表示：

( $x, y, w$  为列向量),  $W$  是  $D \times d$  维  $W = (w_1, w_2, \dots, w_d)$



或者

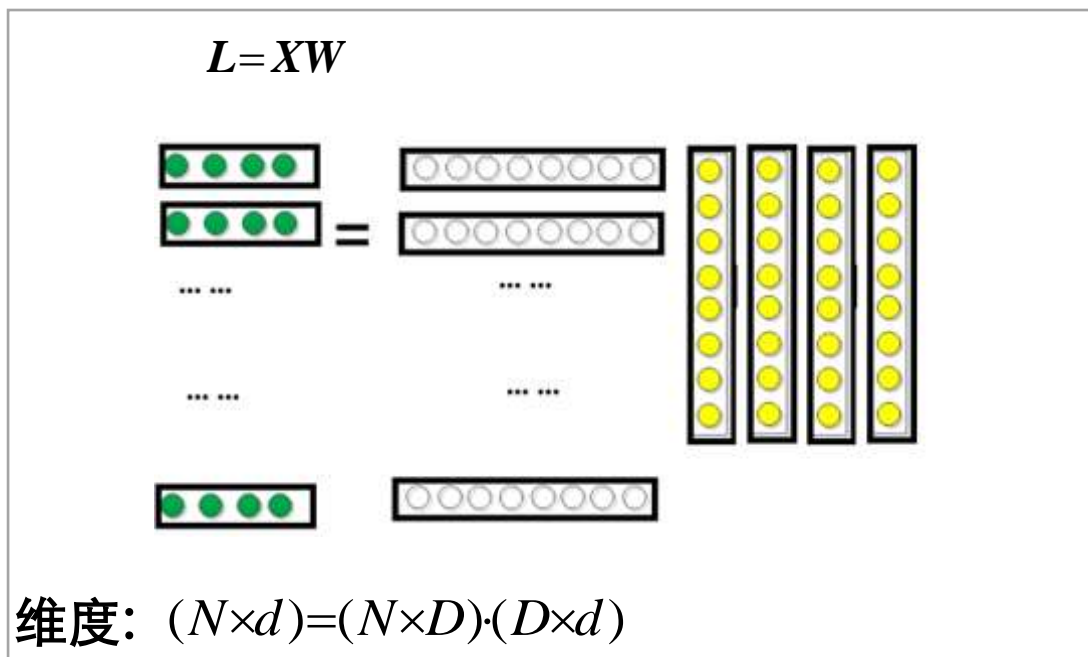


# 特征提取

## 线性变换

### 特征提取目标：学习变换矩阵

给定  $X=(x_1,x_2,...,x_n)^T$ ，通过某种降维准则，学习变换矩阵  $W=(w_1,w_2,...,w_d)$ 。



$W$  是  $d$  个基向量, 向量  $x$  被映射到基向量线性张成的子空间内。

# 特征提取

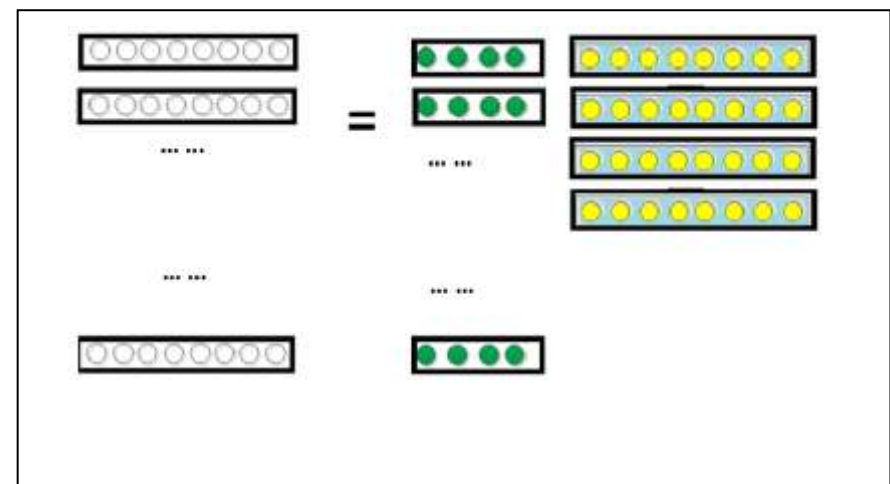
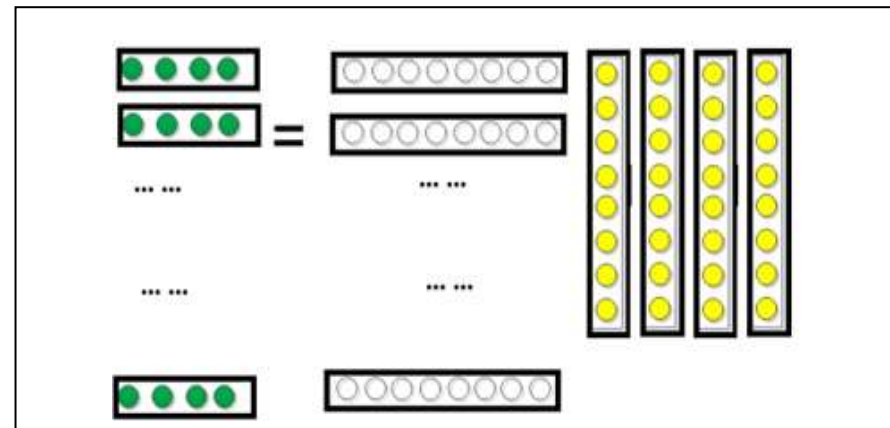
## 线性变换

