HW2

asandstar@github

2022 年 10 月 26 日

1 现有下表所示训练数据，给定感知机模型 $z = sgn(w_1x + w_2y + b)$ 参数取值为 $w_1 = 1, w_2 = 1, b = -1$ 步长 $\eta = 0.5$，试学习感知机权重，给出过程

x	y	z
-0.89	-1.62	-1
-0.30	0.96	+1
1.49	0.25	+1
0.06	-0.68	-1
-1.13	0.14	-1
0.78	0.96	+1

设 $\boldsymbol{w} = (w_1, w_2, b), \boldsymbol{X}^\top = (x, y, 1)$, 模型 $z = sgn(\boldsymbol{w} \cdot \boldsymbol{X})$

其中已知 $w_1 = 1, w_2 = 1, b = -1$, 所以 $z = sgn(x + y - 1)$

选择一个数据点 (x_i, y_i, z_i)，如果分类错误则更新参数 $\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta \boldsymbol{X}_i z_i$,

(1) 对 $\boldsymbol{X_1}^\top = (-0.89, -1.62, -1), \boldsymbol{w} = (1, 1, -1)$,

代入得 $z = sgn(-0.89 + (-1.62) - 1) = -1$, 分类正确

(2) 对 $\boldsymbol{X_2}^\top = (-0.30, 0.96, +1), \boldsymbol{w} = (1, 1, -1)$,

代入得 $z = sgn(-0.30 + 0.96 - 1) = -1 \neq +1$, 分类错误

更新参数 $\boldsymbol{w} = (1, 1, -1) + 0.5 \times (-0.30, 0.96, 1) \times (+1) = (0.85, 1.48, -0.5)$

(3) 对 $\boldsymbol{X_1}^\top = (-0.89, -1.62, -1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times (-0.89) + 1.48 \times (-1.62) - 0.5) = -1$, 分类正确

(4) 对 $\boldsymbol{X_2}^\top = (-0.30, 0.96, +1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times (-0.30) + 1.48 \times 0.96 - 0.5) = +1$, 分类正确

(5) 对 $\boldsymbol{X_3}^\top = (1.49, 0.25, +1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times 1.49 + 1.48 \times 0.25 - 0.5) = +1$, 分类正确

(6) 对 $\boldsymbol{X_4}^\top = (0.06, -0.68, -1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times 0.06 + 1.48 \times (-0.68) - 0.5) = -1$, 分类正确

(7) 对 $\boldsymbol{X_5}^\top = (-1.13, 0.14, -1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times (-1.13) + 1.48 \times 0.14 - 0.5) = -1$, 分类正确

(8) 对 $\boldsymbol{X_6}^\top = (0.78, 0.96, +1), \boldsymbol{w} = (0.85, 1.48, -0.5)$,

代入得 $z = sgn(0.85 \times 0.78 + 1.48 \times 0.96 - 0.5) = +1$, 分类正确

综上，感知机权重为 $\boldsymbol{w} = (w_1, w_2, b) = (0.85, 1.48, -0.5)$

2 现有以下数据

$$x = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 0 \\ 2.5 & 2 & 0 \\ -2 & -2 & 1 \\ -2.7 & -1 & 1 \\ -3 & -2.2 & 1 \end{bmatrix}$$

其中每行为一个样本, 前 2 列为属性, 第 3 列表示类别, 试应用 LDA 算法计算投影后的数据点

LDA 的思想：给定训练样例集，设法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新样本的类别。

设 μ_i 为第 i 类的均值向量, 例如 $\mu_0 = (\mathbf{x_0}, \mathbf{y_0})^\top$ 为类别 0 的样本中心点

$\boldsymbol{w}^\top = (w_0, w_1)$

$\boldsymbol{w}^\top \mu_0, \boldsymbol{w}^\top \mu_1$ 为两类样本的中心在直线上的投影

$\boldsymbol{w}^\top \Sigma_0 \boldsymbol{w}, \boldsymbol{w}^\top \Sigma_1 \boldsymbol{w}$ 为投影所有的点到直线后，两类样本的协方差

