# 线性判别分析（LDA）

- Linear Discriminant Analysis

  - 用途：数据预处理中的降维，分类任务

  - 历史：Ronald A. Fisher在1936年提出了线性判别方法

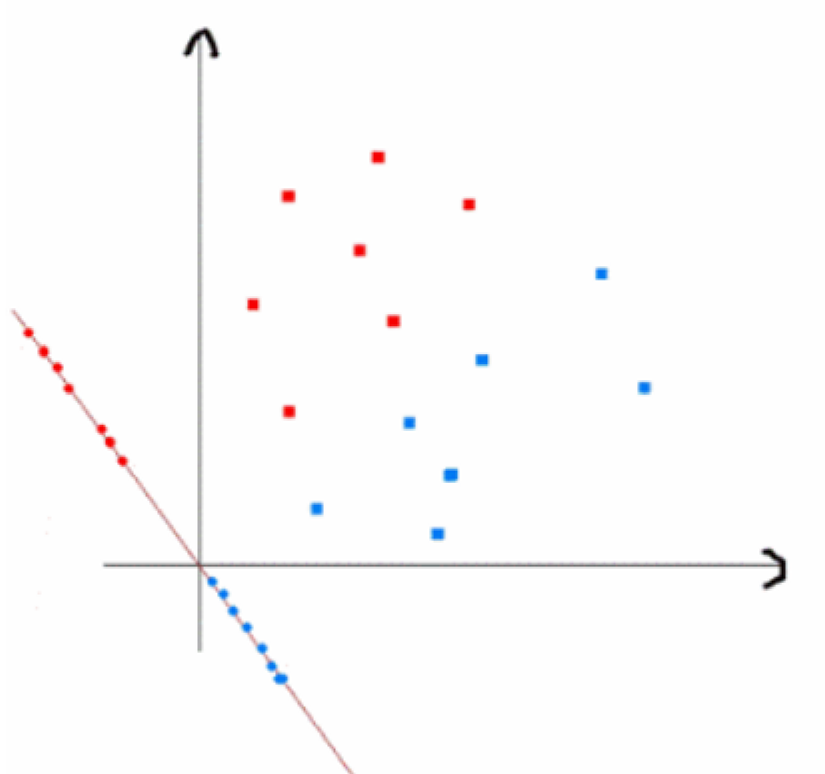
  - 目标：LDA关心的是能够最大化类间区分度的坐标轴成分

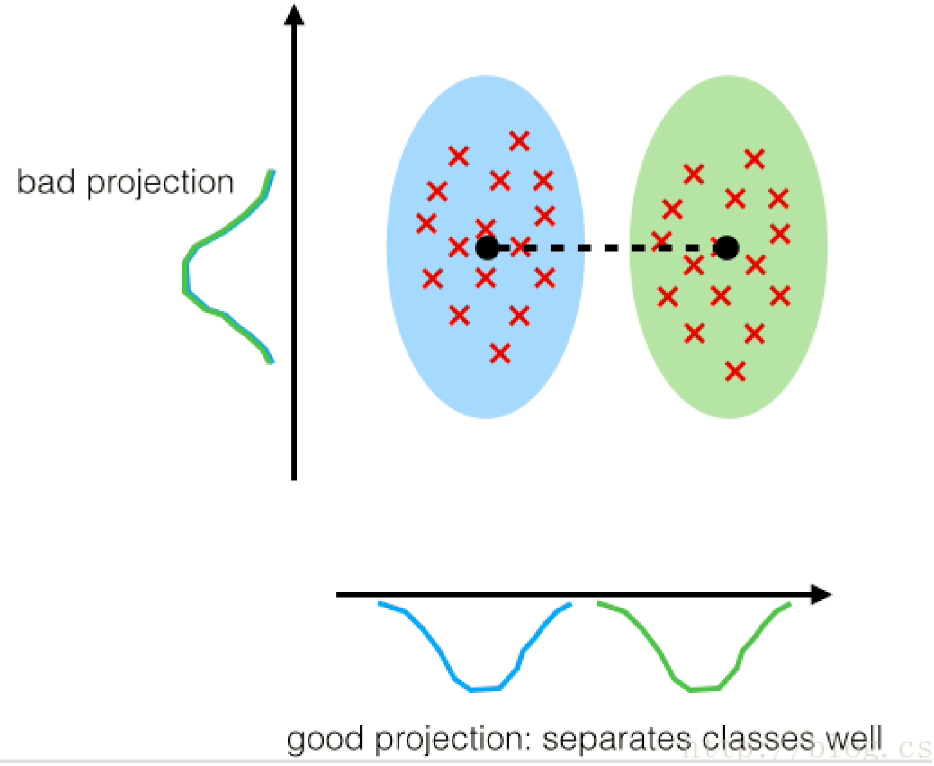
    将特征空间（数据集中的多维样本）投影到一个维度更小的 k 维子空间中，同时保持区分类别的信息

