* Support Vector Machine及其变体相关机理的总结

1. 传统SVM

在介绍SVM机理之前，我们选择先对其前体最大边界分类器(Maximal Margin Classifier) 进行介绍，可以说，SVM是最大边界分类器的一个泛化。

* 1. Maximal Margin Classifier

最大边界分类器，这是一种适用于线性数据的二分类学习器，对于一个给定的数据集，我们可以找到不止一个超平面来作为分类依据，一个很自然的选择就是maximal margin hyperplane(AKA optimal separating hyperplane),也就是距离训练集数据距离最远的超平面（也就是把距离作为margin)。这种方法虽然有效，但是当特征数量p较大的时候还是比较引起过拟合的问题的。

下图就展示了一个在二维空间中的maximal的例子，其中加粗直线便是最大边界超平面(maximal margin hyperplane)我们可以发现有3个样本（加粗的O,X）是在与分类界限的平行线上的，这三个样本我们就叫做支撑向量（support vectors）

图示

描述已自动生成

构建Maximal Margin Classifier的流程如下：

其中M就代表样本距离分割面的最小距离

既然我们要引入SVC，则说明MMC分类器有短板，在相当一些情况 MMC无法找到一个超平面完成分类任务。

* 1. Support Vector Classifier

SVC是MMC分类器的一种衍生，相较MMC来说，其不要求分类器完美分割，而给予了分类界限一定的缓冲空间这样下来对异常值有更好地容忍性，有更好地区分能力，如下图所示。

地图的截图

中度可信度描述已自动生成

其具体的实现算法过程如下:

C是一个非负的超参数，M是margin的距离我们希望越大越好，而 是我们对于模型误分类的容忍度。相比之前的分类器，我们只有一点区别，其实很简单，就是我们对于模型分类器分类情况的指针，如果那么样本分类完全正确，如果那么样本在margin的错误一遍，如果那么样本就被分类错误。

根据以上，我们就可以想到超参数C可以理解为我们对于模型错误程度的容忍度，因为C其实就是所有的和。如果C=0，那么模型就和之前的模型没区别，实际运用中，我们可以用交叉验证来确定最优的C值。

* 1. Support Vector Machine

SVM在前者基础上可以处理非线性问题，为此，可以考虑增大特征空间，比如加入二次项，三次项来解决非线性问题。

如果我们利用SVC的原始特征来学习，则特征为，如过把特征空间增加到二维则是，学习过程因此变为如下所示：

SVM扩大特征空间的方法有一个很特别的名称，叫做kernal（核函数），核函数可以有效的在高维空间上拟合数据而有不会使得模型学习的计算量过大。

下面对一些接触到的核函数进行介绍

* + - 高斯核函数

高斯核函数为

展开后得到

上式的三项中，如果仅存在前面两项，则并没有进行升维操作，我们将前两项忽略，重点关注第三项。

根据泰勒展开式

将第三项展开得到

第三项可以展开写成两个向量相乘的形式

因此，我们知道高斯核函数将原始特征空间上升到无穷维。

* + - 多项式核函数

多项式核函数为

假设存在一个维度转换函数 ，原始向量为转换后得到在此基础上求内积得到

当参数c和d确定时，新维度数也就确定了。例如上面的例子中，新维度数是6维。整体来说展开后有多少项，新特征空间维度便是多少。有c的存在，高维空间中还保留了原始低维空间中的维度。使得结果同时具有高低次项，维度更为丰富。而c=0时，高维空间中的低次项没有了，全是高次项。

* + - 线性核函数

线性核函数为

对应的支持向量机实际上就是线性支持向量机（linearSVC），因此，线性支持向量机实际上可以看做非线性支持向量机的一个特例。

* + - 径向基核函数（rbf核函数）

径向基核函数为

其中可手动设置需满足大于0。

gamma默认值为

为范数在平面几何的向量来说就是模，然而一旦维度很高就不太可以描述了，可能是衡量距离的一个东西，那么这个代表什么意思呢？简单来说就是一个圆而已，在二维平面，就是圆心，x是数据，这个数据距离圆心的距离，因为和数据的位置和大小无关，只和到圆心的半径有关，况且同一半径圆上的点到圆心是相等的因此取名为径向，代入映射函数就是径向基函数。其升维过程与高斯核函数类似。

素材来源：

An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (Nello Cristianini, John Shawe-Taylor).pdf

<https://blog.csdn.net/weixin_43249038/article/details/125833362>