支持向量机

**6.1 间隔与支持向量**

分类的基本想法是：基于给定的含有个样本的训练集，在样本空间中找到一个划分超平面，将不同类别的样本分开。如何确定这样的超平面？

**假设训练集线性可分**，直观上看应该寻找位于两类样本正中间的划分超平面。超平面可通过如下方程描述：

 （1）

其中为法向量，为位移项。将由确定的超平面记为。

下面计算样本空间中**任意一点到超平面的距离**：

取超平面上任意一点，则有：

 （2）

点到超平面的距离为向量在垂直方向上的投影。设向量与向量的夹角为，则向量在向量上的投影的长度为：



于是点到超平面的距离为：

 （3）

由（2）式知，代入（3）式，得

 （4）

现在假设超平面****能将样本正确分类，即对于，若，则；若，则。则总存在缩放变换，使得下式成立：

 （5）

距离超平面最近的几个样本点使上式等号成立，它们被称为支持向量。两个异类支持向量到超平面的距离之和称为间隔：

 （6）

要找到具有最大间隔的划分超平面，就是要找满足（5）中约束的使得最大的参数。即

 （7）

上式等价于

 （8）

这就是支持向量机的基本型。

**6.2 对偶问题**

（8）式是一个凸二次规划问题，考虑到这个问题的特殊结构，对其使用拉格朗日乘子法，可得到与此问题等价的对偶问题。通过求解对偶问题来得到原问题的最优解，这就是线性可分条件下支持向量机的对偶算法。

**为什么要引入对偶问题？个人理解有以下三个原因：**

**●对偶问题将原始问题中的约束转为了对偶问题中的等式约束**

**●改变了问题的复杂度。由求特征向量转化为求比例系数。在原始问题下，求解的复杂度与样本的维度有关，即的维度。在对偶问题下，求解的是，复杂度只与样本数量有关。**

**●可以自然地引入核函数，从而推广到非线性分类问题。**

下面介绍具体思路。

①原问题是有约束的凸优化问题，对这类问题常用拉格朗日乘子法来处理。这里先通过拉格朗日乘子法，将原问题转化为一个与之等价的极小极大问题。

拉格朗日乘子法是一种寻找多元函数在一组约束下的极值的方法。通过引入拉格朗日乘子，可将有个变量与个约束条件的最优化问题转化为具有个变量的无约束优化问题。

假设是定义在上的连续可微函数。考虑约束最优化问题：

 （9）

首先引进广义拉格朗日函数：

 （10）

这里是拉格朗日乘子，**特别要求**。（这点在后面的KKT条件中会详细解释）

考虑的函数：



因为上式中已经确定，所以上式是关于的函数。

**下面通过是否满足约束条件两方面来分析这个函数：**

●如果违反原问题的约束条件，即存在某个使或某个使，则有：



因为若某个使，则可令；若某个使，则可令使，而将其余各各取为0.

●如果满足原问题的约束条件，则有：



因为中间的最大化是确定的过程，可将看成常量，常量的最大值即为本身。

综上所述，有：

那么在满足约束条件下，有：。

**总结：通过拉格朗日法，将有约束的原始问题转换成一个与之等价的无约束问题。**

**—————————————————————————————————————**

**②通过第一步，我们得到了一个与原问题等价的极小极大问题。然而极小极大问题不易求解，可通过求其对偶问题的解来解决原问题。下面推导一下对偶问题。**

对于（9）式的原问题，基于（10）式，其拉格朗日对偶函数定义为：



这里的表示最大下界。对任意和都有：



设为原问题可行域中的点，则有



若原问题最优值为，则对任意和都有。

于是对偶问题给出了原问题最优值的下界，且下界取决于的值。那么基于对偶函数能获得的最优下界是什么？这就引出了优化问题：



**无论原问题如何，对偶问题始终是凸优化问题。因此往往易于求解。**

定义对偶问题的最优解为：



**什么情况下能通过对偶问题求原问题的解**？

首先对于一般的优化问题，弱对偶性成立，即。（证明见《统计学习方法》附录C）；要想通过对偶问题直接求解原问题，需要满足强对偶性，也就是。定理给出了满足强对偶性的要求。

