

凸优化-KKT条件

1. 引言

上一节我们讲了如何构建原问题的对偶问题，首先我们引入拉格朗日函数 $L(x,u,v)$ 将有约束的优化问题转换为无约束的优化问题，然后对原问题的参数求导，获得使拉格朗日函数最小的拉格朗日对偶函数 $g(u,v)$ ，最后使得对偶函数最大的问题则成为原问题的对偶问题

从上面的对偶问题构建方法可以看出获得对偶问题要对原问题参数极小，然后对拉格朗日乘子最大化，这样做的目的是使得转化后的对偶问题更容易求解。又因为对偶问题具有弱对偶性，即 $f^* \geq g^*$ ，因此，对偶问题提供了原问题最优值的下界，**如果可以满足某种条件使得二者相等，即具有强对偶性**，那么这个时候我们就可以通过求解对偶问题的解来间接地获得原问题的解。

因此，上段所说的“在满足某些条件的情况下”，这所谓的“满足某些条件”就是要满足KKT条件。上一节我们也提到过Slater's condition，可以保证对偶问题具有强对偶性。虽然该条件是一个很弱的条件，**但是我们注意到该条件并未给出解的形式或者给出二者之间的联系，即原问题和对偶问题最优解的关系**。同时，Slater's Condition只是满足强对偶性的一个充分条件，但不是必要条件，即原问题的对偶问题具有强对偶性并不一定满足Slater's Condition。而本文所要讲述的KKT条件既是在满足一些有规则的条件下，优化问题能有最优化解法的一个**必要和充分条件。

(回忆) Slater's condition：如果原 (primal) 问题为凸优化问题

(即： f, h_1, \dots, h_m 为凸函数, l_1, \dots, l_r 是仿射函数) , 同时存在至少一个可行解满足 $h_1(x) < 0, \dots, h_m(x) < 0, l_1(x) = \dots = l_r(x) = 0$, 那么该问题的强对偶性成立, 即 $f^* = g^*$ 。

2. KKT条件定义

给定一般优化问题：

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0, i = 1, \dots, m \\ & l_i(x) = 0, j = 1, \dots, r \end{aligned}$$

KKT条件(Karush-Kuhn-Tucker conditions) 是:

- $0 \in \partial f(x) + \sum_{i=1}^m u_i \partial h_i(x) + \sum_{j=1}^r v_j \partial l_j(x) \Rightarrow Stationarity;$
- $u_i \cdot h_i(x) = 0, \text{ for all } i \Rightarrow Complementary Slackness;$
- $h_i(x) \leq 0, l_j(x) = 0, \text{ for all } i, j \Rightarrow Primal Feasibility;$
- $u_i \geq 0 \text{ for all } i \Rightarrow Dual Feasibility.$

综上所述, 对于具有强对偶性的优化问题, 且该优化问题存在至少一个可行解, 那么下列问题等价:

x^*, u^* 和 v^* 是原问题和对偶问题的最优解 (solution) $\Leftrightarrow x^*, u^*$ 和 v^* 满足 KKT 条件

3. KKT条件证明

(1) 必要性 (Necessity) :

如果 x^*, u^* 和 v^* 是原问题和对偶问题的最优解, 且对偶问题保持强对偶性, 即对偶间隙为0, 那么:

$$\begin{aligned}
 f(x^*) &= g(u^*, v^*) \\
 &= \min_x L(x, u^*, v^*) \\
 &= \min_x f(x) + \sum_i^n u_i^* h_i(x) + \sum_j^r v_j^* l_j(x) \\
 &\leq f(x^*) + \sum_i^n u_i^* h_i(x^*) + \sum_j^r v_j^* l_j(x^*) \\
 &\leq f(x^*)
 \end{aligned}$$

第一个不等式成立的原因是, 假设原问题可行解域为 C , 那

么 $\min_{x \in C} L(x, u, v) = L(x^*, u, v) \geq \min_x L(x, u, v)$ 。

根据上面的推导, 由于不等式前后均为 $f(x^*)$, 所以, 所有不等号实际上都是等号。

对于第一个不等式, 等号成立代

表 x^* 是 $\min_x f(x) + \sum_i^m u_i^* h_i(x) + \sum_j^r v_j^* l_j(x)$ 的最优解, 因此, 0 必然

是 $\min_x f(x) + \sum_i^m u_i h_i(x) + \sum_j^r v_j l_j(x)$ 的次梯度，既证stationarity条件。

