### PCA:

1. PCA介绍：主成分分析，降维算法
2. 降维作用:压缩和可视化
3. 优化目标:把所有数据都映射到该项向量(U)上，希望投射距离均方差(垂直的线)尽量的小(与线性回归中误差平方(竖线)和不是一个概念),而投射后的数据分布方差(数据离散程度)尽量的大(S)
4. 协方差计算:计算特征均值(列均值)-->减均值-->转置结果为x’-->1/n\*x’(x’T) ,左对角线相同，表示(方差)离散程度,右对角线表示相关性
5. 协方差矩阵的含义:值为正，表明两个特征正相关；值为负，两个特征负相关；值为0，两个特征不相关.
6. PCA的步骤(在PCA之前要做特征缩放):
7. 计算协方差矩阵
8. 通过SVD求解特征向量和特征值，使其满足AU=SU，S为特征值，U是特征向量.将原A变为除方差位置外(对角线),协方差位置都为0的矩阵.
9. 选取k个特征值较大的特征向量做特征压缩，Z=XUreduc(k)
10. 数据重建X’ new=ZUreduct.T
11. 不是解决过拟合的首选方案(过拟合原因特征过多，使用数据压缩会导致信息流失)
12. 选区降维列数两种方式:直接规定列个数n\_component(成分)=2;用降维后信息保存率0.99