# 线性判别分析（LDA）

- Linear Discriminant Analysis

原理：投影到维度更低的空间中，使得投影后的点，会形成按类别区分，一簇一簇的情况，相同类别的点，将会在投影后的空间中更接近方法





# 线性判别分析（LDA）

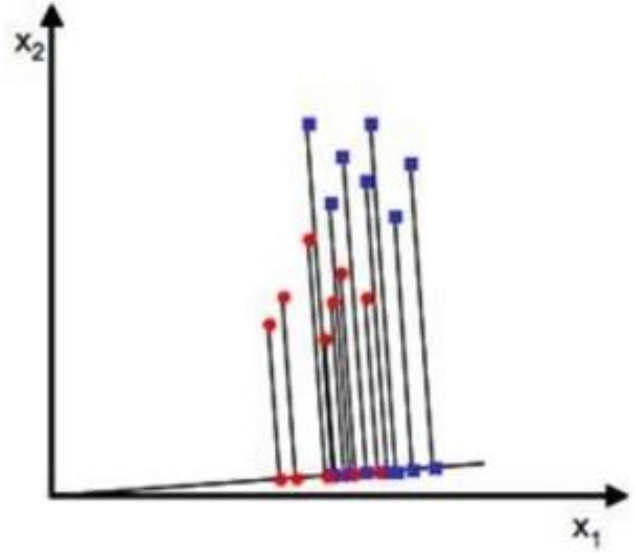
- Linear Discriminant Analysis

  - 监督性：LDA是“有监督”的，它计算的是另一类特定的方向
  - 投影：找到更合适分类的空间
  - 与PCA不同，更关心分类而不是方差

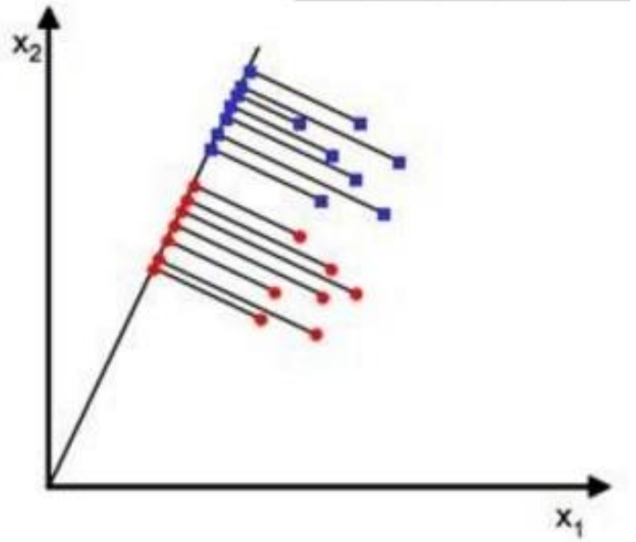
# 线性判别分析（LDA）

✔ 数学原理

原始数据：




变换数据：




目标：找到该投影 $y = w^T x$

# 线性判别分析（LDA）

- Linear Discriminant Analysis

- LDA分类的一个目标是使得不同类别之间的距离越远越好，同一类别之中的距离越近越好

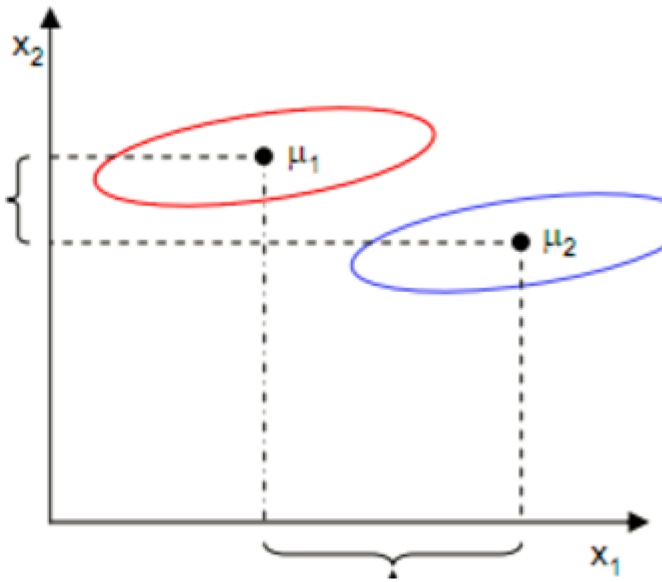
- 每类样例的均值：$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x$

- 投影后的均值：$\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in \omega_i} y = \frac{1}{N_i} \sum_{x \in \omega_i} w^T x = w^T \mu_i$

- 投影后的两类样本中心点尽量分离：$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T(\mu_1 - \mu_2)|$

# 线性判别分析（LDA）

- Linear Discriminant Analysis
  - 只最大化J(w)就可以了？
  - X1的方向可以最大化J(w)，但是却分的不好
  - 散列值：样本点的密集程度，值越大，越分散，反之，越集中
  - 同类之间应该越密集些：$\boxed{\tilde{s}_i^2 = \sum_{y\in\omega_i} (y-\tilde{\mu}_i)^2}$ 映射后的值.



# 线性判别分析（LDA）

- Linear Discriminant Analysis

- 目标函数： $J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$

$\sum (w^T\vec{x} - w^T\vec{u_i})(w^T\vec{x} - w^T\vec{u_i})^T = \sum (w^T\vec{x} - w^T\vec{u_i})(\vec{x}^T w - \vec{u_i}^T w)$

