Support Vector Machine 03 Weak Duality Proof

Chen Gong

16 November 2019

在前面我们已经展示的 Hard Margin 和 Soft Margin SVM 的建模和求解。前面提到的 SVM 有三宝，间隔，对偶，核技巧。前面我们已经分析了间隔，大家对于其中用到的对偶，虽然我们用比较直觉性的方法进行了解释，但是估计大家还是有点懵逼。这节我们希望给到通用性的证明，这里实际上就是用到了约束优化问题。

1 弱对偶性证明

首先，我们需要证明约束优化问题的原问题和无约束问题之间的等价性。

1.1 约束优化问题与无约束问题的等价性

对于一个约束优化问题，我们可以写成：

$$\left\{\begin{array}{l}\min_{x\in\mathcal{R}^p} f(x) \\ s.t. \quad m_i(x)\leq 0,\ i=1,2,\cdots,N \\ \qquad\ \ n_i(x)=0,\ i=1,2,\cdots,N\end{array}\right. \tag{1}$$

我们用拉格朗日函数来进行表示：

$$\mathcal{L}(x,\lambda,\eta)=f(x)+\sum_{i=1}^N\lambda_i m_i+\sum_{i=1}^N\eta_i n_i \tag{2}$$

我们可以等价的表示为：

$$\left\{\begin{array}{l}\min_x \max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta) \\ s.t.\ \lambda_i\geq 0,\ i=1,2,\cdots,N\end{array}\right. \tag{3}$$

这就是将一个约束优化问题的原问题转换为无约束问题。那么这两种表达形式一定是等价的吗？我们可以来分析一下：

如果，x 违反了约束条件 $m_i(x)\leq 0$，那么有，$m_i(x)>0$。且 $\lambda_i>0$，那么很显然 $max_\lambda\ \mathcal{L}=+\infty$。

如果，x 符合约束条件 $m_i(x)\leq 0$，那么很显然 $max_\lambda\ \mathcal{L}\neq+\infty$。

那么：

$$\min_x\max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta)=\min_x\{\max\ \mathcal{L},+\infty\}=\min_x\{\max\ \mathcal{L}\} \tag{4}$$

其实大家可以很明显的感觉到，这个等式自动的帮助我们过滤到了一半 $m_i(x)\geq 0$ 的情况，这实际上就是一个隐含的约束条件，帮我们去掉了一部分不够好的解。

1.2 证明弱对偶性

原问题我们可以写为：

$$\begin{cases} \min_x \max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta) \\ s.t.\ \lambda_i \geq 0,\ i=1,2,\cdots,N \end{cases} \tag{5}$$

而原问题的对偶问题则为：

$$\begin{cases} \min_{\lambda,\eta} \max_x\ \mathcal{L}(x,\lambda,\eta) \\ s.t.\ \lambda_i \geq 0,\ i=1,2,\cdots,N \end{cases} \tag{6}$$

原问题是一个关于 x 的函数，而对偶问题是一个关于 λ,η 的最小化问题，而弱对偶性则可以描述为：对偶问题的解 $\leq$ 原问题的解。为了简化表达，后面对偶问题的解我们用 d 来表示，而原问题的解我们用 p 来表示。那么我们用公式化的语言表达也就是：

$$\min_{\lambda,\eta} \max_x\ \mathcal{L}(x,\lambda,\eta) = d \leq \min_x \max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta) = p \tag{7}$$

在前面我们使用感性的方法证明了 $\max\min\ \mathcal{L} \leq \min\max\ \mathcal{L}$，下面我们给出严谨的证明：

很显然可以得到：

$$\min_x\ \mathcal{L}(x,\lambda,\eta) \leq \mathcal{L}(x,\lambda,\eta) \leq \max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta) \tag{8}$$

那么，$\min_x\ \mathcal{L}(x,\lambda,\eta)$ 可表示为一个与 x 无关的函数 $A(\lambda,\eta)$，同理 $\max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta)$ 可表示为一个与 λ,η 无关的函数 $B(x)$。显然，我们可以得到一个恒等式：

$$A(\lambda,\eta) \leq B(x) \tag{9}$$

那么接下来就有：

$$\begin{aligned} A(\lambda,\eta) &\leq \min\ B(x) \\ \max\ A(\lambda,\eta) &\leq \min\ B(x) \\ \min_{\lambda,\eta} \max_x\ \mathcal{L}(x,\lambda,\eta) &\leq \min_x \max_{\lambda,\eta}\ \mathcal{L}(x,\lambda,\eta) \end{aligned} \tag{10}$$

弱对偶性，证毕!!