常见的两种降维表示途径：

- 投影： $L = XW$

- 矩阵分解：低秩表示  $X = LR$

**思考 1：**矩阵分解  $X = LR$  的等号一定存在吗？



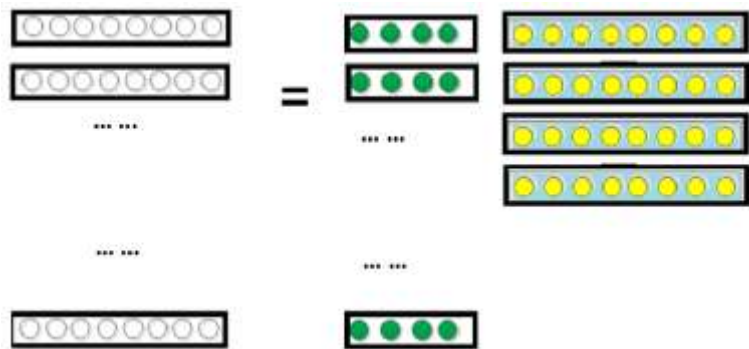


# 特征提取

## 线性变换

**思考 1：**矩阵分解  $X=LR$  的等号一定存在吗？

至少要  $\text{rank}(X)=\text{rank}(L)=\text{rank}(R)$

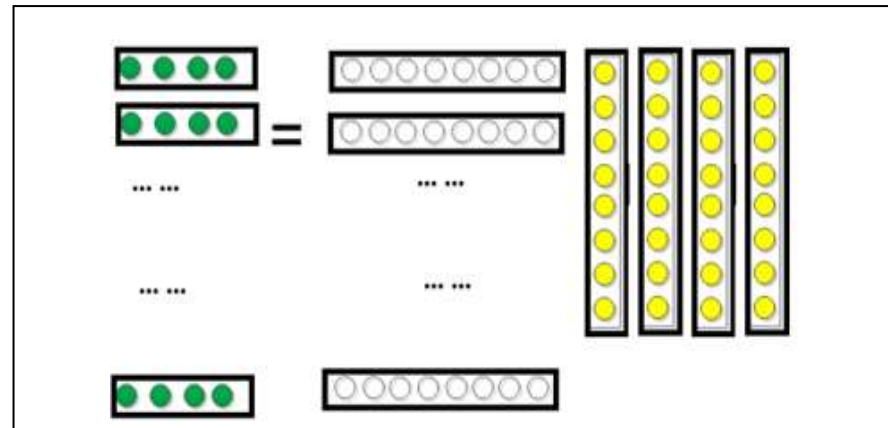


# 特征提取

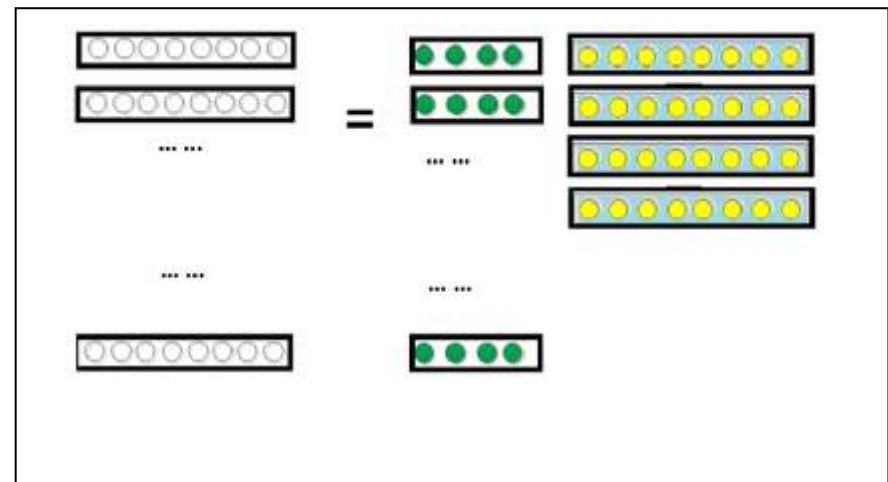
## 线性变换

思考 2：如果  $X = LR$  成立，  
两种低维表示  $L$  什么情况等价？

$$L = XW$$



$$X = LR$$

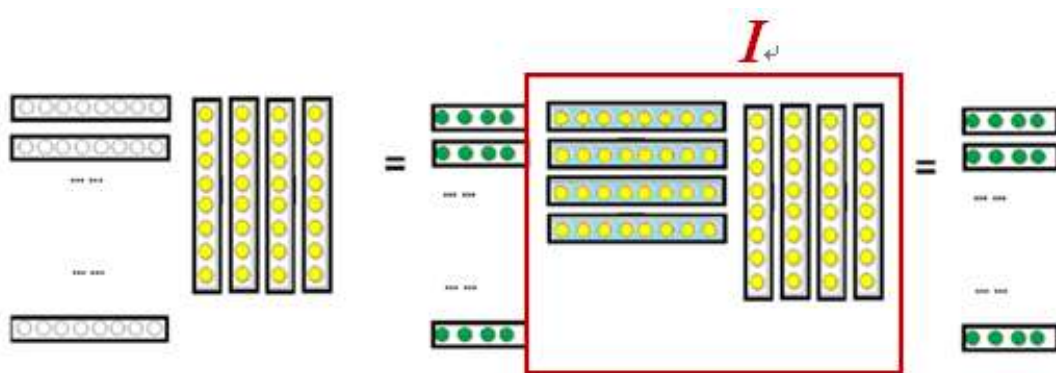


# 特征提取

## 线性变换

如果  $X=LR$  成立，式子两边同时右乘  $W$ :  $XW=LRW$

当  $R$  行满秩、 $W$  列满秩，且当  $RW=I$  (即  $W=R^{-1}$ ) 时， $L=XW$ 。



此时，两个式子的低维表示  $L$  等价。

# 特征提取

## 线性变换

无损压缩时为等号；通常会有损降维：

The diagram shows the equation  $\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{L} \cdot \mathbf{R}$  enclosed in a light gray box. Below the box, three labels with arrows point to specific parts of the equation: 'approximation' points to the  $\approx$  symbol, 'left factor' points to the red  $\mathbf{L}$ , and 'right factor' points to the blue  $\mathbf{R}$ .

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{L} \cdot \mathbf{R}$$

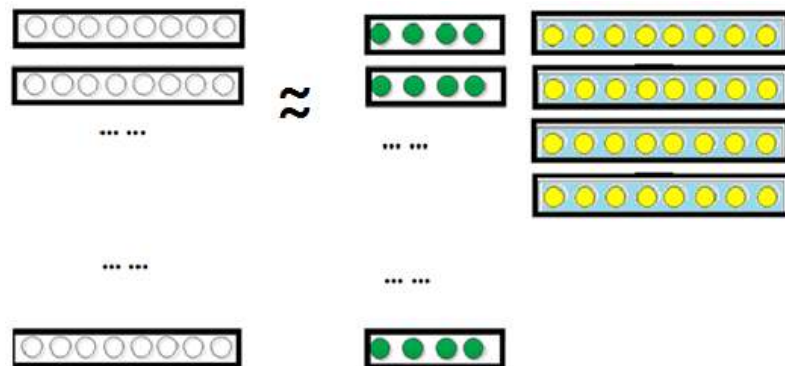
approximation   left factor   right factor

# 特征提取

## 主分量分析(PCA)

PCA 降维表示：

Pattern  $\approx$  low-dimensional representation \* **eigen patterns**



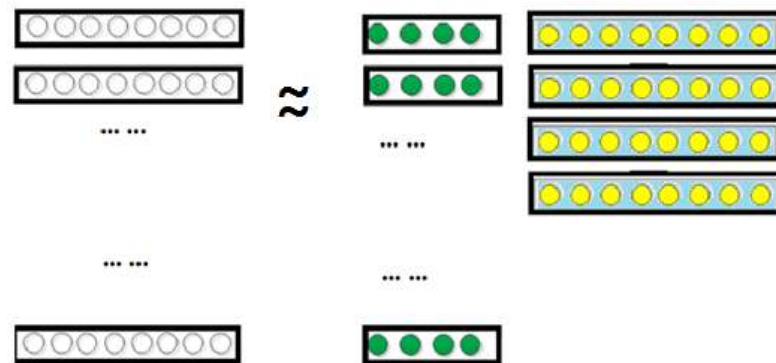
目标：学习 eigen pattern

# 特征提取

## 主分量分析(PCA)

目标函数：均方误差最小原则（求最优重构子空间）

$$\min_{L,R} \|X - LR\|^2 \quad \text{s.t. } RR^T = I$$



# 特征提取

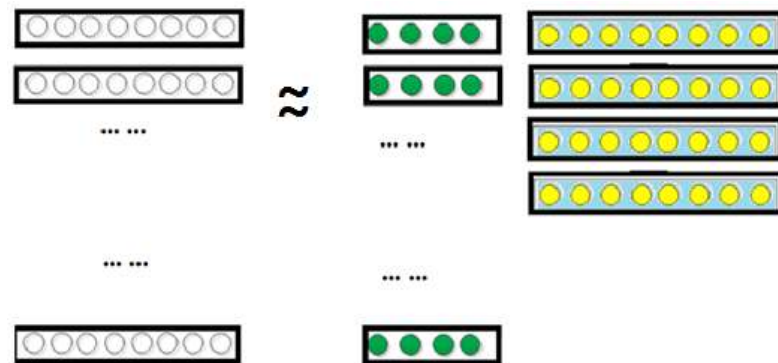
## 主分量分析(PCA)

