

支持向量机

周志健

1 解决的问题

软间隔支持向量机:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^L \xi_i \\ s.t \quad & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, L \\ & \xi_i \geq 0, \quad i = 1, \dots, L \end{aligned} \tag{1}$$

令数据样本增加一个固定为 1 的维度, 得到:

$$[\mathbf{x}, 1] \rightarrow \mathbf{x} \qquad [\mathbf{w}, 1] \rightarrow \mathbf{w} \tag{2}$$

规定线性核, 问题 1 转换为:

$$\begin{aligned} \min_{\mathbf{w}} &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^L \xi_i \\ s.t \quad & y_i \mathbf{w}^T \mathbf{x}_i \geq 1 - \xi_i, \quad i = 1, \dots, L \\ & \xi_i \geq 0, \quad i = 1, \dots, L \end{aligned} \tag{3}$$

2 SMO

2.1 流程图

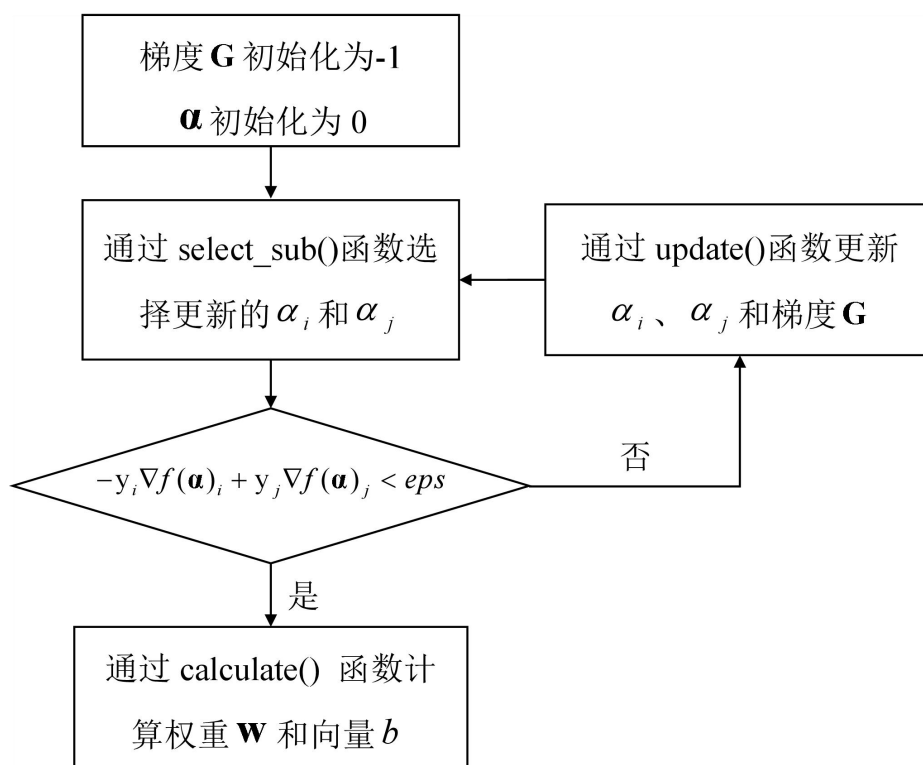


图 1.SMO 流程图

2.2 实验结果

表 1: Iris 数据集分类(SMO)

	Accuracy	Recall	Precision	F1
Setosa vs Versicolor	1.0	1.0	1.0	1.0
Setosa vs Virginica	1.0	1.0	1.0	1.0
Versicolor vs Virginica	0.83	1.0	0.74	0.85

表 1 是手动实现的 SVM 在 Iris 数据集上 One vs One 的表现，均使用默认参数。Iris 数据集共 150 个数据，每类 Iris50 个，测试集比例为 0.3。

表 2: Wifi localization 数据集分类

	Accuracy	Recall	Precision	F1
手动实现	0.91	0.82	0.98	0.89
Sklearn 库	0.89	0.76	1.0	0.87

表 2 是手动实现的 SVM 与 Sklearn 库 SVM 在 Wifi localization 数据集上的表现，使用默认参数。Wifi localization 数据集共 2000 个数据，分割数据集时正反例数据个数相等，测试集比例为 0.3。

3 DCDM

3.1 流程图

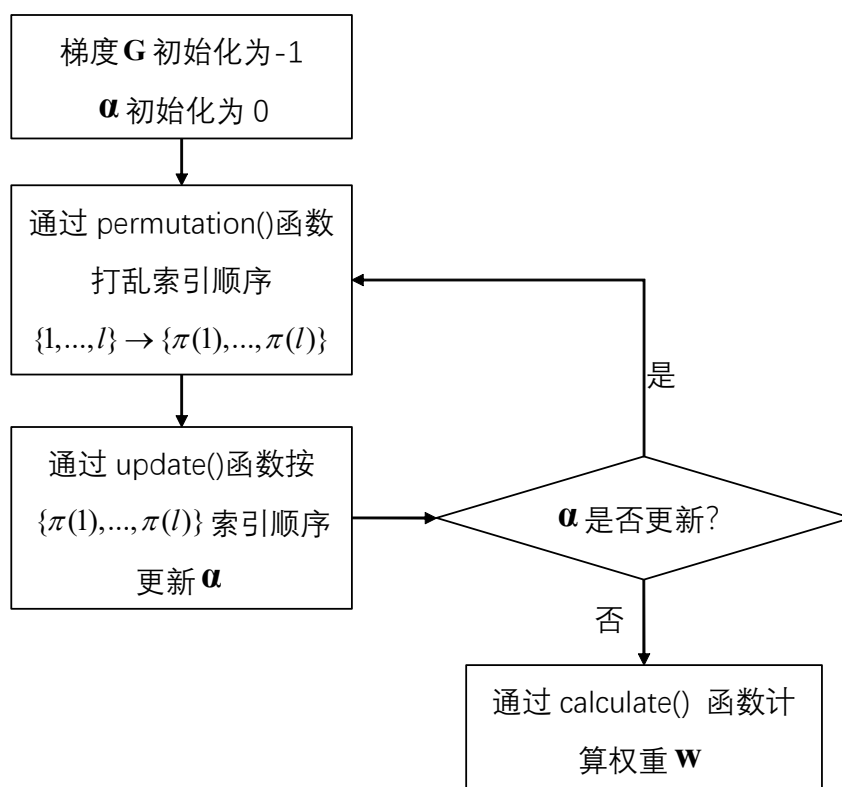


图 2.DCDM 流程图

3.2 实验结果

表 3: Iris 数据集分类(DCDM)

	Accuracy	Recall	Precision	F1
Setosa vs Versicolor	1.0	1.0	1.0	1.0
Setosa vs Virginica	1.0	1.0	1.0	1.0
Versicolor vs Virginica	0.97	1.0	0.93	0.97

表 4: Wifi localization 数据集分类(DCDM)

	Accuracy	Recall	Precision	F1
手动实现	0.99	1.0	0.99	0.99

可能因为算法终止条件为 α 是否更新，所以分类器训练效果较好。