(1) 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 $\boldsymbol{w}^\top \Sigma_0 \boldsymbol{w} + \boldsymbol{w}^\top \Sigma_1 \boldsymbol{w}$ 尽可能小

(2) 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 $||\boldsymbol{w}^\top \mu_0 - \boldsymbol{w}^\top \mu_1||_2^2$

所以最大化的目标为 $J = \dfrac{||\boldsymbol{w}^\top \mu_0 - \boldsymbol{w}^\top \mu_1||_2^2}{\boldsymbol{w}^\top \Sigma_0 \boldsymbol{w} + \boldsymbol{w}^\top \Sigma_1 \boldsymbol{w}} = \dfrac{\boldsymbol{w}^\top (\mu_0 - \mu_1)(\mu_0 - \mu_1)^\top \boldsymbol{w}}{\boldsymbol{w}^\top (\Sigma_0 + \Sigma_1) \boldsymbol{w}}$

定义类内散度矩阵 $\mathbf{S}_w = \Sigma_0 + \Sigma_1$

$= \sum_{x \in X_0} (\boldsymbol{x} - \boldsymbol{\mu_0})(\boldsymbol{x} - \boldsymbol{\mu_0})^\top + \sum_{x \in X_1} (\boldsymbol{x} - \boldsymbol{\mu_1})(\boldsymbol{x} - \boldsymbol{\mu_1})^\top$

定义类间散度矩阵 $\mathbf{S}_b = (\boldsymbol{\mu_0} - \boldsymbol{\mu_1})(\boldsymbol{\mu_0} - \boldsymbol{\mu_1})^\top$

则最大化的目标可化为 S_w 和 S_b 的广义瑞利商 $J = \dfrac{\boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w}}{\boldsymbol{w}^\top \mathbf{S}_b \boldsymbol{w}}$

注意到上式的分子和分母都是关于 w 的二次项, 因此上式的解与 w 的长度无关, 只与其方向有关, 所以令 $\boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w} = 1$

故最大化的目标等价于

$$\min_{w} -\boldsymbol{w}^\top \mathbf{S}_b \boldsymbol{w}$$
$$s.t. \boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w} = 1$$

利用拉格朗日乘子法得 (λ 为拉格朗日乘子)

$$L(w,\lambda) = -w^\top S_b w + \lambda(w^\top S_w w - 1)$$

对上式求导并令其为 0 得

$$\frac{\partial L}{\partial w} = -2S_b w + 2\lambda S_w w = 0$$

因为 $\boldsymbol{w}^\top \mu$ 为标量，所以 $\mathbf{S}_b w = (\mu_\mathbf{0} - \mu_\mathbf{1})(\mu_\mathbf{0} - \mu_\mathbf{1})^\top w$ 中的 $(\mu_\mathbf{0} - \mu_\mathbf{1})^\top w$ 也为标量，即 $\mathbf{S}_b w$ 的方向仅与 $(\mu_\mathbf{0} - \mu_\mathbf{1})$ 有关

不妨令 $\mathbf{S}_b w = \lambda(\mu_\mathbf{0} - \mu_\mathbf{1})$, 代入上述求导所得等式得

$$w = S_w^{-1}(\mu_\mathbf{0} - \mu_\mathbf{1})$$

对 S_w 进行奇异值分解 $S_w = \mathbf{U}\Sigma\mathbf{V}^\top$, 其中 Σ 为实对角矩阵，其对角线上的元素是 S_w 的奇异值

