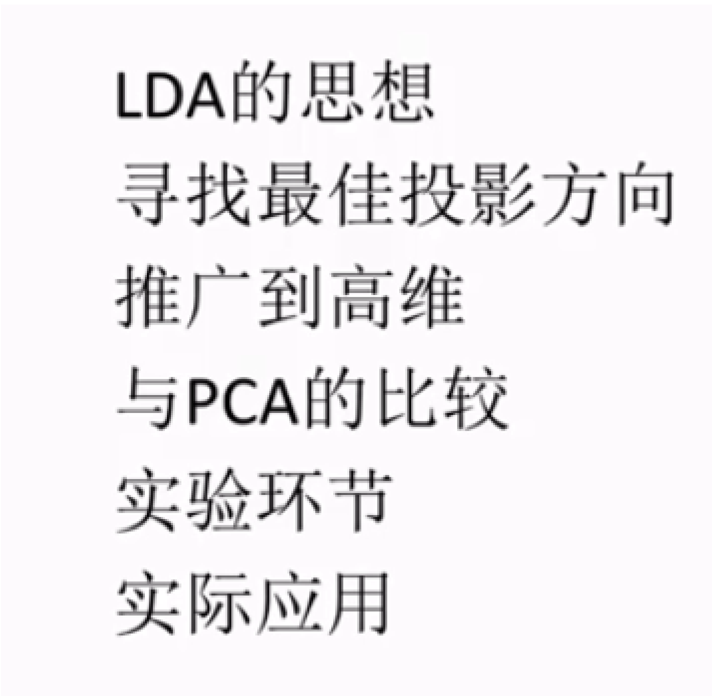
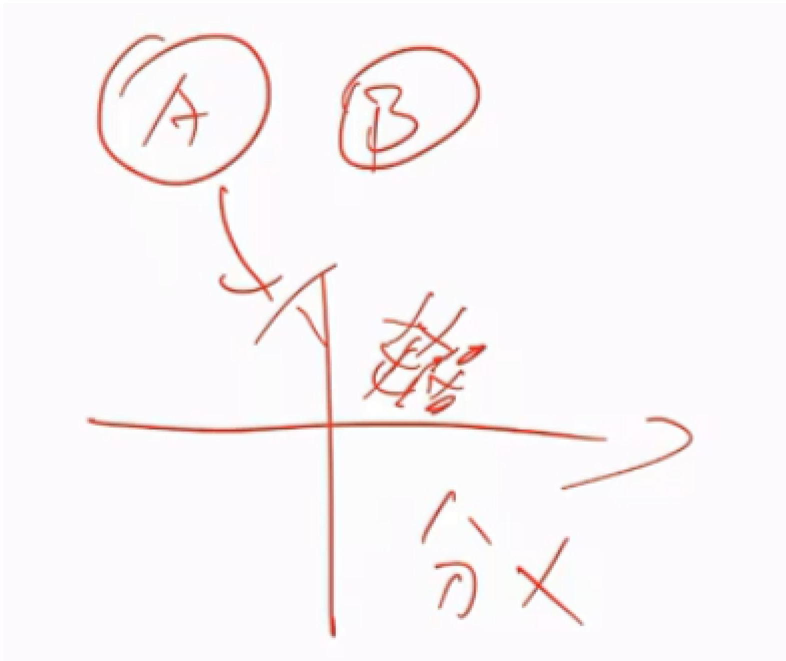
# 10 线性判别分析



