90Zeng

当你的才华撑不起你的野心时，只有静下心来好好学习！纵使命运注定是个打酱油的，也要打一瓶与别人不一样的酱油！

博客园 首页 新随笔 联系 订阅 管理 随笔-95 文章-2 评论-117

公告

昵称：90Zeng
园龄：3年4个月
粉丝：150
关注：11
+加关注

<			2017年12月			>
日	一	二	三	四	五	六
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

搜索

找找看

最新随笔

1. 数理统计知识点归纳
2. Vim中自动在程序起始处添加版权和作者信息
3. Shell高级编程学习笔记(基础篇)
4. 基于Python的机器学习实战：Apriori
5. 基于Python的机器学习实战：AadBoost
6. 基于Python的机器学习实战：KNN
7. C++中构造函数和析构函数的调用顺序
8. Theano教程：Python的内存管理
9. python学习笔记："爬虫+有道词典"实现一个简单的英译汉程序
10. 基于theano的降噪自动编码器（Denoising Autoencoders--DA）

随笔分类

C++(21)
Data mining/数据挖掘(8)
data structure/数据结构(4)
Deep learning/深度学习(13)
Hadoop(3)
Java(10)
Machine learning/机器学习(23)

简易解说拉格朗日对偶（**Lagrange duality**）

引言：尝试用最简单易懂的描述解释清楚机器学习中会用到的拉格朗日对偶性知识，非科班出身，如有数学专业博友，望多提意见！

1.原始问题

假设$f(x), c_i(x), h_j(x)$是定义在R^n上的连续可微函数（为什么要求连续可微呢，后面再说，这里不用多想），考虑约束最优化问题：

$$\min_{x\in R^n} f(x)$$
$$s.t.\ c_i(x) \le 0, i = 1,2,\ldots,k$$
$$h_j(x) = 0, j = 1,2,\ldots,l$$

称为约束最优化问题的原始问题。

现在如果不考虑约束条件，原始问题就是：

$$\min_{x\in R^n} f(x)$$

因为假设其连续可微，利用高中的知识，对$f(x)$求导数，然后令导数为0，就可解出最优解，很easy. 那么，问题来了（呵呵。。。），偏偏有约束条件，好烦啊，要是能想办法把约束条件去掉就好了，bingo! 拉格朗日函数就是干这个的。

引进广义拉格朗日函数（generalized Lagrange function）：

$$L(x,\alpha,\beta) = f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)$$

不要怕这个式子，也不要被拉格朗日这个高大上的名字给唬住了，让我们慢慢剖析！

这里$x = (x^{(1)}, x^{(2)}, \ldots, x^{(n)})^T \in R^n$，$\alpha_i, \beta_j$是拉格朗日乘子（名字高大上，其实就是上面函数中的参数而已），特别要求$\alpha_i \ge 0$.

积分与排名
积分 - 96323
排名 - 3056

最新评论
1. Re:数据结构复习：交换排序原理及C++实现
链表元素的排序无法控制第几位啊，很绝望。
--ssss98
2. Re:基于Python的机器学习实战：Apriori
@抬头看见阳光同样是这个问题啊，只能自己改写一下了...
--lchzh
3. Re:基于Python的机器学习实战：Apriori
多对一的关联规则怎么没有实现呢？
--lchzh
4. Re:一步一步详解ID3和C4.5的C++实现
@haolujun在构建决策树的过程中，如果某一属性已经考察过了那么就从数据集中去掉这一属性，此处不是真正意义上的去掉，而是将考虑过的属性全部标记为110，当然可以是其他数字，只要能和原来训练集中的任......
--90Zeng
5. Re:一步一步详解ID3和C4.5的C++

现在，如果把$L(x,\alpha,\beta)$看作是关于α_i,β_j的函数，要求其最大值，即

$$\max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta)$$

再次注意$L(x,\alpha,\beta)$是一个关于α_i,β_j的函数，经过我们优化（不要管什么方法），就是确定α_i,β_j的值使得$L(x,\alpha,\beta)$取得最大值（此过程中把x看做常量），确定了α_i,β_j的值，就可以得到$L(x,\alpha,\beta)$的最大值，因为α_i,β_j已经确定，显然最大值$\max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta)$就是只和$x$有关的函数，定义这个函数为：

$$\theta_P(x) = \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta)$$

其中

$$L(x,\alpha,\beta) = f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)$$

下面通过x是否满足约束条件两方面来分析这个函数:

- 考虑某个x违反了原始的约束，即$c_i(x) > 0$或者$h_j(x) \neq 0$，那么：

$$\theta_P(x) = \max_{\alpha,\beta:\alpha_i\geq 0} [f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)] = +\infty$$

注意中间的最大化式子就是确定α_i,β_j的之后的结果，若$c_i(x) > 0$，则令$\alpha_i \to +\infty$，如果$h_j(x) \neq 0$，很容易取值β_j使得$\beta_j h_j(x) \to +\infty$

