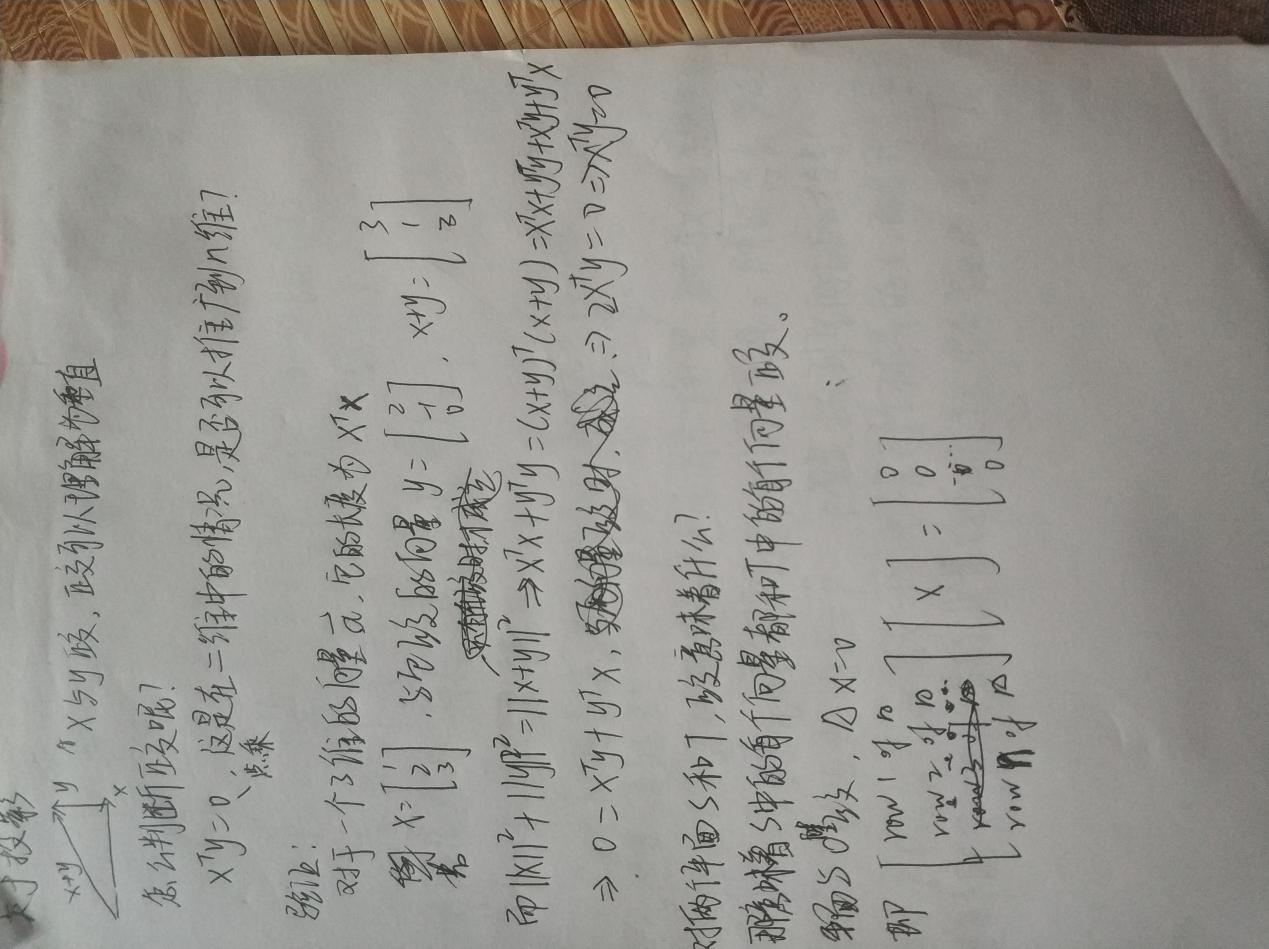
