

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

支持向量机理论首先由Vapnik提出，由[引用]给出第一个描述。Vapnik于1998年创作的“Statistical Learning Theory”一书给出了支持向量机的详细描述。在此，首先讨论一般形式的支持向量机理论。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。

对偶性适用于目标函数可导的带有约束的优化问题。原问题和对偶问题需要满足Karush-Kuhn-Tucker（KKT）条件[引用]。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。对于凸优化问题的结局方案可参考[引用]，在[引用]中Kuhn给出解决不等式约束问题的方法，其中凸优化起到了重要作用。此处将原问题转化为对偶问题进行求解，对偶问题的形式为：

（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：



一般而言，支持向量机的解是稀疏的，即当问题规模增加时，支持向量的数量比问题规模线性增加速度慢。支持向量机的稀疏性于[引用]被首次讨论。在[引用]讨论了支持向量机在解决模式识别问题中出现的稀疏性问题。此外，在该文章中还给出了支持向量个数的下限计算。

样本线性不可分情况下的最优超平面

在处理相对复杂的样本线性不可分的情况时，需要对线性可分的支持向量机理论进行扩展。线性不可分情况即存在不满足约束条件的的点。这种违反约束条件的点有两类情况。第一类，数据点落在分离区域以内，但落在决策面正确的一侧；第二类，数据点落在决策面错误的一侧。这种线性不可分情况下的类之间的分离间隔称为软间隔。

为了处理不可分离的数据点，引入松弛变量到决策面的定义中，松弛变量用于度量一个训练样本对线性可分理想条件的偏离程度。重新定义决策面为：



根据以上定义，有当，样本点落入分离区域的内部，仍在决策面正确的一侧。当时，样本点落入超平面错误的一侧。满足的样本点对应的特征向量为支持向量。

为了使决策面在训练集上的平均误差最小，可以对以下泛函进行最优化：



参数可认为是一个正则系数，与即二范式前添加正则化系数效果相同。越大意味着对训练样本质量具有较高信心，反之则表明训练样本具有较大噪声。

该问题的对偶问题可表示为：



约束条件为：



使用核方法的支持向量机

核方法基于Cover定理：假设空间不是密集分布的，将复杂的模式分类问题非线性地投射到高维空间将比投射到低维空间更可能是线性可分的。（Cover，1965）在[引用]中，对核方法以及支持向量理论给出了综合论述。表示从输入空间中取出的向量，表示一系列非线性函数，即特征函数的集合。特征函数用于从样本数据的原维度映射到更高的维度。那么超平面可定义为：

（1）



根据1节权重可表示为：

（2）

将（2）带入（1），得到决策面方程可表示为：



核函数[引用177]被用于计算输入空间的两个数据点在特征空间的内积。将式（）中的内积用核函数表示为：



此处核函数的选取应遵循Mercer定理（引用）。

使用核函数的支持向量机，在样本线性不可分情况下的目标函数形式为：



约束条件为：



使用核方法的支持向量机构造重点在于特征函数以及核函数的选择。多项式核、径向基函数核有不同形式，但它们均满足Mercer定理。只要满足该定理，核的选择具有一定的自由度。

在不同的应用中，核函数的不同会导致分类效果不同。

支持向量机的常用损失函数综述

不敏感损失函数

不敏感损失函数以鲁棒性作为目标，即使用该损失函数的支持向量机应满足对样本参数小的变化不敏感。因此，应该考虑一个样本产生的偏差对系统整体产生的最差影响。因此，这实际上是一种最小化最大误差思想，该思想被Huber[]提出。最小化最坏情况的影响，在工程方面是一种重要思想。在[引用]中对该思想进行了进一步讨论。考虑噪声的概率密度函数关于原点对称的情况，此时应考虑利用绝对误差作为目标函数进行最小化。因此损失函数具有以下形式：



Vapnik将上式扩展为：



该损失被称为-不敏感损失函数。其中是一个指定的参数。该模型在一个范围内的损失均被置为0。因此在样本的偏差内具有不敏感的性质。满足了最小化偏差内对模型产生的影响。

铰链损失函数

铰链损失函数、二次铰链损失函数、最小二乘损失函数。Cristianini and Shawe-Taylor

(2000), Scholk ¨ opf and Smola (2002), and Suykens et al. (2002)，在文章[Comparison of L1 and L2 Support Vector Machines](http://www.lib.kobe-u.ac.jp/repository/90000225.pdf)中比较了L1和L2型的支持向量机的性质。

铰链损失函数：



二次铰链损失函数：



最小二乘损失函数：



其中，变量和分别表示相应的期望输出和给定输入下的实际计算出的输出。使用这三种损失函数的支持向量机分别称为L1，L2，LS型支持向量机。其中，使用铰链损失的L1型的支持向量机已在节中叙述。L2型支持向量机的目标函数与L1型类似，L2型支持向量机的目标函数可写为：



L2型的支持向量机中的正则化项采用平方的形式，在Deep Learning using Linear Support Vector Machines中提出L2型支持向量机具有可微分的性质，并且对于违反间隔的点施加了更大损失。在该文章中，对比了深度学习中使用softmax和SVM的效果，在实验中表明使用L2型的深度网络具有更好的分类效果。

支持向量机的训练方法

用于解决模式识别和回归问题的支持向量机计算代价都包括一个二次项和一个三次项。具体而言，当选择的正则项很小时，计算代价以增加。当很大时，计算代价以增加[3]。在实际应用时，经常采用二次规划的方式解决SVM的训练问题。

无界支持向量增加了训练的难度

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