SMO中启发式选择变量

- 在SMO算法中，我们每次需要选取一对 α 来进行优化，而通过启发式的选取我们可以更高效的选取优化的变量使得目标函数下降最快，包括对第一个 α_1，和第二个 α_2。所以SMO采取不同的启发式手段；

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 α 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 α_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: $\begin{cases} \alpha_i = 0 \Longleftrightarrow y_i(w^T x_i + b) \geq 1 \\ \alpha_i = C \Longleftrightarrow y_i(w^T x_i + b) \leq 1 \\ 0 < \alpha_i < C \Longleftrightarrow y_i(w^T x_i + b) = 1 \end{cases}$
2. 在遍历了整个训练集并优化了相应的αα后第二轮迭代我们仅仅需要遍历其中的非边界αα. 所谓的非边界 α 就是指那些不等于边界0或者C的 α 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 α 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 α ，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:

第二个变量的选择

第二个变量的选择

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 α_1 之后，我们希望我们选取的第二个变量 α_2 优化后能有较大的变化。根据我们之前推导的式子 $\alpha_2^{new,unclipped} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$ 可以知道，新的 α_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 α_2 来近似最大化步长。

有时候按照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:

1. 在非边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
2. 如果非边界数据集上没有，则在整个数据仅上进行第二个变量的选择
3. 如果仍然没有则重新选择第一个 α_1