对于第二个不等式，等号成立代表 $u_i \cdot h_i(x) = 0$ ，既证complementary slackness条件。

对于primal feasibility和dual feasibility，它们是原问题和对偶问题的约束，满足条件。

综上所述，KKT条件必要性证明完毕，如果 x^* , u^* 和 v^* 是原问题和对偶问题的最优解（solution），同时对偶间隙为0，那么 x^* , u^* 和 v^* 满足KKT条件。

(2) 充分性 (Sufficiency) :

如果存在 x^* , u^* 和 v^* 满足KKT条件，那么：

$$\begin{aligned} g^* &= g(u^*, v^*) \\ &= \min_x f(x) + \sum_i^m u_i^* h_i(x) + \sum_j^r v_j^* l_j(x) \\ &= f(x^*) + \sum_i^m u_i^* h_i(x^*) + \sum_j^r v_j^* l_j(x^*) \\ &= f(x^*) \end{aligned}$$

第一和第二个等式是由定义获得，第三个等式是根据KKT条件中的stationarity条件，即0是函数 $L(x, u, v)$ 的次梯度，因此等式必然成立，最后等式是根据KKT条件中得slackness条件获得。因此， $g^* = f^*$ ，该解使得对偶间隙为0，分别为原问题和对偶问题的最优解。

综上所述，既证如果 x^* , u^* 和 v^* 满足KKT条件，那么 x^* , u^* 和 v^* 是原问题和对偶问题的solution。

4. KKT条件实例：KKT条件在SVM问题中的应用

我们将SVM问题记为：

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2} \|\beta\|_2^2 + C \sum_i^n \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, i = 1, \dots, n \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, i = 1, \dots, n \end{aligned}$$

引入对偶变量 v, w 构建拉格朗日函数：

$$L(\beta, \beta_0, \xi, v, w) = \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^n \xi - \sum_{i=1}^n v_i \xi + \sum_{i=1}^n w_i (1 - \xi - y_i(x_i^T \beta + \beta_0))$$

因为原问题满足Slater's condition, 同时 f 和 g 是可微的, 因此, 该问题满足KKT条件。应用stationarity条件, 即分别对 β, β_0, ξ 求导, 我们可以获得：

- $\beta = \sum_{i=1}^n w_i y_i x_i$;
- $0 = \sum_{i=1}^n w_i y_i$;
- $w = CI - v$;

应用slackness条件, 我们可以获得：

- $v_i \xi_i = 0$;
- $w_i (1 - \xi_i - y_i(x_i^T \beta + \beta_0)) = 0$;

综上所述, 我们获得以下结论：

- $\beta = \sum_{i=1}^n w_i y_i x_i$, 决定支持向量分割平面;
- $w_i \neq 0$, 如果 $y_i(x_i^T \beta + \beta_0) = 1 - \xi_i$, 因为二者乘积必为0;

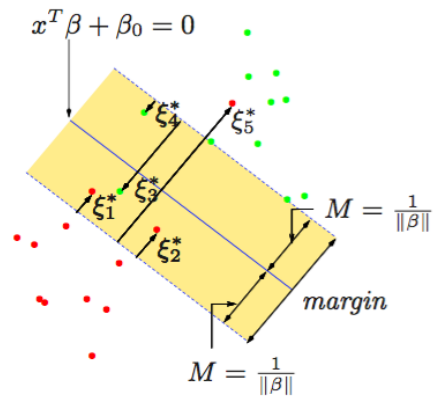
我们将满足 $w_i \neq 0$ 的点称为 支持向量 (support points or support vectors) , 落在分类边界上或以里; 对于 $w_i = 0$ 的点, $y_i(x_i^T \beta + \beta_0) > 1 - \xi_i$, 说明SVM分类器可以很好划分该样本, 即该样本远离分类边界, 实际上使得 $w_i = 0$, 意味着对分类不起作用。

因此, **对于支持向量**：

- 如果 $\xi_i = 0$, 意味着 x_i 落在分类边界上, 此时, $w_i \in (0, C]$, 因为, $v_i \xi_i = 0$, 同时, $w = CI - v$;
- 如果 $\xi_i \neq 0$, 意味着 x_i 落在相反的分类边界上, 此时, $w_i = C$ 。

综上所述, **KKT条件解释了为什么SVM分类器仅由支持向量决定, 跟远离分界面**

的点无关，如下图。所以，KKT条件并不能保证给我们找到最优解的方法，但是可以很好的解释原问题上的一些性质，如支持向量。



blog comments powered by Disqus (<http://disqus.com>)