定理:若是凸函数，是仿射函数，且不等式约束严格可行，则存在，，使是原始问题的解，是对偶问题的解，并且有：

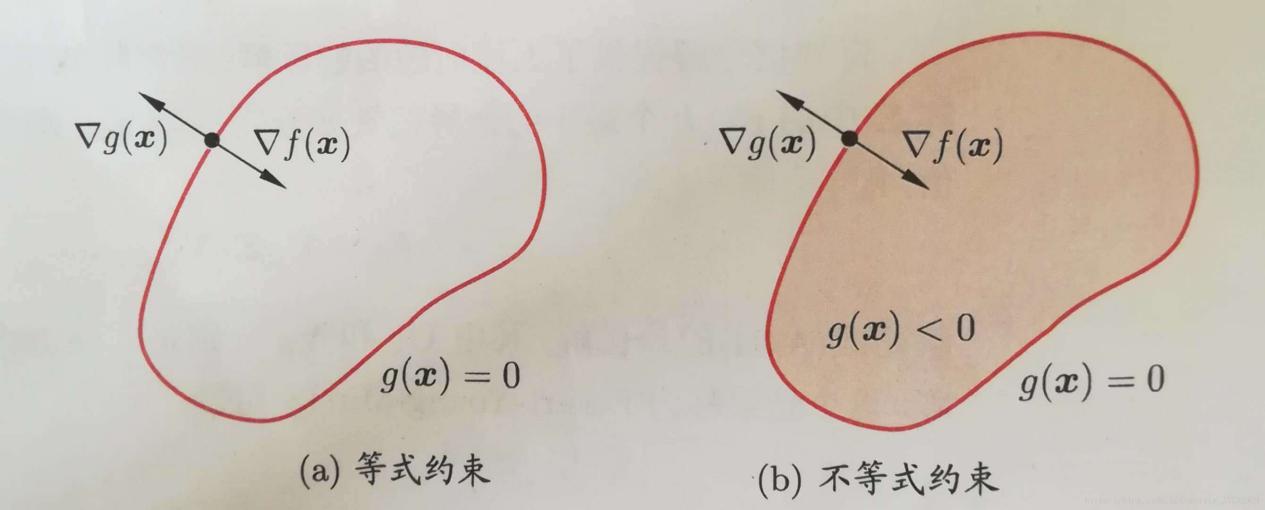


定理：若是凸函数，是仿射函数，且不等式约束严格可行，则和分别是原问题和对偶问题的解的充要条件是和满足下面的KKT条件：



（这里定理还有疑惑，有待探讨。）

**下面解释为什么《机器学习》6.2中的求解过程需要满足（6.13）式的KKT条件。**



●先考虑一个等式约束的最优化问题：使目标函数最小且同时满足的约束。在最优点，**梯度的方向必相同或相反**，即存在，使得



于是通过定义拉格朗日函数：

 （11）

这个函数对的偏导数置零，就得到了上面的式子。于是将原问题转化为无约束优化问题。

●再考虑不等式约束。如上图所示，此时最优点或在的区域中，或在边界上。时，约束不起作用，可直接通过条件来获得最优点；这等价于将置零。时，与等式约束时情况类似，**但此时的方向必定相反，即存在常数使得。**

于是在约束下最小化等价于在如下约束下最小化（11）式的拉格朗日函数：



上式就是KKT条件。**这里就解释了上文的拉格朗日函数里，为什么要求的系数。**

其中最后一个式子称为对偶互补条件。

**总结：用拉格朗日法处理含有不等式约束的目标函数时，这个不等式约束及其对应的拉格朗日乘子需要满足KKT条件。**

———————————————————————————————————————

③通过第二步，我们知道原始问题满足条件，从而转换成对偶问题。这个对偶问题是：

 （12）

解出后，求出即可得到模型：



1. 式的求解需要用到SMO算法。它的基本思路是每次选择两个变量，并固定其它参数。在参数初始化后，不断执行以下两个步骤直到收敛：

●选取一对需更新的变量

●固定以外的参数，求解式（12）获得更新后的。

（SMO算法还没完全理解，有待讨论。）

**6.3 核函数**

在之前的讨论中，我们假设训练集线性可分。然而现实任务中的训练集通常不是线性可分的。对这样的问题，可将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分。如果原始空间维数有限，那么一定存在一个高维特征空间使样本线性可分。

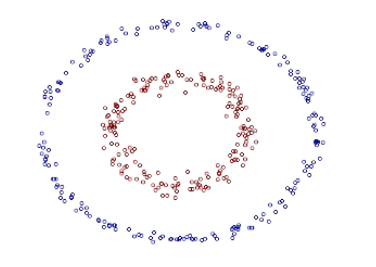
于是用线性分类方法求解非线性分类问题分为两步：首先用一个变换将原空间的数据映射到新空间；然后在新空间里用线性分类方法学习一个分类模型。

