SVM 推导

baiyun

baiyunpeng20@mails.ucas.ac.cn

1 符号定义

对于二分类问题，我们有如下定义：

1. 假定某一个点的特征向量为:$\boldsymbol{X_i} = x_1, x_2, ..., x_n$.
2. 所求的超平面为:$\boldsymbol{W}\boldsymbol{X} + b = 0$.
3. 若点在平面上方，即 $y_i = 1$, 有：$\boldsymbol{W^T}\boldsymbol{X_i} + b >= 1$.
4. 若点在平面上方，即 $y_i = -1$, 有：$\boldsymbol{W^T}\boldsymbol{X_i} + b <= -1$.
5. $\boldsymbol{W}\boldsymbol{X_i} + b = 1$ 和 $\boldsymbol{W}\boldsymbol{X_i} + b = -1$ 被称为支持向量.
6. 支持向量之间的距离为: $\frac{2}{||\boldsymbol{W}||}$.

2 开始推导

我们将平面上点所满足的约束条件加入到其中，可以得到 $y_i(\boldsymbol{W^T}\boldsymbol{X_i} + b) >= 1$。那么我们现在想要让支持向量之间的距离 $\frac{2}{||\boldsymbol{W}||}$ 最大，即：

$$max\ \frac{2}{||\boldsymbol{W}||}$$

$$subject\ to\ y_i(\boldsymbol{W^T}\boldsymbol{X_i} + b) >= 1$$

等价于：

$$min\ \frac{1}{2}||\boldsymbol{W}||$$

又等价于：

$$min\ \frac{1}{2}||\boldsymbol{W}||^2$$

这里，我们使用拉格朗日数乘法来进行求解。我们为第 i 个约束条件设置因子 α_i，因为我们需要求的是最小值，那么我们就需要让约束条件也就是 α_I 与其因子的结果是负的，因为如果这个结果是正的，那么所有的 α_i 取 0，最后的拉格朗日方程就是最小值了。我们让 α_i 取正数，整个方程转化为：

$$g(\boldsymbol{W}, b) = \frac{1}{2}||\boldsymbol{W}||^2 + \sum_i \alpha_i(1 - y_i(\boldsymbol{W^T}\boldsymbol{X_i} + b))$$

对 $\boldsymbol{W}, b$ 分别求偏导，有：

$$\frac{\partial g(\boldsymbol{W}, b)}{\partial \boldsymbol{W}} = ||\boldsymbol{W}|| + \sum_i -\alpha_i y_i \boldsymbol{X_i}$$

$$\frac{\partial g(\boldsymbol{W}, b)}{\partial b} = -\sum_i \alpha_i y_i$$

令偏导为 0，得：

$$||\boldsymbol{W}|| = \sum_i \alpha_i y_i x_i$$

$$\sum_i \alpha_i y_i = 0$$

将 $||\boldsymbol{W}||$ g$((W), b)$ 中，得到：

$$\frac{1}{2}(\sum_i^m \alpha_i y_i x_i \sum_j^m \alpha_j y_j x_j) + \sum_i^m \alpha_i(1 - y_i \sum_j^m \alpha_j y_j x_j x_i - y_i b)$$

继续展开:

$$-\frac{1}{2}\sum_{i}^{m}\sum_{j}^{m}\alpha_i\alpha_j y_{iyj}x_ix_j - b\sum_{i}^{m}\alpha_i y_i + \sum_{i}^{m}\alpha_i$$

在将 $\sum_i \alpha_i y_i = 0$ 带入该式，得到:

$$-\frac{1}{2}\sum_{i}^{m}\sum_{j}^{m}\alpha_i\alpha_j y_{iyj}x_ix_j + \sum_{i}^{m}\alpha_i$$

这里，通过 K-T 定理可以知道必然会有 $\boldsymbol{\alpha}$ 存在，但其求解方式则用到了 SMO 方法，这个方法目前还没有掌握，所以先留坑吧。