两类的SVM (参考：Statistical Learning Theory - Vapnik – 1998 P401)

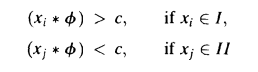
假设给定样本集合



是标记，在二个类别的SVM中，只能取值 -1，1

是样本点，在分类问题中，可以理解为特征向量，例如在文本分类中，假设词典是 ，那么在二维分类问题中，一个文档的特征向量可以定义为一个2\*n的向量 },如果一篇文档的标记为1,则的元素均正数，元素均为0, 其中是该文档种出现的单词；当不是文档中的单词时。

假设标记为1的 形成集合 ,标记为-1的形成集合。训练的目标是找到参数向量满足下面的公式：

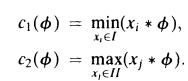


如果这样的存在，我们称该样本集合()是可以被超平面



分开的，类似于三维平面公式 k

假设是单元量，即向量模为1。





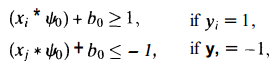
* 是最大化的点。
* 和决定的超平面是<>=分开样本点的最大超平面或最优超平面



定理：最优超平面是唯一的，证明略。

怎样从样本点学习出参数？

一个等价的问题是找到和满足：

 （1-1）

并且最小化

 (1-2)

定理：假设



则

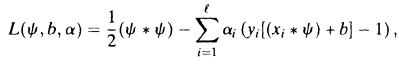


显然是单元向量并且使最大。

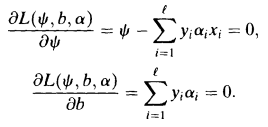
所以求得参数的过程可以通过满足(1-1)，最大化目标函数(1-2)来实现,即求解一个二次规划问题。

什么是支持向量?

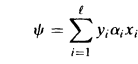
解决上述二次规划的问题通过计算下面拉格朗日函数的鞍点来实现：

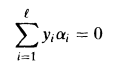
 （1-3）

其中是拉格朗日乘因子。计算（1-3）的鞍点，要是该函数在和上最小，在上最大。所以：

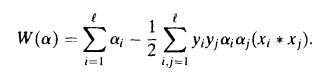


所以：

 （1-4）

 （1-5）

将（1-4）代入（1-3）得到：

 （1-6）

从而将问题转化为求得使(1-6)最大的

所以二次规划的解为：



的值通过最大化得到。

由于最优的和满足Kuhn-Tucker条件：



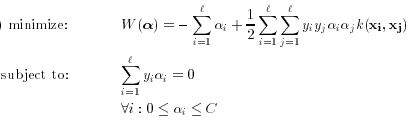
所以非零的满足条件：

 （1-7）

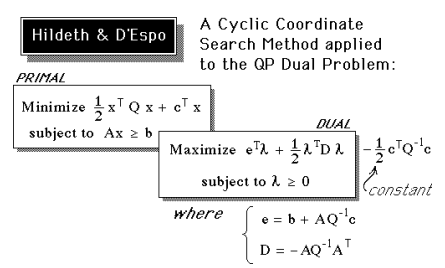
*支持向量是一部分样本点，这些样本点满足(1-7)。从几何意义上，这些样本点离最优超平面最近。*

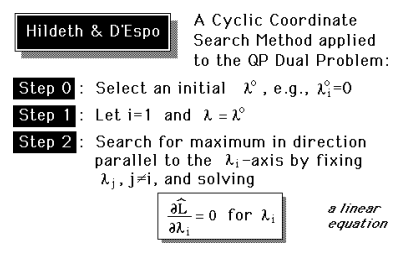
怎样求解？

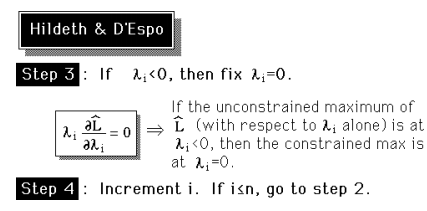
二次规划的问题(1-1)和(1-2)转换为对偶问题（具体参考Statistical Learning Theory - Vapnik – 1998 P411）：

