## SMO中局发式选择变量

在 SMO 算信中, 於门甸坝需要达取一对从来进行优况, 面过最复 式的选取我们可以更高成了的选取优化的效量使得目标的数数 下降最小夫, 包+24第一千日, 和第二千日。 阳叶 SMO 采取不同
而危发或争取分)

12 m 6 1	第一	子爱	3 To	总净
----------	----	----	------	----

第一个效量的选择的外领到	, 与之前随历整行人不同,在这里交行
在整个样在集和维也界样在	真间电行及精力

## 第一个变量的选择

第一个变量的选择为外循环,与之前便利整个  $\alpha$  列表不同,在这里我们在整个样本集和非边界样本集间进行交替:

1. 首先我们对整个训练集进行遍历,检查是否违反KKT条件,如果改点的  $\alpha_i$  和  $x_1$  , $y_i$  违反了KKT条件则说明改点需要进行优化。

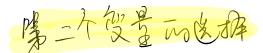
Karush-Kuhn-Tucker(KKT)条件是正定二次规划问题最优点的充分必要条件。针对SVM对偶

问题,KKT条件非常简单: 
$$\begin{cases} \alpha_i = 0 \Longleftrightarrow y_i(w^Tx_i + b) \geq 1 \\ \alpha_i = C \Longleftrightarrow y_i(w^Tx_i + b) \leq 1 \\ 0 < \alpha_i < C \Longleftrightarrow y_i(w^Tx_i + b) = 1 \end{cases}$$

2. 在遍历了整个训练集并优化了相应的 $\alpha\alpha$ 后第二轮迭代我们仅仅需要遍历其中的非边界 $\alpha\alpha$ . 所谓的非边界  $\alpha$  就是指那些不等于边界0或者C的  $\alpha$  值。 同样这些点仍然需要检查是否违反 KKT条件并进行优化.

之后就是不断地在两个数据集中来回交替,最终所有的  $\alpha$  都满足KKT条件的时候,算法中止。

为了能够快速选取有最大步长的  $\alpha$  ,我们需要对所有数据对应的误差进行缓存,因此特地写了个 SVMUtil类来保存svm中重要的变量以及一些辅助方法:



Amber	个变量的选择	•
-	八亚二四三年	ح.
	THE BOUND HAVE A STATE A STATE OF STATE	-

SMO中的第二个变量的选择过程为内循环,	当我们已经选取第一个 $lpha_1$	之后,	我们希望我们选取
的第二个变量 $lpha_2$ 优化后能有较大的变化。	根据我们之前推导的式子		

$$lpha_2^{new,unclipped}=lpha_2^{old}+rac{y_2(E_1-E_2)}{\eta}$$
 可以知道,新的  $lpha_2$  的变化依赖于  $|E_1-E_2|$  ,

当  $E_1$  为正时,那么选择最小的  $E_i$  作为  $E_2$  ,通常将每个样本的  $E_i$  缓存到一个列表中,通过在列表中选择具有  $|E_1-E_2|$  的  $\alpha_2$  来近似最大化步长。

	上述的启发式方式仍不能够是的函数值有足够的下降,这是按下述步骤进行选择:
	界数据集上选择能够使函数值足够下降的样本作为第二个变量 边界数据集上没有,则在整个数据仅上进行第二个变量的选择
	边界数据集工及有,则任金十数据以工进行第二十变量的远择 $% = 1$ 数据实际的 $% = 1$ 数据或证明, $%$
3. XHX-1/3:	が ( 大





