支持向量机（SVM）

2019年10月15日　　18:30

作者：zx青

支持向量机（SVM）

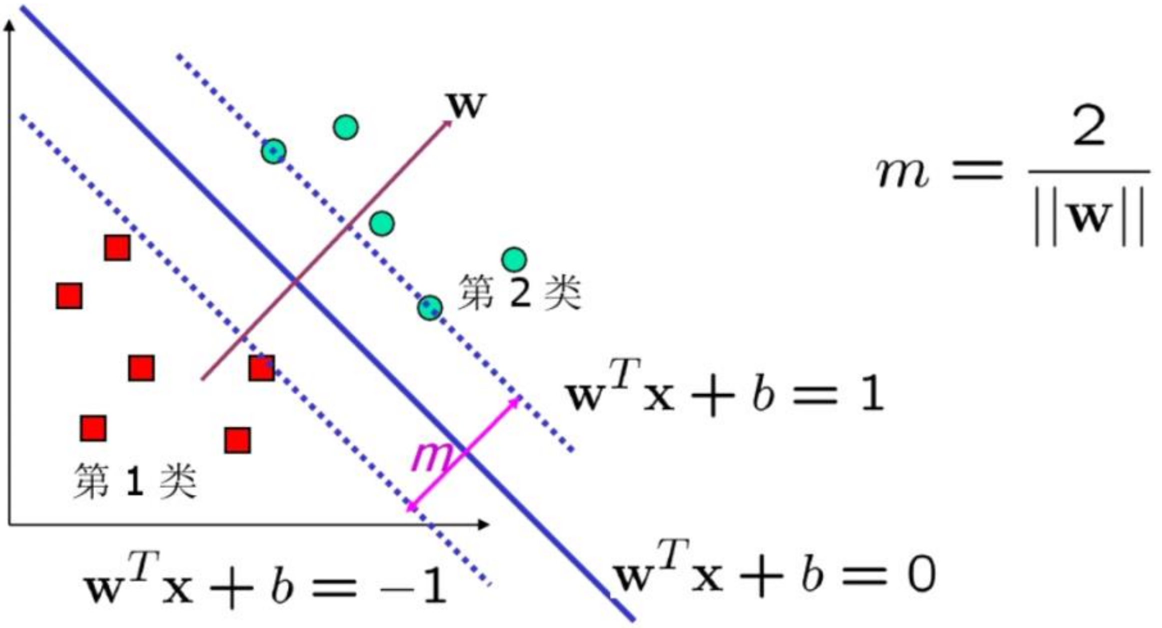
（参考资料：
http://bytesizebio.net/2014/02/05/support-vector-machines-explained-well/
https://www.zhihu.com/question/21094489/answer/86273196）

一、什么是SVM

支持向量机（Support Vector Machine,SVM)是用于分类的一种算法。

SVM主要的思想是：建立一个具有最大间隔的超平面，使得该平面两侧距平面最近的两类样本之间的距离最大化，从而对分类问题提供良好的泛化能力。将数据分隔开的对象称为超平面，也就是分类的决策边界，那个具有最大间隔的超平面对应的两侧虚线所穿过的样本点就是SVM中的支持样本点，称为支持向量(support vector)。支持向量到超平面的距离被称为间隔（margin)

分类的超平面表示为：$w^T x + b = 0$



二、线性可分SVM

（1）问题描述

假设一个特征空间上的训练数据集：$D = \pi r^2\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$，*其中*$x_i \in R^m, y_i \in \{-1,1\}, i = 1,2,\dots,m$*其中*$x_i$*是第i个特征向量*$y_i$*为类标记*，当它等于+1时为正例；为-1时为负例。再假设训练数据集是线性可分的。

线性可分的情况意味着存在超平面使训练样本点中的正类和负类样本分别位于该超平面的两侧。

$w^T x + b = 0$

如果能确定这样的参数对(w,b)的话，就可以构造决策函数来进行识别新样本。

$f(x) = \operatorname{sgn}\left(\left(w^T x\right) + b\right)$

问题是这样的参数对(w,b)有许多。

解决方法：采用最大间隔原则（选择似的训练集D对于线性函数$w^T x + b$的几何间隔取最大值的参数对(w, b)，并由此构造决策函数。

几何间隔：点到超平面的距离。$\gamma = \frac{|w^T x + b|}{\|w\|}$(二维平面中，若点坐标为$(x_0, y_0, z_0)$)点到平面$Ax + By + Cz + D = 0$*，则点到平面的*距离公式：$\mathrm{d} = \left|\frac{Ax_0 + By_0 + Cz_0 + D}{\sqrt{A^2 + B^2 + C^2}}\right|$