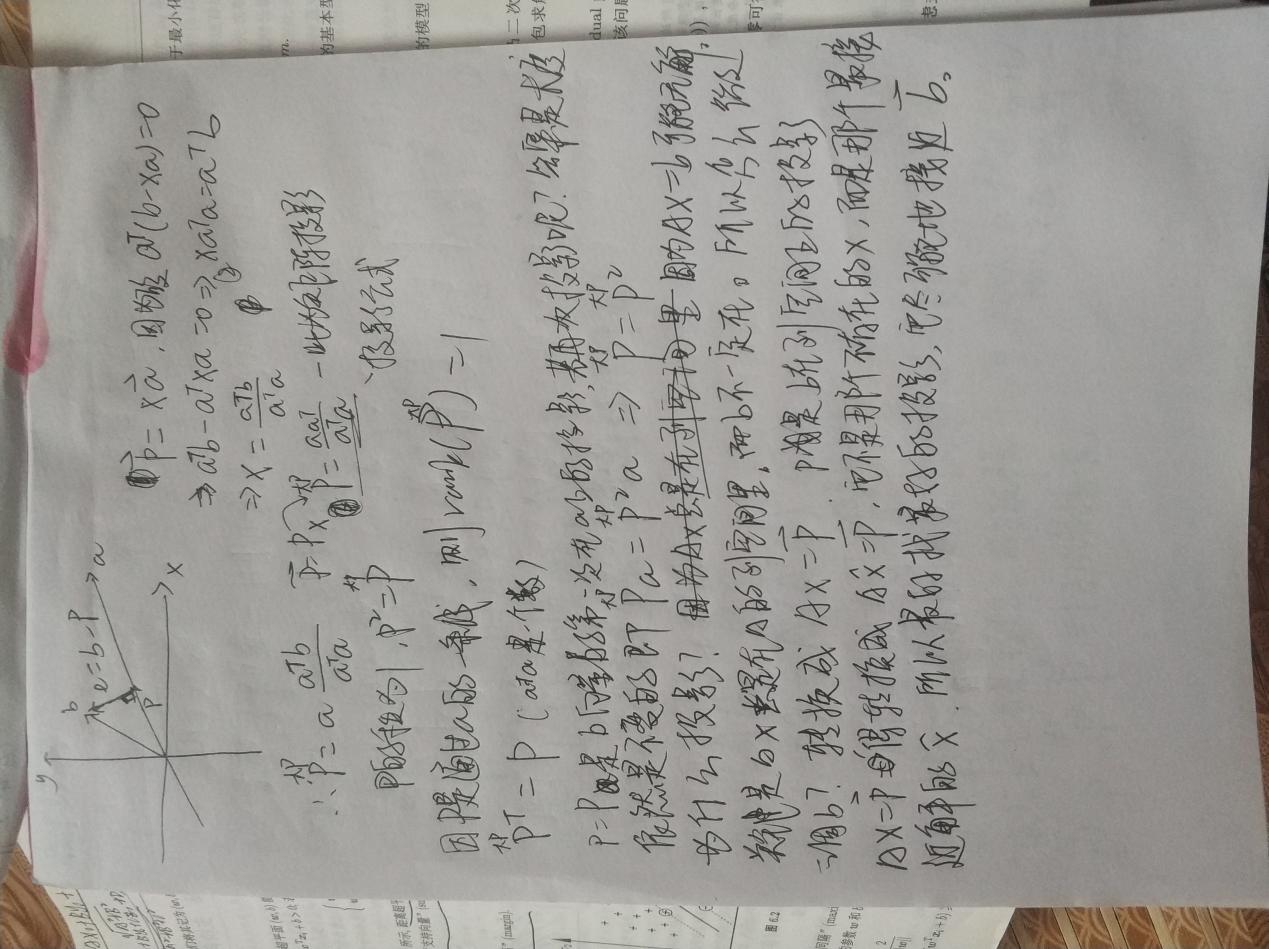