- 考虑x满足原始的约束，则：

实现
请
教：count(examples[0].begin(),examples[0].end(),110)，110是个什么特殊的数字呢？
--haolujun
6. Re:python学习笔记："爬虫+有道词典"实现一个简单的英译汉程序
这个是爬虫，真搞笑，博主技术很渣
--韦驮天
7. Re:简易解说拉格朗日对偶（Lagrange duality）
超赞，楼主解析问题，由浅入深，娓娓道来，值得多看几遍。
--aoguren
8. Re:简易解说拉格朗日对偶（Lagrange duality）
特别好，还没看完，明天继续看！
--郎毛毛
9. Re:简易解说拉格朗日对偶（Lagrange duality）
学习了，多看一遍就多一层理解
--师太跟我吧
10. Re:基于Python的机器学习实战：Apriori
谢谢楼主，您讲解的很清楚。我是新手，看了后对这个算法更加清楚。但是最后的关联规则生成函数好像不完善，只有一个元素推出一个元素，和一个元素推出多个元素的规则，没有多个元素推出多个元素的规则。
--抬头看见阳光

阅读排行榜
1. 简易解说拉格朗日对偶（Lagrange duality）(20285)
2. C++之vector中元素删除(12502)
3. Shell高级编程学习笔记(基础篇)(11135)
4. 基于Python的机器学习实战：Apriori(9138)
5. 基于Python的机器学习实战：KNN(6372)
6. k近邻法的C++实现：kd树(5611)
7. K-means聚类算法原理和C++实现(4943)
8. 基于Python的机器学习实战：AadBoost(4716)
9. hadoop学习笔记之一步一步部署

$\theta_P(x) = \max_{\alpha,\beta:\alpha_i\geq 0} [f(x)] = f(x)$，注意中间的最大化是确定$\alpha_i, \beta_j$的过程，$f(x)$就是个常量，常量的最大值显然是本身.

通过上面两条分析可以得出：

$$\theta_P(x) = \begin{cases} f(x), & x\text{满足原始问题约束} \\ +\infty, & \text{其他} \end{cases}$$

那么在满足约束条件下：

$$\min_x \theta_P(x)$$

$$= \min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta) = \min_x f(x)$$

即 $\min_x \theta_P(x)$ 与原始优化问题等价,所以常用 $\min_x \theta_P(x)$ 代表原始问题，下标 P 表示原始问题，定义原始问题的最优值：

$$p^* = \min_x \theta_P(x)$$

原始问题讨论就到这里，做一个总结：通过拉格朗日这位大神的办法重新定义一个无约束问题（大家都喜欢无拘无束），这个无约束问题等价于原来的约束优化问题，从而将约束问题无约束化！

2.对偶问题

定义关于α, β的函数：

$$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta)$$

注意等式右边是关于x的函数的最小化，x确定以后，最小值就只与α, β有关，所以是一个关于α, β的函数.

考虑极大化$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta)$，即

$$\max_{\alpha,\beta:\alpha_i\geq 0} \theta_D(\alpha,\beta) = \max_{\alpha,\beta:\alpha_i\geq 0} \min_x L(x,\alpha,\beta)$$

这就是原始问题的对偶问题，再把原始问题写出来：

$$\min_x \theta_P(x) = \min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta)$$

形式上可以看出很对称，只不过原始问题是先固定$L(x,\alpha,\beta)$中的x，优化出参数α,β，再优化最优x，而对偶问题是先固定α,β，优化出最优x，然后再确定参数α,β.

定义对偶问题的最优值：

$$d^* = \max_{\alpha,\beta:\alpha_i\geq 0} \theta_D(\alpha,\beta)$$

---

3. 原始问题与对偶问题的关系

定理：若原始问题与对偶问题都有最优值，则

$$d^* = \max_{\alpha,\beta:\alpha_i\geq 0} \min_x L(x,\alpha,\beta) \leq \min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta) = p^*$$

证明：对任意的α,β和x，有

$$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta) \leq L(x,\alpha,\beta) \leq \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta) = \theta_P(x)$$

即

$$\theta_D(\alpha,\beta) \leq \theta_P(x)$$

由于原始问题与对偶问题都有最优值，所以

$$\max_{\alpha,\beta:\alpha_i\geq 0} \theta_D(\alpha,\beta) \leq \min_x \theta_P(x)$$

即

$$d^* = \max_{\alpha,\beta:\alpha_i\geq 0} \min_x L(x,\alpha,\beta) \leq \min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta) = p^*$$

也就是说原始问题的最优值不小于对偶问题的最优值，但是我们要通过对偶问题来求解原始问题，就必须使得原始问题的最优值与对偶问题的最优值相等，于是可以得出

下面的推论：

推论：设 x^* 和 α^*,β^* 分别是原始问题和对偶问题的可行解，如果 $d^* = p^*$，那么 x^* 和 α^*,β^* 分别是原始问题和对偶问题的最优解。

所以，当原始问题和对偶问题的最优值相等：$d^* = p^*$ 时，可以用求解对偶问题来求解原始问题（当然是对偶问题求解比直接求解原始问题简单的情况下），但是到底满足什么样的条件才能使的 $d^* = p^*$ 呢，这就是下面要阐述的 KKT 条件

