

深度学习 week8

一、实现思路

使用Logistic回归和SVM对数据进行分类

1) Logistic

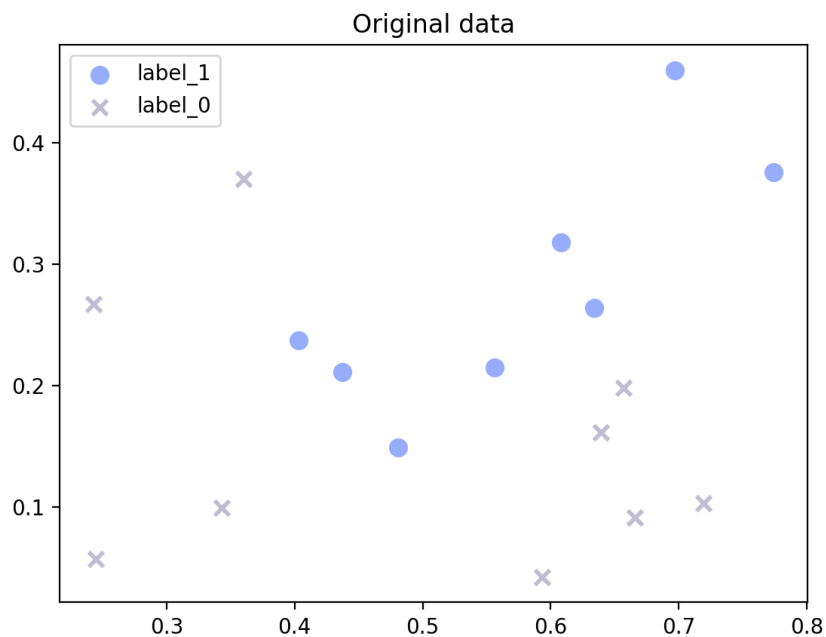
使用 sklearn 中内置

LogisticRegression 函数回

归

给定不同参数 C，预测计算

准确率并绘图



2) SVM

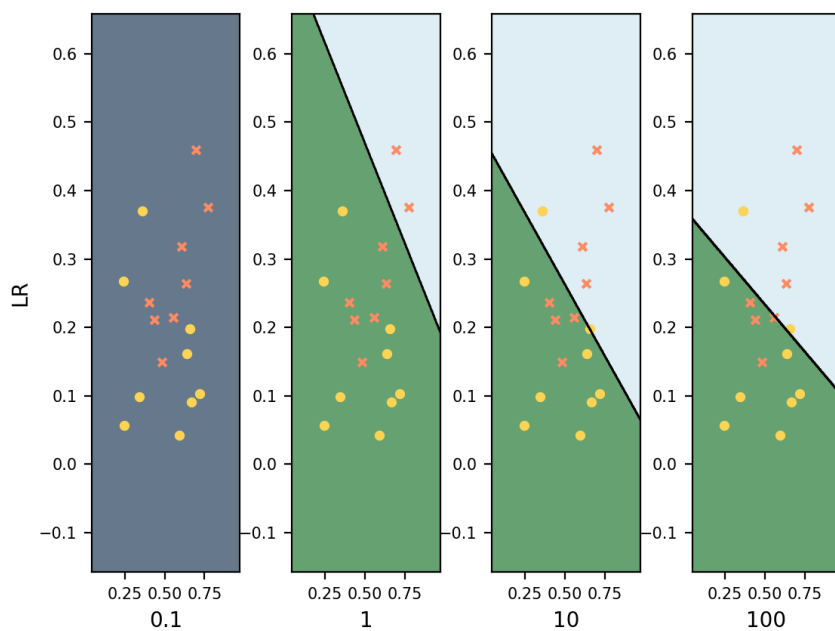
使用 sklearn 中内置 SVM

函数回归

给定不同参数 C，不同的核

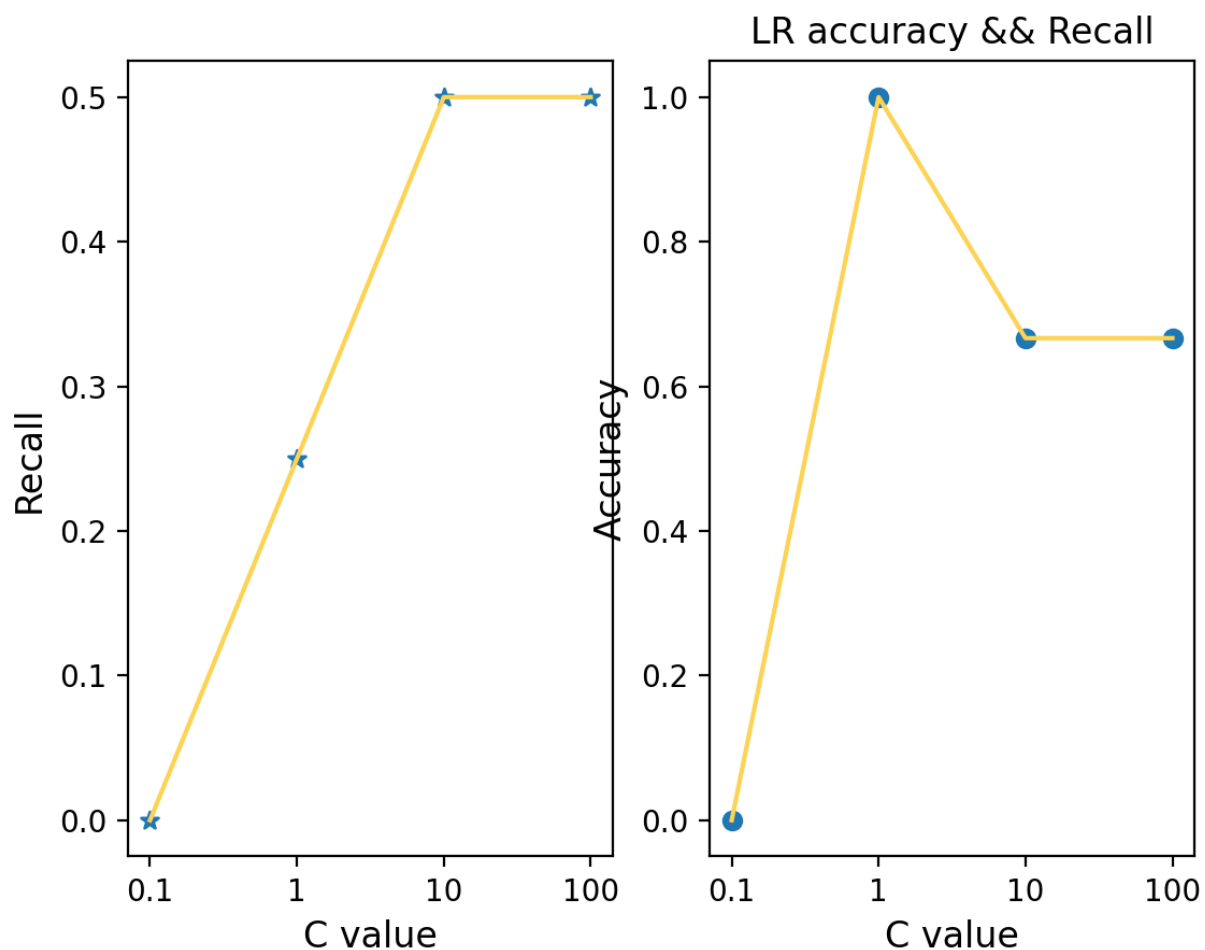
函数

预测计算准确率并绘图



二、运行结果

1) LR:



惩罚因子为：0.1

真实标签为：

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为：

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

准确率为0.529412

不存在1的正确分类

召回率为0.000000

惩罚因子为：10

真实标签为：

[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]

预测结果为：

[1 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0]

