matlab_SVM 视频讲解_2

线性可分模式的最优超平面的详细推导过程:

考虑训练样本$\{x_i, y_i\}_{i=1}^{N}$,其中x_i是输入模式的第 i 个样本, $y_i \in \{-1,+1\}$.

设用于分离的超平面方程是:

$$w \bullet x + b = 0 \qquad (1)$$

其中 w 是超平面的法向量,b 是超平面的常数项.

[虚幻的叫法 w 是可调的权值向量,b 是偏置,上面的是 w 和 b 的数学本质]

现在的目的就是寻找最优的分类超平面,即寻找最优的 w 和 b.设最优的 w 和 b 为w_0和b_0,则最优的分类超平面为:

$$w_0 \bullet x + b_0 = 0 \qquad (2)$$

若得到上面的最优分类超平面,就可以用其来对测试集进行预测了.

设测试集合为$\{t_i\}_{i=1}^{M}$,**则用最优分类超平面预测出的测试集的标签为:**

$$t_{i\_}label = \mathrm{sgn}(w_0 \bullet t_i + b_0) \qquad (3)$$

上面是明确问题目的性,也就是到底要解决什么问题,要做什么.

SVM 的主要思想是建立一个超平面作为决策曲面,使得正例和反例之间的隔离边缘被最大化.

即最优分类超平面等价于求最大间隔

定义 支持向量

满足下面条件的特殊数据点(x_i, y_i)称为支持向量:

$$w \bullet x_i + b = -1, y_i = -1 \qquad (4)$$

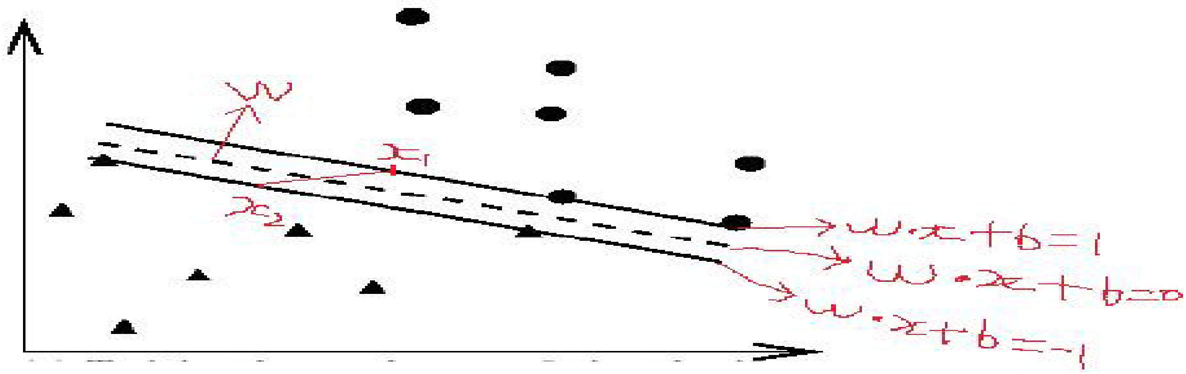
or

$$w \bullet x_i + b = 1, y_i = +1 \qquad (5)$$

支持向量是那些最靠近决策面的数据点,这样的数据点是最难分类的,

因此,它们和决策面的最优位置直接相关.

推导最大间隔:



$w\cdot x_1+b=1 \quad w\cdot x_2+b=-1$

$\Rightarrow \quad w\cdot(x_1-x_2)=2$

$\Rightarrow \quad \frac{w}{\|w\|}(x_1-x_2)=\frac{2}{\|w\|}$

$\max\frac{2}{\|w\|} \Leftrightarrow \min\|w\| \Leftrightarrow \min\frac{\|w\|^2}{2}$

设 x_1, x_2 如图,则正反例的间隔为 $dis=\frac{w}{\|w\|}\bullet(x_1-x_2)$.又

$$w\bullet x_1+b=1 \qquad (6)$$

$$w\bullet x_2+b=-1 \qquad (7)$$

所以 $dis=\frac{w}{\|w\|}\bullet(x_1-x_2)=\frac{2}{\|w\|}$.

则 $\frac{2}{\|w\|}$ 最大化 $\Leftrightarrow$ $\|w\|$ 最小化 $\Leftrightarrow$ $\frac{\|w\|^2}{2}$ 最小化.

且对于任意的 (x_i, y_i) 有[可以证明不详说]:

$$\begin{cases} w\bullet x_i+b\le -1, y_i=-1 \\ w\bullet x_i+b\ge 1, y_i=1 \end{cases} \quad \text{即(结合一下)}\ y_i(w\bullet x_i+b)\ge 1 \qquad (8)$$

so. 寻找最优超平面即正反例间隔最大化问题,最终归结为一个二次最规划问题[原问题 primal pro]:

$$\min_{w} \frac{\|w\|^2}{2}$$

$$\text{s.t} \quad y_i(w \bullet x_i + b) \geq 1, \forall i = 1,2,...N$$

使用 Lagrange 乘子法可以解决上述问题.

首先建立 Lagrange 函数:

$$J(w,b,a) = \frac{1}{2}w^T w - \sum_{i=1}^{N} a_i[y_i(w \bullet x_i + b) - 1] \qquad (9)$$

其中辅助非负变量 a_i 称为 Langrange 乘子.

对 w 和 b 求偏导并置零,有

$$\frac{\partial J}{\partial w} = 0 \Leftrightarrow w = \sum_{i=1}^{N} a_i y_i x_i \qquad (10)$$

$$\frac{\partial J}{\partial w} = 0 \Leftrightarrow \sum_{i=1}^{N} a_i y_i = 0 \qquad (11)$$

再整理 J 最终可以得到原问题的对偶问题(dual pro):

$$\max_{a} Q(a) = J(w,b,a) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j y_i y_j x_i^T x_j$$

$$\text{s.t} \quad \sum_{i=1}^{N} a_i y_i = 0, a_i \geq 0$$

对偶问题完全是根据训练数据来表达的.而且,函数 Q(a)的最大化仅依赖于输入模式点积的集合$\{x_i^T x_j\}$

求解出求偶问题的最优解,设用a_i^*表示最优的 Lagrange 乘子,则此时原问题的最优解为:

$$w_0 = \sum_{i=1}^{N} a_i^* y_i x_i$$

$$b_0 = 1 - w_0 x^{(s)}, y^{(s)} = 1$$

至此原问题就得到了解决!

则判断函数为:

$\mathrm{sgn}(\sum_{i=1}^{N} a_i^* y_i x_i x + b)$,其中 x 为测试集中的样本.