

codeDog123

java-ml(二)

SVM分类Iris.csv数据集

首先说明几个概念：

True Positive (TP): “真正” 指被model预测为正的正样本，即判断为真的正确率

True Negative(TN): “真负” 指被model预测为负的负样本，即判断为假的正确率

False Positive(FP): “假正” 指被模型预测为正的负样本，即误报率

False Negative(FN): “假负” 指被模型预测为负的正样本，即漏报率

精准度 (precision) : $P = TP/(TP+FP)$ 指被分类器判定正例中的正样本的比重

召回率(Recall) : $R = TP/(TP+FN) = 1 - FN/T$ 指的是被预测为正例的占总的正例的比重

准确率(Accuracy) : $A = TP/(TP+FN) = (TP+TN)/(TP+FN+FP+TN)$ 分类器的准确度,能将正的判定为正，负的判定为负。

以下为SVM的训练和预测:



```
public static void main(String[] args) throws IOException{
    Dataset data=FileHandler.loadDataset(new File("D:/test/Iris.csv"),4,"");
    Sampling s=Sampling.SubSampling;
    Classifier c=new LibSVM(); //创建svm分类器
    for(int i=0;i<5;i++){
        Pair<Dataset, Dataset> datas=s.sample(data, (int)(data.size()*0.8),i); //根据初始
数据集分成训练集和测试集
        c.buildClassifier(datas.x()); //训练分类器
        Map pms=EvaluateDataset.testDataset(c, datas.y()); //测试分类器的性能
    }
}
```

导航

[博客园](#)[首页](#)[新随笔](#)[联系](#)[订阅](#) XML[管理](#)

公告

昵称：[codeDog123](#)

园龄：[1年6个月](#)

粉丝：[1](#)

关注：[0](#)

[+加关注](#)

2018年3月						
日	一	二	三	四	五	六
25	26	27	28	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7

统计

随笔 - [31](#)

文章 - [0](#)

评论 - [11](#)

引用 - [0](#)

搜索

常用链接

```
        System.out.println(pms);
    }
    //预测
    double[] values = new double[] { 5.1,3.5,1.4,0.2 };
    Instance instance=new DenseInstance(values);//设置一个样本
    System.out.println(c.classify(instance));//输出该样本的分类结果

}
```



输出为:

{Iris-versicolor=[TP=11.0, FP=0.0, TN=18.0, FN=1.0], Iris-virginica=[TP=9.0, FP=1.0, TN=20.0, FN=0.0], Iris-setosa=[TP=9.0, FP=0.0, TN=21.0, FN=0.0]}

{Iris-versicolor=[TP=10.0, FP=0.0, TN=19.0, FN=1.0], Iris-virginica=[TP=10.0, FP=1.0, TN=19.0, FN=0.0], Iris-setosa=[TP=9.0, FP=0.0, TN=21.0, FN=0.0]}

{Iris-versicolor=[TP=8.0, FP=0.0, TN=21.0, FN=1.0], Iris-virginica=[TP=10.0, FP=1.0, TN=19.0, FN=0.0], Iris-setosa=[TP=11.0, FP=0.0, TN=19.0, FN=0.0]}

{Iris-versicolor=[TP=9.0, FP=0.0, TN=21.0, FN=0.0], Iris-virginica=[TP=9.0, FP=0.0, TN=21.0, FN=0.0], Iris-setosa=[TP=12.0, FP=0.0, TN=18.0, FN=0.0]}

{Iris-versicolor=[TP=10.0, FP=0.0, TN=20.0, FN=0.0], Iris-virginica=[TP=6.0, FP=0.0, TN=24.0, FN=0.0], Iris-setosa=[TP=14.0, FP=0.0, TN=16.0, FN=0.0]}

Iris-setosa

可以看到五次训练结果的准确度分别为:

第一次: $(11+18)/(11+18+1)=0.9666666666666667$

第二次: $(10+19)/(10+19+1)=0.9666666666666667$

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

我的标签

[传统软件工程\(2\)](#)
[结构化方法\(1\)](#)
[面向对象方法\(1\)](#)
[敏捷开发\(1\)](#)
[云计算\(1\)](#)
[struts2\(1\)](#)
[ajax\(1\)](#)
[java的机器学习库\(1\)](#)
[java-ml\(1\)](#)
[Scrum敏捷开发\(1\)](#)
[更多](#)

随笔分类

[C#](#)
[C++](#)
[matlab\(2\)](#)
[常用算法\(6\)](#)
[机器学习\(6\)](#)

