支持向量机概述

支 持 向 量 机 （ Support Vector Machine,SVM ） 是 一 类 按 监 督 学 习 （ supervised learning） 方式对数据进行二元分类的广义线性分类器（generalized linear classifier） ， 其决策边界是对学习样本求解的最大边距超平面（maximum-margin hyperplane） 。与逻辑回归和神经网络相比， 支持向量机， 在学习复杂的非线性方程时提供了一种更为清晰， 更加强大的方式。

硬间隔、 软间隔和非线性 SVM

假如数据是完全的线性可分的，那么学习到的模型可以称为硬间隔支持向量机。换个说法，硬间隔指的就是完全分类准确，不能存在分类错误的情况。软间隔，就是允许一定量的样本分类错误。

算法思想

找到集合边缘上的若干数据（称为支持向量（ Support Vector））， 用这些点找出一个平面（称为决策面） ， 使得支持向量到该平面的距离最大。

背景知识

任意超平面可以用下面这个线性方程来描述：

二维空间点到直线 的距离公式是：

扩展到𝑛维空间后，点到超平面的距离为：

其中

散点图

描述已自动生成

如图所示，根据支持向量的定义我们知道，支持向量到超平面的距离为 𝑑，其他点到超平面的距离大于 𝑑。 每个支持向量到超平面的距离可以写为：

根据支持向量的定义我们知道，支持向量到超平面的距离为 𝑑，其他点到超平面的距离大于 𝑑。

于是我们有这样的一个公式：故：

我们暂且令𝑑为1（之所以令它等于 1，是为了方便推导和优化，且这样做对目标函数的优化没有影响），将两个方程合并，我们可以简写为：

至此我们就可以得到最大间隔超平面的上下两个超平面：