**1、工具：dlib**

描述：包含机器学习算法的C++开源工具包

版本：19.11

语言：C++

网站：http://dlib.net/

支持算法：

1. 深度学习
2. 常规的基于SMO的支持向量机用于分类和回归
3. 用于大规模分类和回归的降阶方法
4. 用于分类和回归的相关向量机
5. 通用多类分类工具
6. **多分类SVM**
7. 解决与结构支持向量机相关的优化问题的工具。
8. 用于序列标记的结构SVM工具
9. 用于解决分配问题的结构SVM工具
10. 用于图像中物体检测的结构SVM工具以及更强大的深度学习工具
11. 用于标记图中节点的结构SVM工具
12. 大规模的SVM-Rank实现
13. 在线内核RLS回归算法
14. **一种在线SVM分类算法**
15. 半确定度量学习
16. 在线核心化质心估计器/新颖性检测器和离线支持向量一类分类
17. 聚类算法：线性或内核k均值，中文Whispers和Newman聚类。
18. 径向基函数网络
19. 多层感知器

**2、工具：libsvm**

描述：一个SVM模式识别与回归的软件包

语言：Java、python、c

网站：https://www.csie.ntu.edu.tw/~cjlin/libsvm/

算法支持：

1. 不同的SVM公式
2. 高效的多分类
3. 交叉验证模型选择
4. 概率估计
5. 各种内核（包括预先计算的内核矩阵）
6. **用于不平衡数据的加权SVM**

**3、工具：weka**

描述：一个Java编写的开源机器学习与数据挖掘软件

语言：Java

网站：https://www.cs.waikato.ac.nz/ml/weka/

**4、工具：scikit-learn**

描述：一个基于Python的机器学习库

语言：python

网站：http://scikit-learn.org/stable/index.html

算法支持：

1. 广义线性模型
2. 线性和二次判别分析
3. 核岭回归
4. **支持向量机**
5. 随机梯度下降
6. 最近邻居
7. 高斯过程
8. 交叉分解
9. **朴素贝叶斯**
10. **决策树**
11. 合奏方法
12. 多类和多标签算法
13. 功能选择
14. 半监督
15. 等渗回归
16. 概率校准
17. 神经网络模型（监督）
18. 高斯混合模型
19. 歧管学习
20. 聚类
21. 双聚类
22. 分解组件中的信号（矩阵分解问题）
23. 协方差估计
24. 新颖性和异常值检测
25. 密度估计
26. 神经网络模型（无监督）

**5、工具：mlpack**

描述：一个基于C++的机器学习库

语言：C++

网站：http://www.mlpack.org/index.html

算法支持：

1. mlpack\_adaboost：使用AdaBoost进行训练和分类，AdaBoost是一个整合分类器
2. mlpack\_allknn：使用树的所有k最近邻居搜索
3. mlpack\_allkfn：所有k-最远的邻居搜索树木
4. mlpack\_allkrann：用树排列近似k-最近邻居搜索
5. mlpack\_approx\_kfn：近似k-最远邻居搜索
6. mlpack\_cf：通过协作过滤生成推荐
7. mlpack\_dbscan：DBSCAN集群
8. mlpack\_decision\_stump：用决策树分类
9. **mlpack\_decision\_tree：用决策树分类**
10. mlpack\_det：密度估计树
11. mlpack\_emst：计算欧几里德最小生成树
12. mlpack\_fastmks：使用树执行快速最大内核搜索
13. mlpack\_gmm\_train：训练一个高斯混合模型
14. mlpack\_gmm\_generate：从GMM生成随机序列
15. mlpack\_gmm\_probability：计算来自给定GMM的一组点的概率
16. mlpack\_hmm\_generate：从隐马尔可夫模型（HMM）生成观测值
17. mlpack\_hmm\_loglik：计算来自HMM的一些观察的对数似然
18. mlpack\_hmm\_train：训练隐马尔可夫模型（HMM）
19. mlpack\_hmm\_viterbi：在一些HMM中找到最可能的隐藏状态，用于某些观察
20. mlpack\_hoeffding\_tree：使用Hoeffding树进行训练和分类，这是一个针对超大型数据集的流式决策树
21. mlpack\_kernel\_pca：执行内核主成分分析
22. mlpack\_kfn：所有k-最远的邻居用树搜索
23. mlpack\_kmeans：执行k-means聚类
24. mlpack\_knn：所有与树木的最近邻居搜索
25. mlpack\_krann：用树排列近似k-最近邻居搜索
26. mlpack\_lars：最小角度回归
27. mlpack\_linear\_regression：简单的最小二乘线性回归
28. mlpack\_local\_coordinate\_coding：本地坐标编码
29. mlpack\_logistic\_regression：通过逻辑回归进行训练或分类
30. mlpack\_lsh：使用局部敏感散列的近似k近邻搜索
31. mlpack\_mean\_shift：平均移位聚类
32. **mlpack\_nbc：用朴素贝叶斯分类器训练或分类**
33. mlpack\_nca：邻域分量分析
34. mlpack\_nmf：非负矩阵分解
35. mlpack\_pca：主要组件分析
36. mlpack\_perceptron：用感知器训练或分类
37. mlpack\_preprocess\_binarize：二进制数据集的特征
38. mlpack\_preprocess\_imputer：估算数据集的缺失值
39. mlpack\_preprocess\_describe：数据集的生成描述性统计
40. mlpack\_preprocess\_split：将数据集分成训练和测试集
41. mlpack\_radical：RADICAL（独立分量分析）
42. **mlpack\_random\_forest：随机森林分类器**
43. mlpack\_range\_search：用树搜索范围
44. mlpack\_softmax\_regression：使用softmax回归进行训练或分类
45. mlpack\_sparse\_coding：字典学习的稀疏编码

**注：**

1. **以上算法均无偏置SVM的直接实现，scikit-learn可实现nu-SVM； libsvm可实现nu-SVM和C-SVM；dlib可实现SVM。**
2. **Weka中可实现C4.5，除此之外算法均无C4.5的直接实现,scikit-learn采用CART的改进算法，与C4.5的主要不同在于特征选择标准不同和是否为二叉树。mlpack采用信息增益作为特征选择标准。**