PDM 的优化

Teng Zhang

2021 年 1 月 5 日

目前的优化问题为

$$\min_{\boldsymbol{w}} \frac{1}{2m} \sum_{i \in [m]} \max \left\{ 0, 1 - \frac{y_i \boldsymbol{w}^\top \boldsymbol{x}_i}{\bar{\gamma}} \right\}^2, \quad \text{s.t. } \bar{\gamma} > 0, \ \|\boldsymbol{w}\|_2 = 1$$

这个形式是有问题的，两个约束都是多余的。注意间隔均值 $\bar{\gamma} = \boldsymbol{w}^\top \overline{y\boldsymbol{x}}$ 关于 $\boldsymbol{w}$ 也是线性的，所以那个比值项是齐次的，放缩 $\boldsymbol{w}$ 不影响目标函数值。若求得的 $\boldsymbol{w}$ 使得 $\bar{\gamma} < 0$，则 $-\boldsymbol{w}$ 可使得 $\bar{\gamma} > 0$ 且不改变目标函数值、不破坏第二个约束。同理第二个约束也是多余的，若求得的 $\boldsymbol{w}$ 范数不为 1，可以将其缩放使范数为 1 且不改变目标函数值。

1 重新优化

将两个约束都去掉，问题变成

$$\min_{\boldsymbol{w}} \frac{1}{2m} \sum_{i \in [m]} \max \left\{ 0, 1 - \frac{y_i \boldsymbol{w}^\top \boldsymbol{x}_i}{\bar{\gamma}} \right\}^2$$

由于放缩 $\boldsymbol{w}$ 不影响目标函数值，不妨放缩使 $\bar{\gamma} = 1$，于是问题可进一步写成

$$\min_{\boldsymbol{w}} \frac{1}{2m} \sum_{i \in [m]} \max \left\{ 0, 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i \right\}^2, \quad \text{s.t. } \sum_{i \in [m]} y_i \boldsymbol{w}^\top \boldsymbol{x}_i = m$$

这个形式直接看也是合理的，最优情况是所有样本间隔均为 1，这样损失为零；若有样本间隔 > 1，由于间隔的和为 m，必然存在另一个样本间隔 < 1，这样就会产生损失。

记 $\epsilon_i = \max\{0, 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i\}$，则问题可写成

$$\min_{\boldsymbol{w}, \epsilon_i} \frac{1}{2m} \sum_{i \in [m]} \epsilon_i^2, \quad \text{s.t. } \epsilon_i \geq 0, \ \epsilon_i \geq 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i, \ \sum_{i \in [m]} y_i \boldsymbol{w}^\top \boldsymbol{x}_i = m$$

其中约束 $\epsilon_i \geq 0$ 是多余的，因为若求得的 $\epsilon_i < 0$，令其为零不破坏约束 $\epsilon_i \geq 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i$ 且可以使目标函数值变小，因此不存在这种情况。最终的问题为

$$\min_{\boldsymbol{w}, \epsilon_i} \frac{1}{2m} \sum_{i \in [m]} \epsilon_i^2, \quad \text{s.t. } \epsilon_i \geq 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i, \ \sum_{i \in [m]} y_i \boldsymbol{w}^\top \boldsymbol{x}_i = m \tag{1}$$

目标函数是二次的，约束都是线性的，因此这是一个凸二次规划，直接调用成熟的 QP 求解器即可。

记 $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m] \in \mathbb{R}^{d \times m}$，$\mathbf{Y} = \operatorname{diag}\{y_1, y_2, \ldots, y_m\}$，$\boldsymbol{\epsilon} = [\epsilon_1; \epsilon_2; \ldots; \epsilon_m]$，式 (**??**) 的矩阵形式为

$$\min_{\boldsymbol{w},\boldsymbol{\epsilon}} \ \frac{1}{2m} \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}, \quad \text{s.t. } \boldsymbol{\epsilon} \geq \boldsymbol{e} - \mathbf{Y}\mathbf{X}^\top \boldsymbol{w}, \ \boldsymbol{e}^\top \mathbf{Y}\mathbf{X}^\top \boldsymbol{w} = m$$

记 $\boldsymbol{u} = [\boldsymbol{w}; \boldsymbol{\epsilon}] \in \mathbb{R}^{d+m}$，写成 QP 的标准形式是

$$\min_{\boldsymbol{u}} \ \frac{1}{2} \boldsymbol{u}^\top \begin{bmatrix} \mathbf{0} & \\ & \frac{\mathbf{I}}{m} \end{bmatrix} \boldsymbol{u}, \quad \text{s.t. } [-\mathbf{Y}\mathbf{X}^\top, -\mathbf{I}] \boldsymbol{u} \leq -\boldsymbol{e}, \ [\mathbf{X}\mathbf{Y}\boldsymbol{e}; \mathbf{0}]^\top \boldsymbol{u} = m$$

这样就可以直接掉包了。

对偶函数为

$$L = \frac{1}{2m} \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} - \boldsymbol{\beta}^\top (\boldsymbol{\epsilon} - \boldsymbol{e} + \mathbf{Y}\mathbf{X}^\top \boldsymbol{w}) - \mu (\boldsymbol{e}^\top \mathbf{Y}\mathbf{X}^\top \boldsymbol{w} - m)$$

令 L 关于 $\boldsymbol{w}$ 的偏导数为零

$$\frac{\partial L}{\partial \boldsymbol{w}} = -\mathbf{X}\mathbf{Y}\boldsymbol{\beta} - \mu \mathbf{X}\mathbf{Y}\boldsymbol{e} = \mathbf{0} \Longrightarrow \mathbf{X}\mathbf{Y}\boldsymbol{\beta} + \mu \mathbf{X}\mathbf{Y}\boldsymbol{e} = \mathbf{0}$$

