机器学习笔记（八）支持向量机(SVM)

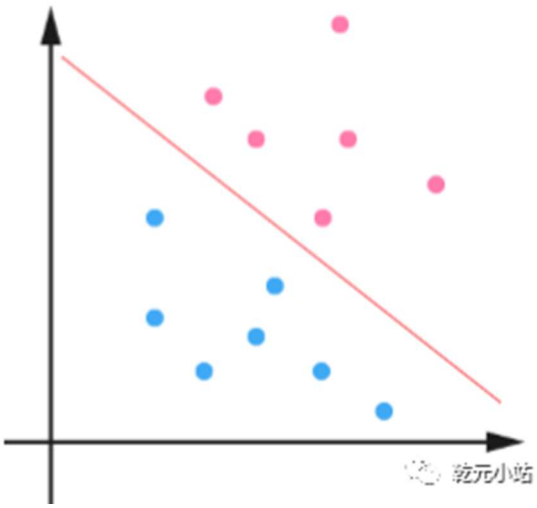
玖五乾元

（一）定义

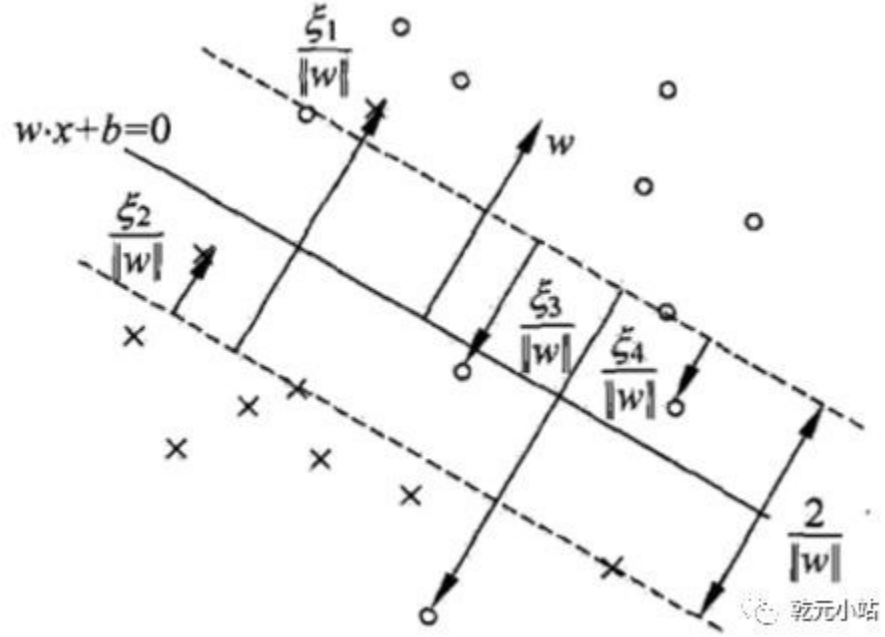
支持向量机(Support vector machines ,SVM)顾名思义是支持向量和机，“支持量”是支持或支撑平面上把两类类别划分开来的向量点，“机”是机器学习领域里对一些算法的称呼。

SVM 是一种二类分类的模型， 分别有线性分类和非线性分类两类模型。其中线性分类模型类似于感知机，但又有别于感知机，SVM 通过支持向量间隔最大化使得其在线性分类场景下分类效果优于感知机，同时还能够支持近似线性分类的场景

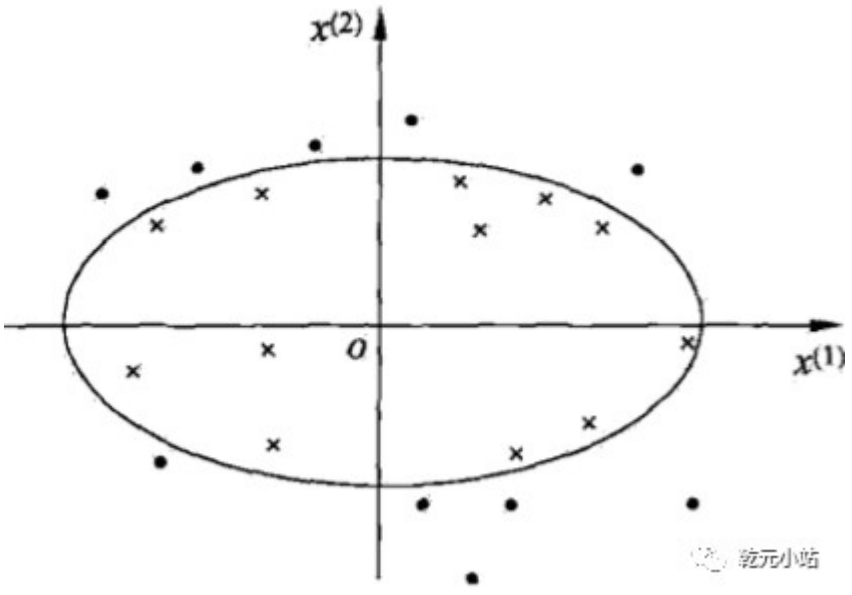
注：1.线性可分，是针对训练数据、测试数据等待处理数据是线性可分的，通俗理解，一个平面上至少存在一条直线能够将数据分成两堆。如下图所示：