再求逆得 $S_w^{-1} = \mathbf{V}\Sigma^{-1}\mathbf{U}^\top$，并求得 w

由题可知被投影的直线 l_1 斜率 $k = \dfrac{w_0}{w_1}$

所以与该直线垂直的直线 l_2 斜率 $k' = -\dfrac{w_1}{w_0}$

l_2 代入数据点，再与 l_1 联立即可求得投影点

具体公式为 $x' = \dfrac{x + y \cdot k'}{(k')^2 + 1}, y' = kx'$

利用 python 求解的代码如下

```python
from statistics import mean
import numpy as np
import matplotlib.pyplot as plt
'''1.导入数据'''
xy = np.array([[1, 2], [3, 1], [2.5, 2], [-2, -2], [-2.7, -1], [-3, 1]])
xy0 = np.array(xy[:3]) #3×2维
xy_0 = np.array(xy[:3])
xy1 = np.array(xy[3:]) #3×2维
xy_1 = np.array(xy[3:])
#数据按题干x摆放，第一列是x,第二列是y
'''2.计算miu0,miu1'''
x0 = mean(xy0[:, 0])
y0 = mean(xy0[:, 1])
x1 = mean(xy1[:, 0])
```

```
    y1 = mean(xy1[:, 1])
    miu0 = [x0, y0]
    miu1 = [x1, y1]
    '''3.计算类内散度矩阵Sw'''
    for i in range(3):
        xy0[i][0] = xy0[i][0] - x0
        xy0[i][1] = xy0[i][1] - y0
        xy1[i][0] = xy1[i][0] - x1
        xy1[i][1] = xy1[i][1] - y1
        #xy0从x中类别为0的矩阵变为x- 的矩阵
    sig0 = sig1 = 0
    for i in range(3):
        sig0 = sig0 + np.outer(xy0[i], xy0[i].T)
        sig1 = sig1 + np.outer(xy1[i], xy1[i].T)
        #某类中每一个i的(x-miu)与(x-miu)^T先外积，再对i求和
    sw = sig0 + sig1 #2×2维
    '''4.Sb·w的方向'''
    miu01 = np.array([x0 - x1, y0 - y1]) #miu0-miu1
    '''5.得到w'''
    swni = np.linalg.inv(sw) #求逆
    w = np.dot(swni, miu01)
    '''6.投影点公式'''
    k = w[0] / w[1]
    kt = -w[1] / w[0]
    xty = (xy[:, 0] + xy[:, 1] * k) / (k**2 + 1)
    yty = k * xty
    xyty = np.array([xty, yty])
    xyty = xyty.T
    print('投影后数据点：\n', xyty)
    '''7.画图'''
    #画出题干数据点、被投影的直线
    plt.scatter(xy_0[:, 0], xy_0[:, 1], c='b', label='+', marker='+')
    plt.scatter(xy_1[:, 0], xy_1[:, 1], c='r', label='-', marker='_')
    plt.plot([-3.5, 3.5], [-3.5 * k, 3.5 * k], label='y')
    #画出投影点与原数据点的连线
    for i in range(6):
        plt.plot([xy[i][0], xyty[i][0]], [xy[i][1], xyty[i][1]])
    plt.xlabel('X1', fontproperties='SimHei', fontsize=15, color='black')
    plt.ylabel('X2', fontproperties='SimHei', fontsize=15, color='black')
```

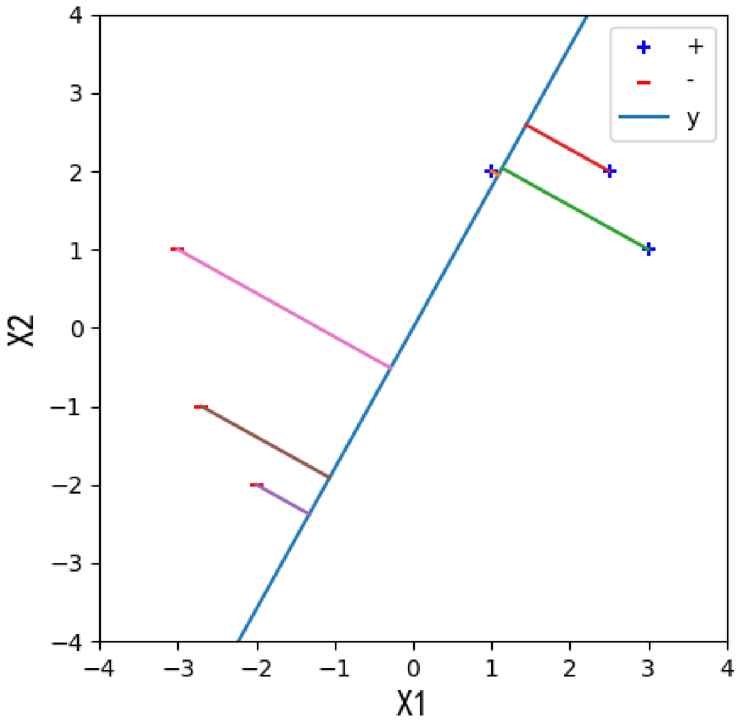
```
    plt.xlim(-4, 4)
    plt.ylim(-4, 4)
    ax = plt.gca()
    ax.set_aspect(1)
    plt.legend()
    plt.show()
```