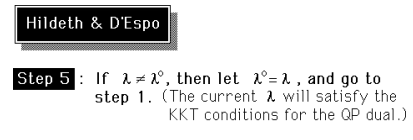
 （1-8）

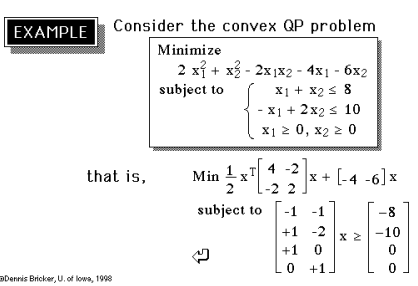
在svm\_multiclass 软件包的新版本中是使用Hildreth and D'Espo 算法来求解。

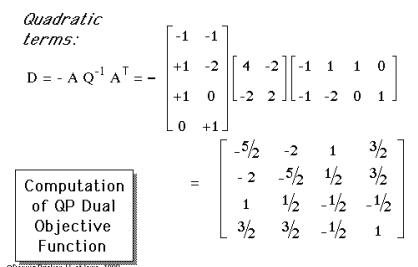


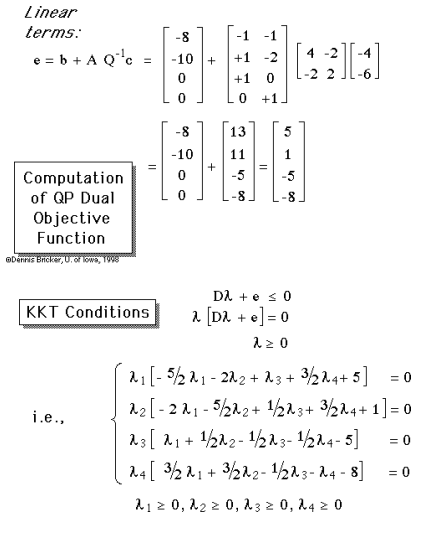


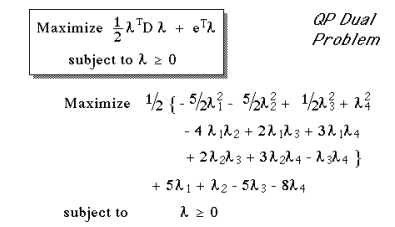


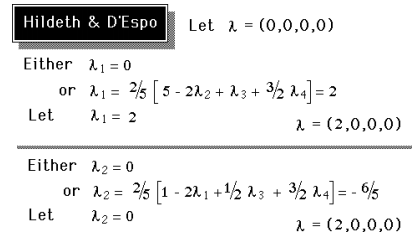


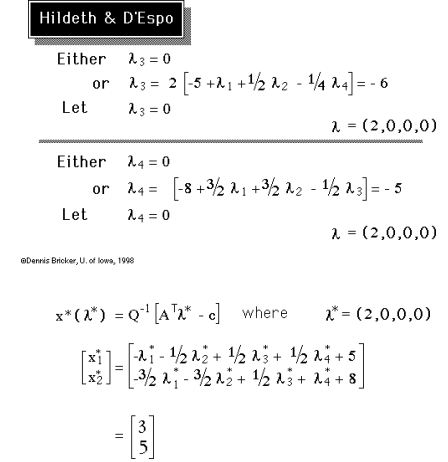


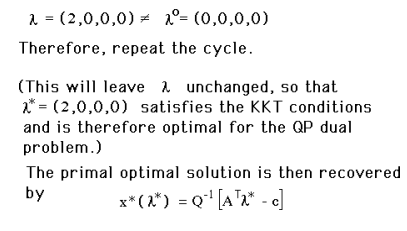












在实际的算法中由于要计算矩阵，的大小是,其中 是样本的大小。当样本很大时，计算和存储会存在问题。在svm\_multiclass中是将二次规划问题（1-8）分为多个部分分别进行优化。即每次都会选取数量相对较少的一部分样本进行优化。最终得到(1-8)的正确解。具体的分解和样本选择算法可以参考 Making Large-Scale SVM Learning Practical (Thorsten Joachims).

多类问题是二类问题的泛化。具体参考Support Vector Learning for Interdependent and Structured Output Spaces。

后面我也会整理一下这两篇论文，方便大家参考。