Support Vector Machine 02 Soft Margin

Chen Gong

15 November 2019

在上一小节中，我们介绍了 Hard-Margin SVM 的建模和求解过程。这个想法很好，但是实际使用过程中会遇到很多的问题。因为，并不一定数据集就可以被很好的分开，而且实际数据没有那么简单，其间有很多的噪声。而 Soft Margin 的基础思想就是允许那么一点点的错误。这样在实际运用中往往可以得到较好的效果。下面我们将进行 Soft Margin SVM 的详细演变过程。

1 Soft Margin SVM

最简单的思路就是在优化函数里面引入一个 loss function。也就是：

$$\min \frac{1}{2}w^Tw + loss\ function \tag{1}$$

那么，我们如何来定义这个 loss function 呢？大致可以分这两种引入的模式：

1. loss = 错误点的个数 = $\sum_{i=1}^{N} I\{y_i(w^Tx_i+b)<1\}$，这个方法非常容易想到，但是我们马上就发现了一个问题，那就是这个函数不连续的，无法进行优化。这种方法非常容易想到。

2. loss：距离。现在我们做如下定义：

1) 如果 $y_i(w^Tx_i+b)\geq 1$，$loss=0$。

2) 如果 $y_i(w^Tx_i+b)<1$，$loss=1-y_i(w^Tx_i+b)$。

那么，我们就可以将 loss function 定义为：

$$loss = \max\{0, 1-y_i(w^Tx_i+b)\} \tag{2}$$

进一步，我们令 $y_i(w^Tx_i+b)=z$，那么：

$$loss_{max} = \max\{0, 1-z\} \tag{3}$$

我们将 loss function 的图像画出来就如下图所示：

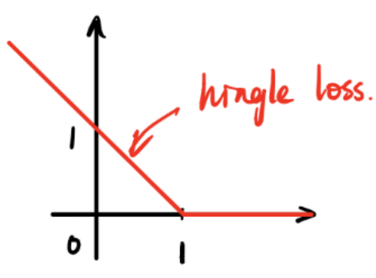

图 1: loss function 的展示图

这个 loss function 已经是连续的了，而且看起来是不是很像书的开着的样子。所以，它有一个非常形象的名字也就是“合页函数” (Hinge loss)。那么到这里，我们的 Soft Margin SVM 可以被定义为:

$$\left\{ \begin{array}{l} \min \ \frac{1}{2}w^Tw + C\sum_{i=1}^{N}\max\{0, 1-y_i(w^Tx_i+b)\} \\ s.t. \quad y_i(w^Tx_i+b) \geq 1 \end{array} \right. \tag{4}$$

但是，这样写显然不是我们想要的形式，我们需要得到更简便一些的写法。我们引入 $\xi_i = 1 - y_i(w^Tx_i+b)$，$\xi_i \geq 0$。我们仔细的想一想 $\max\{0, 1-y_i(w^Tx_i+b)\}$ 和 ξ_i 之间的关系。有了 $\xi_i \geq 0$，我们可以得到其实 $\xi_i \geq 0$ 和 $\max\{0, 1-y_i(w^Tx_i+b)\}$ 实际上是等价的。那么这个优化模型我们可以写成:

$$\left\{ \begin{array}{l} \min \ \frac{1}{2}w^Tw + C\sum_{i=1}^{N}\xi_i \\ s.t. \quad y_i(w^Tx_i+b) \geq 1-\xi_i, \ \xi_i \geq 0 \end{array} \right. \tag{5}$$

在图像上表示即为:

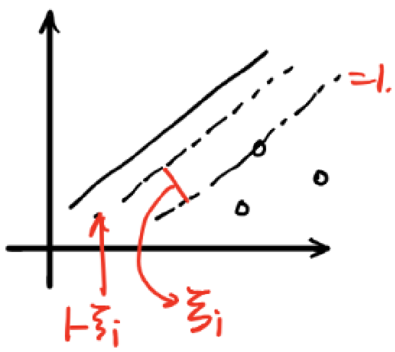

图 2: Soft Margin SVM 模型展示图

在以前的基础上我们增加了一个缓冲区，由于这个缓冲区的存在我们可以允许有点点的误差。而支持向量的区间被放宽到了 $1-\xi_i$。