求得结果为 $w = [3.30924855, 1.84393064]^\top$

投影后数据点：

[[1.08730441 1.95135353]
[1.13595088 2.03865794]
[1.44268362 2.58914223]
[-1.32422388 -2.37654599]
[-1.06487504 -1.91110019]
[-0.28556595 -0.51249688]]

python 作图如下



3 现用梯度下降求解优化问题：

$$x^* = arg\min_x f(x)$$

即迭代公式为

$$x_{i+1} = x_i - \eta \frac{\partial f}{\partial x}$$

若 f 是二阶多项式, 试证明迭代收敛的条件是 $\eta < 2\left(\frac{\partial^2 f}{\partial x^2}\right)^{-1}$

设 f 的表达式为 $f(x) = ax^2 + bx + c$, 则 $\frac{\partial f}{\partial x} = 2ax + b, \frac{\partial^2 f}{\partial x^2} = 2a$
又 $f(x)$ 有最小值，故 $a > 0$
因为 $||\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}|| = \Big|\, ||2ax + b|| - ||2ay + b||\, \Big|$
$\leq ||2a(x - y)||$
$\leq 2a||x - y||$
即存在 $L = 2a$, 使 $||f'(x) - f'(y)|| \leq L||x - y||$
所以 f' 满足 Lipschitz continuous 条件，f 符合 Lipschitz continuous gradient
由 Descent Lemma，$f(x) \leq f(y) + \frac{\partial f}{\partial x}(x - y) + \frac{L}{2}||x - y||^2$
在上式中把 x, y 换成 x_{i_1}, x_i 得 $f(x_{i+1}) \leq f(x_i) + \frac{\partial f}{\partial x}(x_{i+1} - x_i) + \frac{L}{2}||x_{i+1} - x_i||^2$
又因为 $x_{i+1} - x_i = -\eta\frac{\partial f}{\partial x}$,
代入上述 Descent Lemma 得到的不等式得

$$\begin{aligned} f(x_{i+1}) &\leq f(x_i) - \eta||f'(x_i)||^2 + \tfrac{L}{2}||\eta f'(x_i)||^2 \\ &= f(x_i) - (\eta - \tfrac{L}{2}\eta^2)||f'(x_i)||^2 \end{aligned}$$

要使迭代收敛，需要满足 $\eta - \frac{L}{2}\eta^2 > 0$, 解得 $\eta < \frac{2}{L}$
而 $L = 2a = \frac{\partial^2 f}{\partial x^2}$
所以迭代的收敛条件为 $\eta < 2\left(\frac{\partial^2 f}{\partial x^2}\right)^{-1}$
证毕

4 简答: 使用 OvR 和 MvM 将多分类任务分解为二分类任务求解时, 试概述针对类别不平衡性处理的方法。

类别不平衡学习的基本策略是 rescaling(再缩放)。

以线性分类器为例，用 $y = \boldsymbol{w}^\top \boldsymbol{x} + b$ 对 $\boldsymbol{w}^\top$ 进行分类时，是用预测的 y 和某个阈值相比较，实际上 y 是 $\boldsymbol{x}$ 属于某类（正例或反例）的可能性
$\frac{y}{1-y}$ 指正例可能性和反例可能性的比值
阈值 =0.5 表明分类器认为真实正、反例可能性相同相同

