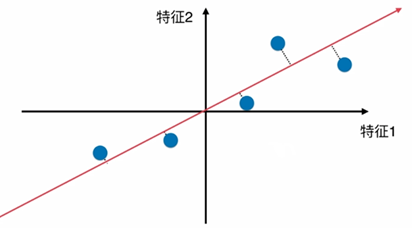
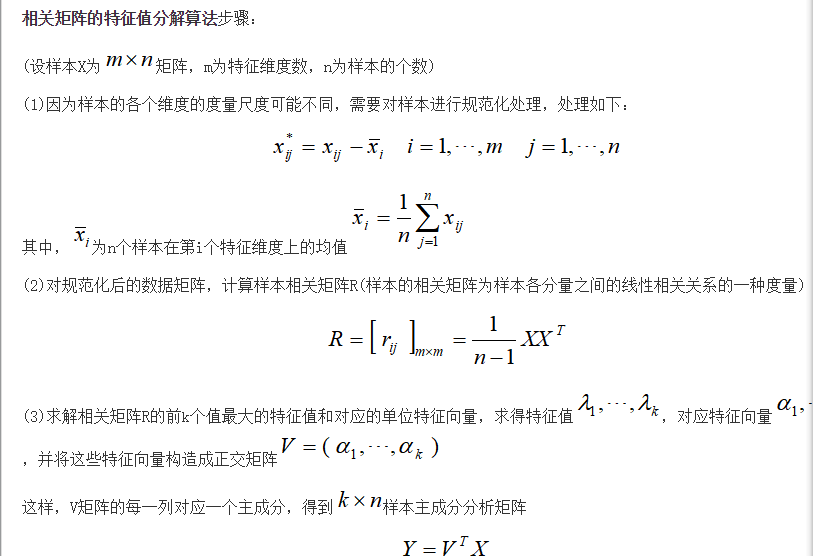
**主成分分析**(principle component analysis,PCA)，用正交变换将由线性相关变量表示的观测数据转换为少数几个由线性无关变量（称主成分）表示的数据的方法，即对数据进行降维处理。



主成分分析的算法：



总结而来：

形成样本矩阵，样本中心化

计算样本矩阵的协方差矩阵

对协方差矩阵进行特征值分解，选取最大的p个特征值对应的特征向量组成投影矩阵

对原始样本矩阵进行投影，得到降维后的新样本矩阵

一般而言，对于度量单位不同的指标或取值范围彼此差异非常大的指标，不直接由其协方差矩阵出发进行主成分分析，而应该考虑将数据标准化。

对原始数据进行标准化处理后倾向于各个指标的作用在主成分的构成中相等。对于取值范围相差不大或度量相同的指标进行标准化处理后，其主成分分析的结果仍与由协方差出发求得的结果有较大区别。其原因是由于对数据进行标准化的过程实际上也就是抹杀原始变量离散程度差异的过程，标准化后的各变量方差相等，均为1，而实际上方差也是对数据信息的重要概括，也就是说，对原始数据进行标准化后抹杀了一部分重要信息，因此才使得标准化后变量在对主成分构成中的作用趋于相等。由此看来，对同度量或取值范围在同量级的数据，还是直接从协方差矩阵求解主成分为宜。