# 10.1 本集内容简介





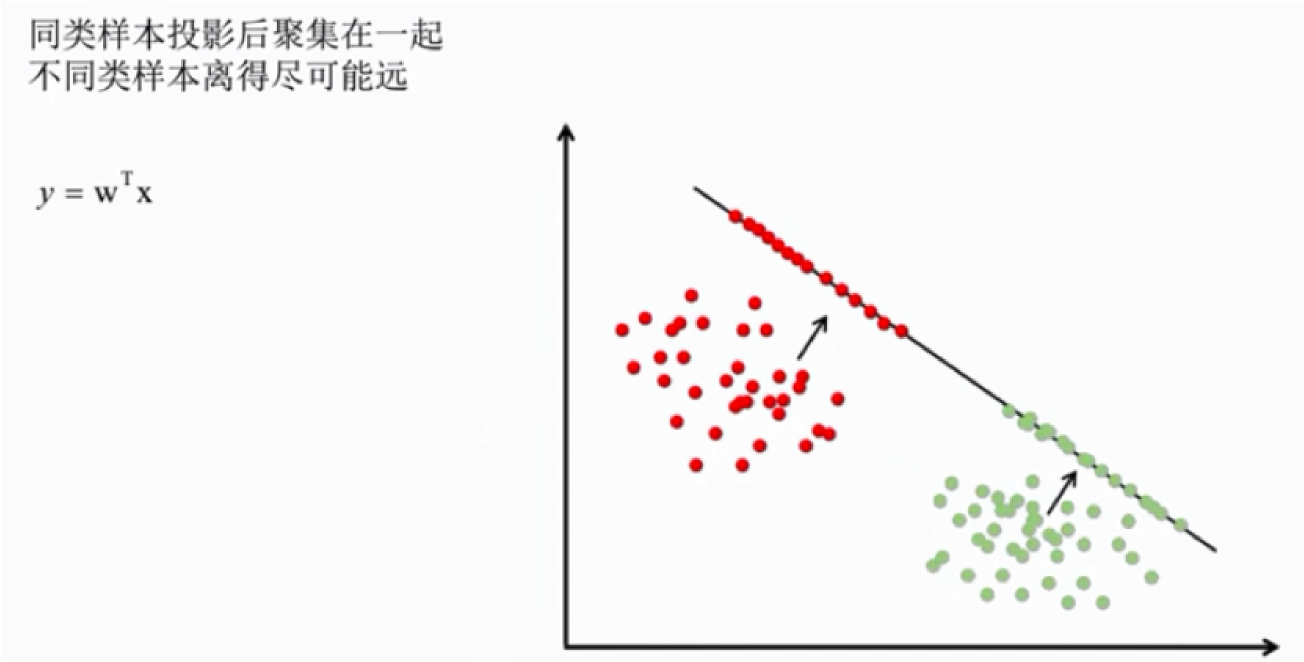

# 10.2 LDA的基本思想

流行学习和监督学习

阐述思想

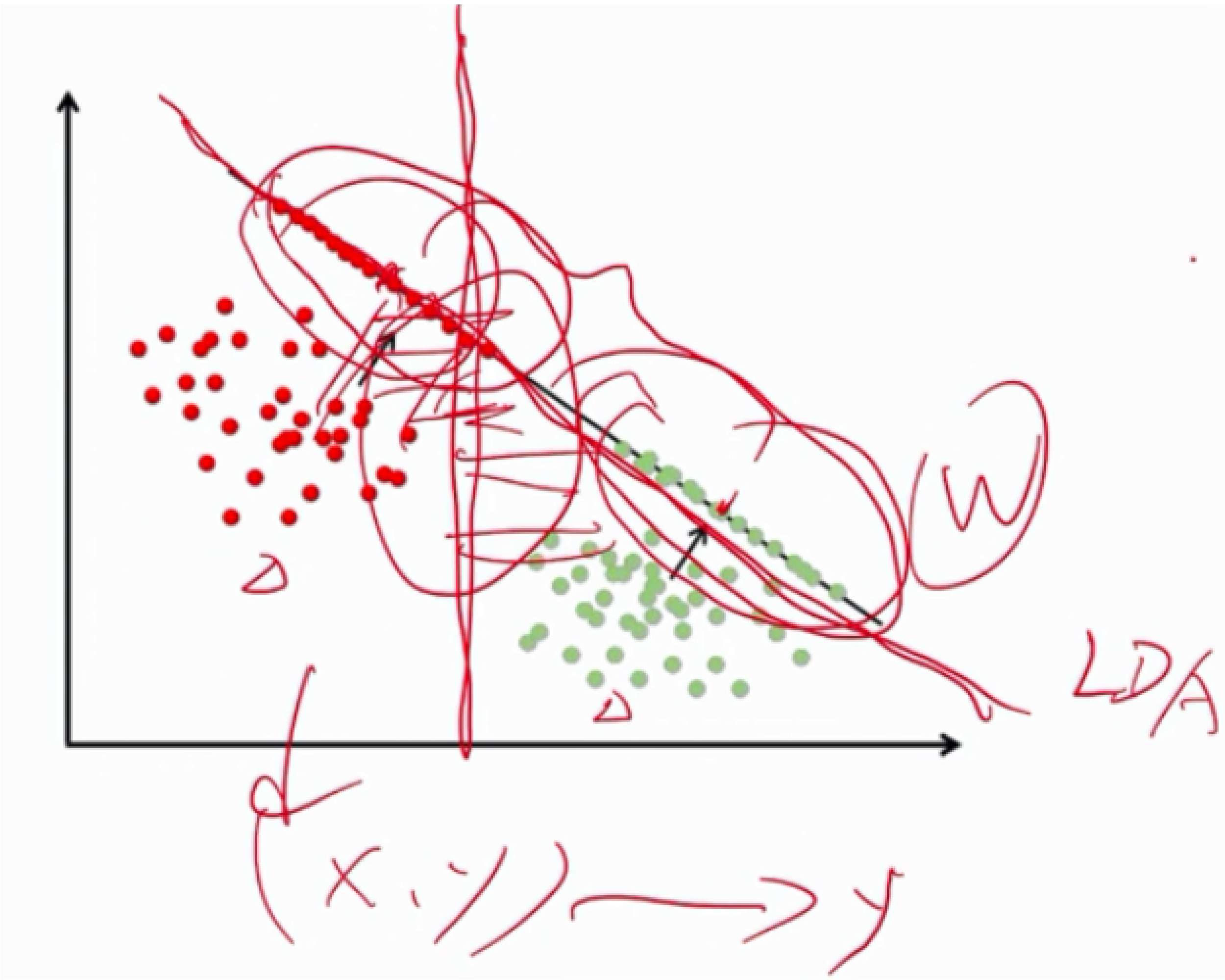
```
主成分分析，流形降维算法都是无监督学习，整个计算过程中没有利用样本标签值，它们投影的结果对分类未必有利
能不能有一种算法，对数据投影之后，能够更便于分类？

分类要达成的目标
区分不同的类，同一类的样本特征要尽量相似；不同类之间要尽可能不同
线性判别分析就是达成这种目标的一种线性算法

最大化类间差异，最小化类内差异
```

