

支持向量机(下篇)

杨航锋

支持向量机的二次优化模型

在支持向量机(上篇)中推导出了对偶形式下的二次优化模型,为了不失一般性引入核函数,即求解下述问题

$$\begin{aligned} \arg \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1, j=1}^m \alpha_i \alpha_j y^{(i)} y^{(j)} k(x^{(i)}, x^{(j)}) - \sum_{i=1}^m \alpha_i \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^m \alpha_i y^{(i)} = 0 \\ \alpha_i \geq 0, i = 1, 2, \dots, m \end{cases} \end{aligned}$$

想知道上式详细推导过程的读者可以去[ML-NOTE](#)下载,当然你也可以在我的[知乎专栏](#)进行阅读。

SMO算法求解模型

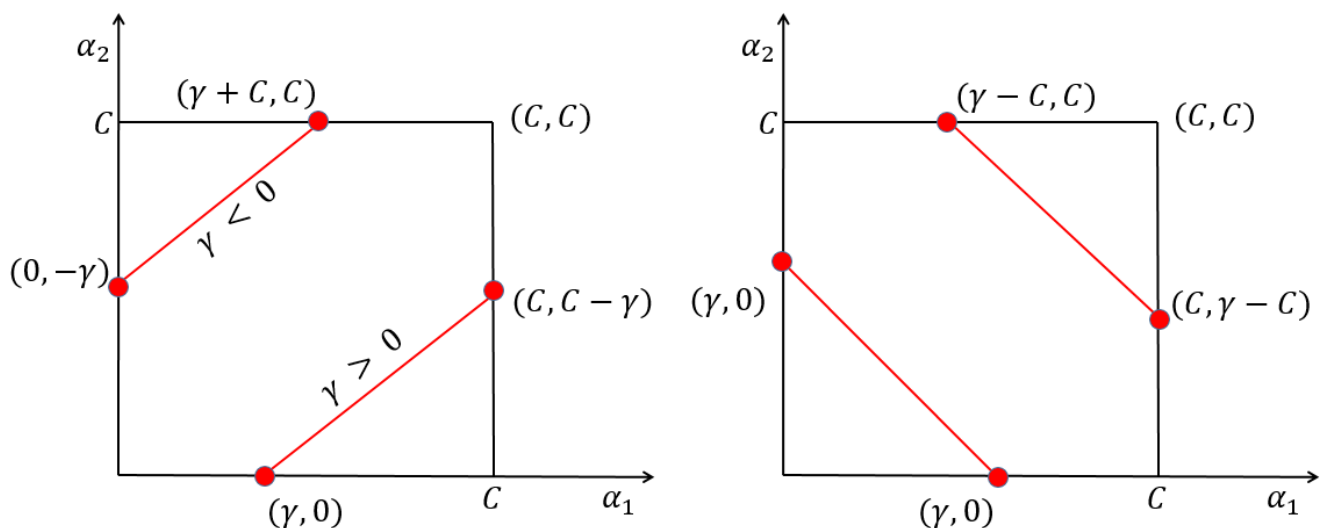
SMO算法的基本思想

在SMO算法迭代过程中,每一次需要选择一对变量 $\{(\alpha_i, \alpha_j) | 1 \leq i \leq m, 1 \leq j \leq m, i \neq j\}$ 其他 $m-2$ 个变量看成常数项,因为在支持向量机模型的约束条件中需要满足 $\sum_{i=1}^m \alpha_i y^{(i)} = 0$, 因此如果改变其中的一个 α_i 那么至少需要改变一个 α_j 才能满足约束条件,故SMO算法将一个复杂的二次优化问题转化为求解若干个比较简单的二变量优化问题。为了方便推导不妨选择 α_1, α_2 作为初次迭代对象,记 $k_{i,j} = k(x^{(i)}, x^{(j)})$, $T = \sum_{i=3}^m \alpha_i$ 则支持向量机的二次优化模型可以改写成

$$\begin{aligned} \arg \min_{\alpha_1, \alpha_2} \quad & \frac{1}{2} k_{1,1} \alpha_1^2 + \frac{1}{2} k_{2,2} \alpha_2^2 + y^{(1)} y^{(2)} k_{1,2} \alpha_1 \alpha_2 - (\alpha_1 + \alpha_2) + y^{(1)} \alpha_1 \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1} \\ & + y^{(2)} \alpha_2 \sum_{i=3}^m y^{(i)} \alpha_i k_{i,2} - T \\ \text{s.t.} \quad & \begin{cases} \alpha_1 y^{(1)} + \alpha_2 y^{(2)} = - \sum_{i=3}^m \alpha_i y^{(i)} = \gamma \\ \alpha_i \geq 0, i = 1, 2, \dots, m \end{cases} \end{aligned}$$