目标函数：均方误差最小原则（求最优重构子空间）

$$\min_{L,R} \|X - LR\|^2 \quad \text{s.t. } RR^T = I$$

如果将因子  $R$  由投影  $W$  表示，

$$RW = I, \quad RR^T = I \quad \Rightarrow \quad R = W^T$$



# 特征提取

## 主分量分析(PCA)

目标函数：均方误差最小原则（求最优重构子空间）

$$\min_{L,R} \|X - LR\|^2 \quad \text{s.t. } RR^T = I$$

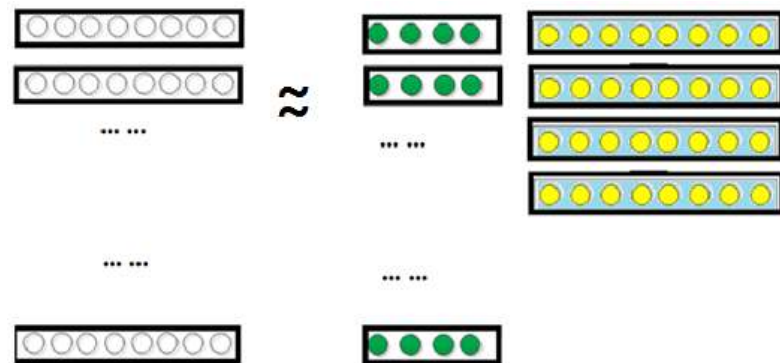
如果将因子  $R$  由投影  $W$  表示,

$$RW = I, \quad RR^T = I \quad \Rightarrow \quad R = W^T$$

推导  $RW = I$ ,

$$W = R^*(\text{右伪逆}) = R^T (RR^T)^{-1}$$

$$W = R^T$$



投影向量关系：  $RR^T = I$ ，则  $W^T W = I$ ，  $R = W^T$



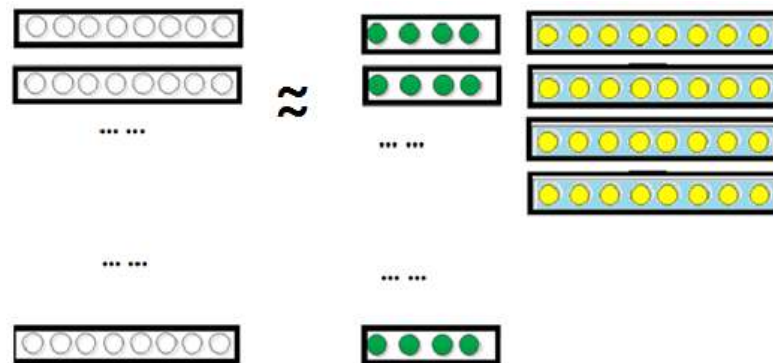
# 特征提取

## 主分量分析(PCA)

### 重写目标函数

$$\min_{L, W} \|X - LW^T\|^2 \quad \text{s.t. } W^T W = I$$

$$\|X - LW^T\|^2 = \sum_i \|x_i^T - l_i^T W^T\|^2$$



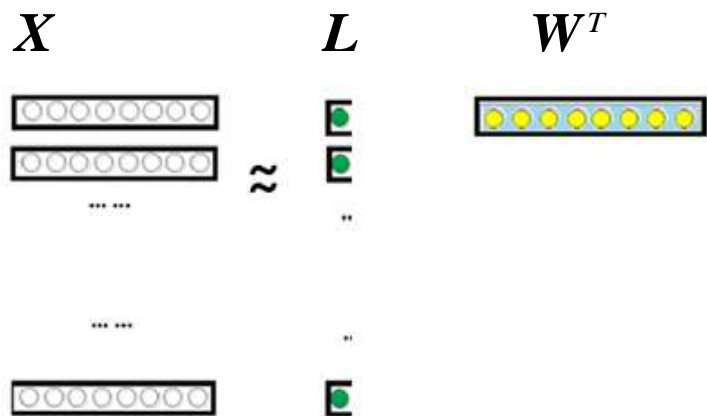
求解该优化问题，可以先固定  $W$ ，  
求出  $L$  关于  $W$  和  $X$  的表示，将该问题转化为只与  $W$  相关的优化问题。

# 特征提取

## 主分量分析(PCA)

如何理解重构？

一个投影方向情况：



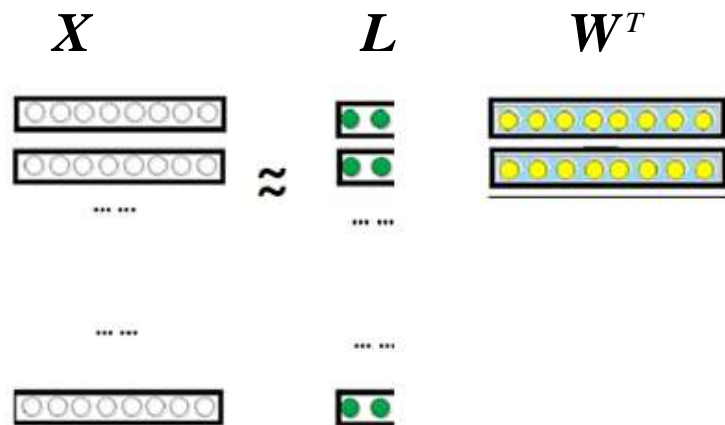
$$\|X - L W^T\|^2 = \sum_i \|x_i^T - l_i^T W^T\|^2$$

# 特征提取

## 主分量分析(PCA)

如何理解重构？

2 个投影方向情况：



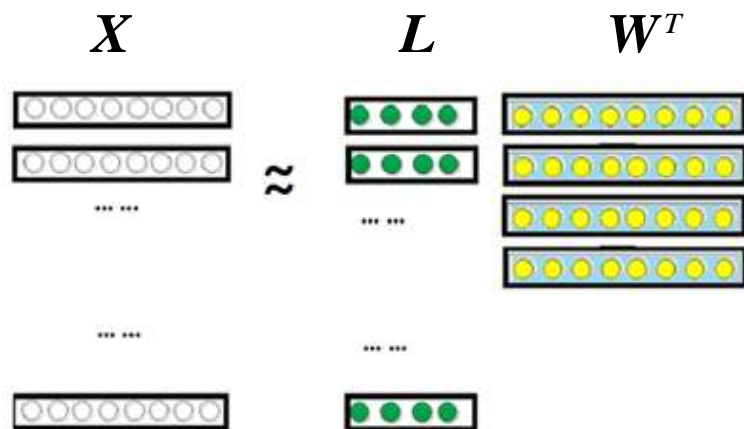
$$\|X - LW^T\|^2 = \sum_i \|x_i^T - l_i^T W^T\|^2$$

# 特征提取

## 主分量分析(PCA)

如何理解重构？

多个投影方向情况：

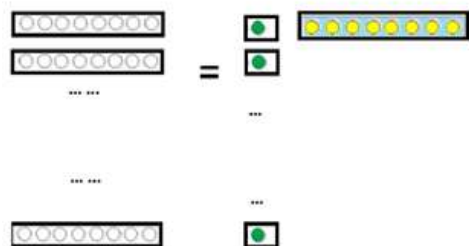


$$\|X - L W^T\|^2 = \sum_i \|x_i^T - l_i^T W^T\|^2$$

# 特征提取

## 主分量分析(PCA)

一般性讨论一个投影方向情况：

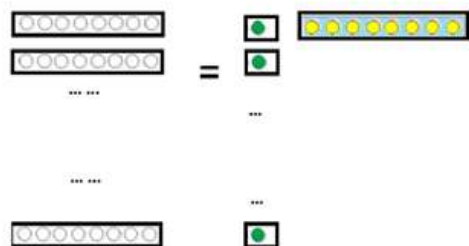


$$\sum_i \|x_i^T - l_i^T W^T\|^2$$

# 特征提取

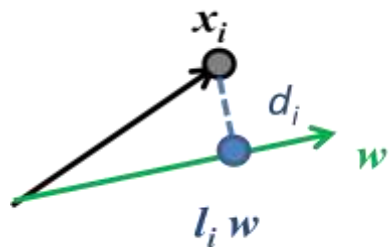
## 主分量分析(PCA)

一般性讨论一个投影方向情况：



$$\sum_i \|x_i^T - l_i^T W^T\|^2$$

当固定  $w$  时，每个样本  $l_i = x_i^T w$  时，重构误差  $\|x_i^T - l_i w^T\|^2$  最小，



(几何比较直观，该结论也可以通过求解  $\min_{l_i} \|x_i^T - l_i w^T\|^2$ )

# 特征提取

## 主分量分析(PCA)

目标函数的等价形式：

$$\begin{aligned}\min_{L, \mathbf{w}} \sum_i \|\mathbf{x}_i^T - l_i \mathbf{w}^T\|^2 &= \min_{\mathbf{w}} \sum_i d_i^2 \\&= \min_{\mathbf{w}} \sum_i \left( \|\mathbf{x}\|^2 - \|l_i \mathbf{w}^T\|^2 \right) = \max_{\mathbf{w}} \sum_i \|l_i \mathbf{w}^T\|^2 \\&= \max_{\mathbf{w}} \sum_i l^2 \|\mathbf{w}\|^2 \\&= \max_{\mathbf{w}} \sum_i l^2 \\&= \max_{\mathbf{w}} \sum_i \mathbf{w}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w} & \Rightarrow & \max_{\mathbf{w}} \sum_i \|\mathbf{w}^T \mathbf{x}_i\|^2 \\&= \max_{\mathbf{w}} \mathbf{w}^T \left( \sum_i \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{w}\end{aligned}$$

