法律声明

课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920

公众号

微 信

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

deepshare.net
深度之眼

软间隔与支持向量回归公式推导

Derivation of soft margin and SVR

本节大纲

Outline

先修内容：西瓜书6.4、6.5

1.软间隔SVM

2.支持向量回归

软间隔SVM

Soft margin SVM

deepshare.net
深度之眼

严格线性可分时的硬间隔SVM：

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

然而实际场景中，严格线性可分的数据并不总是存在，那么就需要适当放松约束条件。软间隔SVM允许某些样本不满足函数间隔值大于等于1这个约束，也即允许函数间隔值小于1，当然，这样的样本尽量越少越好，因此优化问题可以改写为

线性不可分时的软间隔SVM：

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

其中

$$\ell_{0/1}(z) = \begin{cases} 0, \text{ if } z \geq 0 \\ 1, \text{ otherwise} \end{cases}$$

软间隔SVM

Soft margin SVM

由于 $\ell_{0/1}$非凸、不连续，数学性质不太好，不易优化求解，因此需要考虑用其他损失函数来近似替代 $\ell_{0/1}$ 软间隔SVM选的替代损失函数是合页（hinge）损失函数

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

合页损失可以理解为：待考察变量z如果大于等于1就不惩罚，但是如果小于1就需要惩罚，但是惩罚的程度取决于自身比1小的程度，显然合页损失很适合用来考察软间隔SVM里面的函数间隔值。

软间隔SVM

Soft margin SVM

将优化问题中的 $\ell_{0/1}$ 损失函数更换为合页损失函数可得

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

为了引入松弛因子，令

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

则上述优化问题即可改写为

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

支持向量回归

SVR

SVR和SVM唯一的联系就是优化目标函数很相像，但是其建模思路差别很大。SVR实际就是找一个以 $f(\boldsymbol{x})$ 为中心，宽度为 2ϵ 的间隔带，使得此间隔带尽可能多地把所有样本都包裹住。

由外到内的思路：假设这条间隔带能包裹住所有样本，那么有两种方式可以做到，一种是 ϵ 取任意大，那么以任意 $f(\boldsymbol{x})$ 为中心的间隔带都能把所有样本包裹住，但是显然这样是没有意义的。另一种就是找尽可能小的 ϵ 使得其能把所有样本包裹住，那么能把所有样本包裹住的间隔带的 ϵ 最小是多少呢？可以通过以下优化问题来求解

$$\begin{aligned}\min_{\boldsymbol{w},b,\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & -\epsilon \leqslant y_i - f(\boldsymbol{x}_i) \leqslant \epsilon\end{aligned}$$

支持向量回归

SVR

但是由于实际采样得到的样本会存在极个别偏离大多数样本很远的噪音，那么此时如果按照以上优化目标来进行求解的话也会导致 ϵ偏大，因此，通常都会考虑人为固定 ϵ 的取值。显然，一旦固定 ϵ后上述优化问题就会无法求解，因此需要给两边添加松弛因子

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi},\hat{\boldsymbol{\xi}}} \quad & \epsilon + \textstyle\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \\
\text{s.t.} \quad & -\epsilon - \xi_i \leqslant y_i - f(\boldsymbol{x}_i) \leqslant \epsilon + \hat{\xi}_i \\
& \xi_i \geq 0 \\
& \hat{\xi}_i \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi},\hat{\boldsymbol{\xi}}} \quad & \textstyle\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \\
\text{s.t.} \quad & -\epsilon - \xi_i \leqslant y_i - f(\boldsymbol{x}_i) \leqslant \epsilon + \hat{\xi}_i \\
& \xi_i \geq 0 \\
& \hat{\xi}_i \geq 0
\end{aligned}
$$

支持向量回归

SVR

为了和SVM的优化问题保持形式一致，从而方便继承SVM自带的一套东西，可以考虑加一个L2正则项，也即

$$\begin{aligned}\min_{\boldsymbol{w},b,\boldsymbol{\xi},\hat{\boldsymbol{\xi}}} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2+\sum_{i=1}^{m}(\xi_i+\hat{\xi}_i)\\ \text{s.t.} \quad & -\epsilon-\xi_i \leqslant y_i-f(\boldsymbol{x}_i) \leqslant \epsilon+\hat{\xi}_i\\ & \xi_i \geq 0\\ & \hat{\xi}_i \geq 0\end{aligned}$$

通常还会考虑给松弛因子添加一个正则化常数C，同时约束条件也可以进行恒等变形，进而也就得到了常见的SVR优化问题

$$\begin{aligned}\min_{\boldsymbol{w},b,\boldsymbol{\xi},\hat{\boldsymbol{\xi}}} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}(\xi_i+\hat{\xi}_i)\\ \text{s.t.} \quad & y_i-f(\boldsymbol{x}_i) \leqslant \epsilon+\hat{\xi}_i\\ & f(\boldsymbol{x}_i)-y_i \leqslant \epsilon+\xi_i\\ & \xi_i \geq 0\\ & \hat{\xi}_i \geq 0\end{aligned}$$

此即为式6.45

结 语

在这次课程中，我们学习了西瓜书

软间隔与支持向量回归的公式推导

那么在下次课程中，我们将会学习西瓜书

朴素贝叶斯

deepshare.net
深度之眼

联系我们:

电话：18001992849

邮箱：service@deepshare.net

Q Q： 2677693114

公众号

客服微信