投影到一维空间，在图中，类间的差异最大化了。




两个类间的距离被最大化了。

图中竖线的投影效果就不好。类之间有重叠。



# 10.3 寻找最佳投影方向

问题的关键是如何得到最佳投影矩阵
首先考虑一维的情况

整个样本集 $\mathbf{x}_i$ $n$

分属两个类 $C_1$ $D_1$ $n_1$ $C_2$ $D_2$

向一维空间投影 $y = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

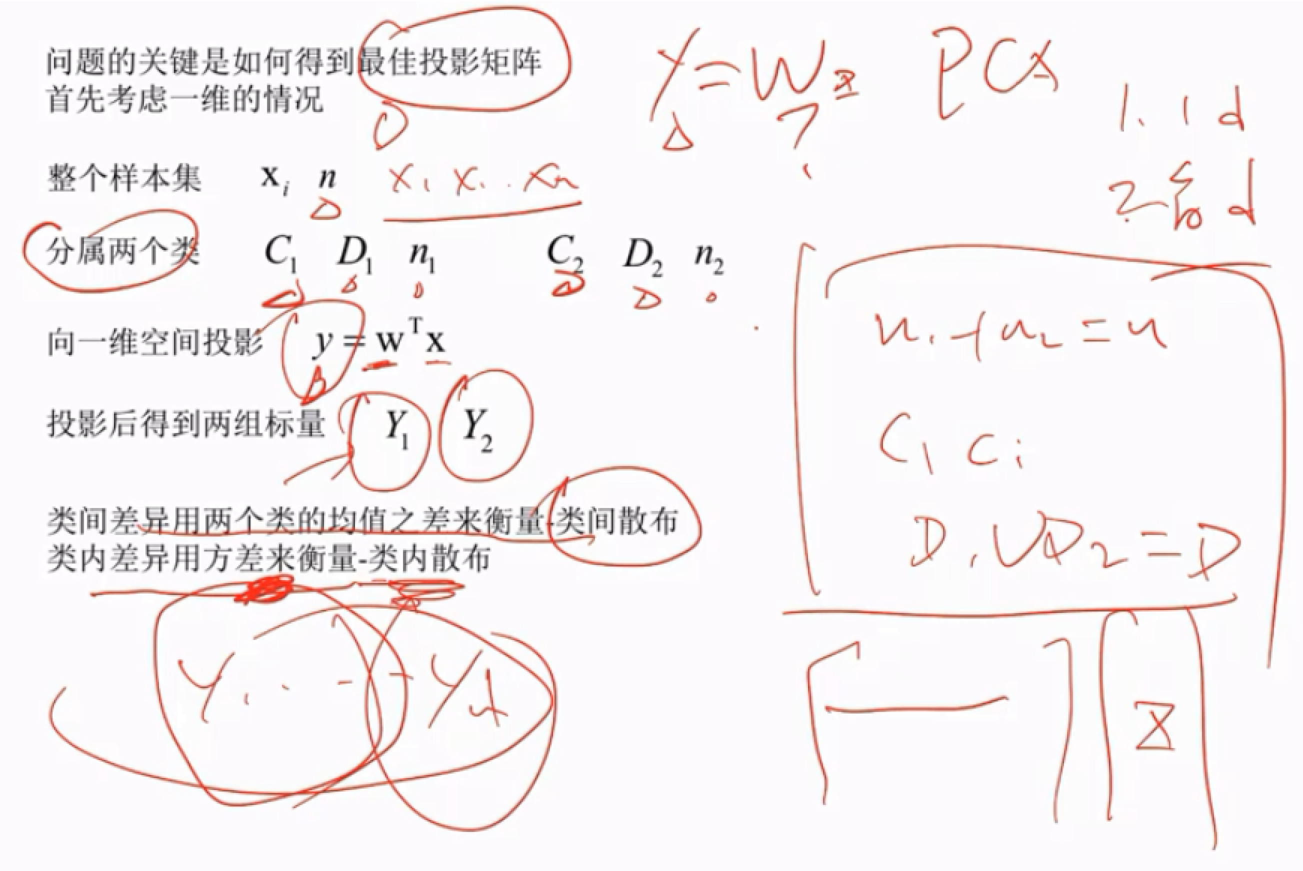
投影后得到两组标量 $Y_1$ $Y_2$

类间差异用两个类的均值之差来衡量-类间散布
类内差异用方差来衡量-类内散布

- 均值
- 方差

分两步进行，

先一维，再到多维。


问题的关键是如何得到最佳投影矩阵
首先考虑一维的情况
整个样本集 $\mathbf{x}_i$ $n$
分属两个类 $C_1$ $D_1$ $n_1$ $C_2$ $D_2$ $n_2$
向一维空间投影 $y = \mathbf{w}^T\mathbf{x}$
投影后得到两组标量 $Y_1$ $Y_2$
类间差异用两个类的均值之差来衡量-类间散布
类内差异用方差来衡量-类内散布
PCA

