6、已知两类样本，

。

，试用PCA变化做一维特征提取，给出新的一维特征值

答：



求R的特征值、特征矢量：





选择对应的向量作为变换矩阵



以此类推：得到

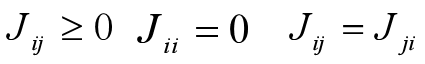
7、特征选择的判据作用是什么？判据构造的原则是什么？请列举有哪些不同类型的判据，并简单说明至少三种特征选择方法。

答：

作用：特征选择的任务是找出一组对分类最好的特征

原则：

1. 与错误率有单调关系，

2. 判据具有“距离”的某些特性。 

3. 对独立的特征有可加性：

4. 增加特征时判据不减小：

方法：

**基于全局寻优的分支定界法**

该方法是一种从包含所有候选特征开始，逐步去掉不被选中的特征的自顶向下的方法，具有回溯的过程，能够考虑到所有可能的组合。其基本思想是：设法将所有可能的特征组合构建成一个树状结构，按照特定的规律对树进行搜索，使得搜索过程尽可能早的达到最优解而不必遍历整棵树。

要做到这一点必须要求准则判据对特征具有单调性，且当处理高维数据时，算法复杂度较高，所以很难广泛使用。

**基于启发式搜索的方法**

很多情况下，即使采取分支定界法，全局寻优的方法计算量可能仍然很大，因此提出了一些基于启发式搜索的次优选择算法：单独最优特征组合，序列前向选择（SFS）及广义的SFS（GSFS），序列后向选择(SBS)及广义的SBS(GSFS)，增l去r选择及广义的增l去r选择方法，浮动搜索方法及浮动的广义后向选择方法(FGSBS)等。

**随机搜索方法**

该类方法在计算中将特征选择问题与遗传算法、模拟退火算法、粒子群优化算法，随机森林或一个随机重采样过程结合起来，以概率推理和采样过程作为算法基础，基于分类估计的有效性，在算法运行过程中对每个特征赋予一定的权重，再根据给定的或自适应的阈值对特征的重要性进行评价。例如，Relief及其扩展算法就是一种典型的根据权重选择特征的随机搜索方法，它能有效的去掉无关特征，但不能去除冗余特征，且只能用于两类分类问题。

8、假设一个抛硬币实验，假设正面向上设为U的概率为ρ，反面向上设为D概率为（1−ρ）. 我们进行了3次实验，得到两次正面，一次反面，即序列为′UUD′。

1. 使用最大似然估计求ρ 的估计值

（2）假设ρ的先验概率是服从 ，用最大后验概率估计ρ 的值.

（3）假设ρ的先验概率是服从 ，求贝叶斯估计的ρ的分布。

答：

1.  



ρ=0舍，取ρ=0.67

（2）





（3）