令 L 关于 $\boldsymbol{\epsilon}$ 的偏导数为零

$$\frac{\partial L}{\partial \boldsymbol{\epsilon}} = \frac{\boldsymbol{\epsilon}}{m} - \boldsymbol{\beta} = \mathbf{0} \Longrightarrow \boldsymbol{\epsilon} = m \boldsymbol{\beta}$$

故对偶问题为

$$\max_{\boldsymbol{\beta},\mu} \ -\frac{m}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} + \boldsymbol{\beta}^\top \boldsymbol{e} + \mu m, \quad \text{s.t. } \mathbf{X}\mathbf{Y}\boldsymbol{\beta} + \mu \mathbf{X}\mathbf{Y}\boldsymbol{e} = \mathbf{0}, \ \boldsymbol{\beta} \geq \mathbf{0}$$

记 $\boldsymbol{\alpha} = [\boldsymbol{\beta}; \mu] \in \mathbb{R}^{m+1}$，写成 QP 的标准形式是

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2} \boldsymbol{\alpha}^\top \begin{bmatrix} m\mathbf{I} & \\ & 0 \end{bmatrix} \boldsymbol{\alpha} - [\boldsymbol{e}; m]^\top \boldsymbol{\alpha}, \quad \text{s.t. } [\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Y}\boldsymbol{e}] \boldsymbol{\alpha} = \mathbf{0}, \ \begin{bmatrix} -\mathbf{I} & \\ & 0 \end{bmatrix} \boldsymbol{\alpha} \leq \mathbf{0}$$

原问题的变量个数为 $d+m$，有 m 个线性不等式和 1 个线性等式约束；对偶问题的变量个数为 $m+1$，有 d 个线性等式约束，不等式约束很简单可忽略，因此直接调包求解的话，可能对偶问题效率会稍微高点，但不绝对，跟 m 和 d 的值有关系。此外，由于两个问题都有大量形式复杂的约束，所以目前看不存在像 SVM、ODM 那样高效的基于坐标下降的解法。

2 核化

传统核方法都有 $\boldsymbol{w}$ 的二次项，再结合损失函数 $y\boldsymbol{w}^\top \boldsymbol{x} \geq 1 - \epsilon$ 这样的一次项，对偶函数一求导就能得到最优的 $\boldsymbol{w}$ 是由 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ 线性表出的，即所谓的表示定理，这样回代到二次项中就出现了 $\boldsymbol{x}_i^\top \boldsymbol{x}_j$ 的内积形式，之后就可以引入核函数了。

PDM 目前的形式是无法直接核化的，因为目标函数中没有 $\boldsymbol{w}$ 的二次项，而间隔损失函数和 $\boldsymbol{w}$ 的二次项似乎不兼容 (开头已述)。我目前想到是利用 $\mathbf{K}\boldsymbol{\alpha}=\mathbf{0} \Longleftrightarrow \mathbf{K}^\top\mathbf{K}\boldsymbol{\alpha}=\mathbf{0}$ 这个结论，于是对偶问题中的等式约束可以改写为

$$[\mathbf{X}\mathbf{Y},\mathbf{X}\mathbf{Y}\boldsymbol{e}]^\top[\mathbf{X}\mathbf{Y},\mathbf{X}\mathbf{Y}\boldsymbol{e}]\boldsymbol{\alpha}=\mathbf{0}$$

这样就有了 $\mathbf{X}^\top\mathbf{X}$ 这样的形式，且 $\mathbf{X}$ 只以 $\mathbf{X}^\top\mathbf{X}$ 的形式出现，因此可以引入核函数。求得 $\boldsymbol{\beta}$ 和 μ 后，根据 $\boldsymbol{\epsilon}=m\boldsymbol{\beta}$ 可以得到 $\boldsymbol{\epsilon}$，利用互补松弛条件

$$\beta_i(\epsilon_i-1+y_i\boldsymbol{w}^\top\phi(\boldsymbol{x}_i))=0$$

对于大于零的 β_i，可以求得 $\boldsymbol{w}^\top\phi(\boldsymbol{x}_i)=y_i-y_i\epsilon_i=y_i-y_im\beta_i$，这些都是有损失的样本，即“支持向量”。设所有的支持向量为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_t$，假如表示定理成立，设

$$\boldsymbol{w}=a_1\phi(\boldsymbol{x}_1)+a_2\phi(\boldsymbol{x}_2)+\cdots+a_t\phi(\boldsymbol{x}_t) \tag{2}$$

两边依次左乘 $\phi(\boldsymbol{x}_1)^\top,\ldots,\phi(\boldsymbol{x}_t)^\top$ 可以得到 t 个方程，求解线性方程组可以得到 $a_1,\ldots,a_t$，最后预测

$$\boldsymbol{w}^\top\phi(\boldsymbol{z})=(a_1\phi(\boldsymbol{x}_1)+a_2\phi(\boldsymbol{x}_2)+\cdots+a_t\phi(\boldsymbol{x}_t))^\top\phi(\boldsymbol{z})$$

所以最终问题就是表示定理是否成立，表示定理要求目标函数中的正则项 $\Omega(\|\boldsymbol{w}\|)$，其中 $\Omega(\cdot)$ 是单调增函数，PDM 没有正则项，可以理解为这里的 $\Omega(\cdot)$ 恒为零函数，但零函数也是单调增函数，所以表示定理成立？