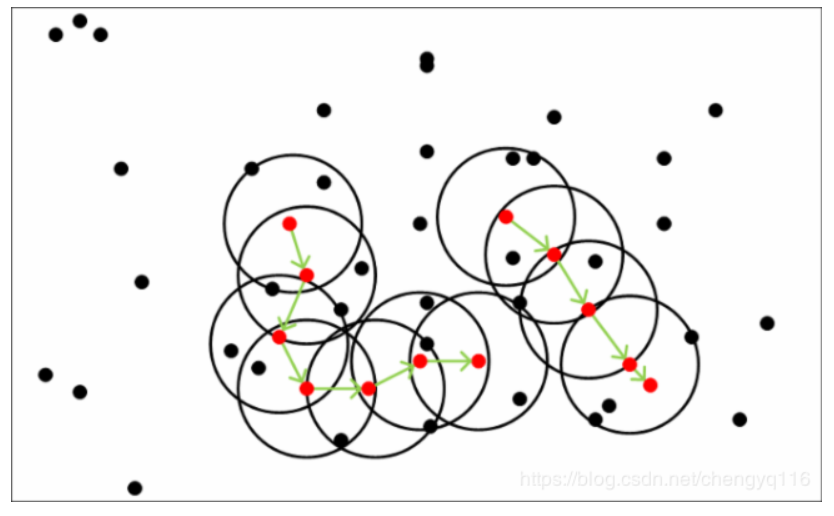
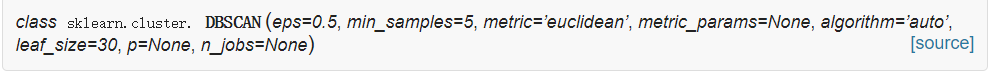
* DBSACN



某些样本可能到两个核心对象的距离都小于 Eps，但是这两个核心对象由于不是密度直达，不属于同一个聚类簇。一般来说，此时 DBSCAN 采用先来后到，先进行聚类的类别簇会标记这个样本为它的类别。DBSCAN 的算法不是完全稳定的算法。

