机器学习作业 1-线性判别分析 LDA

2150248-姚天亮-自动化

一、作业要求

用 Python 编程实现线性判别分析 LDA，并给出下面数据集上的结果及说明。

二、原理说明

线性判别分析(Linear Discriminant Analysis， LDA)是一种有监督的降维和分类技术。LDA 的目标是在保留类别间的尽可能多的差异的同时，降低类别内部的差异，从而获得较好的分类效果。希望最大化投影后类别间投影点的离散程度，同时最小化投影后同类别内部投影点的离散程度。

给定训练样例集，将样例投影到一条直线上，使得同类样例的投影点尽可能接近、异类样例的投影点尽可能远离;在对新样本进行分类时，将其投影到同样的这条直线上，再根据投影点的位置来确定样本的类别。

给定数据集$D=\{(x_i,y_i)\}_{i=1}^{m}, y_i\in\{0,1\}$,令$X_i$，$\mu_i$，$\Sigma_i$分别表示第 i∈{0, 1}类示例的集合、均值向量、协方差矩阵。若将数据投影到直线 w 上，则两类样本的中心在直线上的投影分别为$w^T\mu_0$和$w^T\mu_1$；若将所有样本点都投影到直线上，则两类样本的协方差分别$w^T\Sigma_0 w$和$w^T\Sigma_1 w$，由于直线是一维空间，因此上述协方差与投影均为实数。

欲使同类样例的投影点尽可能接近，可以让同类样例投影点的协方差尽可能小，即$w^T\Sigma_0 w+w^T\Sigma_1 w$尽可能小;而欲使异类样例的投影点尽可能远离，可以让类中心之间的距离尽可能大，即 $||w^T\mu_0-w^T\mu_1||_2^2$尽可能大.同时考虑二者，则可得到欲最大化的目标:

$$J=\frac{||w^T\mu_0-w^T\mu_1||_2^2}{w^T\Sigma_0 w+w^T\Sigma_1 w}$$

定义"类内散度矩阵"

$$\begin{aligned}\mathbf{S}_w &= \mathbf{\Sigma}_0+\mathbf{\Sigma}_1\\ &=\sum_{\boldsymbol{x}\in X_0}(\boldsymbol{x}-\boldsymbol{\mu}_0)(\boldsymbol{x}-\boldsymbol{\mu}_0)^{\mathrm{T}}+\sum_{\boldsymbol{x}\in X_1}(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}}\end{aligned}$$

以及"类间散度矩阵"

$$\mathbf{S}_b=(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)^{\mathrm{T}}$$

则 J 可以改写为

$$J=\frac{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}}$$

该式即为最大化目标。

注意到 J 的分子和分母都是关于 w 的二次项因此 J 的解与 w 的长度无关,只与其方向有关。不失一般性，令 $w^T\mu_0 w=1$，则 J 等价于

$$\begin{aligned}\min_{\boldsymbol{w}} \quad & -\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}=1\end{aligned}$$

由拉格朗日乘子法，上式等价于

$$\mathbf{S}_b\boldsymbol{w}=\lambda\mathbf{S}_w\boldsymbol{w}$$

其中λ是拉格朗日乘子.注意到S_bw 的方向恒为$\mu_0-\mu_1$，不妨令

$$\mathbf{S}_b\boldsymbol{w}=\lambda(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)$$

代入知

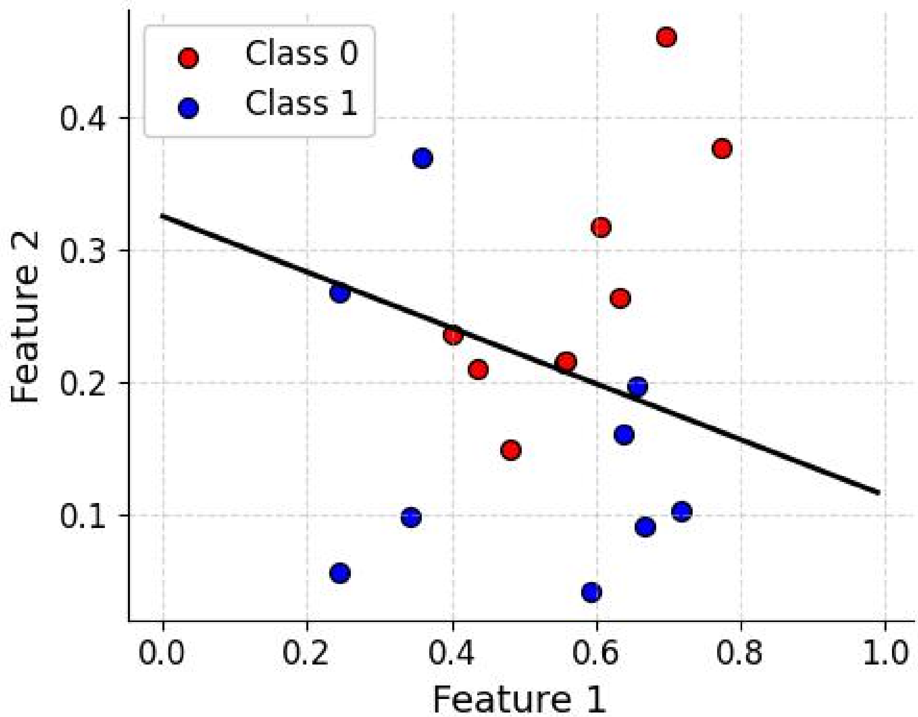
$$\boldsymbol{w}=\mathbf{S}_w^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)$$

考虑到数值解的稳定性，在实践中通常是对S_w进行奇异值分解，即$S_w=U\Sigma V^T$，这里 Σ 是一个实对角矩阵，其对角线上的元素是S_w的奇异值，然后，再由$S_w^{-1}=V\Sigma^{-1}U^{-1}$得到$S_w^{-1}$，同时，LDA 可从贝时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.。

可以将 LDA 推广到多分类任务中.假定存在 N 个类，且第 i 类示例数为 mi 叫.我们先定义"全局散度矩阵"

$$\begin{aligned}\mathbf{S}_t &= \mathbf{S}_b+\mathbf{S}_w \\ &= \sum_{i=1}^{m}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}\end{aligned}$$

三、实验结果



```
LDA Coefficients (Weights):
 [[ -2.29448249 -10.89707109]]

LDA Mean vectors:
 [[0.57133333 0.27122222]
 [0.49533333 0.15344444]]

Classification Threshold:  0.728906743264478
Linear Discriminant Analysis Score: 0.7222222222222222
```

心得

通过这次实践，我深刻理解了 LDA 的原理和目标。LDA 旨在最大化类别间的差异，最小化类别内部的差异，从而获得良好的分类效果。这种思路非常巧妙，充分利用了数据的标签信息。

在实现过程中，我还了解到 LDA 背后的数学原理，包括类内散布矩阵、类间散布矩阵以及广义特征值问题的求解。这些数学概念虽然抽象，但在算法中发挥着关键作用。

另一方面，我也意识到 LDA 的局限性。LDA 假设数据服从高斯分布，并且只能学习线性判别边界。因此，对于非线性可分的数据，LDA 可能无法取得良好的效果。此时，我们需要使用其他更加复杂的非线性降维和分类方法。

这次作业让我对 LDA 算法有了更深入的理解，同时也激发了我去学习其他机器学习算法的兴趣。我期待在后续的学习中，能够掌握更多实用的算法和技术，为解决实际问题做好充分准备。