**支持向量机问题讨论总结**

吴锐 李林 李震

1. **Svm 的基本原理与假设是什么？**

假设两类数据线性可分，存在无数个分离超平面可以将两类数据正确分开。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，（间隔最大化超平面是唯一的）。

SVM是一种二类分类模型。

1. 当训练样本线性可分时，通过硬间隔最大化，学习一个线性分类器
2. 当训练数据近似线性可分时，引入松弛变量（也叫惩罚变量），通过软间隔最大化，学习一个线性分类器
3. 当数据线性不可分时，通过使用核函数技巧以及软间隔最大化来解决分类问题
4. **SVM为什么采用间隔最大化？**
5. 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。
6. 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。
7. **几何间隔与函数间隔分别代表着什么？**

**函数间隔：**

对于一个训练样本我们定义它到超平面的函数间隔为:

我们希望函数间隔越大越好, 即:

若则样本分类正确。

对于整个训练集, 我们的函数间隔定义为:

函数间隔越大, 代表我们对于分类的结果非常确定. 我们希望函数间隔越大越好. 但是，只要我们成比增加w,b就可以使得最小间隔变大。

为了解决这个问题, 我们就需要加上一些限制条件,所以，需要将w的大小固定，如||w||=1，使得函数间隔固定。这时的间隔也就是几何间隔 .

**几何间隔：**

几何间隔就是点到超平面的距离。

在二维空间中，几何间隔就是点到直线的距离。在三维及以上空间中，就是点到超平面的距离。

当||w||=1几何间隔就是函数间隔.这个时候, 如果任意放大||w||,几何间隔是不会改变的, 因为||w||也会随着被放大.

为什么要将求解SVM的原始问题转变为其对偶问题

寻找存在约束问题时的最优解时，约束的存在减小了目标搜索的范围，但却使得问题更加复杂，为了使问题更易于求解，把目标函数和约束全部融入一个新的问题，即拉格朗日函数，通过这个函数寻找最优解。

对偶问题事实上把SVM从依赖d个维度（数据集的维度d，维度d太高影响运算）转变为依赖N个数据点，考虑到最后计算时只有支持向量才有意义，所以这个计算量实际上比N小很多

1. **为什么SVM要引入核函数？常见核函数有哪些？**

当样本在原始空间线性不可分时，可以将样本从原始空间映射到一个更高维的特征空间，使得样本在这个新特征空间线性可分。

核函数本质是两个函数的内积

常见的核函数：

1、线性核函数，主要用于线性可分情况，一般情况下，对于特征数量相对于样本数量非常多时，适合线性核函数

2、多项式核函数，对于正交归一化后的数据，可以优先选择此函数

3、径向基核函数，灵活性强，在不确定使用什么核函数时，可以优选选择此核函数

4、sigmoid核函数，SVM使用此核函数，相当于一个两层的感知机

在学习预测中，只定义核函数，而不是显式的定义映射函数。因为特征空间维数可能很高，甚至可能是无穷维，因此直接计算是比较困难的。相反，直接计算比较容易（即直接在原来的低维空间中进行计算，而不需要显式地写出映射后的结果）。

1. **SVM RBF核函数的具体公式？**

这个核会将原始空间映射为无穷维空间。不过，如果 σ 选得很大的话，高次特征上的权重实际上衰减得非常快，所以实际上（数值上近似一下）相当于一个低维的子空间；反过来，如果 σ 选得很小，则可以将任意的数据映射为线性可分——当然，这并不一定是好事，因为随之而来的可能是非常严重的过拟合问题。不过，总的来说，通过调控参数σ ，高斯核实际上具有相当高的灵活性，也是使用最广泛的核函数之一。

1. **为什么SVM对缺失数据敏感？**

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。

1. **SVM如何处理多分类问题？**

**一对多**，就是对每个类都训练出一个分类器，由svm是二分类，所以将此而分类器的两类设定为目标类为一类，其余类为另外一类。这样针对k个类可以训练出k个分类器，当有一个新的样本来的时候，用这k个分类器来测试，那个分类器的概率高，那么这个样本就属于哪一类。这种方法效果不太好，bias比较高。

**一对一**,针对任意两个类训练出一个分类器，如果有k类，一共训练出C(2,k) 个分类器，这样当有一个新的样本要来的时候，用这C(2,k) 个分类器来测试，每当被判定属于某一类的时候，该类就加一，最后票数最多的类别被认定为该样本的类。

1. **为什么要引入松弛变量和软间隔最大化？**

原因：一些离群点或者噪声影响分界面

线性不可分意味着有样本点不满足约束条件­­­,为了解决这个问题，对每个样本引入一个松弛变量,使得约束条件变为：

目标函数变为

目标函数的两层含义：

* Margin尽量大
* 误分类的样本尽量少、

引入松弛变量，就允许某些样本点的函数间隔小于1,。C为离群点的权重，值越大，说明离群点对目标函数影响越大，这时候间隔会变小。最小化目标函数有两层含义，使得1/2||w||尽量小，即间隔尽量大，同时分类错误的个数尽量小，C为调和二者的系数。