SMO算法求解过程

在SVM中 α 的约束条件为 $0 \leq \alpha \leq C$, 已知 $\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = \gamma$



$$y^{(1)} \neq y^{(2)} \Rightarrow \alpha_1 - \alpha_2 = \gamma$$

$$y^{(1)} = y^{(2)} \Rightarrow \alpha_1 + \alpha_2 = \gamma$$

如图所示 α_1 、 α_2 之间的关系被限制在正方形框中的线段中，所以二变量的优化问题实际上仅仅是一个变量的优化问题。记上一轮迭代得到的解为 α_1^{old} 、 α_2^{old} ，沿着约束方向 α_2 没有经过约束的解为 $\alpha_2^{new, unr}$ ，该轮迭代结束后的解为 α_1^{new} 、 α_2^{new} 。不妨假设 L 、 H 分别为 α_2^{new} 的下界和上界，因此有

$$L \leq \alpha_2^{new} \leq H$$

成立，那么怎么求解 L 、 H 呢？有已知条件当 $y^{(1)} \neq y^{(2)}$ 时有(根据上图分析)

$$\begin{cases} \alpha_1^{new} - \alpha_2^{new} = \gamma \\ \alpha_1^{old} - \alpha_2^{old} = \gamma \end{cases} \Rightarrow \alpha_1^{new} = \alpha_2^{new} + (\alpha_1^{old} - \alpha_2^{old})$$

$$\begin{cases} 0 \leq \alpha_1^{new} \leq C \\ 0 \leq \alpha_2^{new} \leq C \end{cases} \Rightarrow \begin{cases} 0 \leq \alpha_2^{new} + (\alpha_1^{old} - \alpha_2^{old}) \leq C \\ 0 \leq \alpha_2^{new} \leq C \end{cases}$$

$$\Rightarrow \begin{cases} \alpha_2^{old} - \alpha_1^{old} \leq \alpha_2^{new} \leq C + (\alpha_2^{old} - \alpha_1^{old}) \\ 0 \leq \alpha_2^{new} \leq C \end{cases}$$

所以

$$\begin{cases} L = \max\{0, \alpha_2^{old} - \alpha_1^{old}\} \\ H = \min\{C, C + \alpha_2^{old} - \alpha_1^{old}\} \end{cases}$$

同理可得，当 $y^{(1)} = y^{(2)}$ 时有

$$\begin{cases} L = \max\{0, \alpha_2^{old} + \alpha_1^{old} - C\} \\ H = \min\{C, \alpha_2^{old} + \alpha_1^{old}\} \end{cases}$$

也就是说当对二元函数 $\phi(\alpha_1, \alpha_2)$ 求导得到无约束情况下的解 $\alpha_2^{new, unr}$ ，则最终 α_2^{new} 应该为

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{new, unr} > H \\ \alpha_2^{new, unr}, & L \leq \alpha_2^{new, unr} \leq H \\ L, & \alpha_2^{new, unr} < L \end{cases}$$

现在问题就转化为怎么求解 $\alpha_2^{new, unr}$ 了。

对 $\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = -\sum_{i=3}^m \alpha_i y^{(i)} = \gamma$ 两边同乘以 $y^{(1)}$ 可得 $\alpha_1 (y^{(1)})^2 + \alpha_2 y^{(1)} y^{(2)} = \gamma y^{(1)}$ 又因为 $(y^{(1)})^2 = 1$ 故 $\alpha_1 = (\gamma - \alpha_2 y^{(2)}) y^{(1)}$, 然后将 α_1 回带到原二次优化函数中消去 α_1 可得

$$\begin{aligned} \arg \min_{\alpha_2} \phi(\alpha_2) = & \frac{1}{2} k_{1,1} (\gamma - \alpha_2 y^{(2)})^2 + \frac{1}{2} k_{2,2} \alpha_2^2 + y^{(2)} k_{1,2} (\gamma - \alpha_2 y^{(2)}) \alpha_2 - ((\gamma - \alpha_2 y^{(2)}) y^{(1)} + \alpha_2) \\ & + (\gamma - \alpha_2 y^{(2)}) \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1} + y^{(2)} \alpha_2 \sum_{i=3}^m y^{(i)} \alpha_i k_{i,2} - T \end{aligned}$$

记 $u = \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1}$ 、 $v = \sum_{i=3}^m y^{(i)} \alpha_i k_{i,2}$, 常数 T 不影响 $\phi(\alpha_2)$ 的驻点因而省略, 则

$$\begin{aligned} \arg \min_{\alpha_2} \phi(\alpha_2) = & \frac{1}{2} k_{1,1} (\gamma - \alpha_2 y^{(2)})^2 + \frac{1}{2} k_{2,2} \alpha_2^2 + y^{(2)} k_{1,2} (\gamma - \alpha_2 y^{(2)}) \alpha_2 - ((\gamma - \alpha_2 y^{(2)}) y^{(1)} + \alpha_2) \\ & + (\gamma - \alpha_2 y^{(2)}) u + y^{(2)} \alpha_2 v \end{aligned}$$

