## 7 支持向量机

### 概述

支持向量机的基本模型是定义在特征空间上的间隔最大的线性分类器，间隔最大使它有别于感知机；核技巧的加持，使它成为实质上的非线性分类器。

支持向量机的**学习策略**就是间隔最大化，可形式化为一个求解凸二次规划（convex quadratic programming）的问题，也等价于正则化的合页损失函数的最小化问题。

当训练数据线性可分时，通过硬间隔最大化，学习一个线性的分类器，即**线性可分支持向量机**，又称为硬间隔支持向量机；

当训练数据近似线性可分时，通过软间隔最大化，也学习一个线性的分类器，即**线性支持向量机**，又称为软间隔支持向量机；

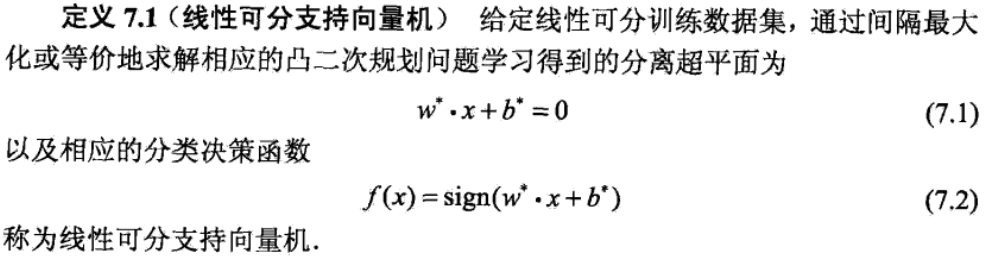
当训练数据线性不可分时，通过使用核技巧及软间隔最大化，学习**非线性支持向量机**。

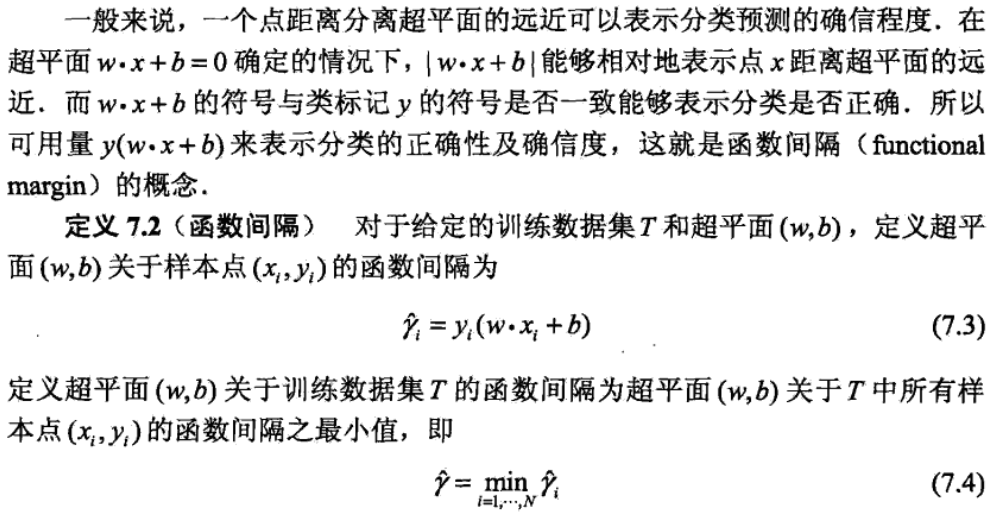
当输入空间为欧氏空间或离散集合、特征空间为希尔伯特空间时，核函数（kernel function）表示将输入从输入空间映射到特征空间得到的特征向量之间的内积。通过使用核函数可以学习非线性支持向量机，等价于隐式地在高维的特征空间中学习线性支持向量机。这样的方法称为**核技巧**。核方法（kernel method）是比支持向量机更为一般的机器学习方法。