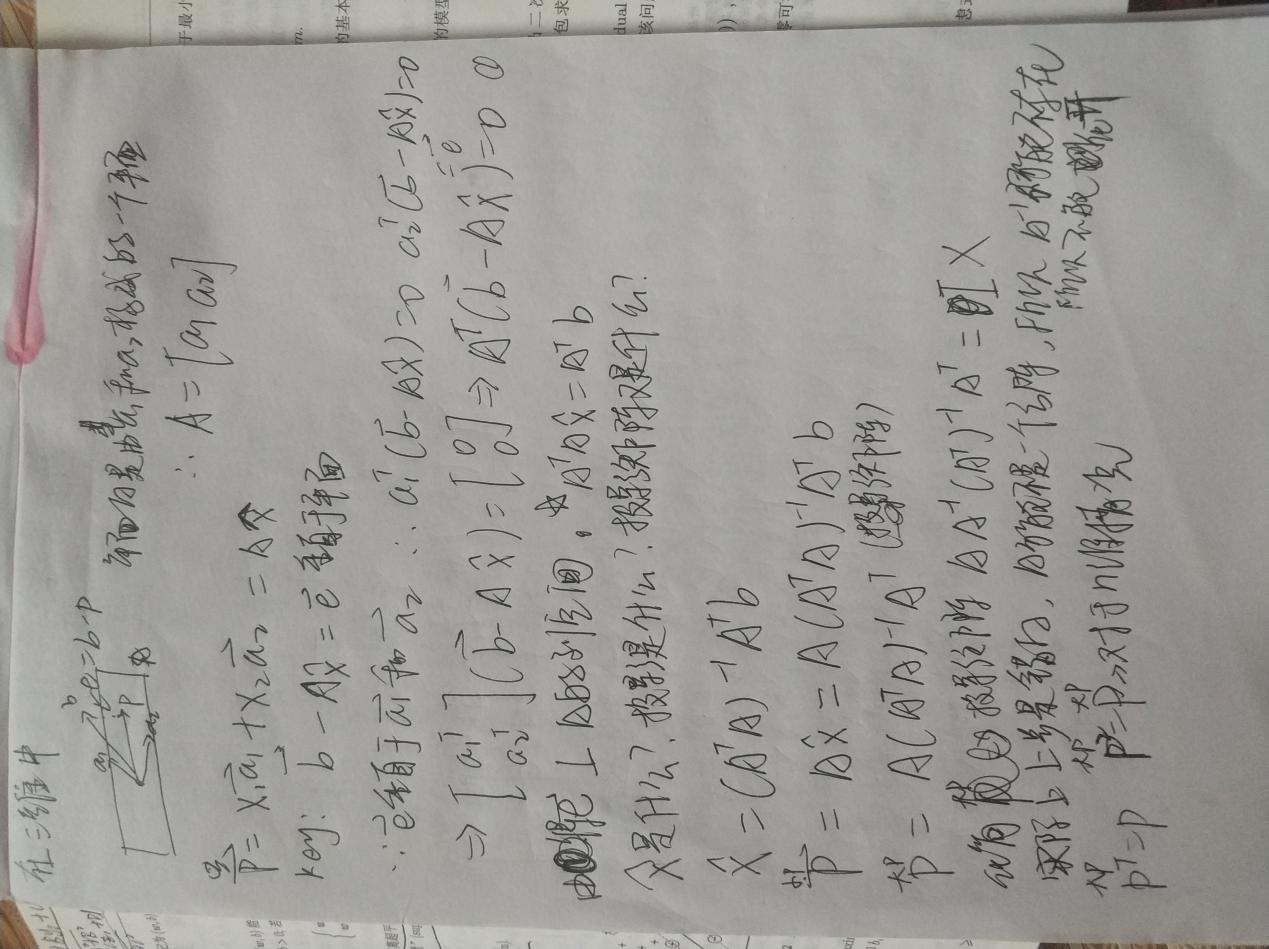
