第8章 支持向量机

目录

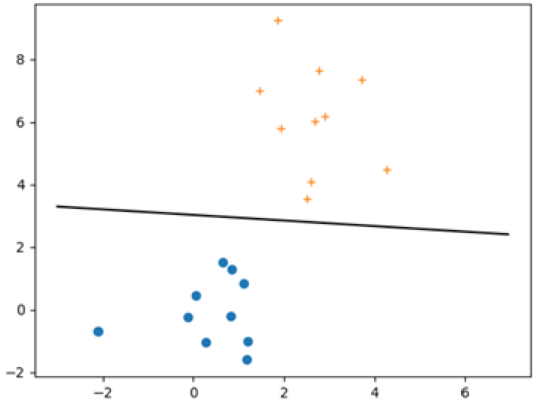
CONTENTS

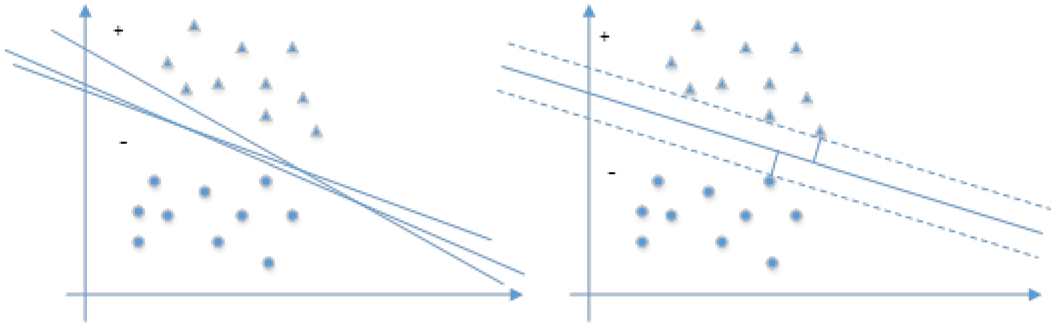
8.1 SVM简介

支持向量机是Cortes和Vapnik于1995年提出的一种基于统计学习的二分类模型。它是一种监督学习方法，在学习过程中通过最大化分类间隔使得结构风险最小化。

图8-1所示是两类线性可分的样本数据分布及划分的示例。图中的线段就是对样本分隔的超平面。

从图8-2可以看出，能将不同样本分开的超平面有很多，但只有一条超平面位于两类样本的“正”中间，这个超平面通常用一个方程d(X)=0来表示，d(X)被称为判决函数或决策函数。





8.1 SVM简介

1. 感知机模型

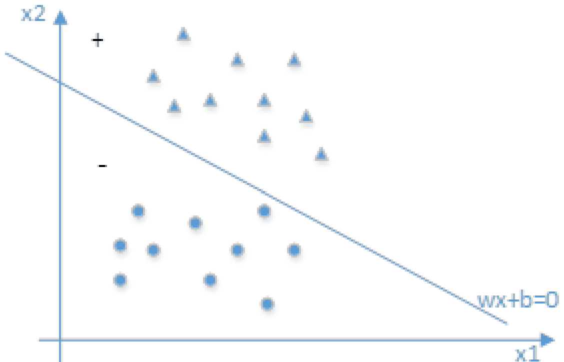
假设输入样本空间 ，输出空间是Y={+1,-1}，输入样本 表示样本的特征向量，即输入空间的样本点；输出 表示样本的类别。从输入样本空间到输出样本空间的函数可表示为：$f(x)=sign(w\cdot x+b)$

该函数称为感知机（Perceptron），其中，$w\in R^n$ 称为权值（Weight）或权值向量（Weight Vecto），$b\in R$ ，称为偏置（Bias），sign为符号 $\mathrm{sign}(x)=\begin{cases}+1 & x\geqslant 0\\ -1 & x<0\end{cases}$