准确率为0.647059

精确度为0.666667

召回率为0.500000

惩罚因子为：1

真实标签为：

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为：

[1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

准确率为0.647059

精确度为1.000000

召回率为0.250000

惩罚因子为：100

真实标签为：

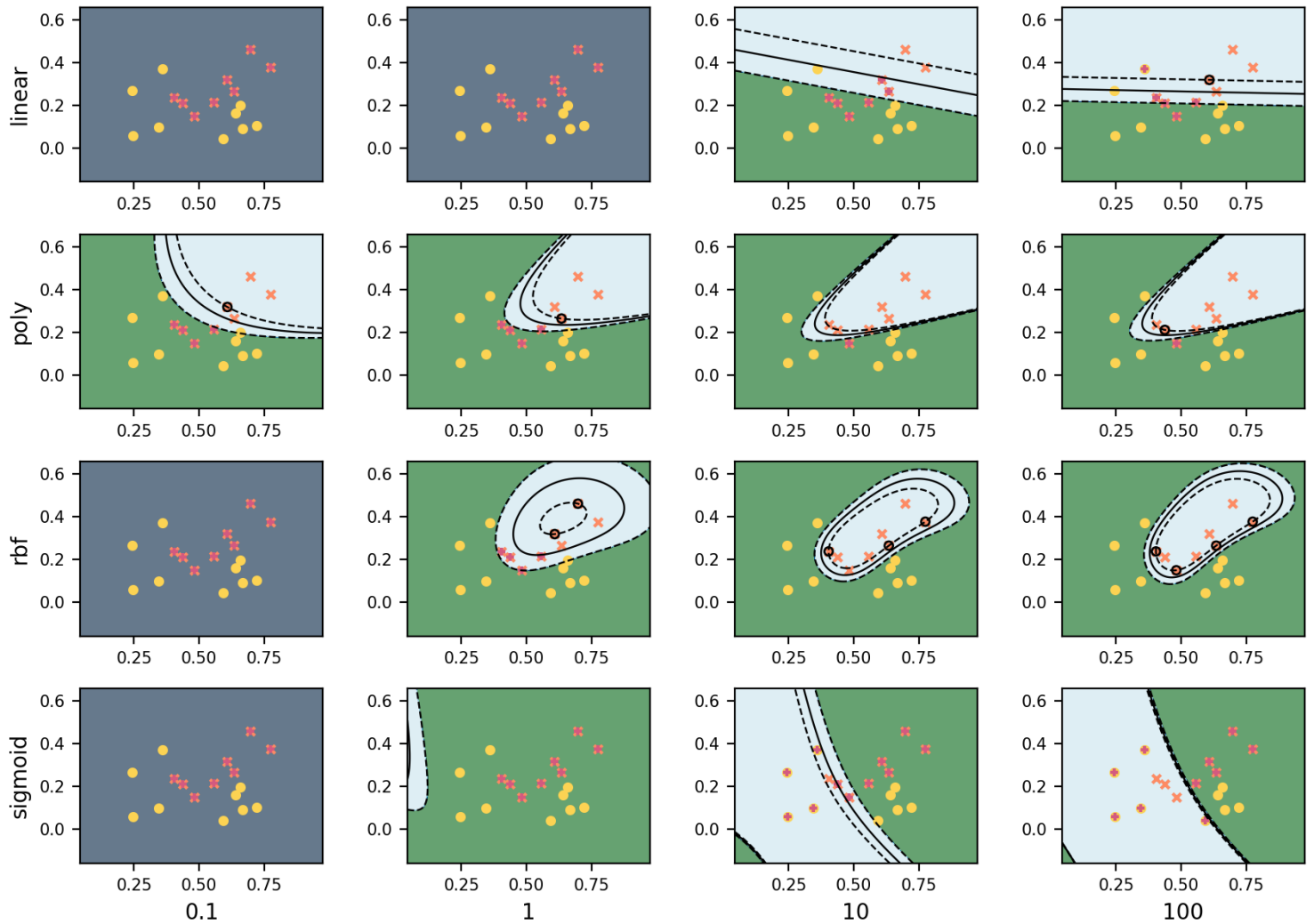
[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为：