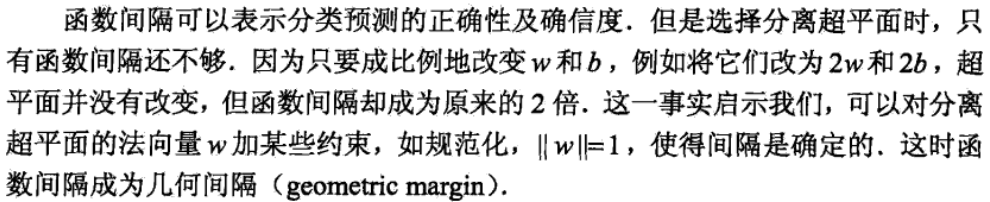
### 线性可分支持向量机与硬间隔最大化

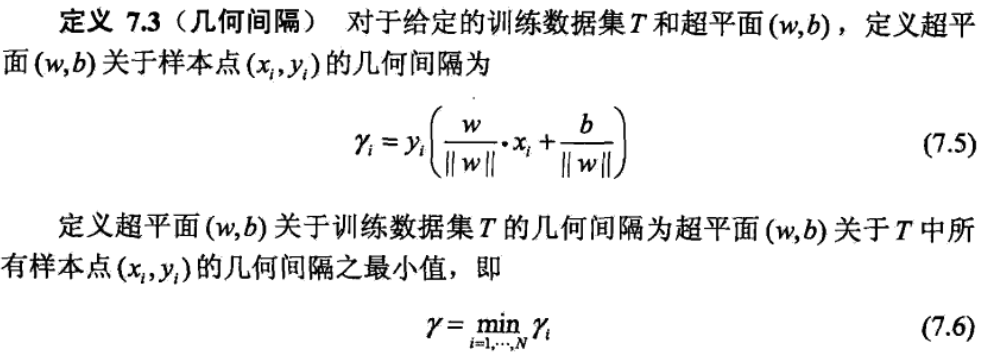
输入都由输入空间转换到特征空间，支持向量机的学习是在特征空间进行的。

一般地，当训练数据集线性可分时，存在无穷个分离超平面可将两类数据正确分开。**感知机**利用误分类最小的策略，求得分离超平面，不过这时的解有无穷多个。**线性可分支持向量机**利用间隔最大化求最优分离超平面，这时，解是唯一的。