## 类间散步

类间散布

$$\mathrm{m}_i = \frac{1}{n_i}\sum_{\mathrm{x}\in D_i}\mathrm{x}$$

$$\widetilde{\mathrm{m}}_i = \frac{1}{n_i}\sum_{\mathrm{x}\in D_i}\mathrm{w}^T\mathrm{x} = \mathrm{w}^T\mathrm{m}_i$$

$$\left|\widetilde{\mathrm{m}}_1 - \widetilde{\mathrm{m}}_2\right| = \left|\mathrm{w}^T(\mathrm{m}_1 - \mathrm{m}_2)\right|$$

类内散布

$$\tilde{s}_i^2 = \sum_{y\in Y_i}\left(y - \tilde{m}_i\right)^2$$

总类内散布

$$\tilde{s}_1^2 + \tilde{s}_2^2$$

优化目标

$$L(\mathrm{w}) = \frac{\left(\widetilde{\mathrm{m}}_1 - \widetilde{\mathrm{m}}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

投影到

## 类内散步

类内散布

$$\tilde{s}_i^2 = \sum_{y\in Y_i}\left(y - \tilde{m}_i\right)^2$$

总类内散布

$$\tilde{s}_1^2 + \tilde{s}_2^2$$

优化目标

$$L(\mathrm{w}) = \frac{\left(\widetilde{\mathrm{m}}_1 - \widetilde{\mathrm{m}}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

用方差来描述，但是没有取平均。

## 类间散步

类内散布矩阵

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T$$

$$S_W = S_1 + S_2$$

$$\tilde{s}_i^2 = \sum_{x \in D_i} \left(w^T x - w^T m_i\right)^2$$

$$= \sum_{x \in D_i} w^T (x - m_i)(x - m_i)^T w$$

$$= w^T S_i w$$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_W w$$

两个类的散步矩阵之和

## 类内散步矩阵

类间散步矩阵公式

$$\left(\tilde{m}_1 - \tilde{m}_2\right)^2 = \left(w^T (m_1 - m_2)\right)^2 = w^T (m_1 - m_2)(m_1 - m_2)^T w$$

$$S_B = (m_1 - m_2)(m_1 - m_2)^T$$

$$\left(\tilde{m}_1 - \tilde{m}_2\right)^2 = w^T S_B w$$

优化目标变为

$$L(w) = \frac{\left(\tilde{m}_1 - \tilde{m}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \qquad L(w) = \frac{w^T S_B w}{w^T S_W w}$$

## 优化目标有冗余

优化目标有冗余

$$L(\mathrm{w})=\frac{\mathrm{w}^{\mathrm{T}}\mathrm{S}_B\mathrm{w}}{\mathrm{w}^{\mathrm{T}}\mathrm{S}_W\mathrm{w}}$$

为了消掉冗余，加上一个约束条件

$$\mathrm{w}^{\mathrm{T}}\mathrm{S}_W\mathrm{w}=1$$

优化的目标变为

$$\max\ \mathrm{w}^{\mathrm{T}}\mathrm{S}_B\mathrm{w}$$
$$\mathrm{w}^{\mathrm{T}}\mathrm{S}_W\mathrm{w}=1$$

优化目标函数用拉格朗日乘数法来求解。

用拉格朗日乘数法求解

$$L=\mathrm{w}^{\mathrm{T}}\mathrm{S}_B\mathrm{w}+\lambda\left(\mathrm{w}^{\mathrm{T}}\mathrm{S}_W\mathrm{w}-1\right)$$

$$\mathrm{S}_B\mathrm{w}+\lambda\mathrm{S}_W\mathrm{w}=0$$

$$\mathrm{S}_B\mathrm{w}=\lambda\mathrm{S}_W\mathrm{w} \qquad \mathrm{S}_W^{-1}\mathrm{S}_B\mathrm{w}=\lambda\mathrm{w}$$

