法律声明

- 本课件包括演示文稿、示例、代码、题库、视频和声音等内容，小象学院和主讲老师拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意及内容，我们保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

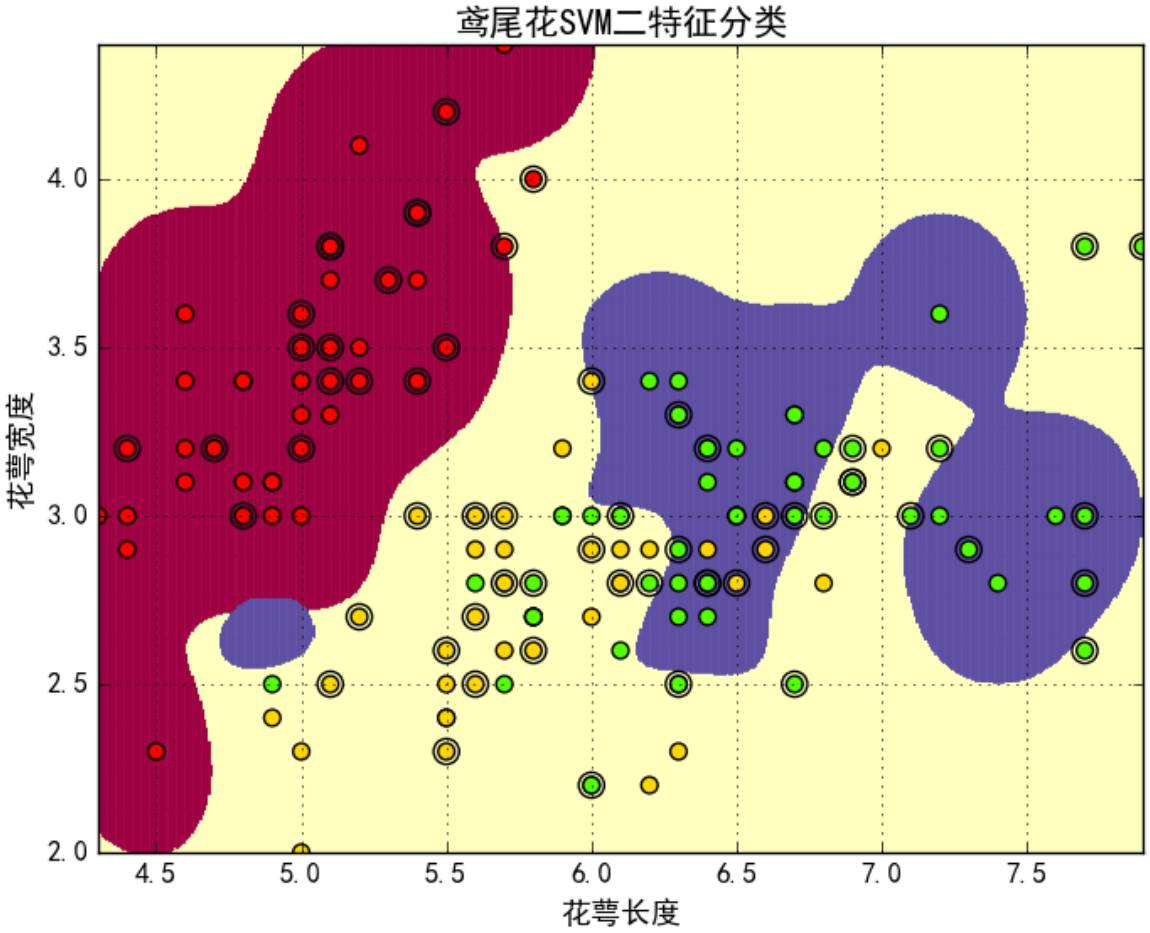
支持向量机实践

邹博

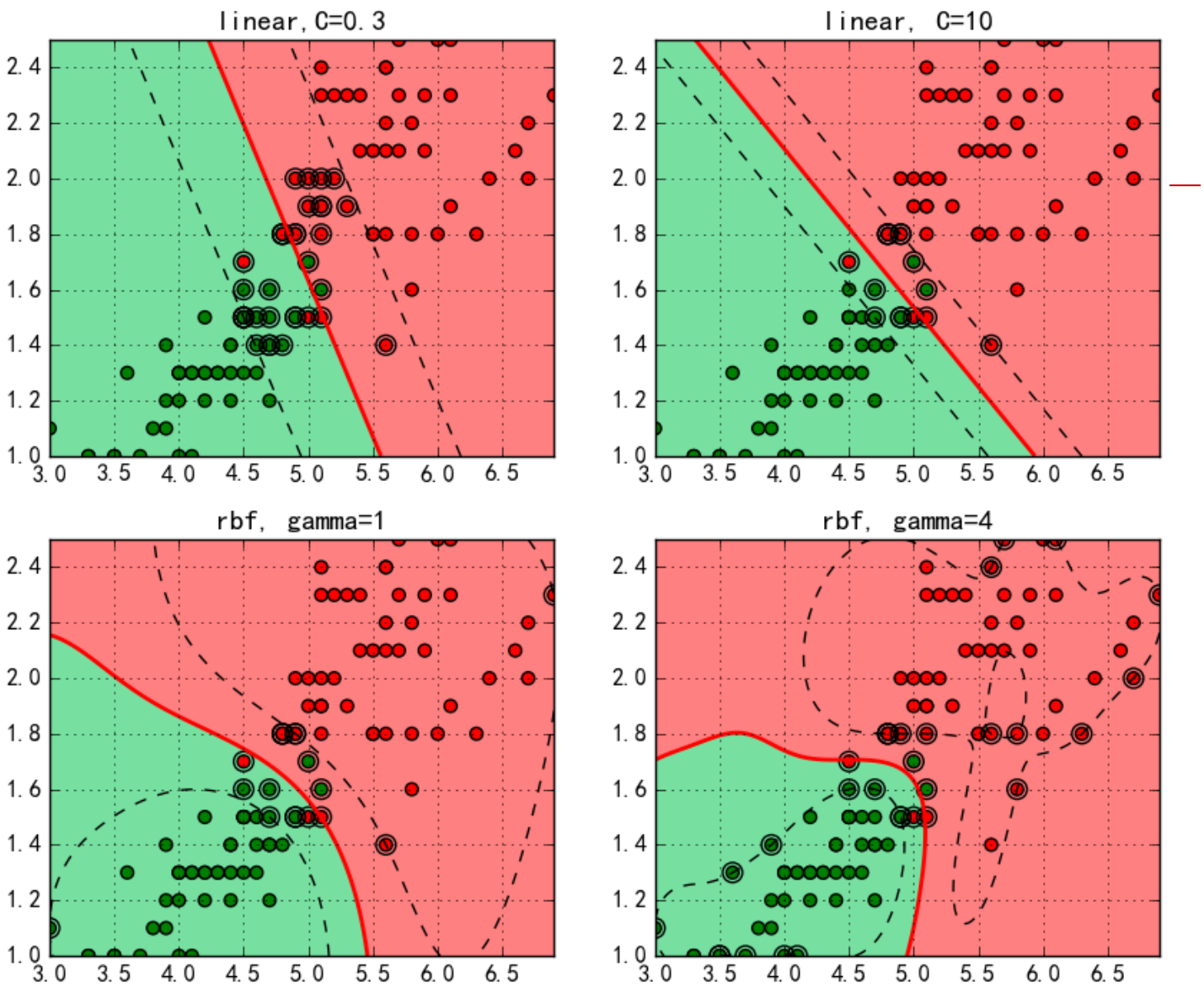