[1 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0]

准确率为0.647059

2) SVM



核函数为: linear

惩罚因子为: 0.1

真实标签为:

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为:

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

准确率为0.529412

不存在1的正确分类

核函数为: linear

惩罚因子为: 10

真实标签为:

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为:

[1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0]

准确率为0.705882

精确度为0.800000

召回率为0.500000

核函数为: linear

惩罚因子为: 1

真实标签为:

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为:

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

准确率为0.529412

不存在1的正确分类

核函数为: linear

惩罚因子为: 100

真实标签为:

[1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]

预测结果为:

[1 1 1 1 1 1 0 0 0 0 1 0 0 0 1 0 1]

准确率为0.823529

精确度为0.777778

召回率为0.875000

核函数为: poly
惩罚因子为: 0.1
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
准确率为0.705882
精确度为0.800000
召回率为0.500000

核函数为: poly
惩罚因子为: 10
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0]
准确率为0.941176
精确度为1.000000
召回率为0.875000

核函数为: rbf
惩罚因子为: 0.1
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
准确率为0.529412
不存在1的正确分类
召回率为0.000000

核函数为: rbf
惩罚因子为: 10
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
准确率为1.000000
精确度为1.000000
召回率为1.000000

核函数为: poly
惩罚因子为: 1
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
准确率为0.823529
精确度为1.000000
召回率为0.625000

核函数为: poly
惩罚因子为: 100
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0]
准确率为0.941176
精确度为1.000000
召回率为0.875000

核函数为: rbf
惩罚因子为: 1
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 1 0 0 0 1 0 0 0 0 1 1 0 0]
准确率为0.941176
精确度为0.888889
召回率为1.000000

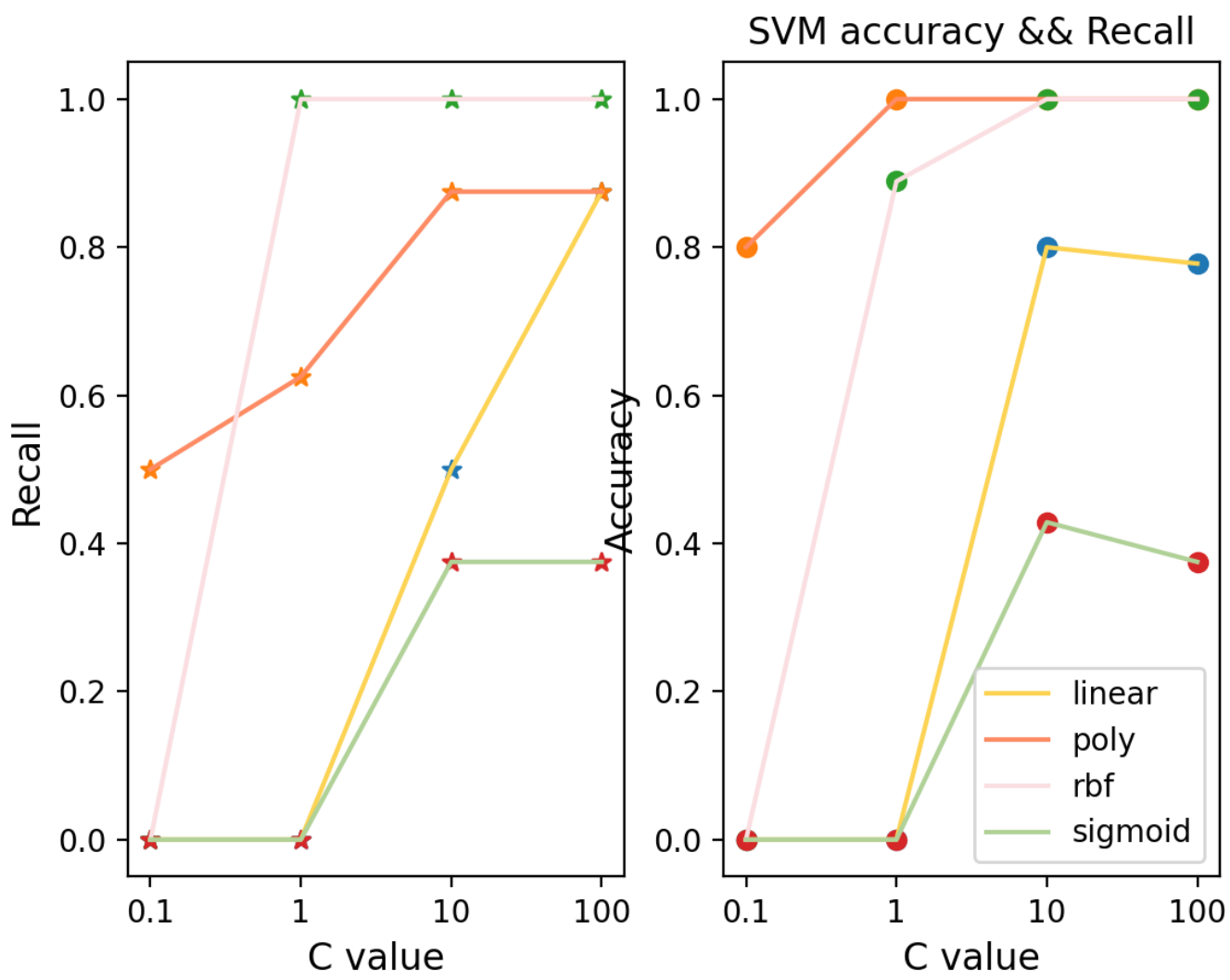
核函数为: rbf
惩罚因子为: 100
真实标签为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
预测结果为:
[1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
准确率为1.000000
精确度为1.000000
召回率为1.000000

