PCA：主成分分析

设有m条n维数据：

1. 将原始数据按列组成n行m列矩阵X
2. 将X的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值
3. 求出协方差矩阵$C=\frac{1}{m}XX^T$
4. 求出协方差矩阵的特征值及对应的特征向量
5. 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k（目标维数）行组成矩阵P
6. $Y=PX$即为降维到k维后的数据

实例：将数据$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$降到 1 维

（所给数据已经进行中心化处理，否则第一步应中心化，即$x_i=x_i-\frac{1}{n}\sum_{i=1}^{n}x_i$

计算协方差矩阵：$C=\underbrace{\frac{1}{m}}_{\text{这里有的书上是}\,m-1}XX^T=\frac{1}{5}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$

这里有的书上是$m-1$

计算特征值：$\left|C-\lambda E\right|=\begin{vmatrix} \frac{6}{5}-\lambda & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5}-\lambda \end{vmatrix}=0$　　　解得：$\lambda_1=\frac{2}{5},\lambda_2=2$

计算特征向量：属于特征值$\lambda_1=\frac{2}{5}$的一个特征向量为$P_1=(-1,1)^T$；属于特征值$\lambda_1=2$的一个特征向量为$P_2=(1,1)^T$

降维处理：$Y=\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}=\begin{pmatrix} -\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

选择最大的特征值对应的特征向量的单位向量，即$\lambda=2$对应的$P_2=(1,1)^T$的单位向量