很多情况下，数据的维度是比较高的，在这种情况下估计概率密度模型比较困难；实际上，模式识别或者机器学习的目标是在特征空间中找到两类（或多类）的分类面，其中并不是那么关心概率密度模型。Fisher判别分析（LDA）就是直接给出一种判别函数，只要估计其参数就可以了；其基本思想是将高维特征样本投影到低维空间，但是这样会损失掉一些信息，并可能导致在高维可分的数据在低维上变得不可分，好处就是降低数据量。

假设一个维输入向量，通过式投影到一维空间；若我们在上设置一个阈值，然后把的样本分为类，的样本分为类，则就成了一个分类面。一般而言，一维投影会导致相当多的信息丢失，因此在原始的维空间能够完美地分离开的样本可能在一维空间中相互重叠，但是通过调整权向量，我们可以选择类别之间分开最大的一个投影。

二分类问题：

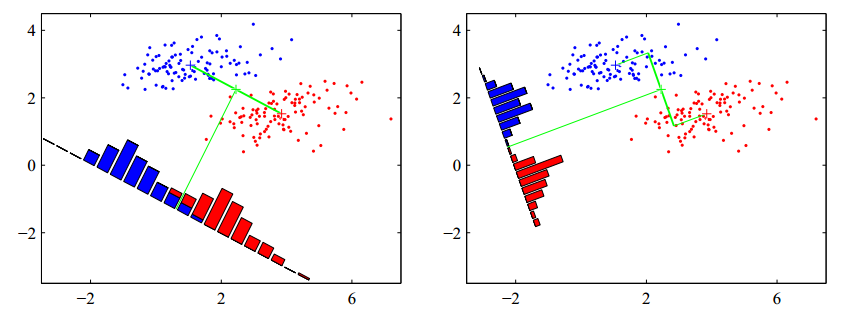
假设类中有个点，类中有个点：

两类的均值向量为：



要使得投影后的区分度尽可能大，则类间均值之间的距离要尽可能的大，即：





如上图所示，两个类别在上能完全分离，但是在投影后的一维空间上（均值连线）有部分重叠，如果类概率分布的协⽅差矩阵与对角化矩阵差距较⼤，那么这种问题就会出现。 Fisher提出的思想是最大化⼀个函数，这个函数能够让类均值的投影分开得较⼤，同时让每个类别内部的方差较小，从而最小化了类别的重叠。

类别各自的类内方差：



其中，若将整个数据集的总的类内方差定义为，Fisher准则根据类间方差和类内方差的比值定义：



其中是类间协方差矩阵：；

是类内协方差矩阵： ；



先对分母进行归一化，消除整数倍变化带来的影响：



通过拉格朗日方法求解：



即是的特征向量，又：



又的放大缩小不影响上式：



多分类问题：

假设输入空间的维度大于类别数量，我们引入个线性“特征”，其中。。

类内协方差矩阵：



其中：



其中是类别中样本的数量。

为了找到类间协方差矩阵的推广，首先考虑整体的协方差矩阵：



其中是全体数据的均值：



整体协方差可以分解为类内协方差，加上另一个矩阵，它可以看做是类间协方差矩阵。





其中：

在投影的维空间中定义类似矩阵：



其中： ， 。



，与之前一样，可得： 。



即先求出的特征值，再取前个特征向量组成矩阵。

由于由个矩阵的和组成，每个矩阵都是两个向量的外积，因此秩等于1，这些矩阵中只有个是相互独立的，因此的秩最大等于，因此最多有个非零特征值，因此。

2017.07.25