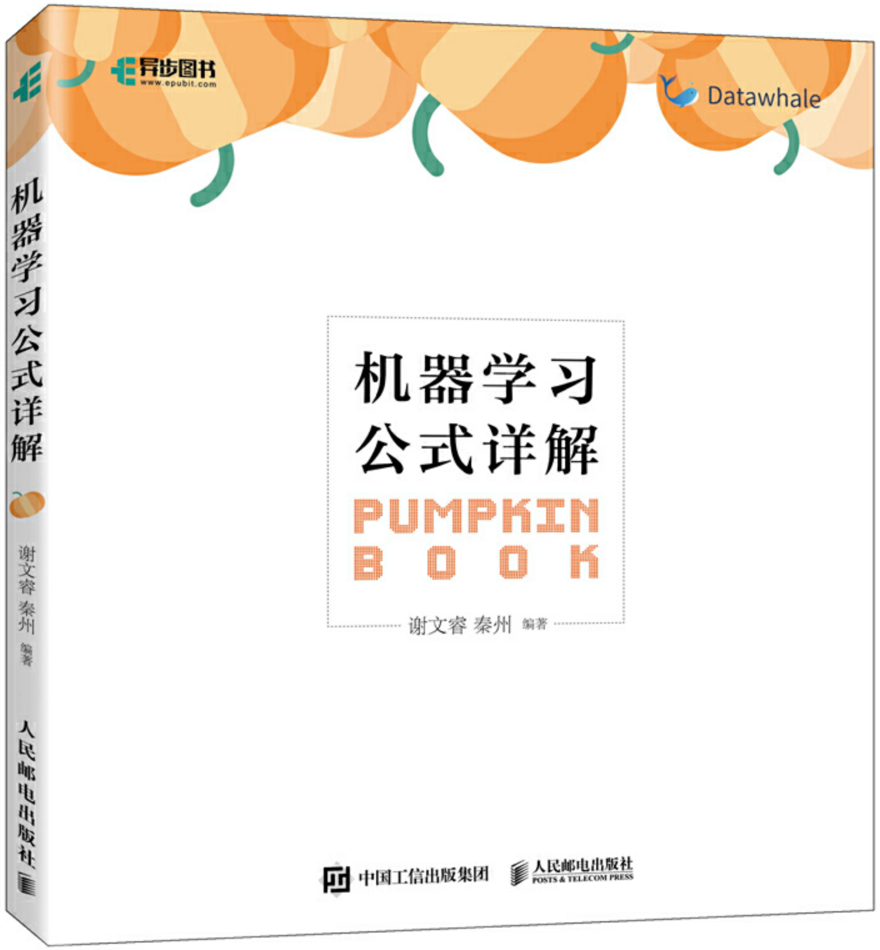

《机器学习公式详解》（南瓜书）

第6章 软间隔与支持向量回归

本节主讲：谢文睿

本节大纲

西瓜书对应章节：6.4、6.5

1. 算法原理
2. 软间隔
3. 支持向量回归

算法原理

在现实任务中，线性不可分的情形才是最常见的，因此需要允许支持向量机犯错

软间隔

从数学角度来说，软间隔就是允许部分样本（但要尽可能少）不满足下式中的约束条件

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geqslant 1, \quad i = 1, 2, \ldots, m \end{aligned}$$

因此，可以将必须严格执行的约束条件转化为具有一定灵活性的“损失”，合格的损失函数要求如下：

- 当满足约束条件时，损失为0
- 当不满足约束条件时，损失不为0，
- （可选）当不满足约束条件时，损失与其违反约束条件的程度成正比

只有满足以上要求，才能保证在最小化（min）损失的过程中，保证不满足约束条件的样本尽可能的少。

软间隔

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}\left(y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) - 1\right)$$

其中，$\ell_{0/1}$是“0/1损失函数”

$$\ell_{0/1}(z) = \begin{cases} 1, & \text{if } z < 0 \\ 0, & \text{if } z \geqslant 0 \end{cases}$$

$C > 0$是一个常数，用来调节损失的权重，显然当$C \to +\infty$时，会迫使所有样本的损失为0，进而退化为严格执行的约束条件，退化为硬间隔，因此，本式子可以看作支持向量机的一般化形式。

软间隔

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}\left(y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) - 1\right)$$

由于$\ell_{0/1}$非凸、非连续，数学性质不好，使得上式不易求解，因此常用一些数学性质较好的“替代损失函数”来代替$\ell_{0/1}$，软间隔支持向量机通常采用的是hinge（合页）损失来代替$\ell_{0/1}$

$$\text{hinge损失：}\ \ell_{\text{hinge}}(z) = \max(0, 1-z)$$

替换进上式可得

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \max\left(0, 1 - y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right)\right)$$

