pca

主成分分析（PCA）是一种统计方法，用于简化数据集的复杂性，同时保留其主要特征。PCA 的主要作用和应用过程可以分为以下几个步骤：

1. 标准化数据：首先，数据集中的每个特征（或变量）都需要标准化，以确保它们具有相同的比例。这通常涉及减去均值并除以标准差。
2. 计算协方差矩阵：然后，计算标准化数据的协方差矩阵。协方差矩阵捕获了不同变量间的相关性。
3. 计算特征值和特征向量：对协方差矩阵进行特征分解以获得其特征值和相应的特征向量。特征向量表示数据的主要方向，而特征值表示这些方向的重要性。
4. 选择主要成分：根据特征值的大小选择前几个最重要的特征向量。这些特征向量代表了数据的“主要成分”。
5. 变换到新的子空间：使用选定的主要成分将原始数据转换到新的子空间。这通常涉及将原始数据矩阵与选定的特征向量矩阵相乘。

PCA 的应用范围非常广泛，包括：

1. 数据降维：在机器学习和数据挖掘中，用于减少数据集的维数，同时保留最重要的信息。
2. 噪声过滤：消除数据中的噪声并提高数据质量。
3. 特征提取：在模式识别和信号处理中用于提取重要特征。
4. 数据可视化：将多维数据降维到二维或三维，以便于可视化和分析。

1. 标准化数据：$X_{std} = \frac{X-\mu}{\sigma}$
2. 计算协方差矩阵：$\mathrm{Cov}(X_{std}) = \frac{1}{n-1} X_{std}^T X_{std}$
3. 特征分解：$\mathrm{Cov}(X_{std}) = V \Lambda V^T$，其中，$V$ 是特征向量，Λ 是特征值对角矩阵。
4. 选择主要成分：基于特征值选择前 k 个特征向量
5. 数据转换：$X_{pca} = X_{std} V_k$

pca

	d_1	d_2	d_3
1:	1	2	3
2:	4	5	6
3:	7	8	9

每行为一个数据

算每列(维度d_1, d_2, d_3)的特征向量，比较选出较大的，之后将较小的维度去掉，假设去d_3

	d_1	d_2
1:	1	2
2:	4	5
3:	7	8

则数据变成2维