8.1 SVM简介

判决函数

若f(x)>0，则属于正例；若f(x)<0，则属于负例。在特征空间 中，令判决函数g(x)$w \cdot x + b$，线性方程 是一个超平面S，其中w是超平面的法向量，b是超平面的截距。这个超平面将特征空间划分为两部分，这两部分的样本点分别被分成正、负两例，超平面S就是分离超平面。



8.1 SVM简介

2　模型参数学习

任意一点(x, y)到直线 $Ax+By+C=0$ $\frac{|Ax+By+C|}{\sqrt{A^2+B^2}}$ 距离为，

因此二维样本 $w^T x+b=0$ y)到线性方程 $\frac{|w^T x+b|}{\|w\|_2}$ $\|w\|_2=\sqrt[2]{w_1^2+w_2^2+\ldots+w_n^2}$ 为

，其中，

对于误分样本点(xi, yi)， $w^T x_i+b>0$ 。， $y_i=-1$ 时，有 $w^T x_i+b<0$ ；当 $y_i=1$ 时，

$-y_i(w^T x_i+b)>0$ 。因此， $-\frac{1}{\|w\|_2}|y_i(w^T x_i+b)|$

于是误分样本点到超平面S的距离为：

若 $-\frac{1}{\|w\|_2}\sum_{i=1}^{N} y_i(w^T x_i+b)$ 误分 $\mathrm{Loss}(w,b)=-\frac{1}{\|w\|_2}\sum_{i=1}^{N} y_i(w^T x_i+b)$ 则误分样本点到超平面S的总距离为：

8.1 SVM简介

采用随机梯度下降法（Stochastic Gradient Descent，SGD）学习参数w和b：

$$\frac{\partial \mathrm{Loss}(w,b)}{\partial w} = -\sum_{i=1}^{N} y_i x_i \qquad \frac{\partial \mathrm{Loss}(w,b)}{\partial b} = -\sum_{i=1}^{N} y_i$$

参数w和b的迭代更新公式：

$$w \leftarrow w - \eta \frac{\partial \mathrm{Loss}(w,b)}{w} = w + \eta y_i x_i$$

$$b \leftarrow b - \eta \frac{\partial \mathrm{Loss}(w,b)}{b} = b + \eta y_i$$

8.1 SVM简介

输入：训练数据集D={(x1,y1),(x2,y2),...,(xN,yN)}、迭代次数、学习率 ，

其中：$x \in X$ ，Y={+1,-1}。

过程：

（1）初始化参数：$w=(0,0,\dots,0)^T$ ，b=0。

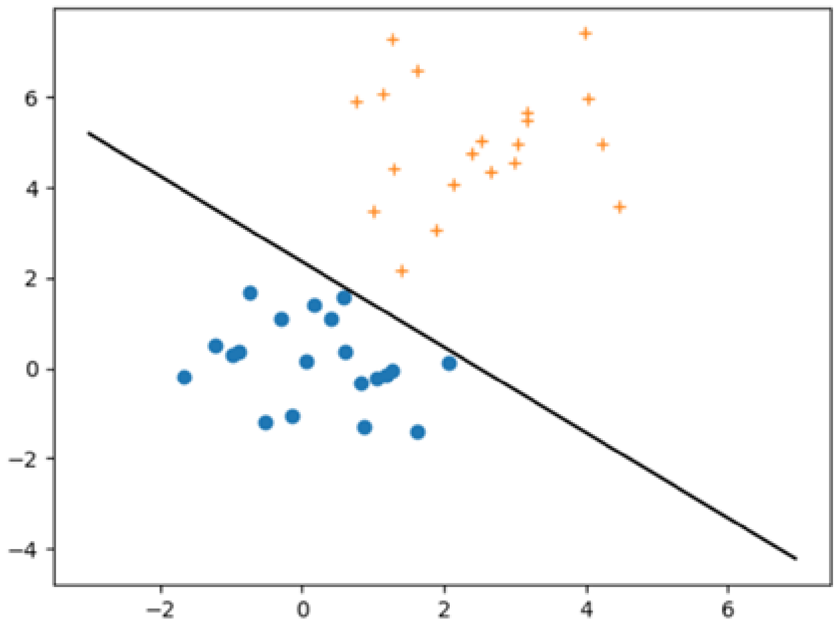
（2）对于j=1,2,...,N，当 $X=\{(x_i,y_i)\mid y_i(w\cdot x_i+b)\leqslant 0\}$为空集，即没有误分样本点，则结束循环，否则转到第（3）步执行。