令表示将映射后的特征向量，于是在特征空间中划分超平面的模型可表示为：



其对偶问题是：





上式的求解涉及到计算，就是**先将原始空间映射到高维空间，再在高维空间中进行内积的计算**。然而由于特征空间维数很高，甚至可能是无穷维，因此直接计算通常很困难。比如对于上图的两类数据，需要对一个二维空间做映射，选择的新空间是原始空间的所有一阶和二阶的线性组合，得到五个维度。如果原始空间是三维，则得到一个19维的新空间。这个数目呈爆炸性增长，使内积的计算非常困难。 于是我们想到，**可不可以直接在低维空间中进行计算，而不需要显式地写出映射后的结果**？这就是核函数。下面给出核函数的定义：

设是输入空间，是特征空间，如果存在一个从到的映射：，使得对所有，函数满足条件：



则称为核函数，为映射函数。这里不要把核函数和映射函数相混淆。核技巧的想法是，只定义核函数，而不显式地定义映射函数。通常直接计算比较容易，而通过和计算比较困难。

**接下来结合定义和上图，具体解释一下核函数**。用表示这个二维平面的两个坐标，我们知道二次曲线可以写成：

 （12）

据此我们可以构造一个五维空间，其中的坐标分别为 

于是（12）式在新的坐标系下可以写成：



这正是一个超平面的方程，也就是说将原始的二维样本空间映射到五维空间后，样本变得线性可分。这是核方法处理非线性问题的基本思想。

设两个向量，将它们按上述方法映射到五维空间后，得到：



于是映射之后的内积为：

 （13）

现在我们引入一个核函数：

 （14）

可以看到（13）（14）式实际上是等价的，但是推导过程有区别：

**●前者先映射到高维空间，再在高维空间中进行内积的计算**

**●后者直接在低维空间中计算，不需要显式地写出映射后的结果。**

**当维度变得很高时，（13）式难以计算，而（14）式不受影响。于是问题解决，避开了在高维空间中的内积计算。**

刚刚我们引入了一个核函数，在具体问题中如何选取合适的核函数就成了最重要的问题。在实际应用中往往依赖领域知识直接选择核函数，然后通过实验验证核函数的有效性。

**总结：**

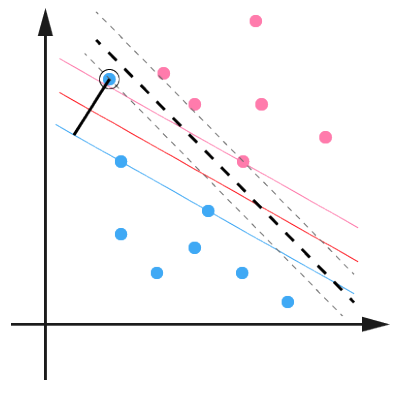
●现实应用中经常遇到线性不可分的情况，通过将原始空间映射到合适的高维空间，样本便线性可分，问题解决。

●再进一步，如果对于此类问题一律采取上面的方法，则当映射后的特征空间维数很高甚至是无穷维时，内积难以计算。

●于是出现了核函数。它先在低维空间中进行计算，却能将分类效果表现在高维，这样就避免了高维空间中的复杂计算。

**6.4 软间隔与正则化**

之前的讨论中，我们假设样本在样本空间或特征空间中线性可分。然而现实任务中很难确定合适的核函数使得训练样本在特征空间中线性可分。即使找到了，也可能是过拟合造成的结果，或者是被异常值所影响。



如上图右上角的蓝点，是蓝点样本中的一个异常值。由于它的存在，之前的分类超平面是红色实线，现在变成了黑色虚线。若这个蓝点再往右侧偏离，可能会导致无法求出将两类样本完全分隔的超平面。为了解决这个问题，允许某些样本不满足约束。考虑到异常值问题，现在的约束条件变为



这里的是松弛变量，也就是允许样本点偏移的距离。当然要对加以限制，否则任意超平面都符合条件。在最大化间隔的同时，不满足约束的样本应该尽可能少，于是最终的目标函数写成：

****

之后的求解过程与上文类似，不再赘述。

参考链接

1. 拉格朗日对偶问题：<https://blog.csdn.net/blackyuanc/article/details/67640844>
2. 支持向量机详解：<https://blog.csdn.net/v_JULY_v/article/details/7624837>

3.核函数：<https://blog.csdn.net/bitcarmanlee/article/details/77604484>