至此, $\phi(\alpha_2)$ 就是一个一元函数了, 要求驻点只需令其导数为 0 即可

$$\begin{aligned} \phi'(\alpha_2) = & -k_{1,1} y^{(2)} (\gamma - \alpha_2 y^{(2)}) + k_{2,2} \alpha_2 + k_{1,2} y^{(2)} \gamma - 2k_{1,2} \alpha_2 \\ & + y^{(1)} y^{(2)} - 1 - u y^{(2)} + v y^{(2)} \\ = & (k_{1,1} - 2k_{1,2} + k_{2,2}) \alpha_2 - k_{1,1} y^{(2)} \gamma + k_{1,2} y^{(2)} \gamma + y^{(1)} y^{(2)} \\ & - 1 - u y^{(2)} + v y^{(2)} \\ = & 0 \end{aligned}$$

在求解上式之前需要做一些准备工作:

1. 优化前的解 α_1^{old} 、 α_2^{old} 满足下面等式约束条件

$$\alpha_1^{old} y^{(1)} + \alpha_2^{old} y^{(2)} = \gamma$$

2. 假设支持向量机的超平面方程为 $f(x) = w^T x + b$, 带入 w 的表达式则

$$f(x) = \sum_{i=1}^m \alpha_i y^{(i)} k(x^{(i)}, x) + b, \quad f(x^{(i)}) \text{ 表示样本 } x^{(i)} \text{ 的预测值, } y^{(i)} \text{ 表示样本 } x^{(i)} \text{ 的真实值,}$$

定义 $E^{(i)}$ 为预测值和真实值之差, 即

$$E^{(i)} = f(x^{(i)}) - y^{(i)}$$

3. 由于 $u = \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1}$ 、 $v = \sum_{i=3}^m y^{(i)} \alpha_i k_{i,2}$, 所以

$$u = f(x^{(1)}) - \sum_{j=1}^2 \alpha_j y^{(j)} k_{j,1} - b$$

$$v = f(x^{(2)}) - \sum_{j=1}^2 \alpha_j y^{(j)} k_{j,2} - b$$

下面接着求 $\phi(\alpha_2)$ 的驻点，并带入准备工作中的式子化简可得

$$\begin{aligned} (k_{1,1} - 2k_{1,2} + k_{2,2})\alpha_2^{new,unr} &= y^{(2)}(y^{(2)} - y^{(1)} + \gamma k_{1,1} - \gamma k_{1,2} + u - v) \\ &= y^{(2)} \left[y^{(2)} - y^{(1)} + \gamma k_{1,1} - \gamma k_{1,2} + (f(x^{(1)}) - \sum_{j=1}^2 y^{(j)} \alpha_j k_{1,j} - b) \right. \\ &\quad \left. (f(x^{(2)}) - \sum_{j=1}^2 y^{(j)} \alpha_j k_{1,j} - b) \right] \\ &= y^{(2)} \left[y^{(2)}(k_{1,1} - 2k_{1,2} + k_{2,2})\alpha_2^{old} + (f(x^{(1)} - y^{(1)}) - (f(x^{(2)} - y^{(2)})) \right] \\ &= (k_{1,1} - 2k_{1,2} + k_{2,2})\alpha_2^{old} + y^{(2)}(E^{(1)} - E^{(2)}) \end{aligned}$$

最后可以得到 $\alpha_2^{new,unr}$ 的表达式为

$$\alpha_2^{new,unr} = \alpha_2^{old} + \frac{y^{(2)}(E^{(1)} - E^{(2)})}{k_{1,1} - 2k_{1,2} + k_{2,2}}$$

因此根据 $\alpha_2^{new,unr}$ 和 α_2^{new} 的关系式就可以迭代更新 α_2^{new} ，然后再利用 α_1^{new} 和 α_2^{new} 的线性关系 $\alpha_1^{old} y^{(1)} + \alpha_2^{old} y^{(2)} = \alpha_1^{new} y^{(1)} + \alpha_2^{new} y^{(2)}$ 就可以更新 α_1^{new} 。

在每次更新完两个 α_i 、 α_j 之后，需要重新计算出阈值 b 因为 b 的值关系到 $f(x)$ 和下次迭代时误差 $E^{(i)}$ 的计算。简单回顾一下支持向量机中的KKT条件

$$\begin{cases} \alpha_i = 0 \iff y^{(i)}(w^T x^{(i)} + b) \geq 1, \\ \alpha_i = C \iff y^{(i)}(w^T x^{(i)} + b) \leq 1, \\ 0 < \alpha_i < C \iff y^{(i)}(w^T x^{(i)} + b) = 1, \end{cases}$$