分类器决策规则：$\dfrac{y}{1-y} > 1$ 时，预测为正例

训练集中正、反例数目不同时，令 m^+ 为正例数，m^- 为反例数, 故观测几率可以表示为 $\dfrac{m^+}{m^-}$

假设训练集是真实样本总体的无偏采样，则观测几率代表真实几率

一旦分类器的预测几率高于观测几率，判定为正例，即 $\dfrac{y}{1-y} > \dfrac{m^+}{m^-}$ 时，预测为正例

但 $\dfrac{y}{1-y} > 1$ 是分类器使用的决策依据，所以对预测值进行调整，令

$$\frac{y'}{1-y'} = \frac{y}{1-y} \times \frac{m^+}{m^-}$$

实际情况是，因为“训练集是真实样本总体的无偏采样”的假设往往不成立，不能用观测几率代表真实几率，现在主要有三类做法。

（1）“欠采样” (undersampling)

去除训练集里的反类样例中一些反例使得正、反例数目接近，再学习

（2）“过采样” (oversampling)

增加训练集里的正类样例中一些正例使得正、反例数目接近，再学习

（3）“阈值移动” (threshold-moving)

基于原始训练集进行学习, 并在用训练好的分类器进行预测时，把式子 $\dfrac{y'}{1-y'} = \dfrac{y}{1-y} \times \dfrac{m^+}{m^-}$ 嵌入到其决策过程中

5 请写出求解支撑矢量机优化问题的 SMO 算法流程, 其中在固定 λ_{i1} 和 λ_{i2} 以外的参数时，推导下面原优化问题的闭式解 λ_{i1}^* 和 λ_{i2}^*

$$\max_{\lambda_i \geq 0} \left\{ \sum_i \lambda_i - \frac{1}{2}(\sum_i \sum_j \lambda_i \lambda_j y_i y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j)) \right\}$$

$s.t. \sum_i \lambda_i y_i = 0$

因为存在约束 $\sum_i \lambda_i y_i = 0$，如果固定 $\lambda_{i1}, \lambda_{i2}$ 之外其他变量，则 $\lambda_{i1}, \lambda_{i2}$ 可由其他变量导出

$$\lambda_{i1} y_1 + \lambda_{i2} y_2 = -\sum_{i \neq 1,2} \lambda_i y_i = \varsigma$$
$$0 \leq \lambda_i \leq C, i = 1, 2$$

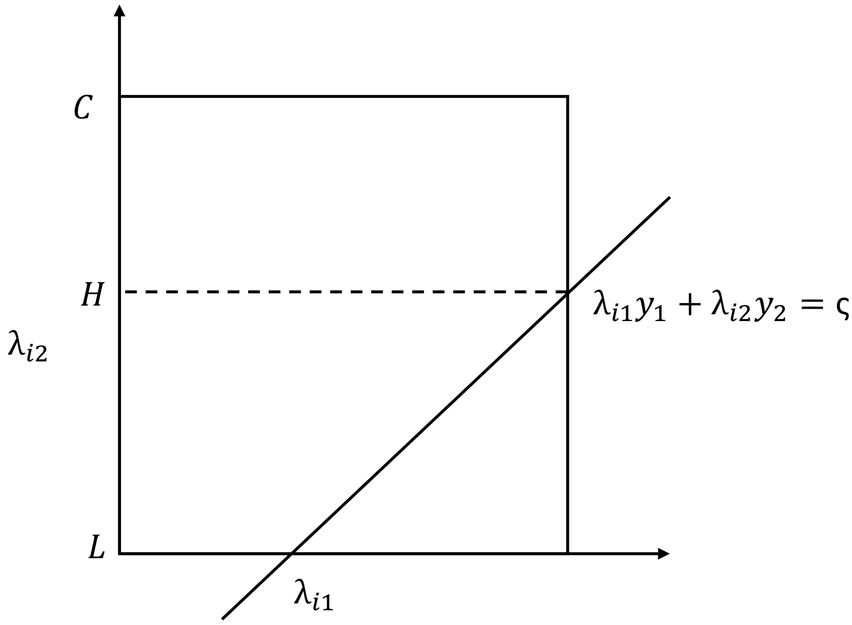
故题干最优化子问题，省略不含 $\lambda_{i1}, \lambda_{i2}$ 的常数项后，可写成：