（3）任意取X中的样本点(xi,yi)更新参数。

输出：感知机模型参数w和b，并利用$f(x)=\mathrm{sign}(w\cdot x_i+b)$ 计算分类的准确率。

8.1 SVM简介

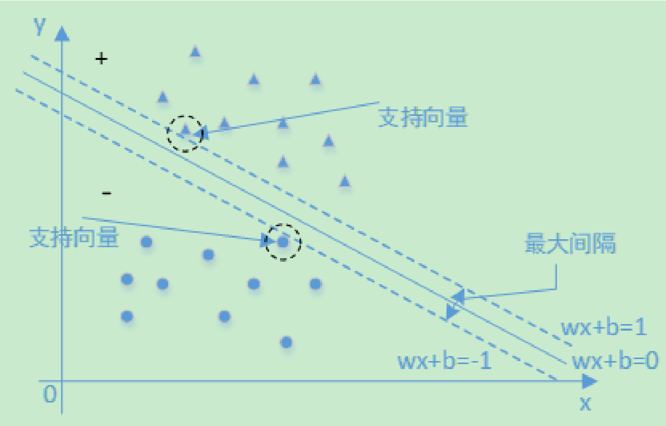
利用以上感知机算法对样本进行分类，其散点图及分类结果如图8-4所示。



8.1 SVM简介

1. 间隔最大化

在对样本数据分类时，超平面离数据点的间隔越大，产生误差的可能性就会越小，也就是分类的确信度越大。因此，为了使分类的确信度尽可能高，需要让选择的超平面尽可能地最大化这个间隔。以最大间隔把两类样本分开的超平面，称之为最大间隔超平面。分类问题中的最大间隔、支持向量表示如图8-5所示。



8.1 SVM简介

支持向量就是离最大间隔超平面最近的样本点，根据前面得到的支持向量到超平面的距离为

$$\begin{cases} \dfrac{w^T x+b}{\|w\|_2} \geqslant d, & y=1 \\ \dfrac{w^T x+b}{\|w\|_2} \leqslant -d, & y=-1 \end{cases}$$

将上式进行变换，进而有：

$$\begin{cases} \dfrac{w^T x+b}{\|w\|_2 d} \geqslant 1, & y=1 \\ \dfrac{w^T x+b}{\|w\|_2 d} \leqslant -1, & y=-1 \end{cases}$$

$$y(w^T x+b) \geqslant 1$$

$$d=\frac{y(w^T x+b)}{\|w\|_2}$$

8.1 SVM简介

SVM算法的目标就是最大化这个几何间隔d：

$$\max \frac{2y(w^T x+b)}{\|w\|_2}, \quad s.t., y_i(w^T x_i + b) \geqslant 1$$

间隔最大化问题就是求最优化问题：

$$\min \frac{\|w\|_2^2}{2}, s.t., y_i(w^T x_i + b) \geqslant 1$$

8.1 SVM简介

这是一个凸二次规划问题，不容易求解，可用拉格朗日乘子法对其对偶问题进行求解。对上面的公式构造拉格朗日函数：

$$L(w,b,\lambda)=\frac{\|w\|_2^2}{2}-\sum_{i=1}^{N}\lambda_i(y_i(w^Tx_i+b)-1)$$

其中，$\lambda_i \geqslant 0$ 。原问题与对偶问题有相同的解：

$$\min_{w,b}\max_{\lambda}L(w,b,\lambda)\Leftrightarrow\max_{\lambda}\min_{w,b}L(w,b,\lambda)$$

调整w和b，使拉格朗日函数取最小值。

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}\lambda_i y_i x_i=0$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{N}\lambda_i y_i=0$$