机器学习与数据科学

Machine Learning and Data Science

主讲: 李春光

www.pris.net.cn/teacher/lichunguang

模式识别与智能系统实验室

信息与通信工程学院 网络搜索教研中心

北京邮电大学

专题 六：支持向量机与统计学习理论
数学基础知识补充-I

- **内容提要**
  - 有约束最优化问题的最优性条件
  - Lagrange对偶问题
    - 弱对偶定理
    - 强对偶定理
    - 鞍点定理

有约束优化问题的最优性条件

- 几何条件：
  - 可行方向的集合与下降方向的集合相交为空集
    - 在局部最优点处，算法在寻找下降方向时发现已无路可走，即算法无法从可行集中找到下降方向

- x* 是一个最优点，当且仅当它是一个可行点且对于所有可行向量 u 如下条件成立：

$$\nabla f\left(\mathbf{x}^*\right)^T\left(\mathbf{u}-\mathbf{x}^*\right)\geq 0$$

有约束最优化问题最优性条件举例

- 无约束最优化问题: $\min_{\mathbf{x}\in R^n} f(\mathbf{x})$
  - 由于 **u – x*** 是任意向量，因此得到: $\nabla f(\mathbf{x}^*)^T = 0$
- 线性等式约束最优化问题： $\min_{\mathbf{x}\in R^n} f(\mathbf{x}) \text{ s.t. } A\mathbf{x} = \mathbf{b}$
  - 设**x*** 是可行解**(i.e. Ax*=b)** 且存在对偶证书**(dual certificate)**向量**v**满足：

$$A\mathbf{x}^* = \mathbf{b}, \quad \nabla f(\mathbf{x}^*) + A^T\mathbf{v} = 0$$

- 非负约束最优化问题： $\min_{\mathbf{x}\in R^n} f(\mathbf{x}) \text{ s.t. } \mathbf{x} \geq \mathbf{0}$
  - 令**u =2x*, u=0,** 则得出最优性条件

$$\nabla f(\mathbf{x}^*)^T \mathbf{x}^* = 0, \quad \nabla f(\mathbf{x}^*) \geq 0, \quad \mathbf{x}^* \geq 0$$

KKT最优性条件

- 为了在计算中检验最优性条件，需要建立最优性条件的代数表示:
  1. **Lagrangian函数关于原变量的一阶条件**
  2. **原变量的可行条件**
  3. **对偶变量的可行条件**
  4. **针对不等式约束的互补松弛条件**
     - 不等式约束取等号与对偶变量取等号之间的“互补”

求解线性支持向量机问题的KKT条件

- 原问题:

$$\min_{\mathbf{w},b}\frac{1}{2}\mathbf{w}^T\mathbf{w},\quad \text{s.t. } \left(\mathbf{w}^T\mathbf{x}_i+b\right)y_i\geq 1,\ \ i=1,\dots,N.$$

- Lagrange函数

$$L\left(\mathbf{w},b,\alpha\right)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)-1\right),\ \text{where}\ \ \alpha_i\geq 0,\ i=1,\dots,N.$$

  - **KKT 条件是1阶必要条件:**

$$\begin{cases}\boxed{\nabla_{\mathbf{w}}L\left(\mathbf{w},b,\boldsymbol{\alpha}\right)=0,\quad \nabla_b L\left(\mathbf{w},b,\boldsymbol{\alpha}\right)=0}\\ y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)-1\geq 0,\ \ i=1,\dots,N\\ \alpha_i\geq 0,\ i=1,\dots,N\\ \alpha_i\left[y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)-1\right]=0,\ i=1,\dots,N\end{cases}\Rightarrow \boxed{\begin{cases}\mathbf{w}=\sum_{i=1}^{N}\alpha_i y_i\mathbf{x}_i,\\ \sum_{i=1}^{N}\alpha_i y_i=0.\end{cases}}$$

求解线性支持向量机问题的KKT条件

- 原问题:

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}, \quad \boxed{\text{s.t. } \left(\mathbf{w}^T\mathbf{x}_i + b\right) y_i \ge 1, \quad i = 1,\ldots,N.}$$

- Lagrange函数

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left(y_i \left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1 \right), \quad \boxed{\text{where } \alpha_i \ge 0, \ i = 1,\ldots,N.}$$

  - **KKT 条件是1阶必要条件:**

$$\begin{cases} \nabla_{\mathbf{w}} L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0, \quad \nabla_b L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0 \\ \boxed{y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1 \ge 0, \quad i = 1,\ldots,N} \\ \boxed{\alpha_i \ge 0, \ i = 1,\ldots,N} \\ \alpha_i\left[y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1\right] = 0, \ i = 1,\ldots,N \end{cases}$$

求解线性支持向量机问题的KKT条件

- 原问题:

