

支持向量机近期进展综述

1. 一般形式的支持向量机（抄书，按自己理解写）周日写完

线性可分的最优超平面

不可分模式的最优超平面

使用核方法的支持向量机

1. 一般形式的支持向量机训练方式（按照自己理解写）周日写完

是一种凸优化（a为参数）

梯度下降法

SMO（附带码）

1. 支持向量机的优势与劣势 周一

不适用于大规模数据集（小规模数据集效果更好）

SVM分类器致力于实现更多的类间分离，而不是利用类内训练数据中的结构。

1. 支持向量机的种类（看论文总结）周二

单类支持向量机

支持向量机应用（看论文总结）周三

Cancer Genomics

1. 一般形式的支持向量机

本质上来讲，支持向量机的作用是在样本空间内建立一个超平面作为其分类的决策平面。并且这个平面可以使得正反样本之间的隔离边缘最大化。

1.1样本线性可分情况下的最优超平面

把训练样本记作，其中为训练样本集中的第个元素，为样本的目标输出。考虑一般的二分类情况，。分别代表两个类的输出值。且两个类是线性可分的。那么用于分离两类样本的超平面可表示为：

（1）

其中，表示权值向量，表示偏置值。那么将样本点带入该超平面方程中有：（2）

式（1）中确定的超平面与最近的数据点的距离间隔即分离边缘。支持向量机的目标是找到一个用于决策的超平面，使得超平面分离边缘最大。

设，分别为最大时的权值向量和偏置的最优值，并记

（3）

根据[1]中的度量方法，可表示为：

（4）

其中，是在最优超平面上的投影。是到最优超平面的距离，的符号由与最优超平面的方向决定，若在最优超平面正面则为正号，反之为负号。

将（4）带入（3），得：

（5）

考虑一个支持向量，，则有：

（6）

1. 带入（5）得:

（7）

因此有：

（8）

因此，可将寻找最优超平面得问题转化含约束优化问题：

（9）

建立拉格朗日函数求解问题：

（10）

且由最优化条件有：

（11）

其中为拉格朗日乘子。约束最优问题得解由拉格朗日函数的鞍点决定。

这里，注意到优化问题为凸函数，并且约束条件是关于的线性约束。因此可将原问题转化为对偶问题进行求解。即：

（12）

将原问题逐项展开得：

（13）

由（11）得：

（14）

由（13），（14）得：



约束条件为：



确定第个最优得拉格朗日乘子后，最优权值向量可写作



其中为支持向量得个数，即最优非零拉格朗日乘子的个数。最优偏置可表示为：



样本线性不可分情况下的最优超平面

对于

1. Pattern, Duda RO Hart PE.. Classification and Scene Analysis[M]. John Wiley and Sons, New York, 1973.