（2）优化目标：

$$\begin{cases} \max\limits_{w,b} \gamma \\ s.t. \quad y_i\left(w^T x_i + b\right) \geq 1 \ ,\forall i \end{cases} \Rightarrow \begin{cases} \max\limits_{w,b} \dfrac{1}{\|w\|} \\ s.t. \quad y_i\left(w^T x_i + b\right) \geq 1 \ ,\forall i \end{cases}$$

即

$$\begin{cases} \min\limits_{w,b} \quad \dfrac{1}{2}\|w\|^2 \\ s.t. \quad y_i\left(w^T x_i + b\right) \geq 1 \ ,\forall i \end{cases}$$

（这里 *转化为* $\frac{1}{2}\|w\|^2$ *是为了求导方便）*

（3）优化问题求解：

使用Lagrange乘子法将其转化为对偶问题，于是引入Lagrange函数

$$\mathcal{L}(\omega, b, \alpha) = \frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{m} \alpha_i [y^{(i)}(\omega^T x^{(i)} + b) - 1]$$

其中 α_i为拉格朗日乘子，且 $\alpha_i \geq 0$。

优化目标转化为：

$\min\limits_{w,b} \max\limits_{\alpha_i \geq 0} L(w, b, \alpha)$

该问题的对偶问题为：

$\max\limits_{\alpha_i \geq 0} \min\limits_{w,b} L(w, b, \alpha)$

求解步骤：先求优化函数对于w,b的极小值，然后再求拉格朗日乘子α_i 的极大值。

❶求$\min_{w,b} L(w, b, \alpha)$ *对*w, b*分别求偏导数令其为0得：*

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{m} a_i y_i x_i$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m} a_i y_i = 0$$

❷令$\varphi(\alpha) = \min_{w,b} L(w, b, \alpha)$,, *代入上面求出的关系，得：*

$$\varphi(\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} a_i [y_i (w^T x_i + b) - 1]$$

$$= \frac{1}{2} w^T w - \sum_{i=1}^{m} a_i y_i w^T x_i - \sum_{i=1}^{m} a_i y_i b + \sum_{i=1}^{m} a_i$$

$$= \frac{1}{2} w^T \sum_{i=1}^{m} a_i y_i x_i - w^T \sum_{i=1}^{m} a_i y_i x_i - \sum_{i=1}^{m} a_i y_i b + \sum_{i=1}^{m} a_i$$

$$= -\frac{1}{2} w^T \sum_{i=1}^{m} a_i y_i x_i - b \sum_{i=1}^{m} a_i y_i + \sum_{i=1}^{m} a_i$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{m} a_i y_i x_i\right)^T \sum_{i=1}^{m} a_i y_i x_i - b * 0 + \sum_{i=1}^{m} a_i$$

$$= -\frac{1}{2}\sum_{i=1}^{m} a_i y_i (x_i)^T \sum_{i=1}^{m} a_i y_i x_i + \sum_{i=1}^{m} a_i$$

$$= \sum_{i=1}^{m} a_i - \frac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j (x_i)^T x_i$$

w求解$\max\limits_{\alpha \geq 0} \ \varphi(\alpha)$

$$\begin{cases} \max\limits_{\alpha \geq 0} \left[\sum_{i=1}^{m} a_i - \dfrac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j (x_i)^T x_i\right] \\ s.t. \sum_{i=1}^{m} a_i y_i = 0, \qquad a_i \geq 0, \mathrm{i} = 1,2,\ldots,\mathrm{m} \end{cases}$$

等价于求解如下极小化问题

$$\begin{cases} \min\limits_{\alpha\geq 0}\ [\frac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j (x_i)^T x_i - \sum_{i=1}^{m} a_i] \\ s.t. \sum_{i=1}^{m} a_i y_i = 0, \qquad a_i \geq 0, \mathrm{i} = 1,2,\ldots,\mathrm{m} \end{cases}$$

上式是一个凸二次规划问题，有唯一的最优解。求得该问题，得α，*则参数对*$(w.b)$*可由下式计算:*

$$w^* = \sum_{i=1}^{m} a_i{}^* y_i x_i \quad , \qquad b^* = -(w^* \sum_{i=1}^{m} a_i{}^* x_i)/(2\sum_{y_i=1}^{m} a_i{}^*)$$

于是可得到如下决策函数:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{m} a_i{}^* y_i (x \cdot x_i{}^T) + b^*\right) = \mathrm{sgn}\left(\sum_{i=1}^{m} a_i{}^* y_i (x \cdot x_i) + b^*\right)$$ （这里的·表示向量内积）