密度可达的核心点及边界点是一类



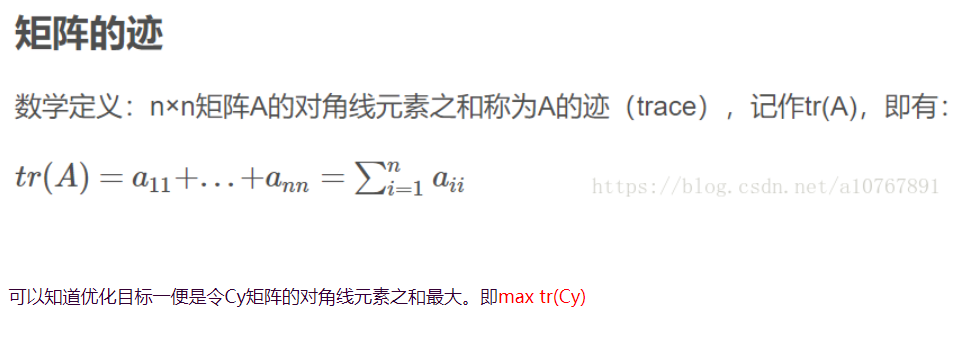
* PCA

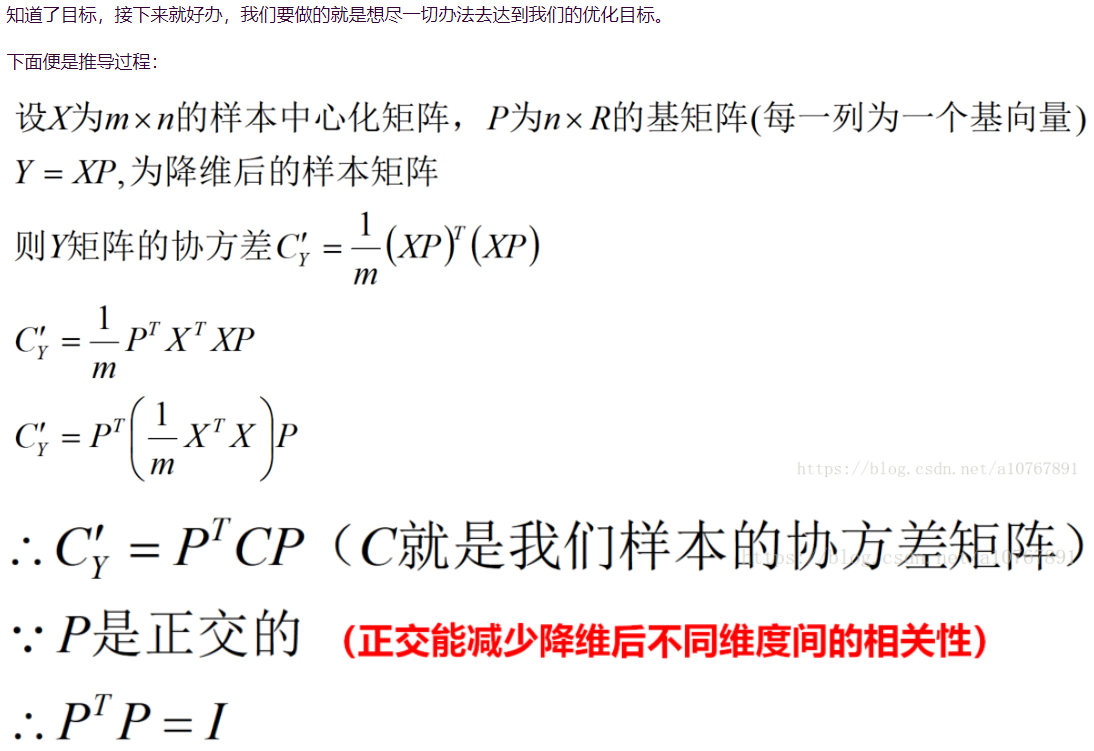
（行是样本，列是特征）

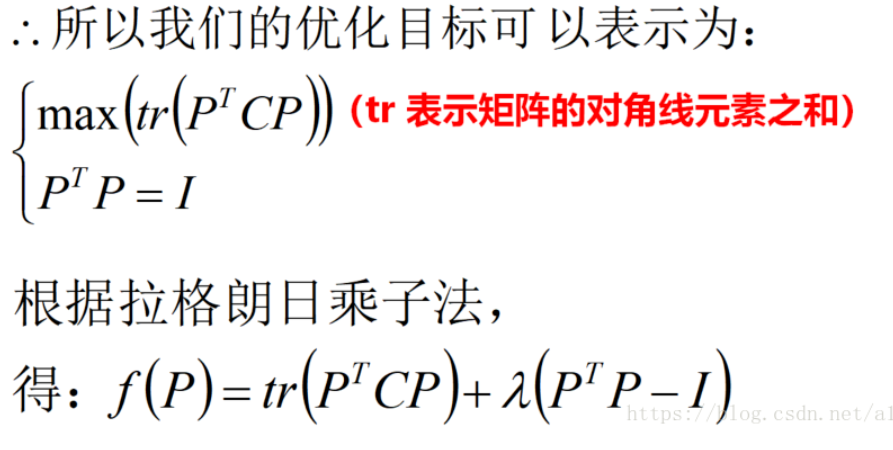
PCA算法的优化目标就是: ① 降维后同一纬度的方差最大

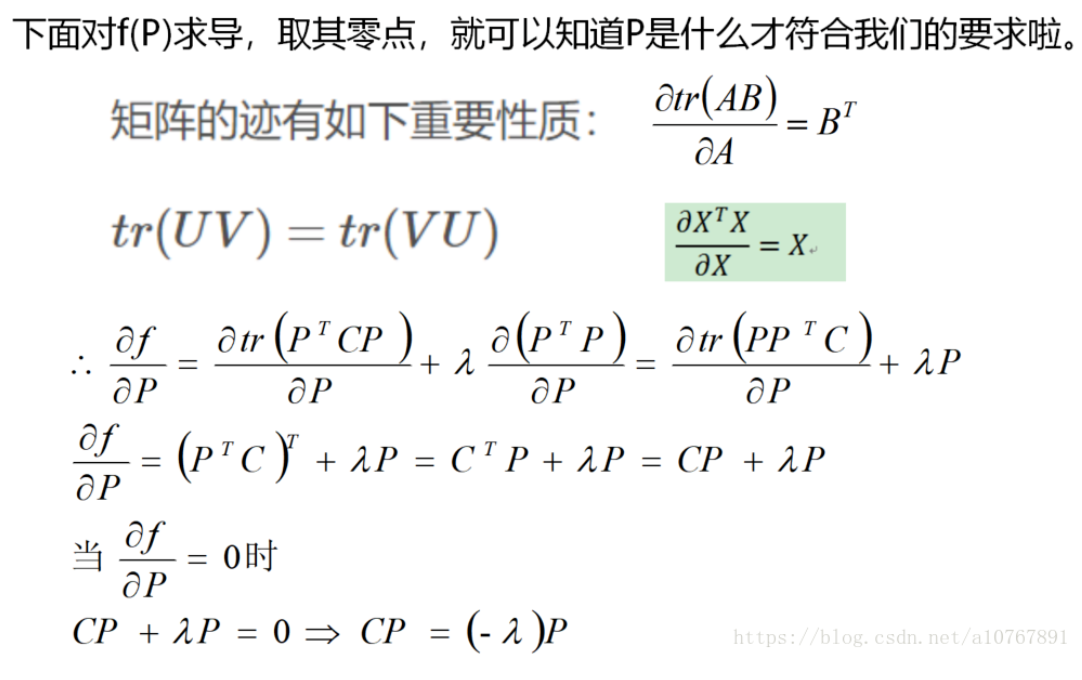
② 不同维度之间的相关性为0

根据线性代数，我们可以知道同一元素的协方差就表示该元素的方差，不同元素之间的协方差就表示它们的相关性。









可以看到，最终求得的结果满足特征向量的关系式，因此由C矩阵的特征向量所构成的基矩阵，就是我们要求的变换矩阵。

由该矩阵降维得到的新样本矩阵可以最大程度保留原样本的信息。