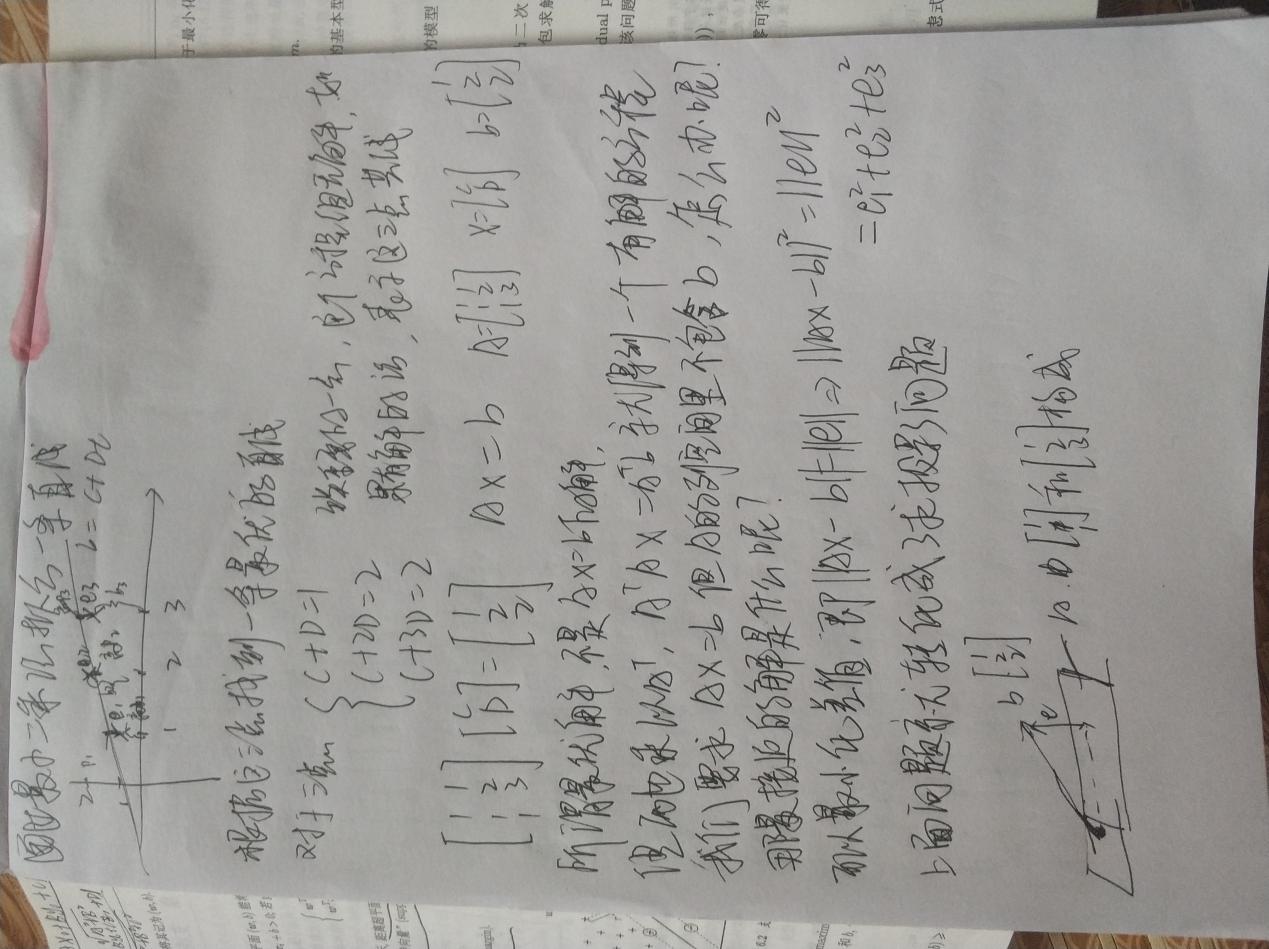