2.近似线性可分，是指一条直线能把绝大多数数据分成两堆，有少数在对方的那一堆里。如下图所示，在实线两侧有 x 跑到 o 里了，也有 o 跑到 x 里了。

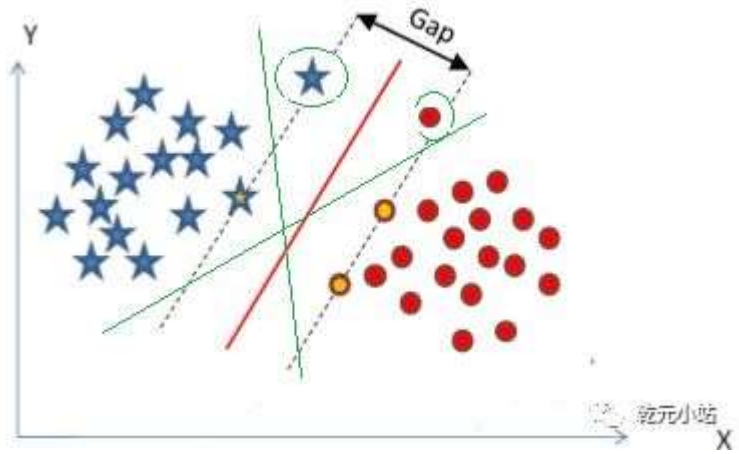


3. 线性不可分，通俗理解是在一个平面上找不到一条直线将数据分成两堆，但数据又明显是两类。如下图所示：



本文着重讲线性可分情况的，其他 两种情况在后续文章中继续探讨，因此，下面所述“模型”，“策略”，“算法”都针对线性可分支持向量机。

在感知机一文中我们知道，其本质是对线性可分训练数据通过寻找误分类最小的超平面 $w \cdot x + b = 0$ 来实现，对数据的分类，实际上这样的平面会有无穷多个，选择不同的 w,b 初值，不同的学习步长，不同的迭代顺序都会得到不同的 w,b 最终值，但都能达到将数据分类的目标。其缺点是这个计算得到的分类超平面对新输入数据(不是训练集合里的数据，这里顺首说一下，机器学习一般先拿一组已知数据用于训练得到模型，再用于对新输入的数据 进行运算)的预测能力不一定是最优的。如下 图所示，其中两条绿色的线都可以将已知的训练数据集准确的分开，但对地于由绿色线圈起来的新数据这两条绿线就会分别做出错误的判断。基于上述原因，SVM 引入间隔最大化来进一步约束感知机的最终结果（间隔有函数间隔和几何间隔，自行脑补相关知识，笼统直观的可以理解为点到线或面的垂直距离），图中红线所代表的就是通过间隔最大化求得的分类超平面，可见他比绿线所代表的感知机方法所求的得分类超平面更优化。



（二）模型

与感知机模型相同，

$$y = f(x) = sign(w \cdot x + b) = \begin{cases} -1, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

w 叫作权值，b 叫作偏置，表示 w 和 x 的内积，sign 是符号函数

w,b 取不同值形成不同的 x 到 y 的映射空间（函数的集合）

（三）策略

a.目标：确定**经验损失函数**并将损失函数最小化。

b. 经验损失函数可选择：

函数间隔：训练数据任意一点 (x_i, y_i) 到 $w \cdot x + b = 0$ 所标识的超平面的函数间隔为 $\hat{\gamma}_i = y_i(w \cdot x_i + b)$，所有点到超平面函数间隔最小为 $\hat{\gamma} = \min_{i=1,2...N} \hat{\gamma}_i$

几何间隔：训练数据任意一点(x_i, y_i)到 $w \cdot x + b = 0$ 所标识的超平面的几何 间隔为

$$\gamma_i = y_i\left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right)$$

，其中，||w||为 w 的 L2 范数，所有点到超平台几休间隔最小值为 $\gamma = \min_{i=1,2\ldots N} \gamma_i$。

由此，**策略**：对已知训练数据集所有点求得几何间隔最大的超平面即为最优化分类，即满足上述 γ 最大超平面。这块有点绕大家好好理解一下,通俗讲，对已知训练数据集有无穷多个 w,b 决定的超平面，对于每个超平面所有点到他的几何间隔都有个最小值，这些最小值中最大的那个超平面就是我们要的解。数学表示为：$\max_{w,b} \gamma$，满足条件

$$y_i\left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right) \geq \gamma, i = 1,2, \ldots, N$$

进一步推导：

根上述两个间隔的定义可见函数间隔和几何间隔之间是||w||倍数关系，代入可得

$$\max_{w,b} \frac{\hat{r}}{\|w\|} \quad \text{满足} \quad y_i(w \cdot x_i + b) \geq \hat{\gamma}, i = 1,2 \ldots N$$

进一步等价推导：

1. 考虑到函数间隔对上面的不等式成立条件不会有影响（把 w,b 成经例放大时，不会引起最小间隔数据点的变化）。所以函数间隔可直接取单位数值 1

2. 求$\frac{1}{\|w\|}$最大等价于求$\frac{1}{2}\|w\|^2$最小（为了数据计算方便）

将上述两项等价替换得最终数学表达式为（即，SVM 线性可策略）：

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i = 1, 2, \ldots$$

（四）算法

至此，对于 SVM 的策略求解转化为纯数学问题求解，至于数学解的存在性不在计述范围呢。该数据问题为一个典型的凸二次规划问题，即约束最优化问题。其求解方法有原始和对偶两种算法，将在下一章中与实例一起介绍

您的支持就是我的动力，请长按下方二维码关注“乾元小站”