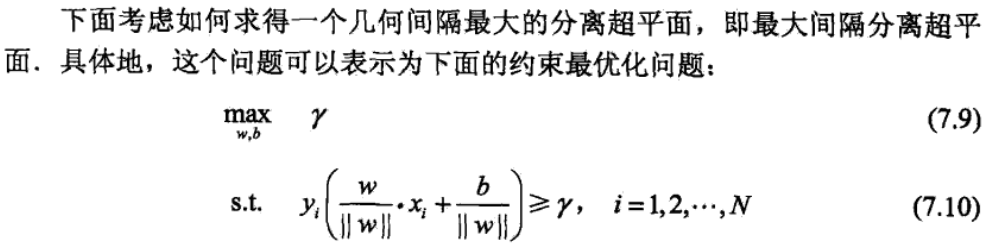


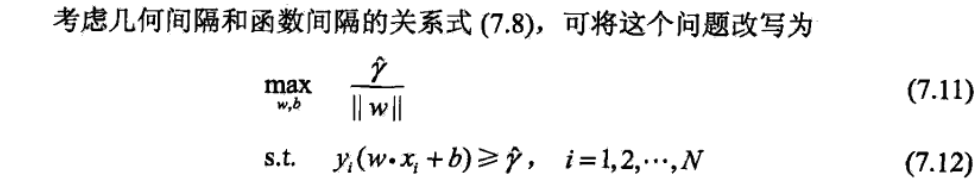




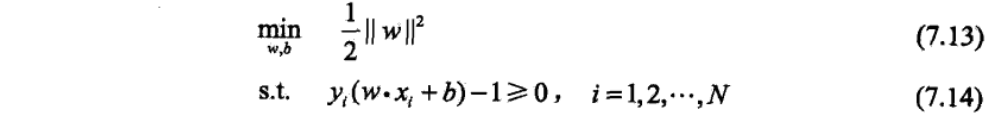


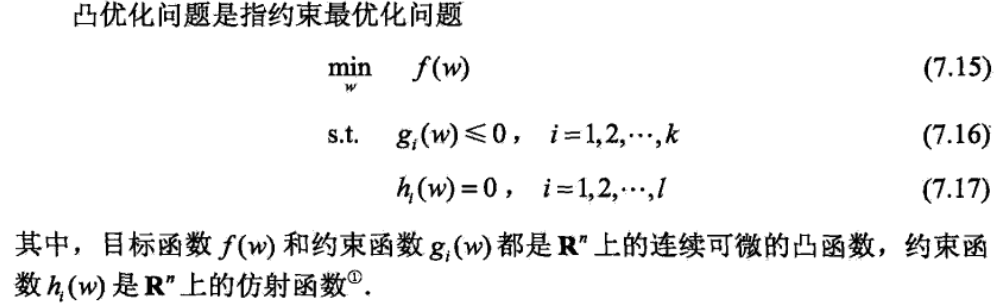
支持向量机学习的基本想法是求解能够正确划分训练数据集并且几何间隔最大的分离超平面。对线性可分的训练数据集而言，线性可分分离超平面有无穷多个（等价于感知机），但是几何间隔最大的分离超平面是唯一的。间隔最大化的直观解释是：对训练数据集找到几何间隔最大的超平面意味着以充分大的确信度对训练数据进行分类。