- 散列值公式展开：$\tilde{s}_i^2 = \sum_{y\in\omega_i}(y - \tilde{\mu}_i)^2 = \sum_{x\in\omega_i}(w^T x - w^T \mu_i)^2 = \sum_{x\in\omega_i} w^T (x - \mu_i)(x - \mu_i)^T w$

- 散列矩阵（scatter matrices）： $S_i = \sum_{x\in\omega_i}(x - \mu_i)(x - \mu_i)^T$

- 类内散布矩阵 Sw = S1+S2： $\tilde{s}_i^2 = w^T S_i w$  $\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_w w$

# 线性判别分析（LDA）

- Linear Discriminant Analysis

目标函数： $J(w) = \dfrac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2}$

分子展开： $(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T\mu_1 - w^T\mu_2)^2 = w^T\underbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}_{S_B}w = w^T S_B w$

$S_B$称作类间散布矩阵

最终目标函数： $J(w) = \dfrac{w^T S_B w}{w^T S_w w}$

# 线性判别分析（LDA）

- Linear Discriminant Analysis

- 分母进行归一化：如果分子、分母是都可以取任意值的，那就会使得有无穷解，我们将分母限制为长度为1

- 拉格朗日乘子法：

$$c(w) = w^t S_B w - \lambda(w^t S_W w - 1)$$

$$\Rightarrow \frac{dc}{dw} = 2S_B w - 2\lambda S_W w = 0$$

$$\Rightarrow S_B w = \lambda S_W w$$

- 两边都乘以Sw的逆：$S_W^{-1} S_B w = \lambda w$ （w就是矩阵$S_W^{-1} S_B$的特征向量了）

# 主成分分析（PCA）

- Principal Component Analysis

  - 用途：降维中最常用的一种手段

  - 目标：提取最有价值的信息（基于方差）

  - 问题：降维后的数据的意义?

