

统计学习方法读书报告

09017244 郑健雄

本周进度:统计学习方法第七章, 负责 7.4.2-7.4.3 讲解。

下周进度:统计学习方法第八章, 并阅读一部分第九章的 EM 算法。

一、 问题与解答

会上讨论的问题:

1.(别人提出)KKT 条件违反严重是如何判断的?

讨论结果: 在 SMO 算法中, KKT 条件被简化成为三个简单条件:

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \geq 1 \quad (7.111)$$

$$0 < \alpha_i < C \Leftrightarrow y_i g(x_i) = 1 \quad (7.112)$$

$$\alpha_i = C \Leftrightarrow y_i g(x_i) \leq 1 \quad (7.113)$$

严重违反 KKT 条件是说针对任意一个条件, 样本点和参数值的不匹配程度, 以第一条规则为例, 如果对于某个样本, 其 a 参数为 0, 而 $y g(x)$ 参数远远小于 1, 则可以认为是严重违反 KKT 条件。对于另外两条规则也是如此。算法以这个不匹配程度作为选择参数的依据。

2. (别人提出)最大化 $1/||w||$ 与 最小化 $0.5||w||^2$ 等价, 为什么一定要取

$0.5||w||^2$, 而不是 $||w||$?

讨论结果: 实际上可以使用 $||w||$ 作为目标函数, 但是由于后续需要进行一系列变换,

$0.5||w||^2$ 在后面的变换中求导数更方便, 得出的形式更简洁, 所以使用 $0.5||w||^2$ 作为目标函数。

3. (别人提出)7.4.2 中新的 b 值的求解结论如何理解。

如果 $\alpha_1^{new}, \alpha_2^{new}$ 同时满足条件 $0 < \alpha_i^{new} < C, i=1,2$, 那么 $b_1^{new} = b_2^{new}$. 如果 $\alpha_1^{new}, \alpha_2^{new}$ 是 0 或者 C , 那么 b_1^{new} 和 b_2^{new} 以及它们之间的数都是符合 KKT 条件的阈值, 这时选择它们的中点作为 b^{new} .

讨论结果: 要解决的对偶问题为

$$\max_{\alpha} \theta_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \quad (1)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, i = 1, \dots, m \quad (2)$$

$$\sum_{i=1}^m \alpha_i y^{(i)} = 0 \quad (3)$$

根据以上的约束关系, 我们可以得出以下公式:

$$\alpha_1 y^{(1)} = - \sum_{i=2}^m \alpha_i y^{(i)} \quad (4)$$

对其进行转化得到:

$$\alpha_1 = -y^{(1)} \sum_{i=2}^m \alpha_i y^{(i)} \quad (5)$$

而根据 SMO 算法的性质，要选择两个参数进行优化，可以表示为

$$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = - \sum_{i=3}^m \alpha_i y^{(i)} = \zeta \quad (6)$$

根据这个约束可以消去一个变量：

$$\alpha_1 = (\zeta - \alpha_2 y^{(2)}) y^{(1)} \quad (7)$$

那么目标函数可以变为：

$$\theta_{\mathcal{D}}(\alpha_1, \alpha_2, \dots, \alpha_m) = \theta_{\mathcal{D}}((\zeta - \alpha_2 y^{(2)}) y^{(1)}, \alpha_2, \dots, \alpha_m) \quad (8)$$

KKT 条件：

$$\alpha_i = 0 \Leftrightarrow y^{(i)}(w^T x^{(i)} + b) \geq 1 \quad (9)$$

$$\alpha_i = C \Leftrightarrow y^{(i)}(w^T x^{(i)} + b) \leq 1 \quad (10)$$

$$0 < \alpha_i < C \Leftrightarrow y^{(i)}(w^T x^{(i)} + b) = 1 \quad (11)$$

根据 KKT 条件，当我们更新完 **a** 参数时，**b** 参数的两个更新等式如下：(a1,a2 符合范围)

$$b_1 = y^{(1)} - \sum_{i=1}^m \alpha_i y^{(i)} \langle x^{(i)}, x^{(1)} \rangle \quad (12)$$

$$b_2 = y^{(2)} - \sum_{i=1}^m \alpha_i y^{(i)} \langle x^{(i)}, x^{(2)} \rangle \quad (13)$$

若 **a1**, **a2** 都在 0 到 C 之间，则求最值点有：

$$\frac{\partial \theta_{\mathcal{D}}}{\partial \alpha_2} = \frac{\partial \alpha_1}{\partial \alpha_2} + 1 - \sum_{i=1}^m y^{(1)} y^{(i)} \frac{\partial \alpha_1}{\partial \alpha_2} \alpha_i \langle x^{(1)}, x^{(i)} \rangle - \sum_{i=1}^m y^{(2)} y^{(i)} \alpha_i \langle x^{(2)}, x^{(i)} \rangle = 0$$

而根据公式 7，有

$$\frac{\partial \alpha_1}{\partial \alpha_2} = -y^{(1)}y^{(2)}$$

将这个等式带入上面的最值求解方程，得出以下方程：

$$y^{(1)} - \sum_{i=1}^m \alpha_i y^{(i)} \langle x^{(i)}, x^{(1)} \rangle = y^{(2)} - \sum_{i=1}^m \alpha_i y^{(i)} \langle x^{(i)}, x^{(2)} \rangle$$

此处可以看出等式两边代表了 b_1 和 b_2 ，说明两个参数相等。

而如果 a_1 或者 a_2 不同时满足 0 到 C 的范围，则根据 KKT 条件，可以得出 b_1 ， b_2 所代表的可以取值区间，然后选择区间中点作为更新参数即可。

二、下周计划安排

下周计划阅读完第八章，并看一部分第九章的内容。

三、读书收获

第七章内容总结：

第七章的主体内容是支持向量机，其主线是线性可分支持向量机→线性支持向量机→非线性支持向量机，三种支持向量机结构相似，并且有简单到复杂。

线性可分支持向量机可以说是线性支持向量机的特例，采用硬间隔最大化，其条件较为苛刻，必须是线性可分数据集才可以使用，应用比较局限。其由少量在边界上的支持向量来确定。

而线性支持向量机引入了松弛变量，使用了软间隔最大化的思路，使得总体线性可分而局部存在噪声点的数据集也可以被分析，模型得到了进一步扩展。支持向量的定义也有所变化。

非线性支持向量机是线性支持向量机的推广和拓展，可以处理比较没有规则的数据集，但是其原理和线性支持向量机是基本一样的，其特殊的地方在于使用了核方法将不规则的数据集映射到特征空间，使其在特征空间符合线性可分，再在特征空间进行线性支持向量机的算法来处理。

为了说明非线性向量机，书中对核方法进行了介绍，给出了正定核充要条件为 Gram 矩阵为半正定矩阵，同时给出了一些常用的核函数，这样就可以在不需要得出映射空间和映射函数的情况下直接得到核函数，得出需要的结果。

最后书中介绍了 SMO 算法，其思路是每次只更新两个参数，直到所有参数满足 KKT 条件，此时得出最优解。该算法的思路比较简单，而且每一个子问题计算起来并不复杂，整体很高效。

第七章收获：阅读本章内容时，发现有些部分由于没有凸优化基础很难理解，准备在阅读过程中再阅读一些凸优化的相关资料进行学习。