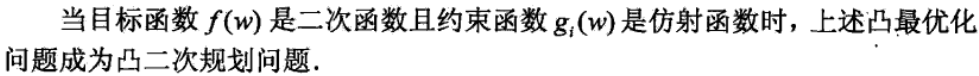


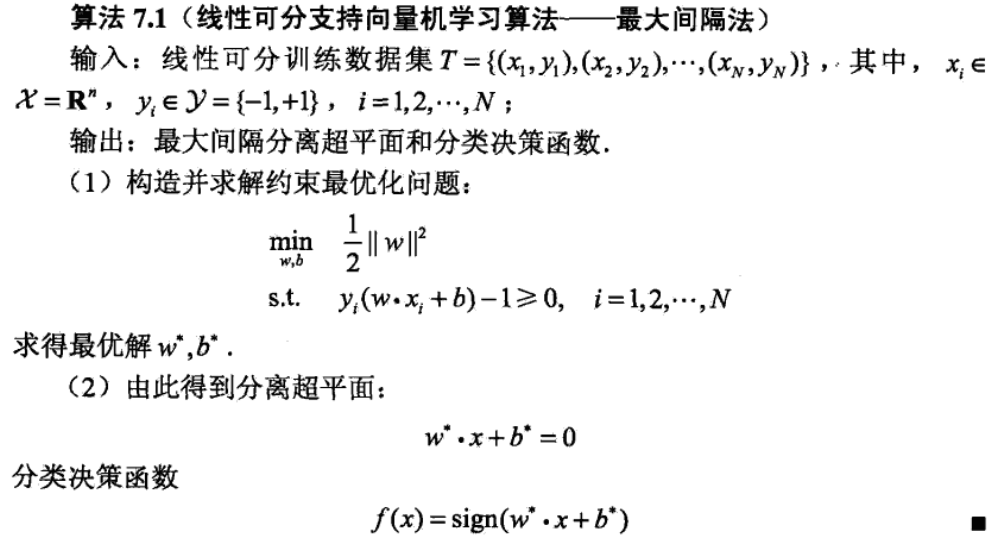


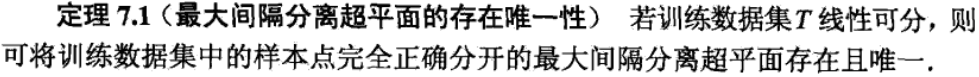
进一步等价为——

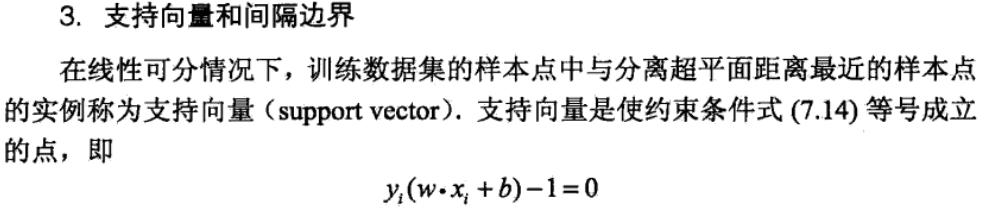


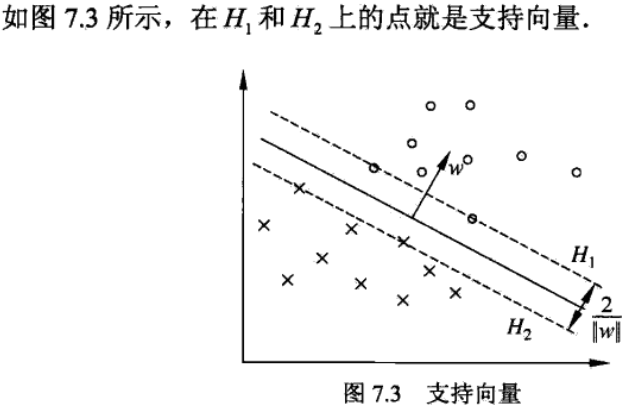


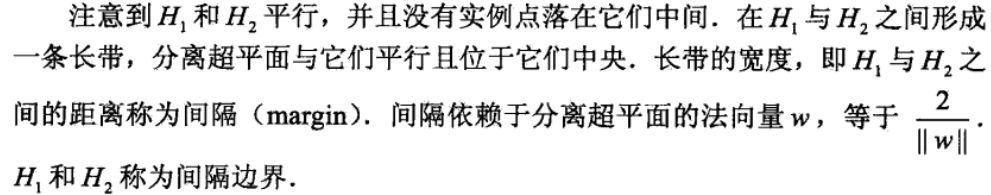






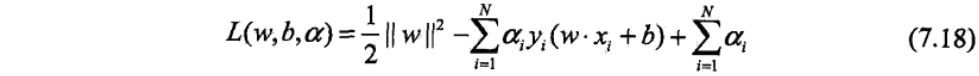




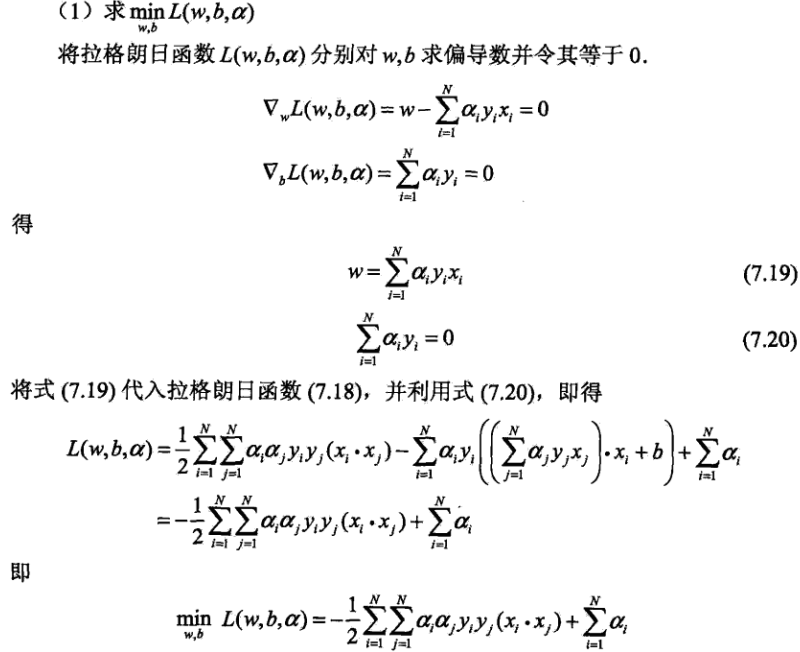


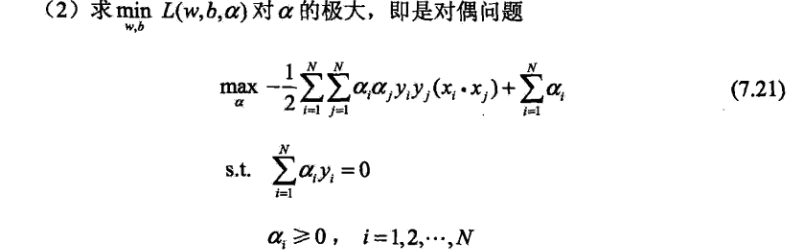
由于支持向量在确定分离超平面中起着决定性作用，所以将这种分类模型称为支持向量机。支持向量的个数一般很少，所以支持向量机由很少的“重要的”训练样本确定。