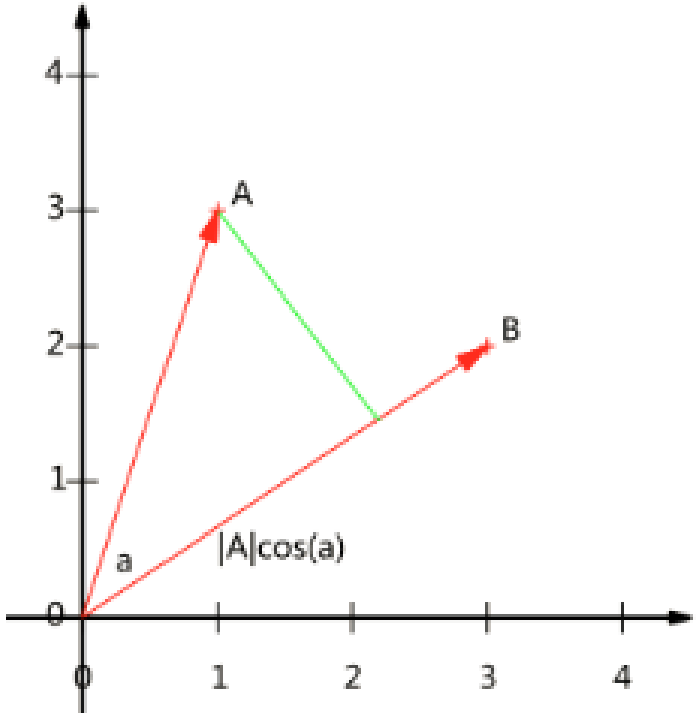
# 主成分分析（PCA）

- 向量的表示及基变换

内积： $(a_1, a_2, \cdots, a_n)^\mathsf{T} \cdot (b_1, b_2, \cdots, b_n)^\mathsf{T} = a_1b_1 + a_2b_2 + \cdots + a_nb_n$