对lambda 和 w 求导

$$\nabla_x\left(x^T A x\right) = 2Ax$$

类间散布矩阵的秩小于等于1 $r(S_B) \le 1$

$$S_W^{-1} S_B w = \lambda w$$

上面这个矩阵的秩也小于等于1

## 10.4 推广到高维和多类的情况

推广到高维

$$y = W^T x$$

类内散布矩阵

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T$$

$$S_W = \sum_{i=1}^{c} S_i$$

类间散布矩阵

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T$$

优化目标

$$\max L(W) = \frac{\mathrm{tr}\left(W^T S_B W\right)}{\mathrm{tr}\left(W^T S_W W\right)}$$

$$S_B w_i = \lambda_i S_W w_i$$

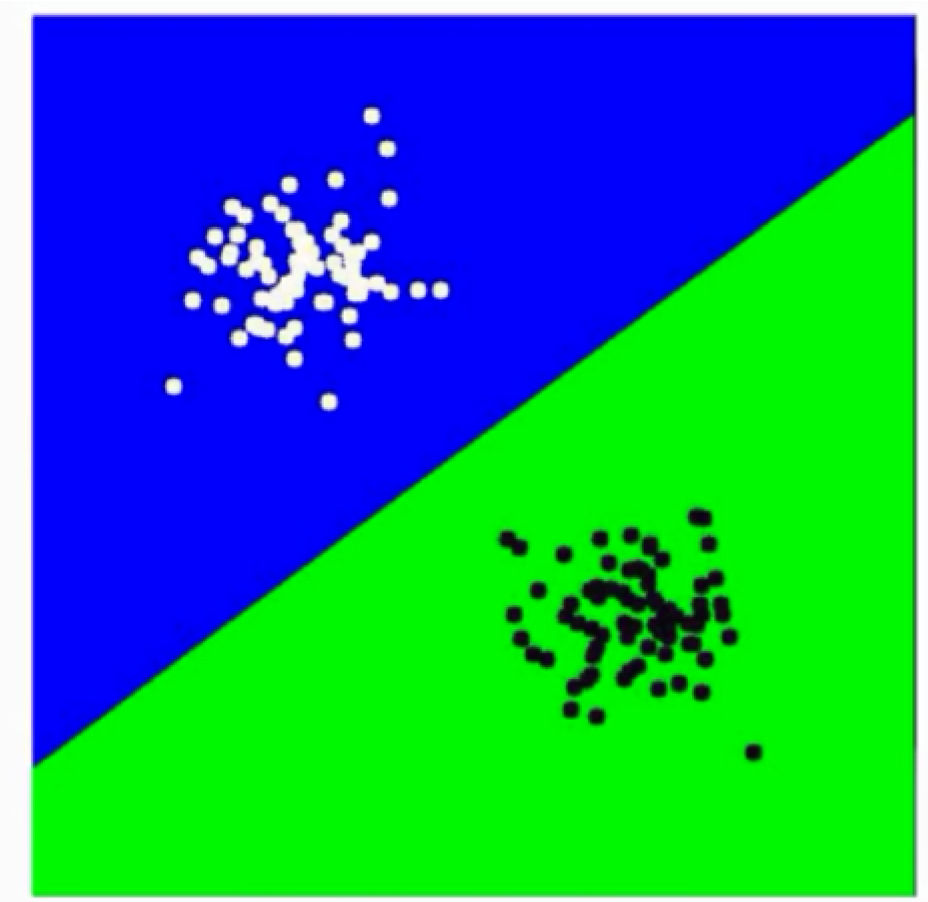
# 10.5 PCA与LDA的比较

与PCA的比较
二者都是求解特征值问题
LDA是有监督学习，PCA是无监督学习
投影的目标不同，LDA是最大化类间差异，最小化类内差异；PCA是最小化重构误差

