**Support Vector Machine**

**1. 如何推导SVM？**

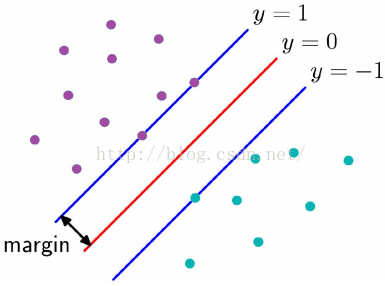
1. 基本概念

支持向量：训练数据集的样本点中与分离超平面距离最近的样本点的实例。

1. 不带松弛变量的SVM推导

假设目标函数为，根据题设有，等比例缩放，则

目标函数：，表示最近点到直线距离尽可能大



建立目标函数

1.总可以通过等比例缩放的方法，使得两类点的函数值都满足

2.约束条件：

3.原目标函数：

4.新目标函数：

5.目标函数变换一下：

6.拉格朗日乘子法

7.原问题是极小极大问题

原问题的对偶问题是极大极小问题

8.将6中的拉格朗日函数分别对求偏导并令其为0

9.计算拉格朗日的对偶函数

10.继续求对的极大

11.整理目标函数：添加负号

12.线性可分支持向量机学习算法

13.分类决策函数

1. 带松弛变量的SVM推导

1.若数据线性不可分，则增加松弛因子，使函数间隔加上松弛变量大于等于1，则约束条件变成

目标函数： （这里是为了保证松弛因子不至于过大）

2.此时的凸优化为

3.拉格朗日函数

4.将三式代入L中，得到

5. 整理，得到对偶问题的最优化问题

求得最优解

6.计算

实践中往往取支持向量的所有值取平均，作为

7.求得分离超平面

8.分类决策函数为

**2. SVM的原理是什么？**

SVM是一种二类分类模型。它的基本模型是在特征空间中寻找间隔最大化的分离超平面的线性分类器。间隔最大是它有别于感知机和线性回归的主要原因。

**3.SVM为什么采用间隔最大化？**

当训练数据线性可分时，存在无穷个分离超平面可以将两类数据正确分开。感知机利用误分类最小策略，求得分离超平面，不过此时的解有无穷多个。线性可分支持向量机利用间隔最大化求得最优分离超平面，这时，解是唯一的。另一方面，此时的分隔超平面所产生的分类结果是最鲁棒的，对未知实例的泛化能力最强。

**4. 为什么SVM要引入核函数？**

当样本在原始空间线性不可分时，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。

**5. 为什么SVM对缺失数据敏感？**

这里说的缺失数据是指缺失某些特征数据，向量数据不完整。SVM没有处理缺失值的策略（决策树有），而SVM希望样本在特征空间中线性可分，所以特征空间的好坏对SVM的性能很重要，缺失特征数据将影响训练结果的好坏。

**6. SVM如何处理多分类问题？**

1. 直接法：直接在目标函数上修改，将多个分类面的参数求解合并到一个最优化问题里面，看似简单但是计算量却非常的大。
2. 间接法：对训练器进行组合，其中比较典型的有一对一，和一对多。

一对多：训练时依次把某个类别的样本归为一类,其他剩余的样本归为另一类，这样k个类别的样本就构造出了个SVM。分类时将未知样本分类为具有最大分类函数值的那类。

一对一：其做法是在任意两类样本之间设计一个SVM，因此个类别的样本就需要设计个SVM。当对一个未知样本进行分类时，最后得票最多的类别即为该未知样本的类别。

**7.常用的核函数有哪些？**

1. 线性核函数：

线性核，主要用于线性可分的情况，我们可以看到特征空间到输入空间的维度是一样的，其参数少速度快，对于线性可分数据，其分类效果很理想，因此我们通常首先尝试用线性核函数来做分类，看看效果如何，如果不行再换别的。

1. 多项式核函数：

多项式核函数可以实现将低维的输入空间映射到高纬的特征空间，但是多项式核函数的参数多，当多项式的阶数比较高的时候，核矩阵的元素值将趋于无穷大或者无穷小，计算复杂度会大到无法计算。