为了求解线性可分支持向量机的最优化问题（7.13-7.14），将它作为原始最优化问题，应用拉格朗日对偶性，通过求解对偶问题得到原始问题的最优解。这样做一是对偶问题往往更容易求解；二是自然引入核函数，进而推广到非线性分类问题。

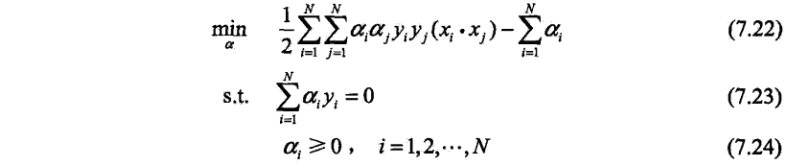


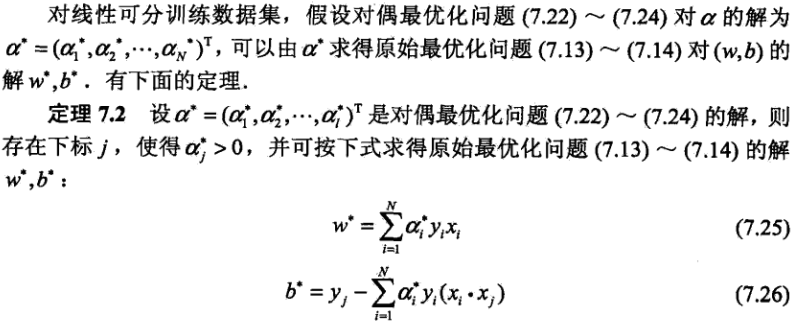
C:\Users\NKU\AppData\Local\Temp\mx3762B.png

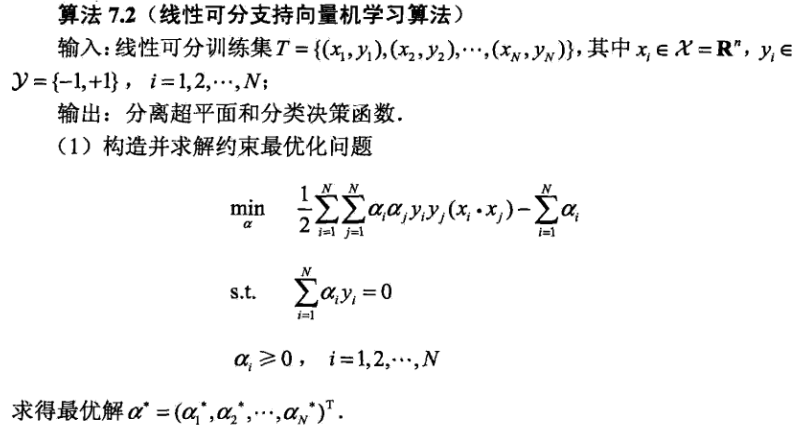


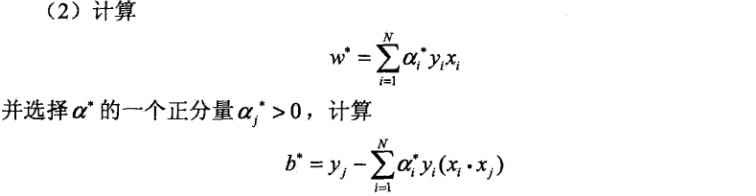


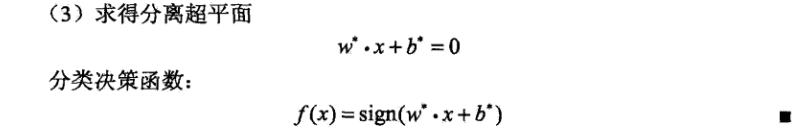
C:\Users\NKU\AppData\Local\Temp\mx33075.png

