

# 线性判别分析 LDA

Li Liang\*

## 1 线性判别分析

### 1.1 二类 LDA 原理

给定数据集  $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$ ,  $y_i \in \{0, 1\}$ , 令  $\mathbf{X}_i, \mu_i, \Sigma_i$  表示第  $i \in \{0, 1\}$  类示例的集合、均值向量和协方差矩阵。若将数据投影到直线  $\omega$  上, 则两类样本的中心在直线上的投影分别为  $\omega^T \mu_0$  和  $\omega^T \mu_1$ , 两类样本在直线上投影的协方差为  $\omega^T \Sigma_0 \omega$  和  $\omega^T \Sigma_1 \omega$  (协方差传播)。

若是同类尽可能近, 则协方差尽可能小; 异类尽可能远则要求类中心距离大。同时考虑两者, 则最大化目标为:

$$\begin{aligned} J &= \frac{\|\omega^T \mu_0 - \omega^T \mu_1\|_2^2}{\omega^T \Sigma_0 \omega + \omega^T \Sigma_1 \omega} \\ &= \frac{\omega^T (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \omega}{\omega^T (\Sigma_0 + \Sigma_1) \omega} \end{aligned} \quad (1)$$

定义类内散度矩阵和类间散度矩阵为:

$$\begin{aligned} S_w &= \sum_0 + \sum_1 \\ &= \sum_{x \in \mathbf{X}_0} (x - \mu_0)(x - \mu_0)^T + \sum_{x \in \mathbf{X}_1} (x - \mu_1)(x - \mu_1)^T \end{aligned} \quad (2)$$

$$S_b = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \quad (3)$$

带入 (1) 式:

$$J = \frac{\omega^T S_b \omega}{\omega^T S_w \omega} \quad (4)$$

---

\*<https://github.com/leeliang/>

上式就是广义瑞利商，广义瑞利商有个性质为： $J$  的最大值为矩阵  $S_w^{-1}S_b$  的最大特征值，而对应的  $\omega$  为  $S_w^{-1}S_b$  的最大特征值对应的特征向量（特征值最大，意味着在对应的特征向量上的变化最大，分类越大），即：

$$S_w^{-1}S_b\omega = \lambda\omega \quad (5)$$

上式就是 Fisher Linear Discrimination 公式，推导过程可参见西瓜书。 $S_w^{-1}S_b$  不一定是对称矩阵，在求它的特征向量时不能使用奇异值分解，这样就只能使用普通的求特征向量的方式。注意到  $S_b\omega = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T\omega$ ，其中  $c = (\mu_0 - \mu_1)^T\omega$  为常数。(5) 式可以重写为：

$$\omega = \frac{c}{\lambda}S_w^{-1}(\mu_0 - \mu_1) \quad (6)$$

注意到 (4) 式分子和分母都是关于  $\omega$  的二次型，因此 (4) 式的解与  $\omega$  的大小无关，至于方向有关。则上式等价于：

$$\omega = S_w^{-1}(\mu_0 - \mu_1) \quad (7)$$

在实践中，通常对  $S_w$  进行奇异值分解，求取  $\omega$ 。

~~LDA 可从贝叶斯决策理论的角度来阐释，并可证明，当两类数据同先验、满足高斯分布且协方差相等时，LDA 可达到最优分类。~~

## 1.2 多类 LDA 原理

给定数据集  $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$ ,  $y_i \in \{0, 1, \dots, k\}$ ，类似地，目标函数为：

$$J = \frac{W^T S_b W}{W^T S_w W} \quad (8)$$

其中， $W$  为一个超平面，

$$\begin{aligned} S_b &= \sum_{1 \leq i, j \leq k} (\mu_i - \mu_j)(\mu_i - \mu_j)^T \\ S_w &= \sum_{j=1}^k S_{wj} = \sum_{j=1}^k \sum_{x \in X_j} (x - \mu_j)(x - \mu_j)^T \end{aligned} \quad (9)$$

按照上式计算  $S_b$  较复杂，LDA 采用一个间接的方式。等价？

首先定义  $X$  的整体散度矩阵:

$$S_t = S_b + S_w = \sum_{j=1}^k (x_j - \mu)(x_j - \mu)^T \quad (10)$$

$\mu$  是所有示例的均值向量。则:

$$S_b = S_t - S_w = \sum_{j=1}^k N_j (\mu_j - \mu)(\mu_j - \mu)^T \quad (11)$$

$N_j$  为第  $j$  个类别的示例个数。

由于  $W^T S_b W$  和  $W^T S_w W$  都是矩阵, 无法使用标量的最优化方法, 一般来说, 常使用的优化目标为:

$$J = \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)} \quad (12)$$

### 1.3 LDA 分类

对二分类问题。由于只有两个类别, 在经过上面的求解后, 最后所有样本将会映射到一维空间中, 一般将两个类别的中心点之间中心点作为分类点。

对于多类的情况主要考虑用来数据降维, 那么降维后的维度是多少呢?  $W$  是一个利用了样本的类别得到的投影矩阵, 是  $S_w^{-1} S_b$  的特征向量, 该特征向量的维数不超过  $S_w^{-1} S_b$  的秩。

$$\begin{aligned} R(S_w^{-1} S_b) &\leq \min(R(S_w^{-1}), R(S_b)) \leq R(S_b) \\ &= R\left(\sum_{j=1}^k N_j (\mu_j - \mu)(\mu_j - \mu)^T\right) \\ &\leq \sum_{j=1}^k N_j R[(\mu_j - \mu)(\mu_j - \mu)^T] \\ &= \sum_{j=1}^k R(\mu_j - \mu) \end{aligned} \quad (13)$$

由于存在线性组合  $\sum_{j=1}^k N_j (\mu_j - \mu) = 0$ , 故上式的秩最大为  $k - 1$ , 即特征向量最多有  $k - 1$  个。