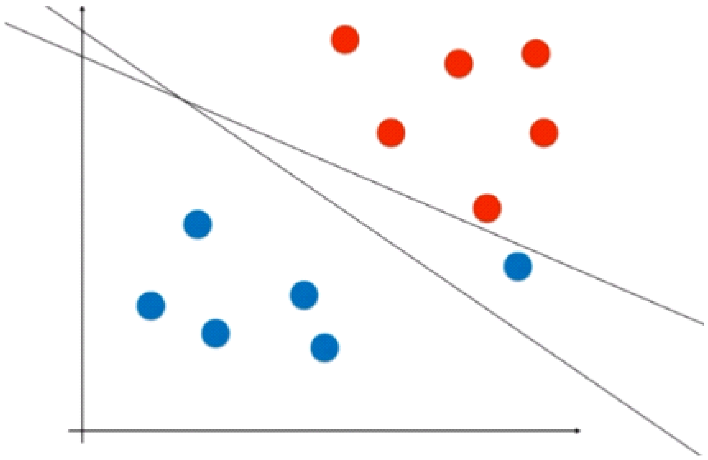
支持向量：称训练集D中的样本x_i为支持向量，如果它对应的$a_i{}^* > 0$。

根据原始最优化问题的KKT条件，有$a_i{}^* \left(y_i\left(w^T x_i + b\right)\right) - 1) = 0$，于是支持向量正好在间隔边界上。

三、线性不可分SVM

（1）问题描述

现在考虑线性不可分的情况，对于训练集D，不存在这样的超平面，使训练集D,不存在这样的超平面，使训练集关于该超平面的几何间隔取正值。如果要用超平面来划分的话，必然有错分的点。但是我们希望使用超平面进行划分，这时应“软化”对间隔的要求，即容许不满足约束条件的样本点存在（软间隔）。



（2）优化目标

$y_i\left(w^T x_i + b\right) \geq 1 - \varepsilon_i, i = 1,1,\ldots,m,$ *其中*ε_i*称为松弛变量。*

为避免ε_i*取太大的值，需要在*目标函数中对它们进行惩罚，于是原始优化问题变为:

$$\begin{cases} \min\limits_{w,b,\varepsilon}\ \frac{1}{2}\|w\|^2 + C\sum_{y_i=1}^{m} \varepsilon_i \\ \text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 - \varepsilon_i, i = 1,1,\ldots,m \\ \qquad \varepsilon_i \geq 0, i = 1,2,\ldots,m \end{cases}$$

其中C>0称为惩罚因子，用于控制目标函数中两项（寻找margin最大的超平面和保证数据点偏差量最小）直接的权重。

（3）优化问题求解:

类似前面引入如下的Lagrange函数:

$$L(w,b,\varepsilon,\alpha,\gamma) = \frac{1}{2}\|w\|^2 + C\sum_{y_i=1}^{m} \varepsilon_i - \sum_{i=1}^{m} a_i(y_i(w^T x_i + b) - 1 + \varepsilon_i) - \sum_{i=1}^{m} r_i \varepsilon_i$$

得到如下的对偶问题:

$$\begin{cases} \min\limits_{\alpha} \ [\frac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j < x_j \ x_i > - \sum_{i=1}^{m} a_i] \\ \sum_{i=1}^{m} a_i y_i = 0 , 0 \le a_i \le C , i \ = 1,2, \dots , n \end{cases}$$

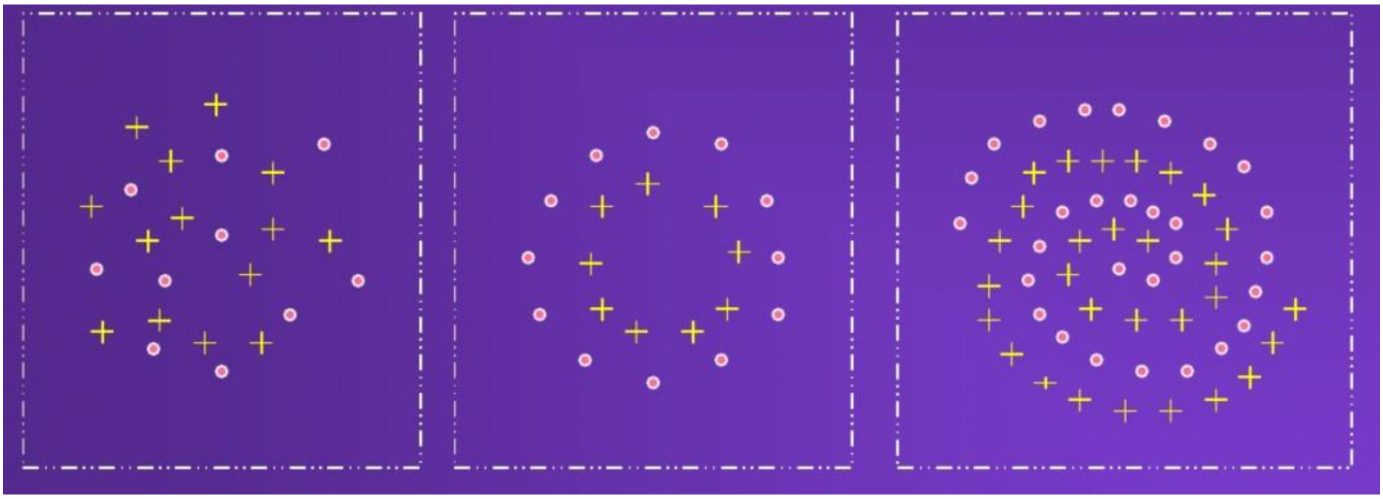
求解对偶问题，可得如下决策函数：

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} a_i{}^* y_i \ < x x_i > + \ b^* \right)$$