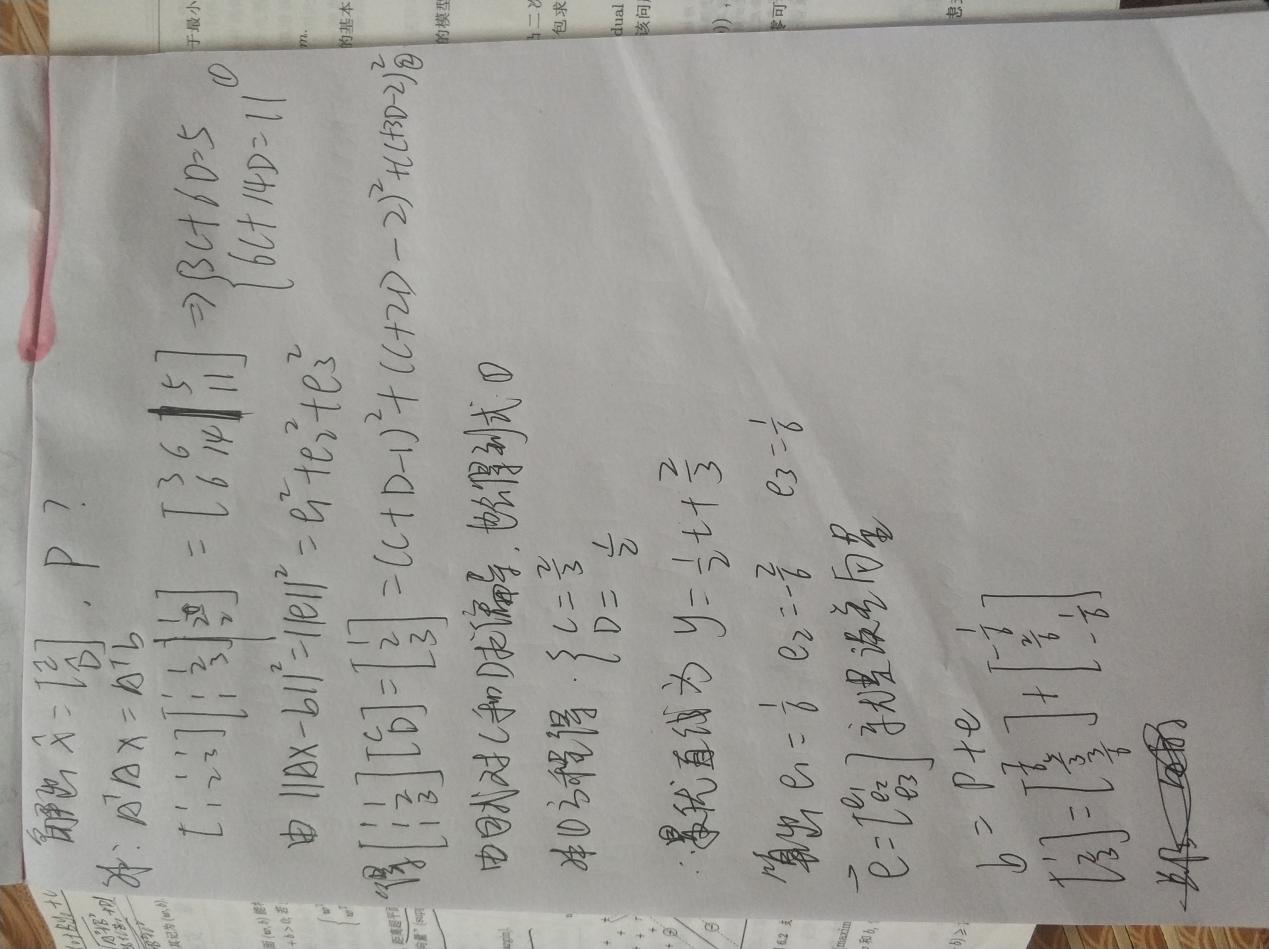
投影