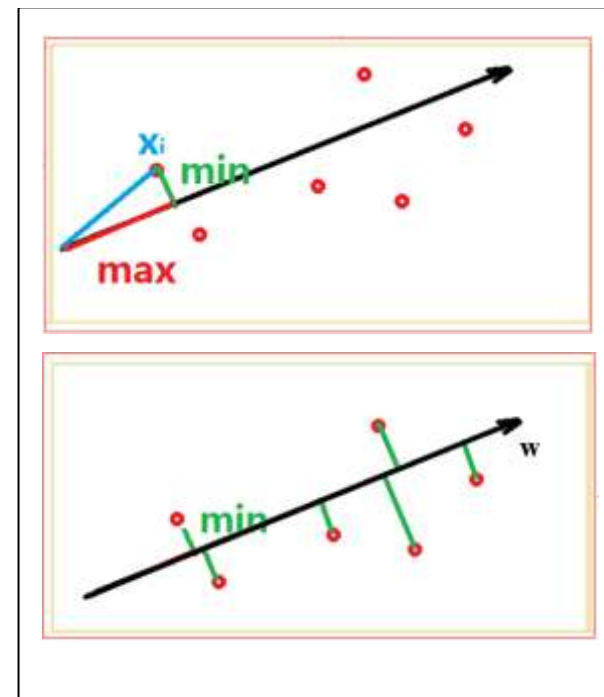
# 特征提取

## 主分量分析(PCA)

目标函数的等价形式：

$$\begin{aligned}\min_{L, w} \sum_i \|x_i^T - l_i w^T\|^2 &= \min_w \sum_i d_i^2 \\&= \min_w \sum_i (\|x\|^2 - \|l_i w^T\|^2) = \max_w \sum_i \|l_i w^T\|^2 \\&= \max_w \sum_i l^2 \|w\|^2 \\&= \max_w \sum_i l^2 \\&= \max_w \sum_i w^T x_i x_i^T w \quad \Rightarrow \quad \max_w \sum_i \|w^T x_i\|^2 \\&= \max_w w^T (\sum_i x_i x_i^T) w\end{aligned}$$

最小误差 等价于 最大投影





# 特征提取

## 主分量分析(PCA)

求解目标函数：

$$\max_{\mathbf{w}} \mathbf{w}^T \mathbf{S} \mathbf{w}$$

$$\text{其中, } \mathbf{w}^T \mathbf{w} = 1, \mathbf{S} = E[\mathbf{x}\mathbf{x}^T] \approx \frac{1}{n} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T = \frac{1}{n} \mathbf{X}^T \mathbf{X}$$

• Lagrange 函数:

$$L = -\mathbf{w}^T \mathbf{S} \mathbf{w} + \lambda(\mathbf{w}^T \mathbf{w} - 1)$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \Longrightarrow \quad (\mathbf{S} - \lambda \mathbf{I}) \mathbf{w} = 0$$

• 等价于求  $\mathbf{S}$  的特征值和对应的特征向量:

$$\mathbf{S} \mathbf{w} = \lambda \mathbf{w}$$

# 特征提取

## 主分量分析(PCA)

特征值对应目标函数值，最大特征值对应特征向量为最优方向：

$$S\mathbf{w}=\lambda\mathbf{w} \quad \Rightarrow \quad \mathbf{w}^T S\mathbf{w}=\lambda$$

目标函数

$$\max_{\mathbf{w}} \mathbf{w}^T S\mathbf{w}$$

# 特征提取

## 主分量分析(PCA)

### $S$ 的特征值分解

$$S\mathbf{w}_j = \lambda_j \mathbf{w}_j, j=1,2,\dots,m,$$

相应的特征值:  $\lambda_1 > \lambda_2 > \dots > \lambda_j > \dots > \lambda_m$ , 特征向量:  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j, \dots, \mathbf{w}_m]$

- $\mathbf{S}\mathbf{W} = \mathbf{W}\mathbf{A}$ , 其中  $\mathbf{A} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_m]$ ,

$$\mathbf{W}^T \mathbf{W} = \mathbf{I}, \quad \mathbf{w}_i^T \mathbf{w}_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}, \quad \mathbf{W}^T = \mathbf{W}^{-1}$$

# 特征提取

## 主分量分析(PCA)

### $S$ 的特征值分解

$$S\mathbf{w}_j = \lambda_j \mathbf{w}_j, j=1,2,\dots,m,$$

相应的特征值:  $\lambda_1 > \lambda_2 > \dots > \lambda_j > \dots > \lambda_m$ , 特征向量:  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j, \dots, \mathbf{w}_m]$

- $\mathbf{SW} = \mathbf{WA}$ , 其中  $\mathbf{A} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_m]$ ,

$$\mathbf{W}^T \mathbf{W} = \mathbf{I}, \quad \mathbf{w}_i^T \mathbf{w}_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}, \quad \mathbf{W}^T = \mathbf{W}^{-1}$$

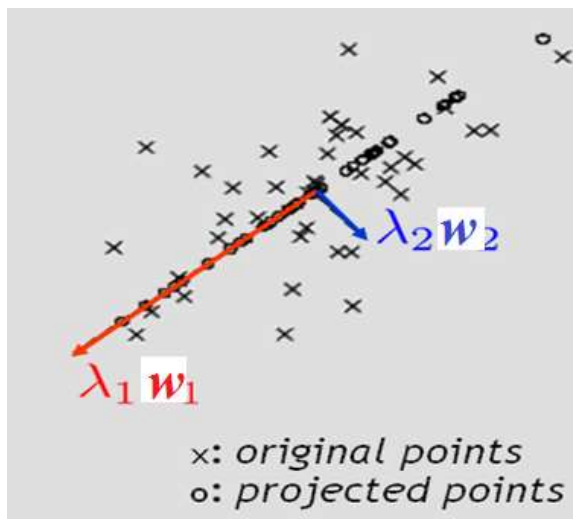
- $\mathbf{W}^T \mathbf{SW} = \mathbf{A}, \quad \mathbf{w}_i^T \mathbf{S} \mathbf{w}_j = \begin{cases} \lambda_j, & i=j \\ 0, & i \neq j \end{cases}$

# 特征提取

## 主分量分析(PCA)

### 注意问题

- $S=(1/N)X^T X$ ,  $(1/N)$  可以省略
- 特征值的意义: 样本在  $w$  方向的投影平均值 (或和) 最大.

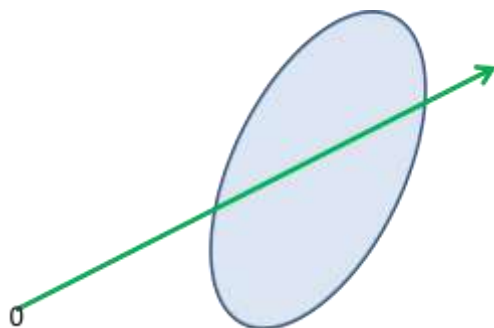


# 特征提取

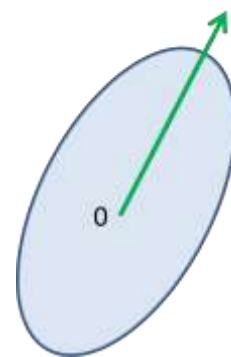
## 主分量分析(PCA)

### 标准化样本

$$x_i = x_i - \bar{x}, \quad \bar{x} = (1/n) \sum_i x_i$$



未标准化



标准化

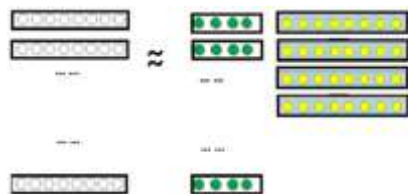
# 特征提取

## 主分量分析(PCA)

注意：特征方向也是特征模式（特征因子 R）

Pattern  $\approx$  low-dimensional representation \* **eigen patterns**

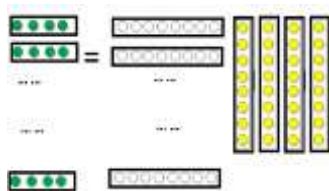
- 矩阵分解：  $X = LR$



特征因子  $R$  就是 投影方向的转置

$$R = W^T$$

- 投影：  $L = XW$



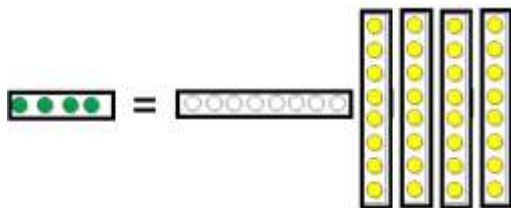
# 特征提取

## 主分量分析(PCA)

**数据的变换、降维、主成分：**

分解出  $m$  个特征方向

$$\begin{aligned} \mathbf{y}^T &= [y_1, y_2, \dots, y_m] \\ &= [\mathbf{x}^T \mathbf{w}_1, \mathbf{x}^T \mathbf{w}_2, \dots, \mathbf{x}^T \mathbf{w}_m]^T, \\ &= \mathbf{x}^T \mathbf{W} \end{aligned}$$