为了使得被优化的样本都满足KKT条件，当 $0 < \alpha_1^{new} < C$ 时则相应的数据点为支持向量且有下式成立

$$y^{(1)}(w^T x^{(1)} + b) = 1 \Rightarrow \sum_{i=1}^m \alpha_i y^{(i)} k_{i,1} + b = y^{(1)}$$

于是新的 b_1^{new} 为

$$b_1^{new} = y^{(1)} - \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1} - \alpha_1^{new} y^{(1)} k_{1,1} - \alpha_2^{new} y^{(2)} k_{2,1}$$

根据 $E^{(1)}$ 的定义可知

$$E^{(1)} = \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1} + \alpha_1^{old} y^{(1)} k_{1,1} + \alpha_2^{old} y^{(2)} k_{2,1} + b^{old} - y^{(1)}$$

$$y^{(1)} - \sum_{i=3}^m \alpha_i y^{(i)} k_{i,1} = -E^{(1)} + \alpha_1^{old} y^{(1)} k_{1,1} + \alpha_2^{old} y^{(2)} k_{2,1} + b^{old}$$

回帶到 b_1^{new} 得到

$$b_1^{new} = -E^{(1)} - y^{(1)} k_{1,1} (\alpha_1^{new} - \alpha_1^{old}) - y^{(2)} k_{2,1} (\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

同理可得, 若 $0 < \alpha_2^{new} < C$ 那么

$$b_2^{new} = -E^{(2)} - y^{(1)} k_{1,2} (\alpha_1^{new} - \alpha_1^{old}) - y^{(2)} k_{2,2} (\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

如果 α_1^{new} 、 α_2^{new} 同时满足条件 $0 < \alpha_i^{new} < C, i = 1, 2$, 那么有 $b_1^{new} = b_2^{new}$; 如果 α_1^{new} 、 α_2^{new} 是 0 或者是 C , 那么 b_1^{new} 、 b_2^{new} 以及它们之间的数都是满足KKT条件的阈值, 这时取

$b^{new} = \frac{b_1^{new} + b_2^{new}}{2}$ 。最后还需要更新对应的 E_i^{new} , 其中 G 代表的所有支持向量 $x^{(j)}$ 的集合

$$E_i^{new} = \sum_G \alpha_j y^{(j)} k_{i,j} + b^{new} - y^{(i)}$$

以上就是SMO算法的完整流程。

SMO算法总结

假设输入空间是 $X \in \mathbb{R}^n$, $Y \in \{+1, -1\}$, 不妨假设含有 m 个样本数据 $(x^{(1)}, y^{(1)})$ 、 $(x^{(2)}, y^{(2)})$ 、 \dots 、 $(x^{(m)}, y^{(m)})$, 其中 $x^{(i)} \in X$ 、 $y^{(i)} \in Y$, 精度为 ε , 输出近似解为 $\hat{\alpha}$;

1. 取初值 $\alpha^0 = 0$, 令 $k = 0$

2. 选取待优化变量 α_1^k 、 α_2^k , 计算出新的 $\alpha_2^{new, unr}$

$$\alpha_2^{new, unr} = \alpha_2^{old} + \frac{y^{(2)} (E^{(1)} - E^{(2)})}{k_{1,1} - 2k_{1,2} + k_{2,2}}$$

3. 更新 α_2^{k+1}

$$\alpha_2^{k+1} = \begin{cases} H, & \alpha_2^{new, unr} > H \\ \alpha_2^{new, unr}, & L \leq \alpha_2^{new, unr} \leq H \\ L, & \alpha_2^{new, unr} < L \end{cases}$$

4. 利用 α_1^{k+1} 和 α_2^{k+1} 的关系求出 α_1^{k+1}

$$\alpha_1^{k+1} = \alpha_1^k + y^{(1)} y^{(2)} (\alpha_2^k - \alpha_2^{k+1})$$

5. 计算 b^{k+1} 和 E_i

6. 在精度 ε 范围内检查是否满足如下的终止条件, 其中 $f(x^{(i)}) = \sum_{j=1}^m \alpha_j y^{(j)} k_{j,i} + b$

$$\begin{cases} \sum_{i=1}^m \alpha_i y^{(i)} = 0, \\ 0 \leq \alpha_i \leq C, i = 1, 2, \dots, m \end{cases}$$

$$\begin{cases} \alpha_i^{k+1} = 0 \iff y^{(i)} f(x^{(i)}) \geq 1, \\ \alpha_i^{k+1} = C \iff y^{(i)} f(x^{(i)}) \leq 1, \\ 0 < \alpha_i^{k+1} < C \iff y^{(i)} f(x^{(i)}) = 1, \end{cases}$$

若满足则转 7；否则令 $k = k + 1$ 转到步骤 2；

7. 取 $\hat{\alpha} = \alpha^{k+1}$ 。