$$\begin{aligned} \min_{\lambda_{i1}, \lambda_{i2}} W(\lambda_{i1}, \lambda_{i2}) = \frac{1}{2}(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1)\lambda_{i1}^2 + \frac{1}{2}(\boldsymbol{x}_2 \cdot \boldsymbol{x}_2)\lambda_{i2}^2 + y_1 y_2 (\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)\lambda_{i1}\lambda_{i2} \\ -(\lambda_{i1} + \lambda_{i2}) + y_1 \lambda_{i1} \sum_{i \neq 1,2} y_i \lambda_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_1) + y_2 \lambda_{i2} \sum_{i \neq 1,2} y_i \lambda_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_2) \end{aligned}$$

$s.t. \quad \lambda_{i1} y_1 + \lambda_{i2} y_2 = \varsigma$

$0 \le \lambda_i \le C, i=1,2$

求解上述两个变量的二次规划问题，先要分析约束条件，然后在该约束条件下求极小
只有两个变量，约束可以用二维空间中的图形表示



因为不等式约束 $0 \le \lambda_i \le C, i=1,2$ 使 $(\lambda_{i1}, \lambda_{i2})$ 在 $[0,C]\times[0,C]$ 内
且等式约束 $\lambda_{i1}y_1 + \lambda_{i2}y_2 = \varsigma$ 使 $(\lambda_{i1}, \lambda_{i2})$ 在平行于 $[0,C]\times[0,C]$ 的对角线的直线上
所以要求是目标函数在一条平行于对角线的线段上的最优值

又因为 $\lambda_{i1}y_1 + \lambda_{i2}y_2 = \varsigma$，其中 ς 为常数, 所以上述两个变量的最优化问题实质上是单变量的最优化问题，不妨设该问题是 λ_{i2} 的最优化问题

设问题的初始可行解为 $\lambda_{i1}, \lambda_{i2}$, 最优解 $\lambda_{i1}^*, \lambda_{i2}^*$, 设在沿着约束方向未考虑不等式 $0 \le \lambda_i \le C, i=1,2$ 约束时 λ_{i2} 的最优解是 $\lambda_{i2}^{*,unc}$
由图可以看出，$L \le \lambda_{i2}^* \le H$, 其中 L, H 是 λ_{i2}^* 所在的平行于对角线的线段端点的界

(1)$y_1 \ne y_2$, 则 $L = max(0, \lambda_{i2} - \lambda_{i1})$, $H = min(C, C + \lambda_{i2} - \lambda_{i1})$

(2)$y_1 \ne y_2$, 则 $L = max(0, \lambda_{i2} + \lambda_{i1} - C)$, $H = min(C, \lambda_{i2} + \lambda_{i1})$

接着，先求沿约束方向未考虑不等式 $0 \le \lambda_i \le C, i=1,2$ 约束时 λ_{i2} 的最优解 $\lambda_{i2}^{*,unc}$
再求考虑不等式 $0 \le \lambda_i \le C, i=1,2$ 约束后的解 λ_{i2}^*

令 $g(x) = \sum_{i=1} \lambda_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b$

$$E_i = g(x_i) - y_i = \left(\sum_{i=1} \lambda_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b \right) - y_i, i=1,2$$

$$v_i = \sum_{j \ne 1,2} y_i \lambda_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) = g(x_i) - \sum_{j=1}^{2} \lambda_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) - b, i=1,2$$