软间隔

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \max\left(0, 1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right)\right)$$

引入松弛变量ξ_i，上述优化问题便和下述优化问题等价

$$\begin{aligned} \min_{\boldsymbol{w},b,\xi_i} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right) \geqslant 1-\xi_i \\ & \xi_i \geqslant 0, i=1,2,\ldots,m \end{aligned}$$

【证明】：令

$$\max\left(0, 1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right)\right) = \xi_i$$

显然$\xi_i \geqslant 0$，当$1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right) > 0 \Rightarrow \xi_i = 1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$，当$1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right) \leqslant 0 \Rightarrow \xi_i = 0$，所以

$$1-y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right) \leqslant \xi_i \Rightarrow y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right) \geqslant 1-\xi_i$$

支持向量回归

相比于线性回归用**一条线**来拟合训练样本，支持向量回归（SVR）而是采用一个以 $f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$为中心，宽度为$2\epsilon$的**间隔带**，来拟合训练样本。

落在带子上的样本不计算损失（类比线性回归在线上的点预测误差为0），不在带子上的则以偏离带子的距离作为损失（类比线性回归的均方误差），然后以最小化损失的方式迫使间隔带从样本最密集的地方（中心地带）穿过，进而达到拟合训练样本的目的。

支持向量回归

因此SVR的优化问题可以写为

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \ell_\epsilon\left(f\left(\boldsymbol{x}_i\right) - y_i\right)$$

其中$\ell_\epsilon(z)$为“ϵ不敏感损失函数”（类比均方误差损失）

$$\ell_\epsilon(z) = \begin{cases} 0, & \text{if } |z| \leqslant \epsilon \\ |z| - \epsilon, & \text{if } |z| > \epsilon \end{cases}$$

$\frac{1}{2}\|\boldsymbol{w}\|^2$为L2正则项，此处引入正则项除了起正则化本身的作用外，也是为了和（软间隔）支持向量机的优化目标保持形式上的一致（在这里不用均方误差也是此目的），这样就可以导出对偶问题引入核函数，C为调节损失权重的常数。

支持向量回归

$$\ell_\epsilon(z)=\begin{cases}0, & \text{if } |z| \leqslant \epsilon \\ |z|-\epsilon, & \text{if } |z|>\epsilon\end{cases}$$

同软间隔支持向量机，引入松弛变量ξ_i，令

$$\ell_\epsilon\left(f\left(\boldsymbol{x}_i\right)-y_i\right)=\xi_i$$

显然$\xi_i \geqslant 0$，并且

当$\left|f\left(\boldsymbol{x}_i\right)-y_i\right| \leqslant \epsilon \Rightarrow \xi_i=0$

当$\left|f\left(\boldsymbol{x}_i\right)-y_i\right|>\epsilon \Rightarrow \xi_i=\left|f\left(\boldsymbol{x}_i\right)-y_i\right|-\epsilon$

所以

$$\left|f\left(\boldsymbol{x}_i\right)-y_i\right|-\epsilon \leqslant \xi_i \Rightarrow\left|f\left(\boldsymbol{x}_i\right)-y_i\right| \leqslant \epsilon+\xi_i$$

$$-\epsilon-\xi_i \leqslant f\left(\boldsymbol{x}_i\right)-y_i \leqslant \epsilon+\xi_i$$

那么SVR的优化问题可以改写为

支持向量回归

$$\min_{\boldsymbol{w},b,\xi_i} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\begin{aligned}\text{s.t.}\quad & -\epsilon-\xi_i \leqslant f\left(\boldsymbol{x}_i\right)-y_i \leqslant \epsilon+\xi_i \\ & \xi_i \geqslant 0, i=1,2,\ldots,m\end{aligned}$$

如果考虑两边采用不同的松弛程度

$$\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\left(\xi_i+\hat{\xi}_i\right)$$

$$\begin{aligned}\text{s.t.}\quad & -\epsilon-\hat{\xi}_i \leqslant f\left(\boldsymbol{x}_i\right)-y_i \leqslant \epsilon+\xi_i \\ & \xi_i \geqslant 0, \hat{\xi}_i \geqslant 0, i=1,2,\ldots,m\end{aligned}$$

预告

下一节：贝叶斯分类

西瓜书对应章节：7.1、7.2、7.3

结束语

欢迎加入【南瓜书读者交流群】，我们将在群里进行答疑、勘误、本次直播回放、本次直播PPT发放、下次直播通知等最新资源发放和活动通知。

加入步骤：

1. 关注公众号【Datawhale】，发送【南瓜书】三个字获取机器人“小豚”的微信二维码
2. 添加“小豚”为微信好友，然后对“小豚”发送【南瓜书】三个字即可自动邀请进群