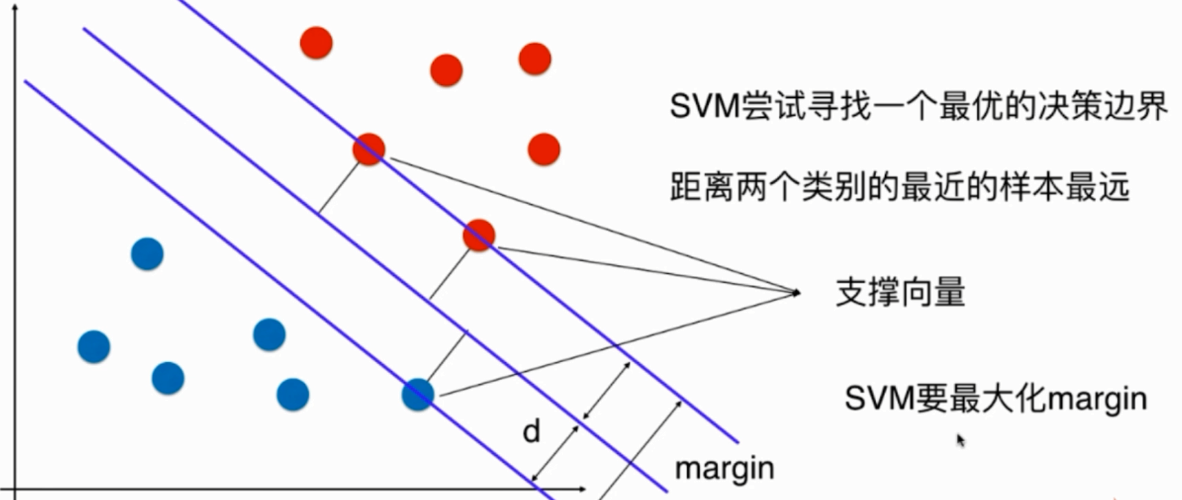
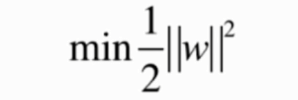
# SVM思想:

SVM分类的思想为在所有不适定的决策边界中找到一条与双方支撑向量中间的距离最大的那个决策边界

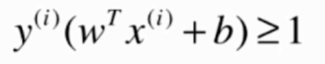


hard margin(严格距离):

margin中间不会有任何的样本点,根据点到直线的距离公式将样本点代入优化为如下的式子,一个有条件的优化公式:

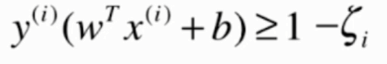


不过,该式子的最小值求极值W条件需满足:

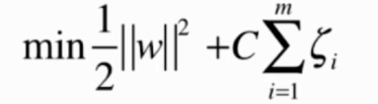


soft margin:

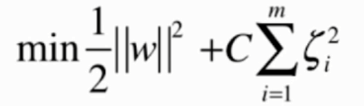
允许margin中间有样本点,这对有特殊值或者误差样本的样本预测效果更好,不会受单个极值影响.采取的方法是对hard margin添加一个eta(截距项) 使得决策边界在一定的范围内根据截距项取值不同而移动.

但是实际算法中,我们需要对eta进行限制,否则会产生将所有样本点都包含的情况,即我们需要添加一个截距项,但是这个截距项又不能太大,我们想到了类似正则化的作用,于是将svm的损失函数进行变形如下:



同样也可以写成L2的形式



以上两个最优化都需要满足soft margin的条件表达式.

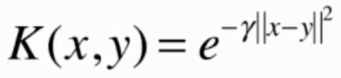
# 核函数:

以线性分类转成非线性分类添加多项式为例,我们可以找到一个核函数,将样本经过该函数后变成一个具有多项式项特征的样本,同样其他核函数也是类似思路,将样本映射到另一类样本的映射函数.一般来说SVM核函数是将低维度样本映射成高维度样本.从而将非线性可分问题转变成线性可分的情况.

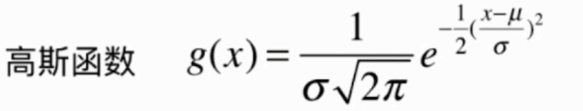
多项式核函数(sklearn中对应poly):



高斯核函数(也称做径向基函数,简称RBF):



其中 里面的 |x - y| 为向量的模.,gama为超参数,gama越小映射的维度越低,伽马越大映射出的维度越高.与高斯分布函数有类似的形态.





其中m为样本个数,所以可以说高斯函数能映射成无穷维,因为维度和样本个数相等.