核函数为: sigmoid
 惩罚因子为: 0.1
 真实标签为:
 [1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]
 预测结果为:
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 准确率为0.529412
 不存在1的正确分类
 召回率为0.000000

核函数为: sigmoid
 惩罚因子为: 1
 真实标签为:
 [1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
 预测结果为:
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 准确率为0.529412
 不存在1的正确分类
 召回率为0.000000

核函数为: sigmoid
 惩罚因子为: 10
 真实标签为:
 [1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]
 预测结果为:
 [0 0 0 0 0 1 1 1 0 0 1 0 0 1 1 0 1]
 准确率为0.470588
 精确度为0.428571
 召回率为0.375000

核函数为: sigmoid
 惩罚因子为: 100
 真实标签为:
 [1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0]
 预测结果为:
 [0 0 0 0 0 1 1 1 0 0 1 1 0 1 1 0 1]
 准确率为0.411765
 精确度为0.375000
 召回率为0.375000



三、结果分析

1) 正则化系数C对于结果的影响

通过SVM和LR的“Accuracy&&Recall”图，可以看出，随着惩罚因子C的增大，SVM+每一种核，以及LR的召回率都上升，SVM+poly，SVM+RBF的准确率也上升，但是SVM+linear,SVM+sigmod,LR的准确率在C:0-10上升，C大于10后下降，这可能是由于数据集非线性，对于非线性分类器（LR，SVM+linear）加大惩罚项也没有办法很好的提升分类效果。

2) 不同核函数对于结果的影响

通过SVM的“Accuracy&&Recall”图，固定C=10，可以看出RBF的分类效果是最好的：recall=1，accuracy=1，其次为poly，recall=0.875，accuracy=1。而linear的分类效果不好，accuracy= 0.8,recall=0.5。

效果最差的为sigmod，recall=0.375，accuracy=0.428，LR的分类效果也不理想，recall=0.5，accuracy=0.644。

并且SVM+RBF的非界限为非线性，SVM+linear 为直线，LR也为直线，这也与核函数的理论推导相一致。

在本题中，对于线性不可分的数据集，RBF的表现优于其他核函数以及LR，基于线性的分类器的表现均不理想。RBF也最常被用于核函数的先验估计。