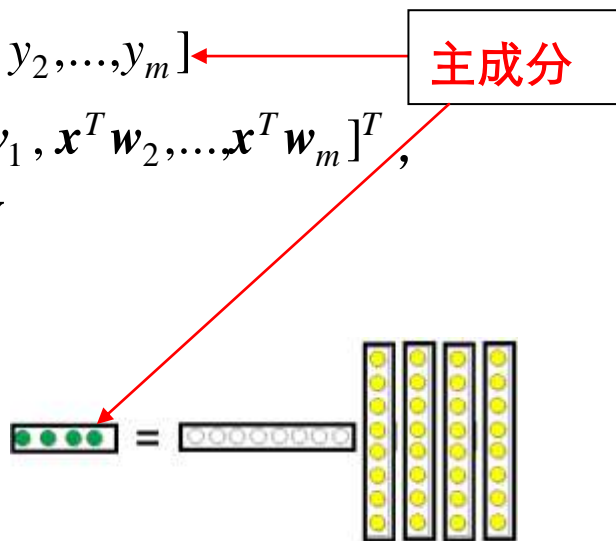
# 特征提取

## 主分量分析(PCA)

数据的变换、降维、主成分：

分解出  $m$  个特征方向

$$\begin{aligned} \mathbf{y}^T &= [y_1, y_2, \dots, y_m] \\ &= [\mathbf{x}^T \mathbf{w}_1, \mathbf{x}^T \mathbf{w}_2, \dots, \mathbf{x}^T \mathbf{w}_m]^T, \\ &= \mathbf{x}^T \mathbf{W} \end{aligned}$$



# 特征提取

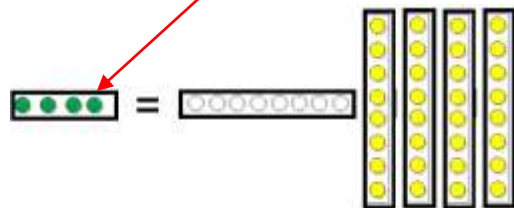
## 主分量分析(PCA)

数据的变换、降维、主成分：

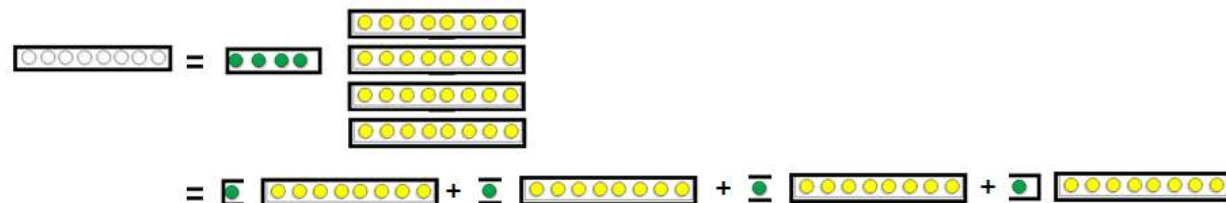
分解出  $m$  个特征方向

$$\begin{aligned} \mathbf{y}^T &= [y_1, y_2, \dots, y_m] \\ &= [\mathbf{x}^T \mathbf{w}_1, \mathbf{x}^T \mathbf{w}_2, \dots, \mathbf{x}^T \mathbf{w}_m]^T, \\ &= \mathbf{x}^T \mathbf{W} \end{aligned}$$

主成分



$$\mathbf{x}^T = \mathbf{y}^T \mathbf{W}^T$$



# 特征提取

## 主分量分析(PCA)

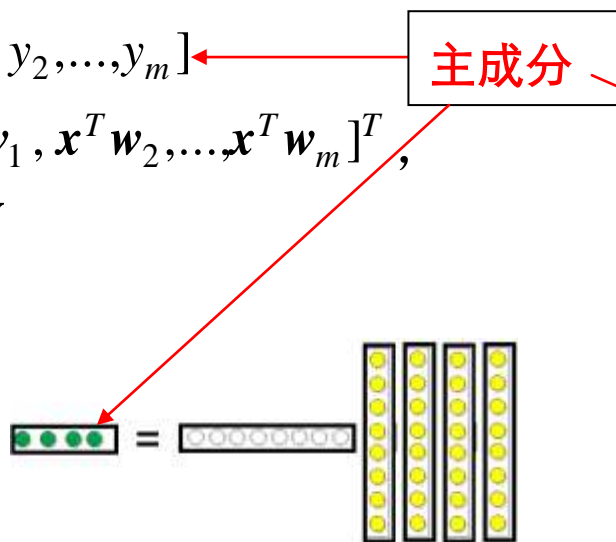
数据的变换、降维、主成分：

分解出  $m$  个特征方向

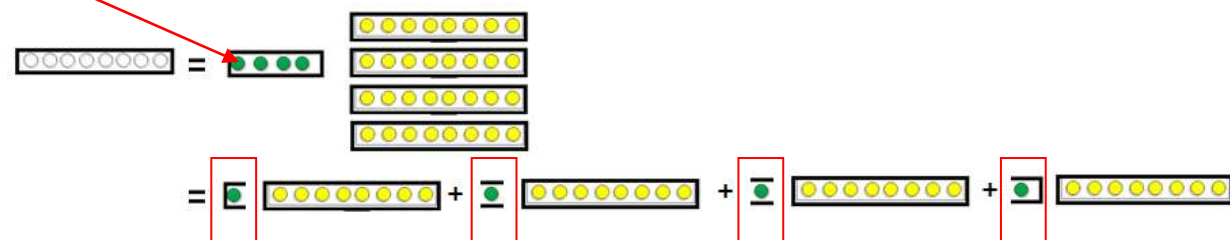
$$\mathbf{y}^T = [y_1, y_2, \dots, y_m]$$

$$= [\mathbf{x}^T \mathbf{w}_1, \mathbf{x}^T \mathbf{w}_2, \dots, \mathbf{x}^T \mathbf{w}_m]^T,$$

$$= \mathbf{x}^T \mathbf{W}$$



$$\mathbf{x}^T = \mathbf{y}^T \mathbf{W}^T$$



# 特征提取

## 主分量分析(PCA)

### PCA 算法流程

(1) 标准化样本

$$x_i = x_i - \bar{x}, \quad \bar{x} = (1/n) \sum_i x_i$$

(2) 求样本的协方差矩阵  $XX^T$  特征值，并降排序  $\lambda_1 > \lambda_2 > \dots > \lambda_m$ ，

对应非零特征向量：  $w_1, w_2, \dots, w_m$ ，

(4) 变换矩阵

Case 1: 无损压缩  $W = (w_1, w_2, \dots, w_m)$

Case 2: 有损压缩  $W = (w_1, w_2, \dots, w_l), l < m$

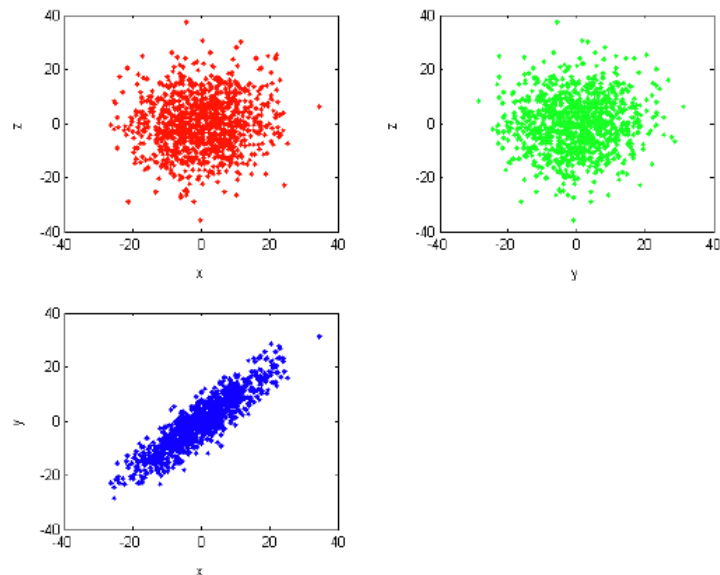
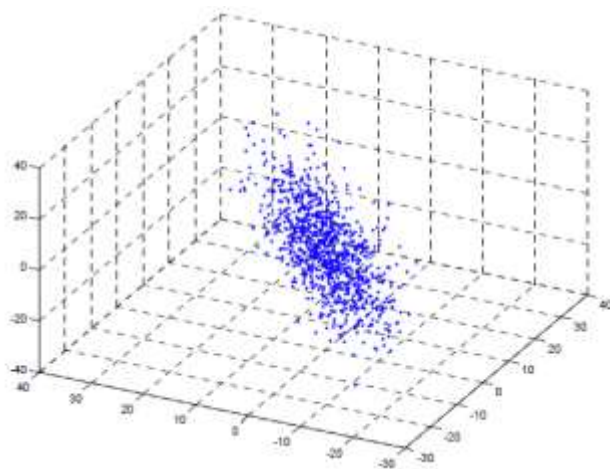
(5) 降维表示