解释： $A \cdot B = |A||B|cos(a)$

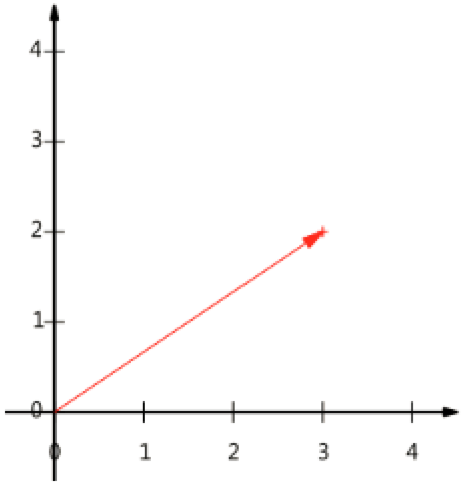
设向量B的模为1，则A与B的内积值等于A向B所在直线投影的矢量长度



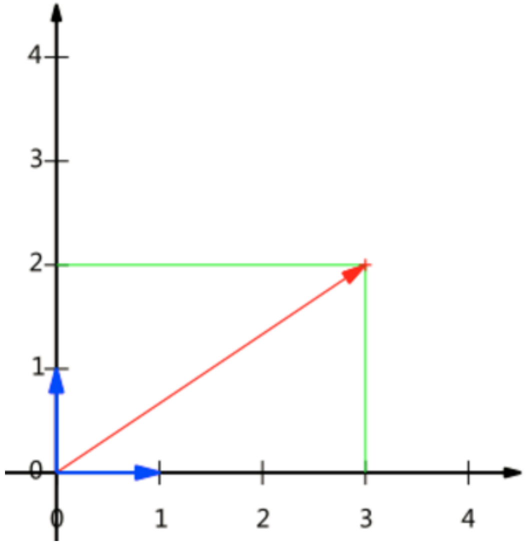
# 主成分分析（PCA）

向量的表示及基变换

向量可以表示为(3,2)
实际上表示线性组合：$x(1,0)^{\mathsf{T}}+y(0,1)^{\mathsf{T}}$




基：(1,0)和(0,1)叫做二维空间中的一组基

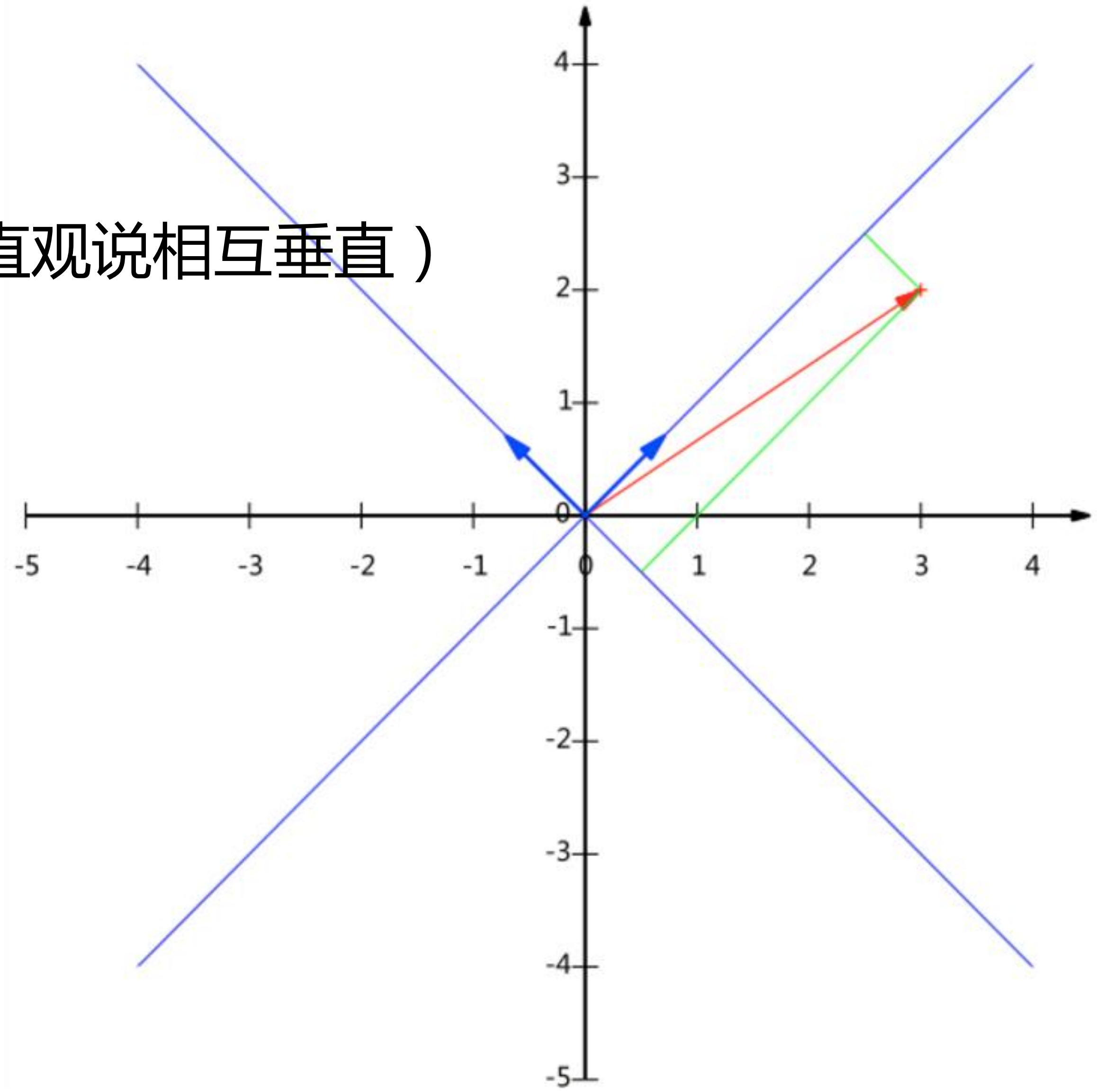


# 主成分分析（PCA）

- 基变换

  - 基是正交的（即内积为0，或直观说相互垂直）

  - 要求：线性无关



# 主成分分析（PCA）

## 基变换

变换：数据与一个基做内积运算，结果作为第一个新的坐标分量，然后与第二个基做内积运算，结果作为第二个新坐标的分量

数据（3，2）映射到基中坐标：$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 5/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