目标函数可写成

$$\begin{aligned} W(\lambda_{i1},\lambda_{i2}) = &\frac{1}{2}(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)\lambda_{i1}^2+\frac{1}{2}(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)\lambda_{i2}^2+y_1y_2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)\lambda_{i1}\lambda_{i2} \\ &-(\lambda_{i1}+\lambda_{i2})+y_1\lambda_{i1}v_1+y_2\lambda_{i2}v_2 \end{aligned}$$

由 $\lambda_{i1}y_1=\varsigma-\lambda_{i2}y_2, y_i^2=1$, 故 $\lambda_{i1}=(\varsigma-\lambda_{i2}y_2)y_1$
代入 $W(\lambda_{i1},\lambda_{i2})$ 得

$$\begin{aligned} W(\lambda_{i2}) = &\frac{1}{2}(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)(\varsigma-\lambda_{i2}y_2)^2+\frac{1}{2}(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)\lambda_{i2}^2+y_2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)(\varsigma-\lambda_{i2}y_2)\lambda_{i2} \\ &-(\varsigma-\lambda_{i2}y_2)y_1-\lambda_{i2}+(\varsigma-\lambda_{i2}y_2)v_1+y_2\lambda_{i2}v_2 \end{aligned}$$

对 λ_{i2} 求导并令其等于零

$$\begin{aligned} \frac{\partial W}{\partial\lambda_{i2}} = &(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)\lambda_{i2}+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)\lambda_{i2}-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)\lambda_{i2}-(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)\varsigma\lambda_{i2} \\ &+(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)\varsigma\lambda_{i2}+y_1y_2-1-y_2v_1+y_2v_2=0 \end{aligned}$$

$((\boldsymbol{x}_1\cdot\boldsymbol{x}_1)+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2))\lambda_{i2}$
$=y_2(y_2-y_1+\varsigma(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)-\varsigma(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)+v_1-v_2)$

$$= y_2\left[y_2-y_1+\varsigma(\boldsymbol{x}_1\cdot\boldsymbol{x}_1)-\varsigma(\boldsymbol{x}_1\cdot\boldsymbol{x}_2)+\left(g(x_1)-\sum_{j=1}^{2}\lambda_jy_j(\boldsymbol{x}_1\cdot\boldsymbol{x}_j)-b\right)\right.$$
$$\left.-\left(g(x_2)-\sum_{j=1}^{2}\lambda_jy_j(\boldsymbol{x}_2\cdot\boldsymbol{x}_j)-b\right)\right]$$

代入 $\varsigma=\lambda_{i1}y_1+\lambda_{i2}y_2$ 得 $((\boldsymbol{x}_1\cdot\boldsymbol{x}_1)+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2))\lambda_{i2}^{*,unc}$
$=y_2(((\boldsymbol{x}_1\cdot\boldsymbol{x}_1)+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2))\lambda_{i2}y_2+y_2-y_1+g(x_1)+g(x_2))$
$=((\boldsymbol{x}_1\cdot\boldsymbol{x}_1)+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2))\lambda_{i2}+y_2(E_1-E_2)$

令 $\eta=((\boldsymbol{x}_1\cdot\boldsymbol{x}_1)+(\boldsymbol{x}_2\cdot\boldsymbol{x}_2)-2(\boldsymbol{x}_1\cdot\boldsymbol{x}_2))$ 得到

$$\lambda_{i2}^{*,unc}=\lambda_{i2}+\frac{y_2(E_1-E_2)}{\eta}$$

为满足不等式约束条件，将上式限制在 $[L,H]$ 间，可得 λ_{i2}^* 表达式

$$\lambda_{i2}^*=\begin{cases} H, & \lambda_{i2}^{*,unc}>H \\ \lambda_{i2}^{*,unc}, & L\le\lambda_{i2}^{*,unc}\le H \\ L, & \lambda_{i2}^{*,unc}<L \end{cases}$$

$$\lambda_{i1}^*=\lambda_{i1}+y_1y_2(\lambda_{i2}-\lambda_{i2}^*)$$