训练样本：  $L = XW$

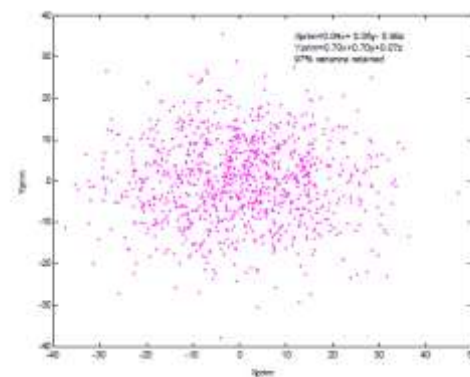
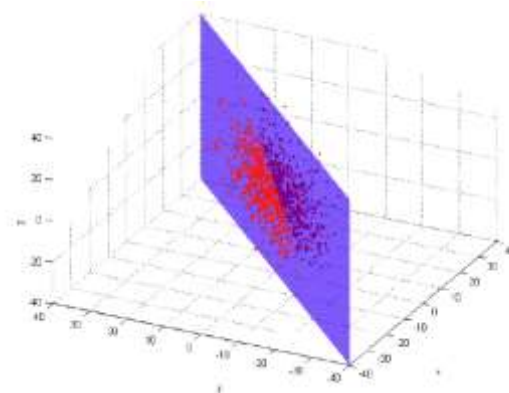
新样本标准化后，降维表示：  $y^T = x^T W$

# 特征提取

## 主分量分析(PCA)



特征选择

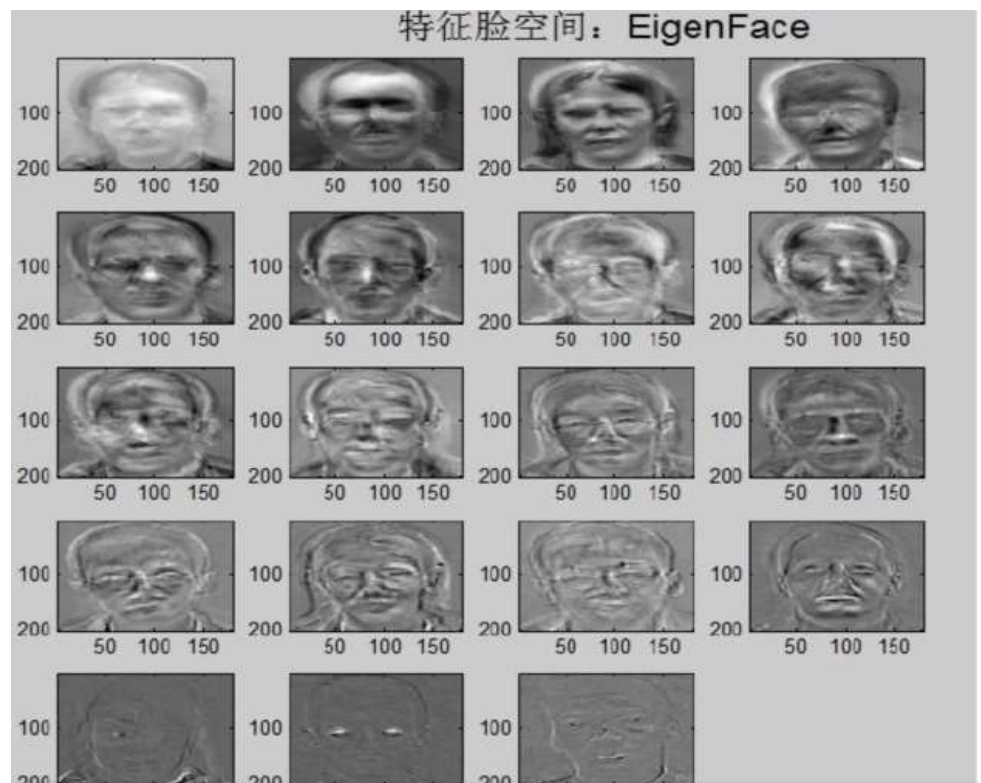


PCA

# 特征提取

## 主分量分析(PCA)

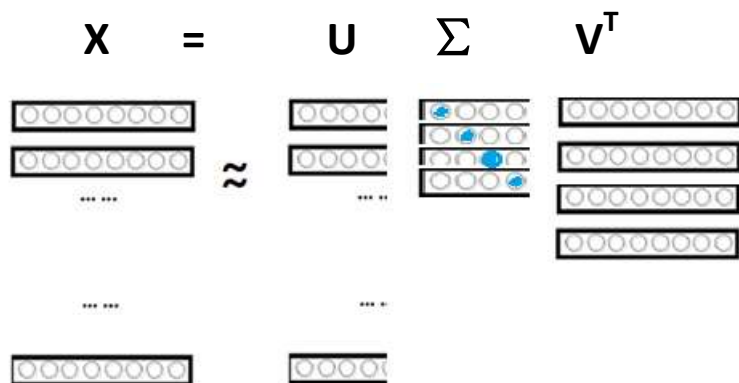
### 人脸识别例子



# 特征提取

## 主分量分析(PCA)

SVD 角度的 PCA :



$X^T X$  与  $XX^T$  特征值分解, 具有相同的非零特征值  $\{\lambda_i\}$ ,

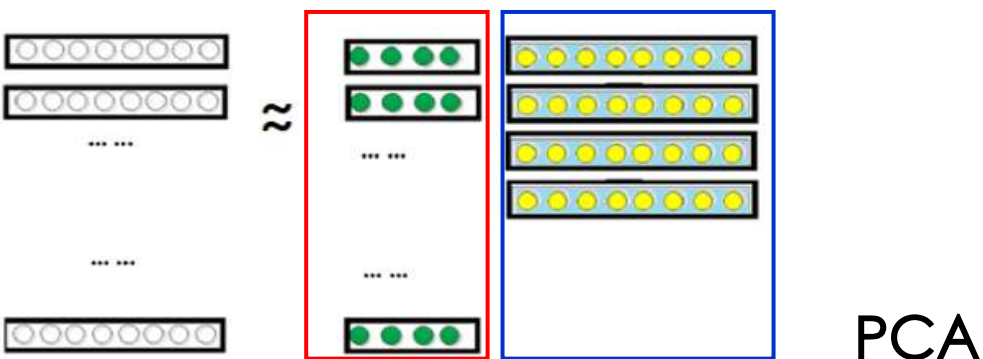
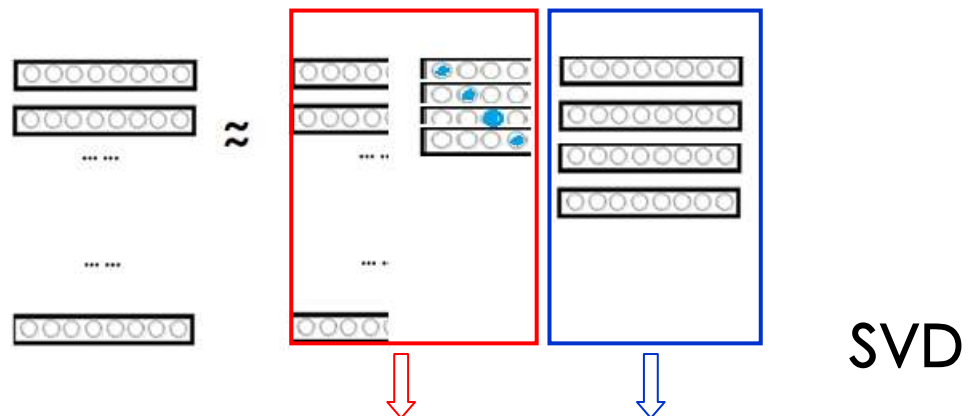
$X^T X$  的特征向量构成矩阵  $V^T$ ,  $XX^T$  的特征向量构成矩阵  $U$

$\Sigma$  仅在对角线上有奇异值, 奇异值与特征值的关系:  $\sqrt{\lambda_i}$

# 特征提取

## 主分量分析(PCA)

SVD 角度的 PCA :