# 主成分分析（PCA）

## 基变换

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \cdots & a_M \end{pmatrix} = \begin{pmatrix} p_1a_1 & p_1a_2 & \cdots & p_1a_M \\ p_2a_1 & p_2a_2 & \cdots & p_2a_M \\ \vdots & \vdots & \ddots & \vdots \\ p_Ra_1 & p_Ra_2 & \cdots & p_Ra_M \end{pmatrix}$$

两个矩阵相乘的意义是将右边矩阵中的每一列列向量变换到左边矩阵中每一行行向量为基所表示的空间中去

# 主成分分析（PCA）

## 协方差矩阵

- 方向：如何选择这个方向（或者说基）才能尽量保留最多的原始信息呢？一种直观的看法是：希望投影后的投影值尽可能分散

- 方差：$Var(a) = \frac{1}{m}\sum_{i=1}^{m}(a_i - \mu)^2$

- 寻找一个一维基，使得所有数据变换为这个基上的坐标表示后，方差值最大

- 协方差（假设均值为0时）：$Cov(a,b) = \frac{1}{m}\sum_{i=1}^{m}a_i b_i$

# 主成分分析（PCA）

- 协方差

  - 如果单纯只选择方差最大的方向，后续方向应该会和方差最大的方向接近重合。

  - 解决方案：为了让两个字段尽可能表示更多的原始信息，我们是不希望它们之间存在（线性）相关性的

  - 协方差：可以用两个字段的协方差表示其相关性 $Cov(a,b)=\frac{1}{m}\sum_{i=1}^{m}a_ib_i$

  - 当协方差为0时，表示两个字段完全独立。为了让协方差为0，选择第二个基时只能在与第一个基正交的方向上选择。因此最终选择的两个方向一定是正交的。

# 主成分分析（PCA）

- 优化目标

  - 将一组N维向量降为K维（K大于0，小于N），目标是选择K个单位正交基，使原始数据变换到这组基上后，各字段两两间协方差为0，字段的方差则尽可能大

  - 协方差矩阵：$X = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{pmatrix}$ $\frac{1}{m}XX^\mathsf{T} = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^m a_i^2 & \frac{1}{m}\sum_{i=1}^m a_i b_i \\ \frac{1}{m}\sum_{i=1}^m a_i b_i & \frac{1}{m}\sum_{i=1}^m b_i^2 \end{pmatrix}$

  - 矩阵对角线上的两个元素分别是两个字段的方差，而其它元素是a和b的协方差。

# 主成分分析（PCA）

- 优化目标

  - 协方差矩阵对角化：即除对角线外的其它元素化为0，并且在对角线上将元素按大小从上到下排列

  - 协方差矩阵对角化：$PCP^{\mathsf{T}} = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$

# 主成分分析（PCA）

## 优化目标

- 实对称矩阵：一个n行n列的实对称矩阵一定可以找到n个单位正交特征向量

$$E = \begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix}$$

- 实对称阵可进行对角化：

$$E^{\mathsf{T}} C E = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

- 根据特征值的从大到小，将特征向量从上到下排列，则用前K行组成的矩阵乘以原始数据矩阵X，就得到了我们需要的降维后的数据矩阵Y

# 主成分分析（PCA）

## PCA实例

- 数据：$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$

- 协方差矩阵：$C = \frac{1}{5}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$

- 特征值：$\lambda_1 = 2, \lambda_2 = 2/5$　特征向量：$c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix}, c_2\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

- 对角化：$PCP^{\mathsf{T}} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 6/5 & 4/5 \\ 4/5 & 6/5 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2/5 \end{pmatrix}$

- 降维：$Y = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -3/\sqrt{2} & -1/\sqrt{2} & 0 & 3/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$