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}, \quad \boxed{\text{s.t. } \left(\mathbf{w}^T\mathbf{x}_i + b\right) y_i \geq 1, \quad i = 1,\dots,N.}$$

- Lagrange函数

$$L(\mathbf{w},b,\alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1\right), \quad \boxed{\text{where } \alpha_i \geq 0, \; i = 1,\dots,N.}$$

  - KKT 条件是1阶必要条件:

$$\begin{cases} \nabla_{\mathbf{w}} L(\mathbf{w},b,\boldsymbol{\alpha}) = 0, \quad \nabla_b L(\mathbf{w},b,\boldsymbol{\alpha}) = 0 \\ y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1 \geq 0, \quad i = 1,\dots,N \\ \alpha_i \geq 0, \; i = 1,\dots,N \\ \boxed{\alpha_i\left[y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1\right] = 0, \; i = 1,\dots,N} \end{cases}$$

专题 六：支持向量机与统计学习理论
数学基础知识补充-I

- **内容提要**
  - 有约束最优化问题的最优性条件
  - **Lagrange**对偶问题
    - 弱对偶定理
    - 强对偶定理
    - 鞍点定理

考虑线性支持向量机问题

- 原问题:

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}, \quad \boxed{\text{s.t. } \left(\mathbf{w}^T\mathbf{x}_i + b\right)y_i \geq 1, \quad i = 1, \ldots, N.}$$

- Lagrange函数

$$L\left(\mathbf{w}, b, \alpha\right) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - 1\right), \boxed{\text{where } \alpha_i \geq 0, \ i = 1, \ldots, N.}$$

原问题 vs. 对偶问题

- **原问题** $\min\limits_{\mathbf{w},b} P(\mathbf{w},b) \quad \text{s.t.} \quad (\mathbf{w},b)\in\Omega$

  其中 $P(\mathbf{w},b)=\frac{1}{2}\mathbf{w}^T\mathbf{w}$

  $$\Omega=\left\{(\mathbf{w},b)\mid \; y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\geq 1,\; \forall i=1,\ldots,N\right\}$$

- **对偶函数与对偶问题**

  $\max\limits_{\boldsymbol{\alpha}} D(\boldsymbol{\alpha}) \;\; \text{s. t.} \quad \alpha_i\geq 0, \;\; i=1,\ldots,N$

  其中 $D(\alpha)=\min\limits_{\mathbf{w},b} L(\mathbf{w},b,\alpha)$

  – **两者有何联系？**

弱对偶理论

- **弱对偶理论**
  - 设**w**和**b**为原问题的可行解，α 为对偶问题的可行解，则有 $D(\alpha) \le P(\mathbf{w}, b)$

在实际求解最优化问题中有何应用价值?

- 证明: (略)
- 推论:
  - 如果(w_0,b_0)为原问题的可行解，α_0为对偶问题的可行解，且$D(\boldsymbol{\alpha}_0) \ge P(\mathbf{w}_0, b_0)$，则$(w_0,b_0)$和$\alpha_0$为分别为原问题和对偶问题的最优解
  - 如果原问题极小值为负无穷，则对偶函数为负无穷；如果对偶函数极大值为正无穷，则原问题没有可行解

对偶间隙(dual gap)

- 根据弱对偶理论，如果原问题的目标函数最优值为p, 对偶问题的目标函数最优值为q, 则必有

$$q \le p$$

  - 如果严格不等号成立，则存在对偶间隙

$$\delta = p - q > 0$$

强对偶理论

- **强对偶理论**
  - 对于凸**(convex)**优化问题，在适当的约束规格**(Constraint Qualification)**下，则有

$$q = p$$

其中

$$q = \max_{\boldsymbol{\alpha} \geq 0} D(\alpha) = \max_{\boldsymbol{\alpha} \geq 0} \min_{\mathbf{w}, b} L(\mathbf{w}, b, \alpha),$$

$$p = \min_{(\mathbf{w}, b) \in \Omega} P(\mathbf{w}, b)$$

  - 即 对偶问题的最大值对应于原问题的最小值

鞍点最优性条件

- **鞍点(saddle point)**

  设 $L(\mathbf{w},b,\alpha)$ 为**Lagrange**函数，如果对于每个 $(\mathbf{w},b)$ 和 $\boldsymbol{\alpha}\geq 0$，有

  $$L(\bar{\mathbf{w}},\bar{b},\alpha)\leq L(\bar{\mathbf{w}},\bar{b},\bar{\alpha})\leq L(\mathbf{w},b,\bar{\alpha})$$

  则称 $(\bar{\mathbf{w}},\bar{b},\bar{\alpha})$ 为 $L(\mathbf{w},b,\alpha)$ 的鞍点

在实际求解最优化问题中有何应用价值?

- 鞍点定理

  1. 鞍点对应最优点
  2. 对于凸优化问题，在**CQ**条件下，最优点对应于一个鞍点