假设我们有m个n维数据，映射至k维，则

最好的映射方式是，在每个基上方差最大，不同基上数据协方差为0。

这样就能保证不同基上的数据是没有相关性的，每个基上的

数据尽可能散开。

假设我们的数据矩阵 $X=\begin{pmatrix} x_{11} & x_{12} & \cdots \\ \vdots & \vdots & \\ x_{n1} & x_{n2} & \end{pmatrix}$

需要做的变换为 $Y=PX$，P是正交矩阵并且为 $k\times n$ 维

注意到 $XX^T=(x_1 \cdots x_m)_{n\times m}\begin{pmatrix} x_1^T \\ \vdots \\ x_m^T \end{pmatrix}_{m\times n}$

$$=\sum_{i=1}^{m} x_i x_i^T = \sum_{i=1}^{m}\begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}(x_{1i} \cdots x_{ni})$$

$$=M \quad M_{ij}=\sum_{k=1}^{m} x_{ik}x_{jk}$$ 若 $i=j$ 则 $M_{ij}=\sum_{k=1}^{m} x_{ik}^2$

若 x_{ik} 为0均值的 则 $\sum_{k=1}^{m} x_{ik}^2$ 实际上是m倍在第i个轴上的方差

设数据X的协方差矩阵为C，Y的协方差矩阵为D

则 $C=\frac{1}{m}XX^T$ 则 $D=\frac{1}{m}YY^T$

$$=\frac{1}{m}(PX)(PX)^T$$

(注：对称矩阵一定可 $$=\frac{1}{m}PXX^TP^T=P\left(\frac{XX^T}{m}\right)P^T$$

特征分解) $$=PCP^T$$

若想D为对角矩阵，则只需对C做正交分解即可

即P是C的特征向量组成的矩阵。

即 $P=($

PCA的关键在于基变换后的Y的协方差矩阵（区别于X的协方差矩阵）。且协方差矩阵 $=\frac{XX^T}{m}$，m为样本列数