# 特征提取

## 线性鉴别分析(LDA)

### 线性鉴别分析的意义：

PCA 不能保证类别区分的有效性，LDA 特征的优点：类内最小、类间最大。

### 线性鉴别准则（回顾第三章）：

$$J_F(w) = \frac{w^T S_b w}{w^T S_w w}$$

注：向量  $w, x, m, y$  都为列向量

# 特征提取

## 线性鉴别分析(LDA)

特征方向的提取：

$$S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w}$$

$\mathbf{w}$  是  $S_w^{-1} S_b$  的特征向量

设矩阵  $S_w^{-1} S_b$  的特征值为  $\lambda_1, \lambda_2, \dots, \lambda_D$

且  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$

选前  $d$  个特征值对应的特征向量  $\mu_1, \mu_2, \dots, \mu_d$

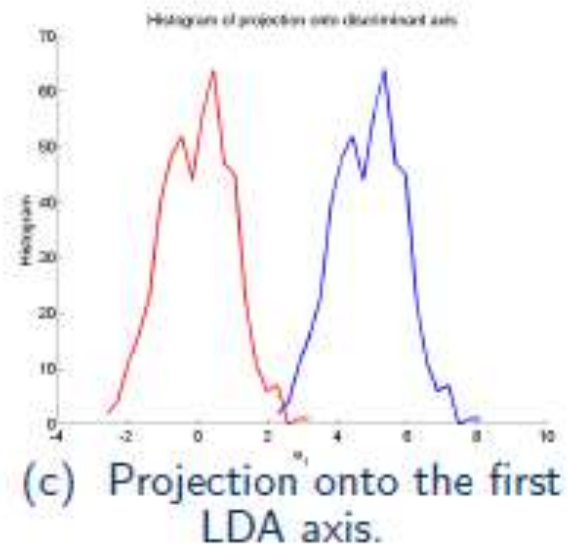
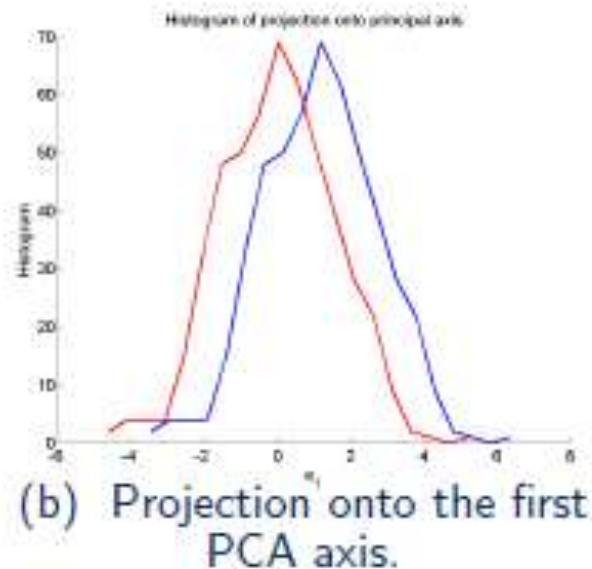
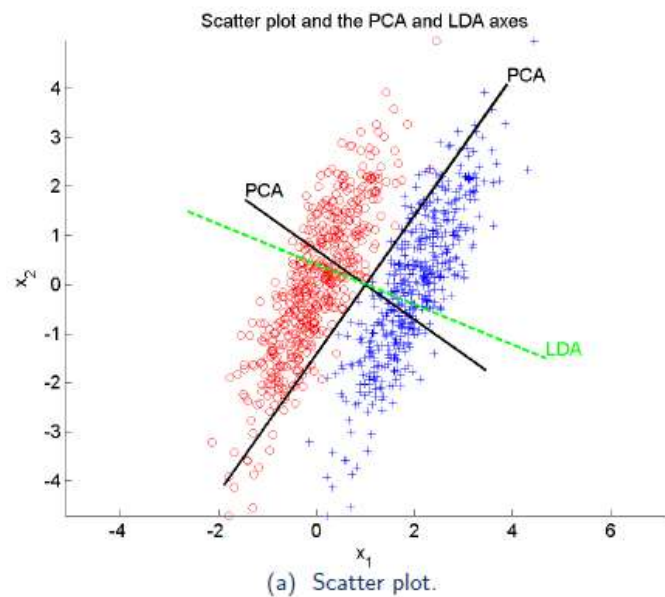
组成矩阵  $W = [\mu_1 \mu_2 \cdots \mu_d]$

(对于多类可以提取多个特征，如果只有两类， $S_b$  的秩为 1，每次只能提取一维特征)

# 特征提取

## 线性鉴别分析(LDA)

### LDA 与 PCA 的比较



# 特征提取

## 核主成分分析 ( KPCA )

### 高维映射后的线性变换

高维映射:  $x \in X \mapsto \phi(x) \in H$ ,  $\phi(x)$  为列向量

变换矩阵:  $\Phi = (\phi(x_1), \phi(x_2), \dots, \phi(x_n))$

线性变换:  $L = W^T \Phi$ ,  $W = (w_1, w_2, \dots, w_d)$

高维映射  $\phi(x)$ ? 线性变换  $W^T$ ?

# 特征提取

## 核主成分分析 ( KPCA )

### 高维空间的主成分分析

$$S\mathbf{w} = \lambda\mathbf{w} \quad \Leftrightarrow \quad \Phi\Phi^T\Phi\boldsymbol{\alpha} = \lambda\Phi\boldsymbol{\alpha}$$

$$\text{其中, } S = \sum_i \phi(\mathbf{x}_i)\phi^T(\mathbf{x}_i) = \Phi\Phi^T, \quad \mathbf{w} = \sum_i \alpha_i \phi(\mathbf{x}_i) = \Phi\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$$

如何可以引入核函数?

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

# 特征提取

## 核主成分分析 (KPCA)

### 核函数引入到主成分分析

两边左乘  $\Phi$  :  $\Phi^T \Phi \Phi^T \Phi \alpha = \lambda \Phi^T \Phi \alpha$

**K   K   K**

$$\Rightarrow K\alpha = \lambda\alpha$$

其中  $K = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{pmatrix}$

# 特征提取

## 核主成分分析 ( KPCA )

### 核主成分分析 KPCA 流程

- (1) 求核矩阵  $K$  的特征值，对应特征向量  $\alpha$  的问题：  $K\alpha = \lambda\alpha$
- (2) 核矩阵  $K$  的特征值降排序  $\lambda_1 > \lambda_2 > \dots > \lambda_m$ ，前  $d$  个特征值对应特征向量  $\alpha_1, \alpha_2, \dots, \alpha_d$
- (3) 高维空间中的投影方向  $w_i = \Phi\alpha_i$   $A = (\alpha_1, \alpha_2, \dots, \alpha_d)$ ，投影矩阵  $W = \Phi A$
- (4) 降维表示

训练集低维表示：

$$L = \Phi^T W = \Phi^T \Phi A = K A$$

新样本  $x$  的低维表示：

$$y = \phi(x)^T \Phi A = K(x, X) A,$$

其中  $K(x, X) = (k(x, x_1), k(x, x_2), \dots, k(x, x_n))$

# 特征提取

## 局部线性变换 (LLE)

**基本思想：**LLE 方法是一种流形学习[Saul 01]，保持样本间的局部线性关系，整体实现非线性映射。

**局部线性结构：**

每一个样本与近邻 k 个样本的局部线性关系

$$\arg \min_w E_w = \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^k w(i,j) \mathbf{x}_{ij} \right\|^2, \quad \sum_{j=1}^k w(i,j) = 1$$

计算降维数据

$$\arg \min_Y E_Y = \sum_{i=1}^n \left\| \mathbf{y}_i - \sum_{j=1}^k w(i,j) \mathbf{y}_j \right\|^2$$



