Math Basis of Support Vector Machine

周陆延

1. 给出支持向量机最大间隔准则原问题的推导，解释分类超平面和支持超平面的含义

   分类超平面为任意将样本划分为两个部分的超平面

   支持超平面则是距离两侧支持向量距离之和最大的分类超平面

   记样本点 x 到超平面 (ω, b) 的距离为

$$r = \frac{|\boldsymbol{\omega}^\top \boldsymbol{x} + b|}{\|\boldsymbol{\omega}\|} \tag{1}$$

   假设超平面能将样本分类则 $\exists\ (\boldsymbol{\omega}, \boldsymbol{b})$ s.t. $\forall i$

$$\begin{cases} \boldsymbol{\omega}^\top \boldsymbol{x}_i + b \geq +1 \text{ if } y_i = +1 \\ \boldsymbol{\omega}^\top \boldsymbol{x}_i + b \leq -1 \text{ if } y_i = -1 \end{cases} \tag{2}$$

   则称刚好使等号成立的样本点为支持向量，两个异类支持向量到超平面的距离和 $\gamma = \frac{2}{\|\omega\|_2^2}$ 为间隔，要使 γ 最大，即

$$\begin{aligned} &\max \frac{2}{\|\boldsymbol{\omega}\|_2^2} = \min \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 \\ &\text{s.t. } y_i(\boldsymbol{\omega}^\top \boldsymbol{x}_i + b) \geq 1 \end{aligned} \tag{3}$$

2. 给出对偶问题的推导，并回答问题：强对偶性在支持向量机中始终成立吗？

   对原问题应用拉格朗日乘数法

$$\begin{aligned} &L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 + \sum_{i=1}^{m} \alpha_i [1 - y_i(\boldsymbol{\omega}^\top \boldsymbol{x}_i + b)] \\ &\text{s.t. } \boldsymbol{\alpha} \geq \mathbf{0} \end{aligned} \tag{4}$$

   首先证明原问题 $\Leftrightarrow$

$$\begin{aligned} &\min_{\boldsymbol{\omega}, b} \max_{\boldsymbol{\alpha}} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) \\ &\text{s.t. } \alpha \geq \mathbf{0} \end{aligned} \tag{5}$$

   证：

$$\begin{cases} \text{if } (\boldsymbol{\omega}, b) \text{ 在可行域内} \quad \max_{\boldsymbol{\alpha}} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 + 0 + 0 \\ \text{if } (\boldsymbol{\omega}, b) \text{ 不在可行域内} \quad \max_{\boldsymbol{\alpha}} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 + \infty + \infty \end{cases} \tag{6}$$

   然后记对偶函数

$$g(\boldsymbol{\alpha}) = \min_{\boldsymbol{\omega}, b} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) \tag{7}$$

   对偶问题即为

$$
\begin{aligned}
&\max_{\boldsymbol{\alpha}} \min_{\boldsymbol{\omega}, b} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) \\
&\text{s.t. } \boldsymbol{\alpha} \geq \mathbf{0}
\end{aligned}
\tag{8}
$$

求偏导得

$$
\begin{aligned}
\frac{\delta L}{\delta \boldsymbol{\omega}} = 0 \Rightarrow \boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \\
\frac{\delta L}{\delta b} = 0 \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0
\end{aligned}
\tag{9}
$$

带回原式即为

$$
\begin{aligned}
\max g(\boldsymbol{\alpha}) &= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i y_i \alpha_j y_j \boldsymbol{x}_i^{\top} \boldsymbol{x}_j + \sum_{i=1}^{m} \alpha_i \left[1 - y_i \left(\left(\sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i^{\top} \right) \boldsymbol{x}_i + b \right) \right] \\
&= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i y_i \alpha_j y_j \boldsymbol{x}_i^{\top} \boldsymbol{x}_j + \sum_{i=1}^{m} \left(\alpha_i - \sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_j^{\top} \boldsymbol{x}_i - \alpha_i y_i b \right) \\
&= -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\top} \boldsymbol{x}_j + \sum_{i=1}^{m} \alpha_i \\
\text{s.t.} &\sum_{i=1}^{m} \alpha_i y_i = 0
\end{aligned}
\tag{10}
$$