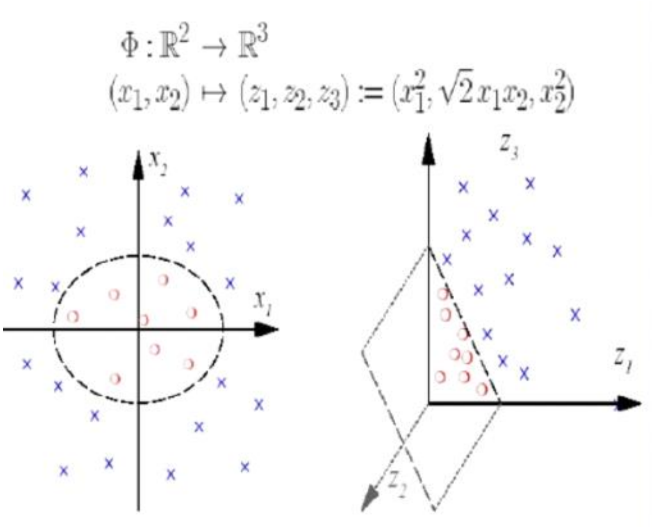
四、非线性可分SVM

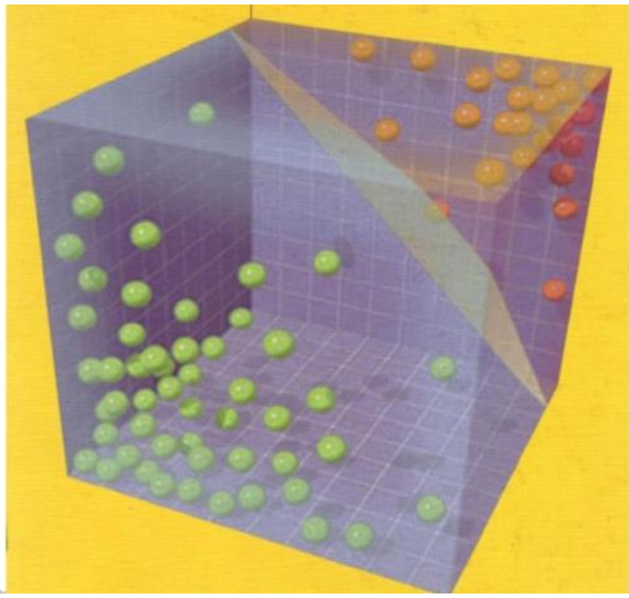
（1）问题描述

对于一般的非线性二分可分情况，对于训练集D，无法寻找到来如前的超平面来划分。



SVM选择一个核函数K(x,z)，通过将数据映射到高维空间来解决在原始空间中线性不可分问题。





引入一个非线性映射$\emptyset(\cdot)$*将其映射到一个高维空间中，数据变的线性可分了：*

核函数： $k\left(x_i, x_j\right) = \ \emptyset\left(x_i\right) \cdot \emptyset(x_j)$

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i k(x_i, x) + \ b \ = \sum_{i=1}^{m} \alpha_i y_i (\emptyset\left(x_i\right) \cdot \emptyset(x) > + \ b$$

（2）优化目标：

其目标优化问题为：

$$\min_{w,b,\varepsilon} \quad \frac{1}{2}\|w\|^2 + C\sum_{y_i=1}^{m}\varepsilon_i$$

s.t. $y_i\left(w^T x_i + b\right) \geq 1 - \varepsilon_i, i = 1,1,\ldots,m$

$\varepsilon_i \geq 0, i = 1,2,\ldots,m$

（3）优化问题求解：

对应的对偶问题为：

$$\begin{cases} \min\limits_{\alpha} \; [\frac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j \, k(x_j \cdot x_i) - \sum_{i=1}^{m} a_i] \\ \sum_{i=1}^{m} a_i y_i = 0\,, 0 \leq a_i \leq C, i = 1,2,\ldots,m \end{cases}$$

求解对偶问题，可得如下决策函数：

$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} a_i{}^* y_i k(x_j \cdot x_i) + b^*\right)$，其中$b^* = y_j - \sum_{i=1}^{m} a_i{}^* y_i k\left(x_j \cdot x_i\right) \quad (0 < a_j{}^* < C)$