随笔档案

[2017年7月 \(1\)](#)
[2017年5月 \(3\)](#)
[2017年4月 \(11\)](#)
[2017年3月 \(9\)](#)
[2017年2月 \(1\)](#)
[2017年1月 \(2\)](#)
[2016年12月 \(1\)](#)
[2016年11月 \(1\)](#)
[2016年10月 \(2\)](#)

最新评论

[1. Re:结构化方法和面向对象方法之比较](#)

请看这个关于OO缺点的文章，再阐述你的看法：

第三次:(8+21)/(8+21+1)=0.9666666666666667

第四次:(9+21)/(9+21)=1

第五次:(10+20)/(10+20)=1

准确度都比较高

最后对于向量{5.1, 3.5, 1.4, 0.2}给出的种类预测为Iris-setosa类。

补充:Classifier的方法有三个:

void	buildClassifier (Dataset data) Create a classifier from the given data set.
java.util.Map<java.lang.Object,java.lang.Double>	classDistribution (Instance instance) Generate the membership distribution for this instance using this classifier.
java.lang.Object	classify (Instance instance) Classify the instance according to this classifier.

注意:classDistribution是返回一个向量，例如

```
double[] values = new double[] { 5.1,3.5,1.4,0.2 };
Instance instance=new DenseInstance(values);//设置一个样本
System.out.println(c.classDistribution(instance));//输出该样本的分类结果
```

输出的是:{Iris-versicolor=0.0, Iris-virginica=0.0, Iris-setosa=1.0},表名Iris-setosa是分类结果。

具体的说明如下

• classDistribution

java.util.Map<java.lang.Object,java.lang.Double> classDistribution(Instance instance)

- SoftwareTeacher
2. Re:结构化方法和面向对象方法之比较
- 膜健神
- zhenw
3. Re:云计算对传统软件工程的影响
- 写得吼
- zhangshaojun96
4. Re:云计算对传统软件工程的影响
- 还没有机会实践过。。。还处在学习hadoop的初级阶段。。。。
- codeDog123
5. Re:敏捷软件工程(agile software development) VS传统软件工程(traditional software development)

。。。。。。作者写的很不错，很详细，让我很快的了解了传统工程学和敏捷开发的异同。。。。。。。。。。
--hlworld

阅读排行榜

instance using this classifier(793)

2. Tomcat配置本地路径到服务器路径下(459)

3. 快排的递归和非递归C++(231)

4. 云计算对传统软件工程的影响(223)

5. 敏捷软件工程(agile software development) VS传统软件工程(traditional software development)(218)

评论排行榜

1. 敏捷软件工程(agile software development) VS传统软件工程(traditional software development)(6)
2. 云计算对传统软件工程的影响(3)
3. 结构化方法和面向对象方法之比较(2)

推荐排行榜

1. 敏捷软件工程(agile software development) VS传统软件工程(traditional software development)(5)
2. 结构化方法和面向对象方法之比较(4)
3. 云计算对传统软件工程的影响(2)

Generate the membership distribution for this instance using this classifier. All values should be in the interval [0,1] Note: The returned map may not contain a value for all classes that were present in the data set used for training. If the map does not contain a value, the value for that class equals zero.

Parameters:

instance - the instance to be classified

Returns:

an array with membership degrees for all the various classes in the data set

分类: [机器学习](#)



codeDog123

关注 - 0

粉丝 - 1

+加关注

« 上一篇 : [java-ml机器学习库\(一\)](#)

» 下一篇 : [encog\(一\)](#)

posted on 2017-04-15 21:05 [codeDog123](#) 阅读(68) 评论(0) [编辑](#) [收藏](#)

0

0

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】超50万VC++源码: 大型组态工控、电力仿真CAD与GIS源码库！

【缅怀】传奇谢幕，回顾霍金76载传奇人生

【推荐】业界最快速.NET数据可视化图表组件

【腾讯云】买域名送解析+SSL证书+建站

【活动】2050 科技公益大会 - 年青人因科技而团聚



最新IT新闻:

- 支付宝可直接兑奖顶呱刮彩票：送100万大奖
 - 马云224亿港币拿下大润发！63岁BOSS兴奋变36岁
 - 爱奇艺IPO发行价正式敲定为18美元 募资22.5亿美元
 - 传魅族将进行千人大裁员，官方称只是末位淘汰
 - 苹果ARKit发布半年 相关应用下载量超1300万次
- » 更多新闻...



最新知识库文章:

- 写给自学者的入门指南
 - 和程序员谈恋爱
 - 学会学习
 - 优秀技术人的管理陷阱
 - 作为一个程序员，数学对你到底有多重要
- » 更多知识库文章...