由支持向量机本身为凸优化问题且满足 Slater 条件（一定 $\exists(\boldsymbol{\omega}, b)$ s.t. 某些点在 margin 内部），得强对偶性在支持向量机中始终成立

3. 根据前面的结果推导SVM的KKT条件

$$
\begin{cases}
\alpha_i \geq 0 \\
y_i(\boldsymbol{\omega}^{\top} \boldsymbol{x}_i + b) \geq 1 \\
\alpha_i(y_i(\boldsymbol{\omega}^{\top} \boldsymbol{x}_i + b) - 1) = 0
\end{cases}
\tag{11}
$$

4. 如果数据非线性可分怎么办？给出修改后的原问题和对偶问题

   使用核技巧将数据映射到高维空间再进行划分

   修改后的原问题：

$$
\begin{aligned}
&\min_{\boldsymbol{\omega}, b} \frac{1}{2} \|\boldsymbol{\omega}\|_2^2 \\
&\text{s.t. } y_i(\boldsymbol{\omega}^{\top} \phi(\boldsymbol{x}_i) + b) \geq 1
\end{aligned}
\tag{12}
$$

   修改后的对偶问题：

$$
\begin{aligned}
&\max_{\boldsymbol{\alpha}} \sum_{i}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&\text{s.t.} \sum_{i=1}^{m} \alpha_i y_i = 0
\end{aligned}
\tag{13}
$$

   其中 $\phi(\boldsymbol{x})$ 是映射后的特征向量， $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 是两个向量在高维空间里的内积

5. 在问题3.4的基础上，若允许少量样本破坏约束，应增加怎样的损失函数，请给出修改后的原问题和对偶问题

应增加形如

$$C\sum_{i}^{m}\ell_{0/1}(y_i(\boldsymbol{\omega}^\top\boldsymbol{x}_i+b)) \tag{14}$$

的项，其中 $\ell_{0/1}$ 是 0/1 损失函数

$$\ell_{0/1}(z)=\begin{cases}1 \text{ if } z<1\\ 0 \text{ otherwise}\end{cases} \tag{15}$$

但考虑到 $\ell_{0/1}(z)$ 较劣的数学性质，实际应用中通常寻找一些代替函数，如：

$$\begin{gathered}\ell_{\text{hinge}}(z)=\max(0,1-z)\\ \ell_{\exp}(z)=\exp(-z)\\ \ell_{\log}(z)=\log(1+\exp(-z))\end{gathered} \tag{16}$$

下面对选用 $\ell_{\text{hinge}}(z)$ 的原问题与对偶问题进行推导

原问题

$$\begin{aligned}&\min\frac{1}{2}\|\boldsymbol{\omega}\|_2^2+C\sum_{i}^{m}\max(0,1-y_i(\boldsymbol{\omega}^\top\boldsymbol{x}_i+b))\\ &\text{s.t. } y_i(\boldsymbol{\omega}^\top\boldsymbol{x}_i+b)\geq 1\end{aligned} \tag{17}$$

引入松弛变量 $\xi_i\geq 0$ 重写 式 17 得到修改后的原问题

$$\begin{aligned}&\min\frac{1}{2}\|\boldsymbol{\omega}\|_2^2+C\sum_{i}^{m}\xi_i\\ &\text{s.t. }\begin{cases}y_i(\boldsymbol{\omega}^\top\boldsymbol{x}_i+b)\geq 1-\xi_i\\ \xi_i\geq 0\end{cases}\end{aligned} \tag{18}$$

对偶问题为

$$\begin{aligned}&\max_{\boldsymbol{\alpha}}\sum_{i}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_i^\top\boldsymbol{x}_j\\ &\text{s.t. }\begin{cases}\sum_{i=1}^{m}\alpha_i y_i=0\\ 0\leq\alpha_i\leq C\end{cases}\end{aligned} \tag{19}$$

其中和 Hard-Margin SVM 的主要区别在于

$$\begin{gathered}\bigtriangledown_{\boldsymbol{\xi}}L(\boldsymbol{\omega},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\mu})=C-\boldsymbol{\alpha}-\boldsymbol{\mu}=\mathbf{0}\\ \text{又}\begin{cases}\alpha_i\geq 0\\ \mu_i\geq 0\end{cases}\\ \Rightarrow 0\leq\alpha_i\leq C\end{gathered} \tag{20}$$