---

4. KKT 条件

定理：对于原始问题和对偶问题，假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数，$h_i(x)$ 是仿射函数（即由一阶多项式构成的函数，f(x)=Ax + b, A是矩阵，x，b是向量）；并且假设不等式约束 $c_i(x)$ 是严格可行的，即存在 x，对所有 i 有 $c_i(x) < 0$，则存在 x^* 和 α^*,β^*，使得 x^* 是原始问题的最优解，α^*,β^* 是对偶问题的最优解，并且

$$d^* = p^* = L(x^*,\alpha^*,\beta^*)$$

定理：对于原始问题和对偶问题，假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数，$h_i(x)$ 是仿射函数（即由一阶多项式构成的函数，f(x)=Ax + b, A是矩阵，x，b是向量）；并且假设不等式约束 $c_i(x)$ 是严格可行的，即存在 x，对所有 i 有 $c_i(x) < 0$，则 x^* 和 α^*,β^* 分别是原始问题和对偶问题的最优解的充分必要条件是 x^* 和 α^*,β^* 满足下面的Karush-Kuhn-Tucker(KKT)条件：

$$\nabla_x L(x^*,\alpha^*,\beta^*) = 0$$
$$\nabla_\alpha L(x^*,\alpha^*,\beta^*) = 0$$
$$\nabla_\beta L(x^*,\alpha^*,\beta^*) = 0$$
$$\alpha_i^* c_i(x^*) = 0, i = 1,2,\dots,k \quad \text{(KKT 对偶互补条件)}$$
$$c_i(x^*) \le 0, i = 1,2,\dots,k$$
$$\alpha_i^* \ge 0, i = 1,2,\dots,k$$

$$h_j(x^*) = 0, j = 1,2,\ldots,l$$

关于KKT 条件的理解：前面三个条件是由解析函数的知识，对于各个变量的偏导数为0（这就解释了一开始为什么假设三个函数连续可微，如果不连续可微的话，这里的偏导数存不存在就不能保证），后面四个条件就是原始问题的约束条件以及拉格朗日乘子需要满足的约束。

特别注意当 $\alpha_i^* > 0$ 时，由KKT对偶互补条件可知：$c_i(x^*) = 0$，这个知识点会在 SVM 的推导中用到.

---

5. 总结

一句话，某些条件下，把原始的约束问题通过拉格朗日函数转化为无约束问题，如果原始问题求解棘手，在满足KKT的条件下用求解对偶问题来代替求解原始问题，使得问题求解更加容易。

请尊重原创知识，本人非常愿意与大家分享 转载请注明出处：http://www.cnblogs.com/90zeng/ 作者：博客园-90Zeng

分类: Machine learning/机器学习

标签: 拉格朗日函数, 原始问题, 对偶问题, KKT条件

好文要顶 关注我 收藏该文


90Zeng
关注 - 11
粉丝 - 150
+加关注

13 1

« 上一篇：机器学习中有关概率论知识的小结
» 下一篇：EM算法原理详解

posted @ 2014-11-09 14:14 90Zeng 阅读(20288) 评论(8) 编辑 收藏

评论列表

#1楼 2015-04-10 22:30 henrysky

KKT条件的第二项感觉怪怪的。这个求导了之后不是在说所有的不等式condition都等于0吗？

支持(0) 反对(0)

#2楼 2016-07-19 17:36 freshing

写的不错，慢慢看

支持(0) 反对(0)

#3楼 2016-07-20 11:18 freshing

.

支持(0) 反对(0)

#4楼 2016-10-26 10:44 092000

“由于原始问题与对偶问题都有最优值，所以d*<p*”，请问这句应该怎么理解呢

支持(0) 反对(0)

#5楼 2017-03-10 09:07 明渊阁

必须是凸函数二者的最优值才相等吗？如果f(x)是凹函数就不行吗，这个有对应的证明吗

支持(0) 反对(0)

#6楼 2017-04-08 03:21 师太跟我吧

学习了，多看一遍就多一层理解

支持(0) 反对(0)

#7楼 2017-05-04 01:15 郎毛毛

特别好，还没看完，明天继续看！

支持(0) 反对(0)

#8楼 2017-06-06 14:48 aoguren

超赞，楼主解析问题，由浅入深，娓娓道来，值得多看几遍。

支持(0) 反对(0)

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 登录 或 注册，访问网站首页。

最新**IT**新闻:
· 马化腾：中国企业需成为新技术的驱动者和贡献者
· 蚂蚁会员积分将清零！69999积分可换iPhone X
· 微信支付商户平台一大波新功能：手机查看“经营数据”
· 因价格昂贵 特斯拉被德国从电动车补贴目录中移除
· 成熟开发者的“元品质”
» 更多新闻...

最新知识库文章:
· 以操作系统的角度述说线程与进程
· 软件测试转型之路
· 门内门外看招聘
· 大道至简，职场上做人做事做管理
· 关于编程，你的练习是不是有效的？
» 更多知识库文章...

Copyright ©2017 90Zeng