零基础学SVM-Support Vector Machine(二)

耳东陈
高校教师，机器学习与计算机视觉

四、线性SVM优化问题求解

4.1 基于拉格朗日乘子法的线性SVM优化问题模型

从第三章开始，我们花了相当长的时间来介绍优化技术。目的是为线性SVM问题的求解打好数学基础。接下来，让我们把目光转回公式(2.14)描述的线性SVM问题的基本模型，为了方便大家阅读，这里重新给出公式（2.14）

$$\min_{\boldsymbol{\omega},\gamma} \frac{1}{2}||\boldsymbol{\omega}||^2$$

$$\text{s. t. } y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma) \geq 1, \;\; i=1,2,\dots,m \qquad (2.14)$$

显然，这是一个具有多个不等式约束条件的优化问题，其拉格朗日函数可以写为：

$$L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha}) = \frac{||\boldsymbol{\omega}||^2}{2} + \sum_{i=1}^{m}\alpha_i(1-y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma)) \quad (3.21)$$

这里 $\boldsymbol{\omega}=[\omega_1,\omega_2,\dots,\omega_d]^T$ ， $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\dots,\alpha_m]^T$ 。该拉格朗日函数最优化的原始问题和对偶问题分别为：

原始问题： $\min\limits_{\boldsymbol{\omega},\gamma}\left[\max\limits_{\boldsymbol{\alpha}:\alpha_j\geq 0} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha})\right]$ （3.22）

对偶问题： $\max\limits_{\boldsymbol{\alpha}:\alpha_j\geq 0}\left[\min\limits_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha})\right]$ （3.23）

根据第三章的推导，我们对拉格朗日对偶问题进行求解，根据公式（3.23）首先求解

$$\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha}) = \min_{\boldsymbol{\omega},\gamma}\left[\frac{1}{2}||\boldsymbol{\omega}||^2 + \sum_{i=1}^{m}\alpha_i(1-y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma))\right] \quad (3.24)$$

为了求拉格朗日函数的极小值，分别令函数 $L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha})$ 对 $\boldsymbol{\omega},\gamma$ 求偏导，并使其等于0。

$$\frac{\partial L}{\partial \boldsymbol{\omega}} = 0 \implies \boldsymbol{\omega} = \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i \quad (3.25)$$

$$\frac{\partial L}{\partial \gamma} = 0 \implies 0 = \sum_{i=1}^{m}\alpha_i y_i \quad (3.26)$$

（这里介绍一下矢量积求导的基本公式，方便大家理解(3.25)的推导过程。假设 $\boldsymbol{f}=\boldsymbol{x}^T\boldsymbol{y}$ ， $\boldsymbol{x}=[x_1,x_2,\dots,x_d]^T$ ， $\boldsymbol{y}=[y_1,y_2,\dots,y_d]^T$ ，则有 $\frac{\partial f}{\partial \boldsymbol{x}}=\boldsymbol{y}$ ，有兴趣大家可以自己推导一下，非常简单!）

这里要特别提到的是，由于参数向量 $\boldsymbol{\omega}$ 是一个d维矢量，因此公式3.24是一个矢量方程，相当于d个标量方程。将公式（3.25）带入到公式（3.24）可以得到：

$$\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\alpha)=\frac{1}{2}\left[\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right]^T\left[\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right]+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

（3.27）

根据多项式乘法的基本规律（所有项和的积等于所有项积的和），不难明白（注意向量 $\boldsymbol{x}_i$ 是列向量）：

$$\left[\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right]^T\left[\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right]=\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \quad (3.28)$$

所以公式(3.27)可以化简为：

$$\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\alpha)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \quad (3.29)$$

将(3.29)带入对偶问题的公式(3.23)，同时考虑公式(3.26)给出的约束，可以将SVM的优化问题转变为：

$$\begin{aligned}&\max_{\alpha}\left[\sum_{i=1}^{m}\alpha_i-\tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\right]\\ &\text{s.t.}\ \ \sum_{i=1}^{m}\alpha_i y_i=0\\ &\qquad \alpha_j\geq 0, i=1,2,\ldots,m.\end{aligned} \quad (3.30)$$

编辑于 2018-04-08

SVM 机器学习 人工智能算法

不要熬夜 回复 耳东陈 (作者) 2018-09-29
非常感谢，受益匪浅！
赞

ZL LI 回复 耳东陈 (作者) 2019-02-23
下一篇啥时候更新哈？
赞

he斯基 2017-10-03
跟上
赞

MG1231 2017-10-05
谢谢老师！但是图好像挂了。。
赞

耳东陈 (作者) 回复 MG1231 2017-10-05
这个帖子目前还没图呢
赞

黄子澄 2017-10-05
陈老师终于又更了😂
赞

禹洋 2017-10-08
催更啊。。。
赞

紫霜 2017-10-09
老师，3.27没看懂，怎么把 λ 去掉的呢
1

紫霜 2017-10-09
我懂了，真的谢谢老师，前面的都看懂了，后面解决α的值，需要SMO算法，这个需要自己了解吗？
赞

耳东陈 (作者) 回复 紫霜 2017-10-09
我会写一下，不过这个算法细节比较多，可能不会写得像前面一样细致。
赞

Lalala 2017-10-18
写的真的很不错，虽然我跳过了对偶同解部分的证明
1

he斯基 2017-10-20
陈老师你好，关于对偶问题那里我还是有点没有看懂 为什么不能直接求min L(w,γ，α)而必须要用对偶转化成 max L(w,γ，α)呢？ @耳东陈
赞

耳东陈 (作者) 回复 he斯基 2017-10-21
这个你可以试着求一下，看看求解过程。
赞

吕祥兴 回复 he斯基 2018-01-06
我在这里（web.mit.edu/6.034/v
们通过计算x_i和x_j的内积来求解，

1

快跑啊小女孩 2017-11-08
吹更吹更
赞

ZJIMPROVE 2017-11-20
终于学完了，断断续续有半个月，非常感谢陈教授。
赞

凉夜如水 2018-02-01
催更催更~谢谢陈老师了
赞

bmexue 2018-02-21
谢谢老师 有些地方以前一直没看懂，这么一理，清楚多了。
赞

马亮 2018-02-28
啥时讲非线性svm
1

Domi 2018-03-05
3.27需要3.26消去λ吧？期待三更
赞

sjacr 回复 Domi 2019-01-11
公式3.26的约束条件使含有截距的那一项是0，自然消去了
赞

流泪的皮皮虾 回复 sjacr 2019-10-13
为什么里面的Wx没有消掉呢？αy是零的话,应该里面全部都可以消掉吧？
赞

我叫陈官富 2018-03-06
和周志华老师的《机器学习》结合起来看，有种豁然开朗的感觉，感谢陈老师的精彩解释～～
1

acblacktea 2018-03-10
楼主有空可以继续更新一下核函数非线性讲解和具体算法实现讲解感激不尽
赞

梅德基里 2018-03-11
3.28的两个i标签相乘怎么变成了i和j标签相乘了呢，i和j分别是带便超平面两侧的点吗，请老师指点一下
赞

ycj 回复 梅德基里 2018-04-15
分开写的话都用i没有歧义，乘在一起就会产生歧义
赞

知友 回复 ycj 2019-01-07
呀？ 什么意思哇，请指教
赞

展开其他 2 条回复

哒哒哒

已赞同 217 ▼ 54 条评论 分享 收藏 ···

老师，核函数什么时候更，坐等中，哈哈哈

1

禾与林 2018-04-05

请问老师什么时候更新SMO算法和非线性SVM，很期待！

赞

1 2 3 下一页