1. 高斯（RBF）核函数：

高斯径向基函数是一种局部性强的核函数，其可以将一个样本映射到一个更高维的空间内，该核函数是应用最广的一个，无论大样本还是小样本都有比较好的性能，而且其相对于多项式核函数参数要少，因此大多数情况下在不知道用什么核函数的时候，优先使用高斯核函数。

**8. SVM原问题和对偶问题关系？为什么引入对偶问题求解？**

SVM对偶问题的获得方法：将原问题的目标函数L和约束条件构造拉格朗日函数，再对中原参数和分别求导，并且三种导数都等于0；再将等于0的三个导数带入原目标函数中，即可获得对偶问题的目标函数。

关系：原问题的最大值相对于对偶问题的最小值。

引入对偶问题的原因：对偶问题往往更易求解，当我们寻找约束存在时的最优点的时候，约束的存在虽然减小了需要搜寻的范围，但是却使问题变得更加复杂。为了使问题变得易于处理，我们的方法是把目标函数和约束全部融入一个新的函数，即拉格朗日函数，再通过这个函数来寻找最优点。

**9. KKT条件是什么？**

设目标函数，不等式约束为，有的教程还会添加上等式约束条件。此时的约束优化问题描述如下：

则我们定义不等式约束下的拉格朗日函数，则表达式为：

其中是原目标函数，是第个等式约束条件，是对应的约束系数，是不等式约束，是对应的约束系数。此时若要求解上述优化问题，必须满足下述条件（也是我们的求解条件）：

这些求解条件就是KKT条件。(1)是对拉格朗日函数取极值时候带来的一个必要条件，(2)是拉格朗日系数约束（同等式情况），(3)是不等式约束情况，(4)是互补松弛条件，(5)、(6)是原约束条件。

**10. SVM降低模型复杂度的方法有哪些？**

1. 加入软间隔（允许某些样本不满足约束）
2. 正则项

**11. 求解SVM过程中的SMO算法具体过程是什么？**

首先，对于对偶问题的二次规划表达如下：

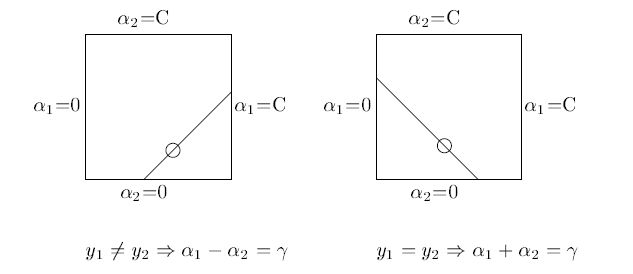
SMO在二次规划的过程中做了两件事：

①选取一对参数

②固定向量的其他参数，将代入上述表达式进行求最优解获得更新后的

SMO不断执行这两个步骤直至收敛。

因为有约束存在，实际上和的关系也可以确定。这两个参数的和或者差是一个常数。



**12. SVM有什么优缺点？**

优点：可用于线性/非线性分类，也可以用于回归；低泛化误差；容易解释；计算复杂度较低。

缺点：对参数和核函数的选择比较敏感；原始的SVM只比较擅长处理二分类问题。

**13. 为什么SMO的优化效率快？**

正如上面提到的，在固定其他参数以后，这就是一个单变量二次规划问题，仅有的约束也是这个变量，显然有闭式解。不必再调用数值优化算法。

KKT条件是对偶问题最优解的必要条件：

除了第一个非负约束以外，其他约束都是根据目标函数推导得到的最优解必须满足的条件，如果违背了这些条件，那得到的解必然不是最优的，目标函数的值会减小。

所以在SMO迭代的两个步骤中，只要中有一个违背了KKT条件，这一轮迭代完成后，目标函数的值必然会增大。总体而言，KKT条件违背的程度越大，迭代后的优化效果越明显，增幅越大。

**14. SMO算法中如何选取？偏移项如何确定？**

先选择最有可能需要优化（违反KKT条件最严重）的，再针对选取最有可能取得较大修正步长的。

因为任意支持向量均满足约束条件中的等号，所以可以取任意支持向量代入求解，但现实中一般使用所有支持向量求解的平均值。