进度：李航《统计学习方法》第七章 支持向量机

1. 线性可分支持向量机与硬间隔最大化

- 学习的目标：首先假设数据集是线性可分的(感知机章节的概念)
- **线性可分支持向量机**：给定线性可分数据集，通过间隔最大化或等价的求解相应的凸二次规划问题得到的分离超平面为$w^* \cdot x + b^* = 0$，以及相应的决策函数$f(x) = sign(w^* \cdot x + b^*)$称为线性可分支持向量机
- **间隔**：对于给定的训练数据集和超平面(w,b)，定义超平面(w,b)关于样本点(x_i,y_i)的函数间隔为 $\gamma_i = y_i(w \cdot x_i + b)$ 定义超平面(w,b)关于T中所有的样本点(x_i,y_i)的函数间隔的最小值，即 $\gamma = min\ \gamma_i$

  函数间隔可以表示分类的正确性和可信度
- **间隔的计算**：它就等于两个异类支持向量的差在 W方向上的投影， W方向是指图6.2所示实线的法线方向。

$$\begin{cases} 1*(w^T x_+ + b) = 1, y_i = +1 \\ -1*(w^T x_- + b) = 1, y_i = -1 \end{cases}$$

  进而可以推出：

$$\begin{cases} w^T x_+ = 1 - b \\ w^T x_- = -1 - b \end{cases}$$

  我们求得了间隔，SVM的思想是使得间隔最大化，也就是，显然，最大化 2/||w|| 相当于最小化 ||w||，为了计算方便，将公式转化成支持向量机的基本型，该基本型是一个凸二次规划问题，可以采用拉格朗日乘子法对其对偶问题求解求解，拉格朗日函数：

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(w^T x_i + b)) \qquad (8)$$

  对W,b求导可得

$$\begin{cases} \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i x_i \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{m} \alpha_i y_i \end{cases}$$

  令其分别为0，可得：

$$\begin{cases} w = \sum_{i=1}^{m} \alpha_i y_i x_i & (9) \\ \sum_{i=1}^{m} \alpha_i y_i = 0 & (10) \end{cases}$$

  将其带入拉格朗日函数（8）中，可得：

$$L(w,b,\alpha)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i x_j$$

$$s.t.\quad \sum_{i=1}^{m}\alpha_i y_i=0,\quad \alpha_i\geq 0,\quad i=1,2,\dots,m \tag{11}$$

原问题就转换为如下关于aa的问题：

$$\max_{\alpha}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i x_j$$

$$s.t.\quad \sum_{i=1}^{m}\alpha_i y_i=0,\quad \alpha_i\geq 0,\quad i=1,2,\dots,m \tag{12}$$

解出 α 之后，根据公式（9）可以求得 w ， 进而求得 b，可以得到模型：

$$f(x)=w^T x+b=\sum_{i=1}^{m}\alpha_i y_i x_i^T x+b \tag{13}$$

该过程的KKT条件为

$$\begin{cases}\alpha_i\geq 0\\ y_i f(x_i)-1\geq 0\\ \alpha_i(y_i f(x_i)-1)=0\end{cases}$$

2. 线性可分支持非线性支持向量机和核函数

对于非线性问题，线性可分支持向量机并不能有效解决，要使用非线性模型才能很好地分类。先看一个例子，如下图，很显然使用直线并不能将两类样本分开，但是可以使用一条椭圆曲线（非线性模型）将它们分开。非线性问题往往不好求解，所以希望能用解线性分类问题的方法求解，因此可以采用非线性变换，将非线性问题变换成线性问题。 对于这样的问题，可以将训练样本从原始空间映射到一个更高维的空间，使得样本在这个空间中线性可分，如果原始空间维数是有限的，即属性是有限的，那么一定存在一个高维特征空间使样本可分。令ϕ(x)表示将 x 映射后的特征向量，于是在特征空间中，划分超平面所对应的的模型可表示为：

f(x)=wTϕ(x)+b （14）

于是有最小化函数：

$$\min_{w,b}\frac{1}{2}\|w\|^2,\quad s.t.\quad y_i(w^T\phi(x_i)+b)\geq 1\ (i=1,2,\dots,m) \tag{15}$$

其对偶问题为：

$$\max_{\alpha}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j)$$

$$s.t.\quad \sum_{i=1}^{m}\alpha_i y_i=0,\quad \alpha_i\geq 0,\quad i=1,2,\dots,m \tag{16}$$

若要对公式（16）求解，会涉及到计算ϕ(xi)Tϕ(xj)ϕ(xi)Tϕ(xj)，这是样本 xi 和 xj映射到特征空间之后的内积，由于特征空间的维数可能很高，甚至是无穷维，因此直接计算ϕ(xi)Tϕ(xj)ϕ(xi)Tϕ(xj)通常是困难的，于是想到这样一个函数：

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j)$$

$$s.t. \ \sum_{i=1}^{m} \alpha_i y_i = 0, \ \alpha_i \geq 0, \quad i = 1, 2, \dots, m \qquad (18)$$

即 xi和 xj在特征空间中的内积等于他们在原始样本空间中通过函数 κ(xi,xj)计算的函数值，于是公式（16）写成如，求解后得到：

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j)$$

$$s.t. \ \sum_{i=1}^{m} \alpha_i y_i = 0, \ \alpha_i \geq 0, \quad i = 1, 2, \dots, m \qquad (18)$$

$$\begin{aligned} f(x) &= w^T \phi(x) + b \\ &= \sum_{i=1}^{m} \alpha_i y_i \phi(x_i)^T \phi(x_j) + b \qquad (19) \\ &= \sum_{i=1}^{m} \alpha_i y_i \kappa(x_i, x_j) + b \end{aligned}$$

这里的函数 κ(xi,xj) 就是核函数，在实际应用中，通常人们会从一些常用的核函数里选择（根据样本数据的不同，选择不同的参数，实际上就得到了不同的核函数）