面临的问题
内类散布矩阵不可逆

# 10.6 实验环节

云端实验室实验运行结果



黑色是一类，白色是一类。找到了最佳决策边界，验证了LDA是线性的。

代码通过c++实现


# 10.7 实际应用

人脸识别

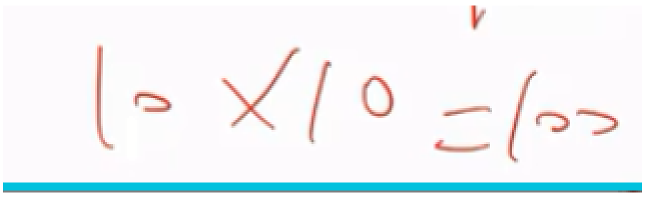
[1] Matthew Turk,Alex Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991.

[2] Peter N Belhumeur J P Hespanha David Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.

LDA一般先做降维，作用其他的算法做分类。

## 用PCA和LDA作人脸识别


假设是一个32*32的矩阵，放到一维向量1024,。



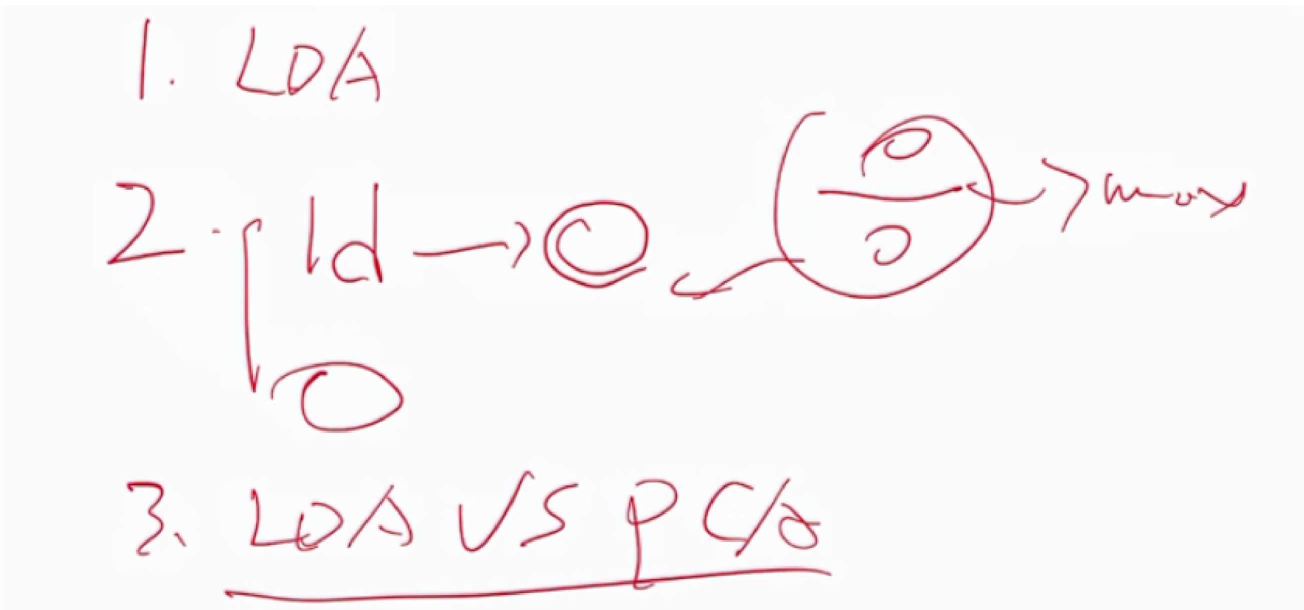

每一行是一个特征点。

先通过PCA降维，再通过LDA降维，

回顾本章，主要讲的是LDA算法，重点在寻找最佳投影方向，这部分用的公式比较多。


##

## 10.8 本集总结




回顾本章，主要讲的是LDA算法，重点在寻找最佳投影方向，这部分用的公式比较多。

要记住，LDA最大化类间差异，最小化类内差异。PCA是最小化重构误差。