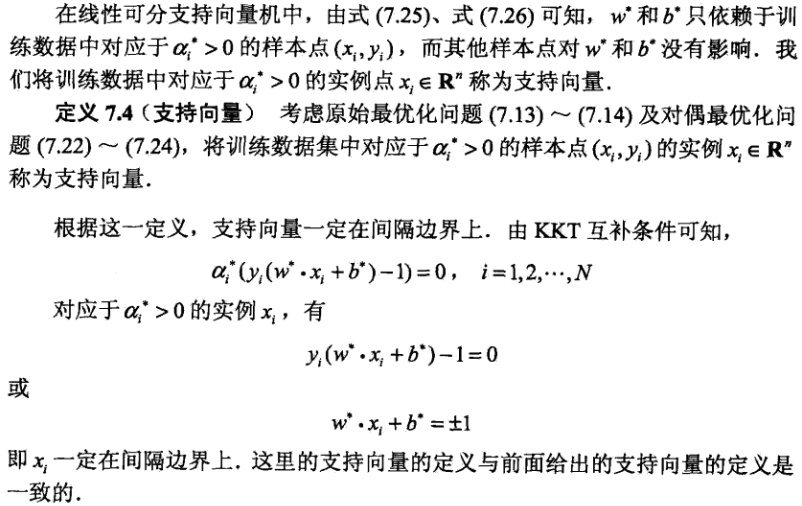






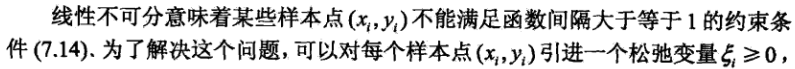


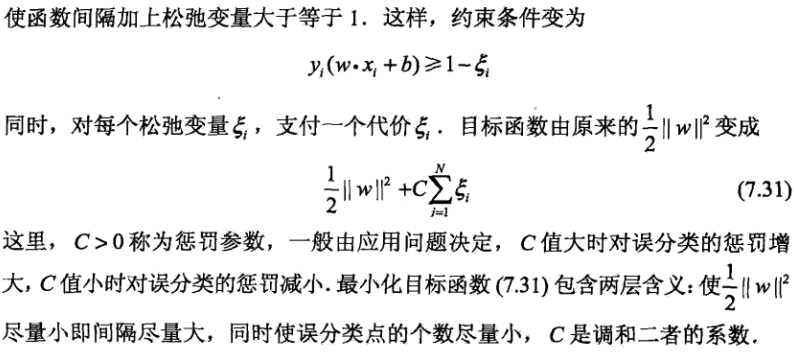


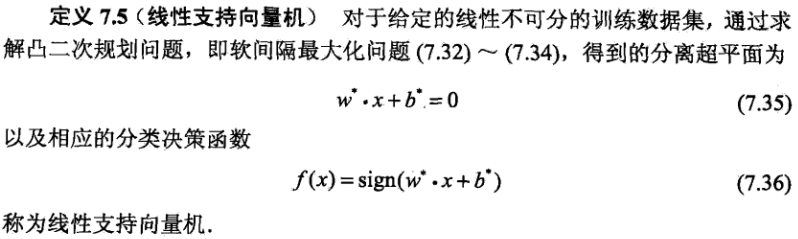


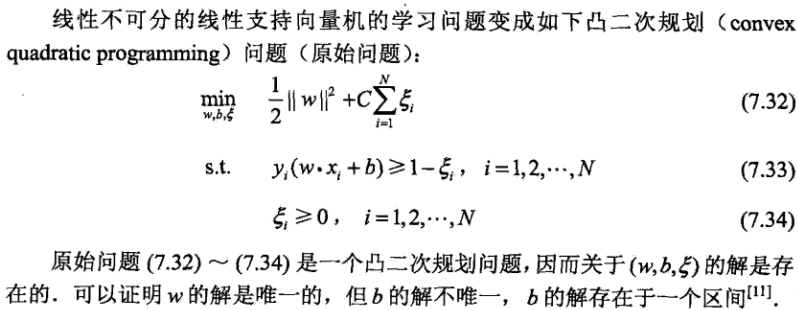
### 线性支持向量机与软间隔最大化

通常情况是，训练数据中有一些特异点（outlier），将这些特异点除去后，剩下大部分的样本点组成的集合是线性可分的。









对偶问题为：