# 特征提取

## 局部线性变换 (LLE)

### 算法流程:

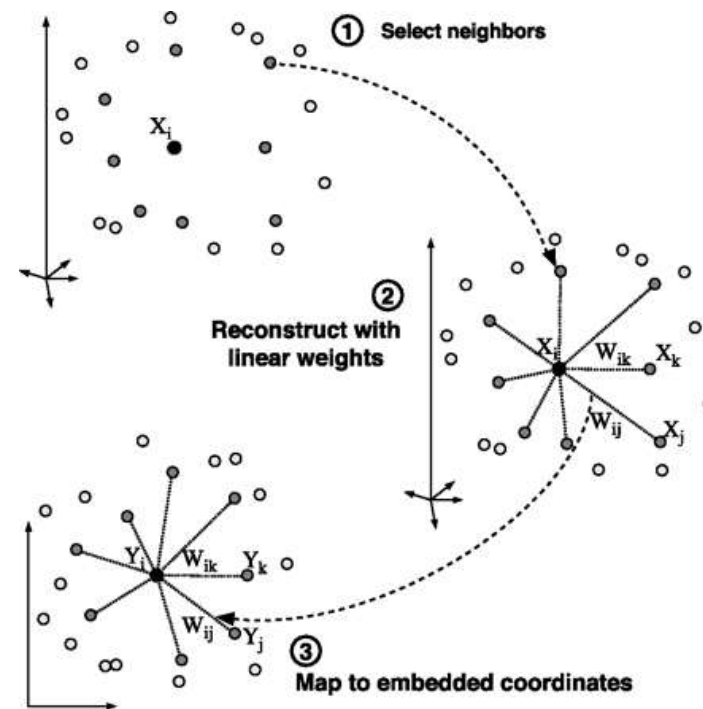
- (1) 对每个点  $x_i$ , 收集它的最近邻  $m$  个实例;
- (2) 求局部线性的权值  $W(i, j)$

$$\arg \min_w E_w = \sum_{i=1}^n \left\| x_i - \sum_{j=1}^k w(i, j) x_{ij} \right\|^2, \quad \sum_{j=1}^k w(i, j) = 1$$

- (3) 根据求得的权值, 计算更新的实例

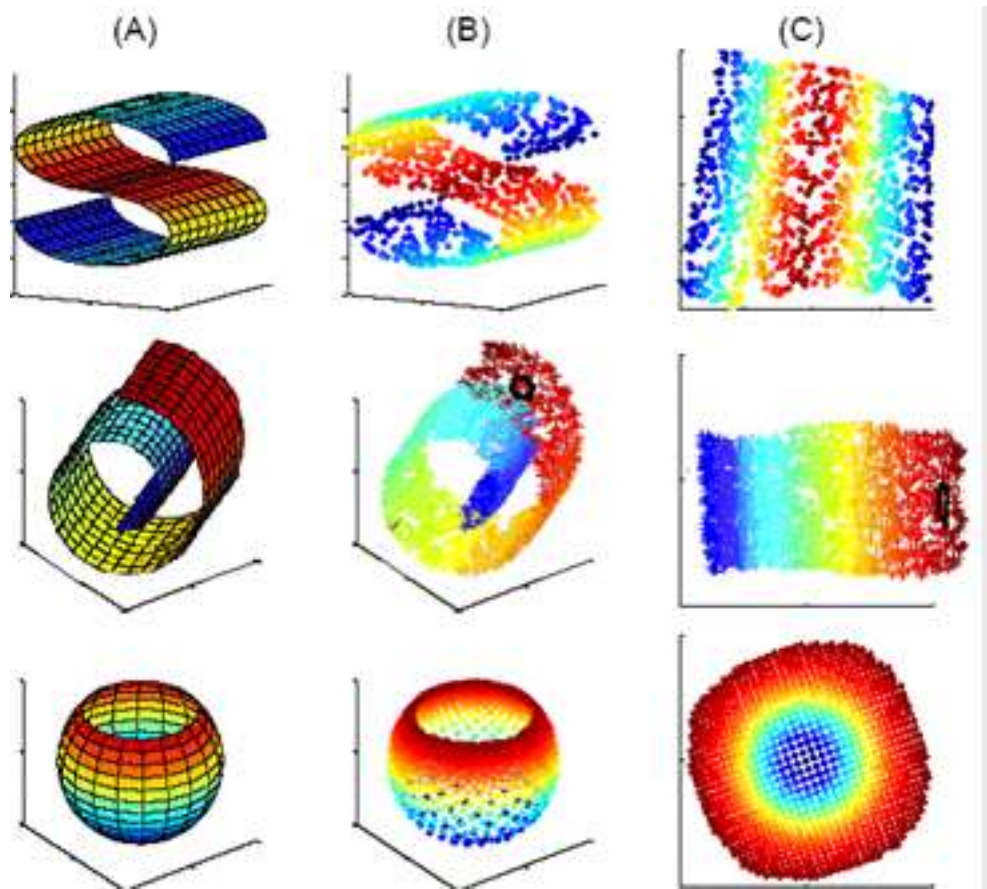
$$\arg \min_Y E_Y = \sum_{i=1}^n \left\| y_i - \sum_{j=1}^k w(i, j) y_j \right\|^2$$

- Performing an eigendecomposition of the matrix  $(I - W)^T(I - W)$ .
- Discarding the eigenvector that corresponds to the smallest eigenvalue.
- Taking the eigenvectors that correspond to the next (lower) eigenvalues. These yield the low-dimensional outputs  $y_i$ ,  $i = 1, 2, \dots, n$ .



# 特征提取

## 局部线性变换 (LLE)

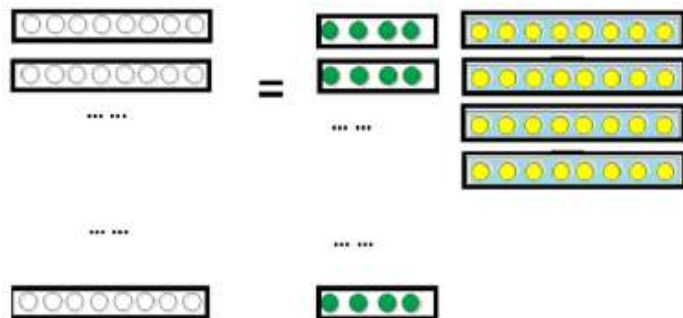


# 特征提取

## 非负矩阵分解

**基本思想：**通过矩阵分解，进行数据降维；分解后的矩阵为非负矩阵

非负矩阵分解：  $X = L \cdot R$



# 特征提取

## 非负矩阵分解

### 不同的目标函数情况

#### (1) 范数误差最小

$$\begin{aligned} \text{minimize} \quad & \|V - WH\|_F \equiv \sum_{i=1}^l \sum_{j=1}^n (V(i, j) - [WH](i, j))^2 \\ \text{subject to} \quad & W(i, k) \geq 0, H(k, j) \geq 0 \quad \forall i, k, j \end{aligned}$$

#### 迭代规则

$$\begin{aligned} H_{au} &\leftarrow H_{au} \frac{(W^T V)_{au}}{(W^T W H)_{au}} \\ W_{ia} &\leftarrow W_{ia} \frac{(V H^T)_{ia}}{(W H H^T)_{ia}} \end{aligned}$$

# 特征提取

## 非负矩阵分解

### 不同的目标函数情况

#### (2) K-L 误差

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^l \sum_{j=1}^n \left( V(i, j) \ln \frac{V(i, j)}{[WH](i, j)} - V(i, j) + [WH](i, j) \right) \\ \text{subject to} \quad & W(i, k) \geq 0, H(k, j) \geq 0 \quad \forall i, k, j \end{aligned}$$

#### 迭代规则

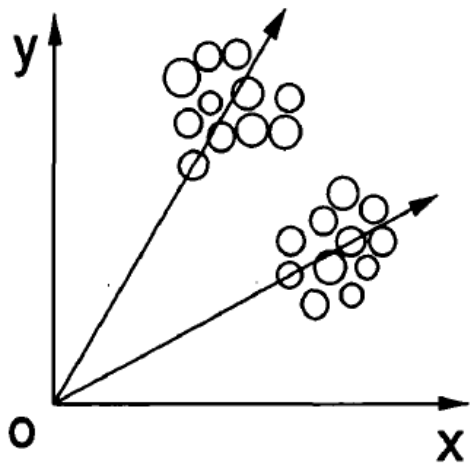
$$\begin{aligned} H_{au} &\leftarrow H_{au} \frac{\sum_i W_{ia} V_{iu} / (WH)_{iu}}{\sum_k W_{ka}} \\ W_{ia} &\leftarrow W_{ia} \frac{\sum_u W_{au} V_{iu} / (WH)_{iu}}{\sum_l H_{al}} \end{aligned}$$

# 特征提取

## 非负矩阵分解

### 几何直观

由非负矩阵分解得到的基向量：



# 小 结

1. 掌握特征选择的基本框架

2. 掌握三种特征选择方法

**Filer, Wrapper, Embedded**

3. 掌握经典的 PCA 特征提取方法;

## 参考文献

1. 周志华，机器学习，清华大学出版社，2016.
2. Duda, R.O. et al. Pattern classification. 2nd, 2003.
3. 边肇祺，张学工等编著，模式识别(第二版)，清华大学，1999。
4. Chris Bishop. Pattern recognition and Machine Learning. Springer, 2006. (PR&ML)