# 內核方法和SVM的扩展

**这份文档的目的是为了回顾“机器学习1-监督学习”中关于支持向量机（SVM）的内容。这份文档也提供了课程内容的扩展的概述，其中包括非二元的分类和支持向量回归。**

我们会用最简单的应用实例来介绍SVM的概念。考虑一个情景，我们有些数据，必须被分成两个不同的组。如果数据是*线性可分的*，或者换言之，可以完全被一个超平面分成两组，我们目标是找到这个超平面的方程，能够把组分得最好。

用问题描述的方式更正式地讲，我们给每个数据点贴上标签分类，为-1或者1,比如。我们的超平面函数定义为，且所有点在的分类中,都有

,

而对的点,有

。

在我们的训练数据中，在超平面和之间不应有任何点，这个区域称为“间隔”。分离平面是函数，对于新的数据点，我们以它们的符号：将它们分类。

(注意w看上去不垂直于分离平面是由于x轴和y轴的比例不同）

让我们分离数据的参数向量w的选择有很多，但是有些明显比另一些好。理想状态下，我们要选择能使间隔最大化的超平面参数。

考虑在间隔的两边的两个点，x+和x-，它们相距尽可能近。在这种情况下，连接这两条直线的向量垂直于定义间隔的超平面。

和,

将这两个等式相减得。因为向量w和x+ - x-是平行的，所以, 其中是向量w的大小/长度。两边都除以得到超平面之间的距离，即。从这里我们看到最大化间隔等同于找出对于训练集中所有点能满足关系的最小的。（回忆一下，当时，，当时，。）

要近似解决这个问题，我们注意到最小化等同于最小化，将这个问题转化为了二次规划优化的问题。拉格朗日乘子把我们的优化问题变成了最大化



的输出，同时满足限制条件所有且。要解释这个公式，需要给一些背景，就是把乘子视为数据点的权重。从限制条件，可知分类的点的权重之和应该等于分类为的点的权重之和。至于，第二项把控了第一项中权重的和不会变得太大。第二项考虑每一对点的分类（如果两个点在同一分组，则，分组不同则为-1）和相似度的度量（由  引发）。

当我们得到最优的拉格朗日乘子，结果是大多数的权等于0*。*只有非0权重的点是对计算w有贡献的点，实际上所有的这样的点落在间隔线上，满足。这些点就是这个模型的支持向量。对于间隔上的某个点，我们能从和获得分离超平面的参数值。

上面介绍了一般的线性分离SVM的计算过程，但现实生活中的数据集通常不会让自己这么容易划分的。这里，我们讨论两种方式来处理非线性可分的数据集，跳过硬间隔的SVM。如果我们的数据基本上是线性可分的，我们可以用软间隔SVM来考虑，放松“所有点都要正确分类”的标准。如果我们的数据是非线性可分的，我们可以考虑用*内核函数*，能找到一个分类之间的非线性分离曲线。通常，我们既要考虑内核函数也要考虑软间隔参数来执行分类任务。

在一个软间隔SVM中，我们不要求数据完全被线性分离，允许一些点被错误分类。我们给每个点一个非负的松弛变量，表示每个点被错误分类的程度：。如果一个点被正确分类到它所在的间隔的一侧，那么。如果一个点被放在了间隔内或者错误的分类区域，那么与点到预期的间隔超平面的距离成正比。

我们的优化问题现在必须平衡我们产生的误差的大小：我们的目标是最小化，其中C是正则化参数，告诉我们想给予错误分类误差的权重。用较小的C值，则我们对误差的惩罚较轻，就增大了间隔的大小。用较大的C就得到较窄的间隔；C趋于无穷大时的极限表示任何错误分类的误差和原来的硬间隔SVM的程度一样。当我们将优化问题转化为最大化



的输出，限制条件保持不变，而其他限制条件现在有一个上界。对于软间隔SVM，我们的支持向量（权为的点)不止包含了间隔超平面上的点，还有那些在间隔内或者错误分类的点。

对于可分但非线性可分的数据，我们可以用一个内核函数来获得非线性的分离曲线。内核函数应该找到我们数据中的一些相似的方面；它也表示了关于数据的结构的领域知识。通常，我们可以把我们想要最大化的函数写成

 。

在我们最初的线性SVM中，我们的内核函数是，得出一个分离超平面。内核函数产生一个分离超空间，而是多项式内核的一般形式。通过内核函数，我们可以将数据投射到变换空间，其中可以找到一个分离超平面，但是在原特征空间上绘图时，结果则是一条非线性的分离曲线。为了计算出新实例的分类，我们现在利用输出的符号



因为大多数的权等于0，即使与线性情况比较，这仍然是相当快的计算方式。

重要的一点是要注意，SVM本来的任务是二元分类。对于两组以上的分类任务，一个通常的策略是使用多个二元分类器来选定对新情况最好的一个分类。例如，我们可以为每一个分类器创建一个分类器，一对多形式的，然后对于新的点，用结果值最大的分类函数进行分类。或者，我们可以建立分类器进行两两的比较，然后选择对于新的点，“赢”了最多两两比较的分类。（此图描画了两两分类的方法。灰线表示一个二元分类器没有作用。注意中间的区域，没有分类器在这里占优势。）

我们也可以将SVM扩展到回归的工作上，或称支持向量回归（SVR）。正如SVM一样，我们将数据投射到一个使用内核函数的SVR任务中，如此它就可以适配一个超平面。不同于给数据分类，超平面可以给出一个数据输出值的估计。此外，间隔和误差的处理也不同。指定一个参数，使得回归超平面中的较小差异不会造成误差成本，比如，当我们尝试最小化



当一个点落在间隔内时，。而值非0的松弛变量则是到点所在的区域的（线性）距离。与标准线性回归中的二次误差函数相比，那里所有估计的差异都不利于函数拟合，而对误差以与估计的平方差来惩罚。

本质上，SVR的方式与SVM差不多。对于训练数据中的每个点，我们有两个松弛变量和,一个是回归超平面中的正的偏差，而另一个是负的偏差。

结果是对于每个点有两个相应的拉格朗日乘子，以及一个对于权重的重新指定的限制条件。得到解后，回归函数有这样的形式

。

像之前一样，大多数权的取值为0,对于那些非0权重的点，至多有和两者其中之一是非0的。