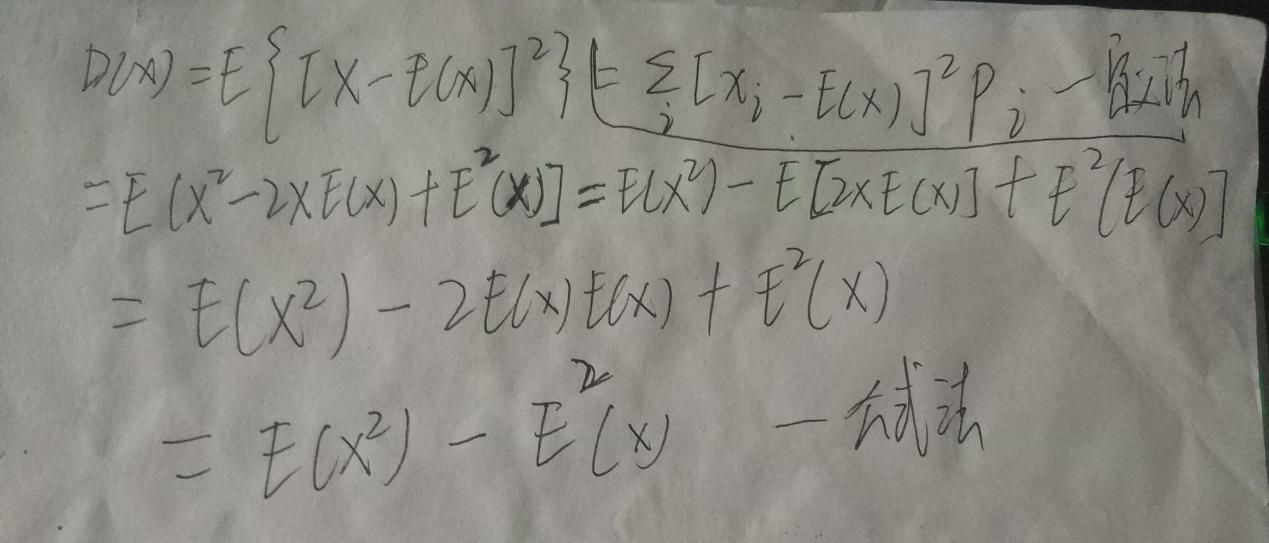








方差公式



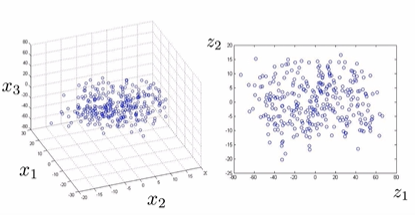
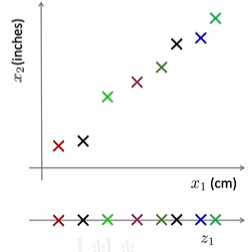
PCA的作用

主成分分析PCA的作用是：

1. 聚类：把复杂的多维数据点，简化为少量数据点，易于分簇
2. 降维：降低高维数据，简化计算，达到数据降维，压缩，降噪的目的

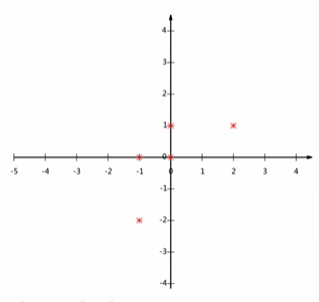
PCA的目的是：

1. 将原有的d维数据集，转换为k维的数据，k<d
2. 新生成的k维数据尽可能多的包含原来d维数据的信息



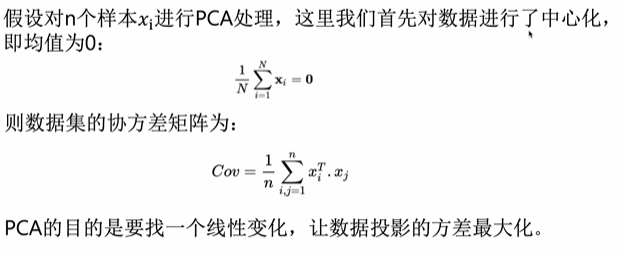
二维降到一维的问题，要在二维平面中选择一个方向，将所有数据都投影到这个方向所在直线上，用投影值表示原始记录。

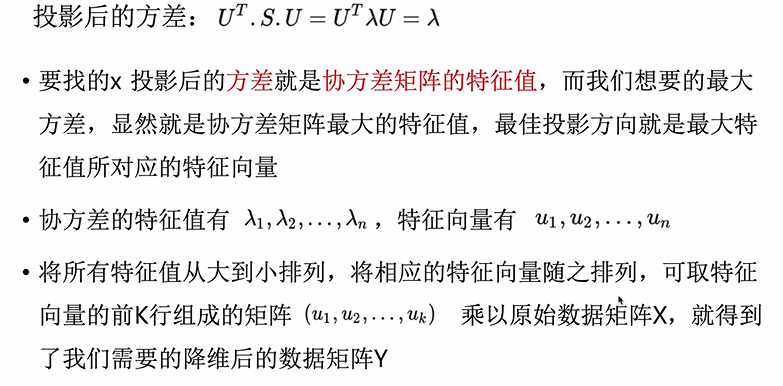
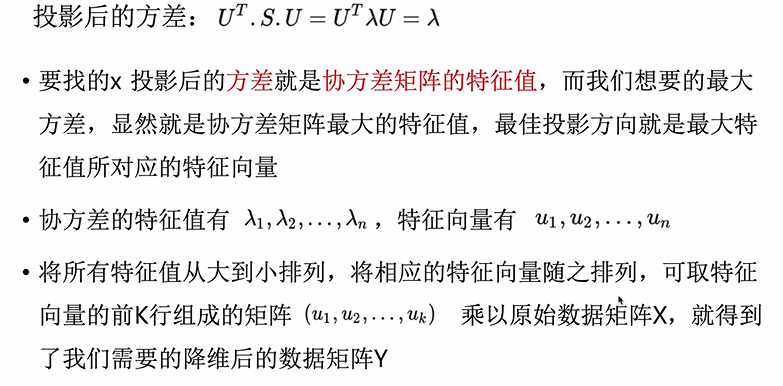
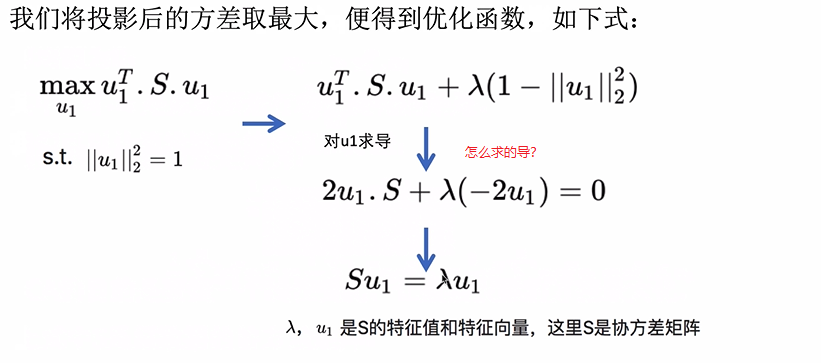
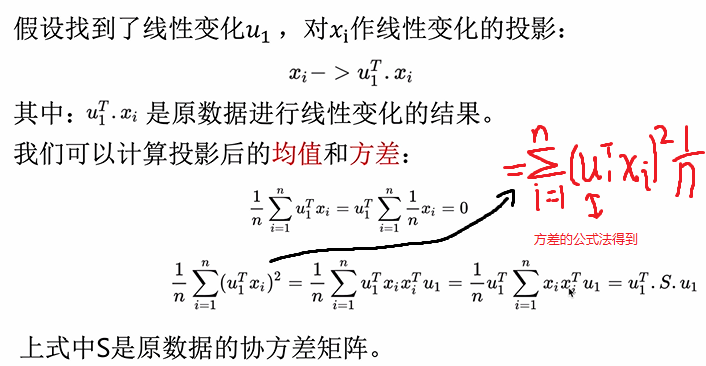
关键是如何选择这个方向，才可能尽量保留最多的原始信息呢？

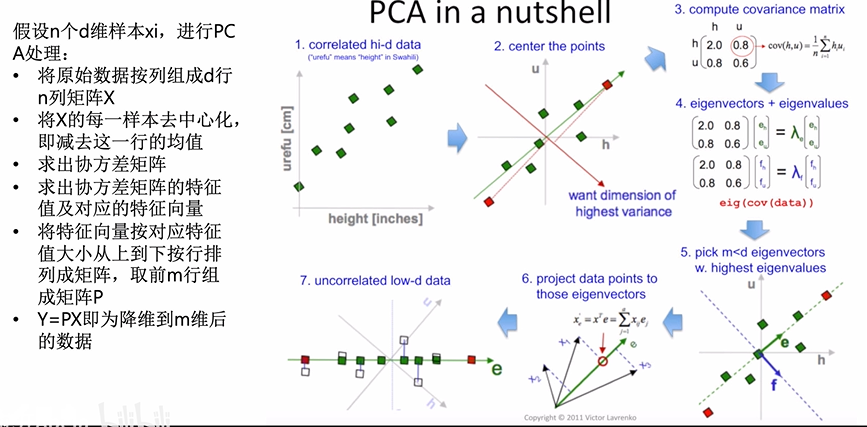
下图中5个数据点，投影到x轴，y轴都不是好的选择

直观上看，投影后的投影值尽可能分散，这样能保留更多的原始数据的结果。这种分散程度，用数学上的方差来表述。

PCA的目标函数





PCA的流程

带核函数的PCA