主要内容

- SVM包的使用
- SVM的调参
- 不平衡数据的处理
- SVM用于手写体数字分类
- 支持向量回归：SVR

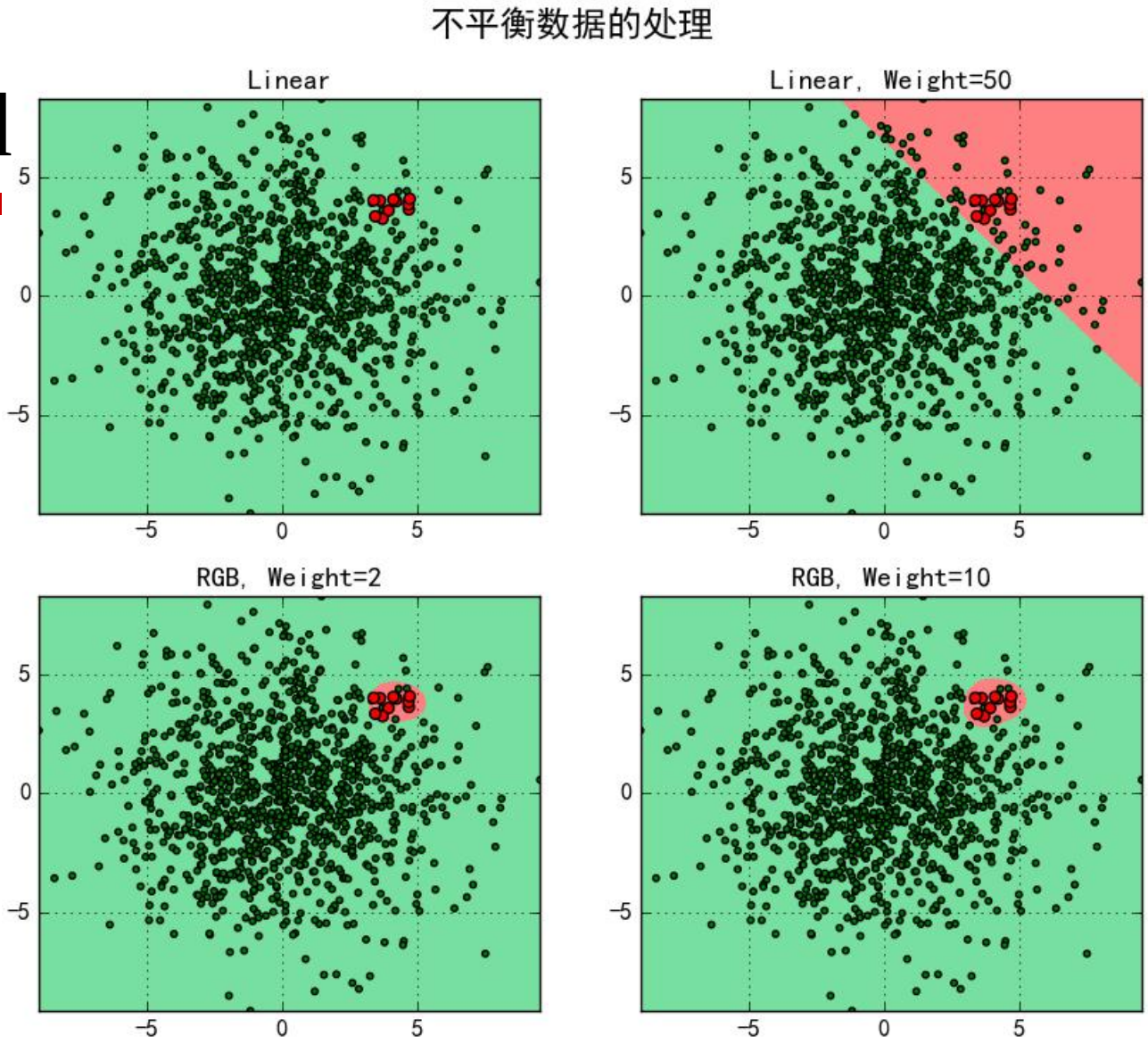
SVM分类



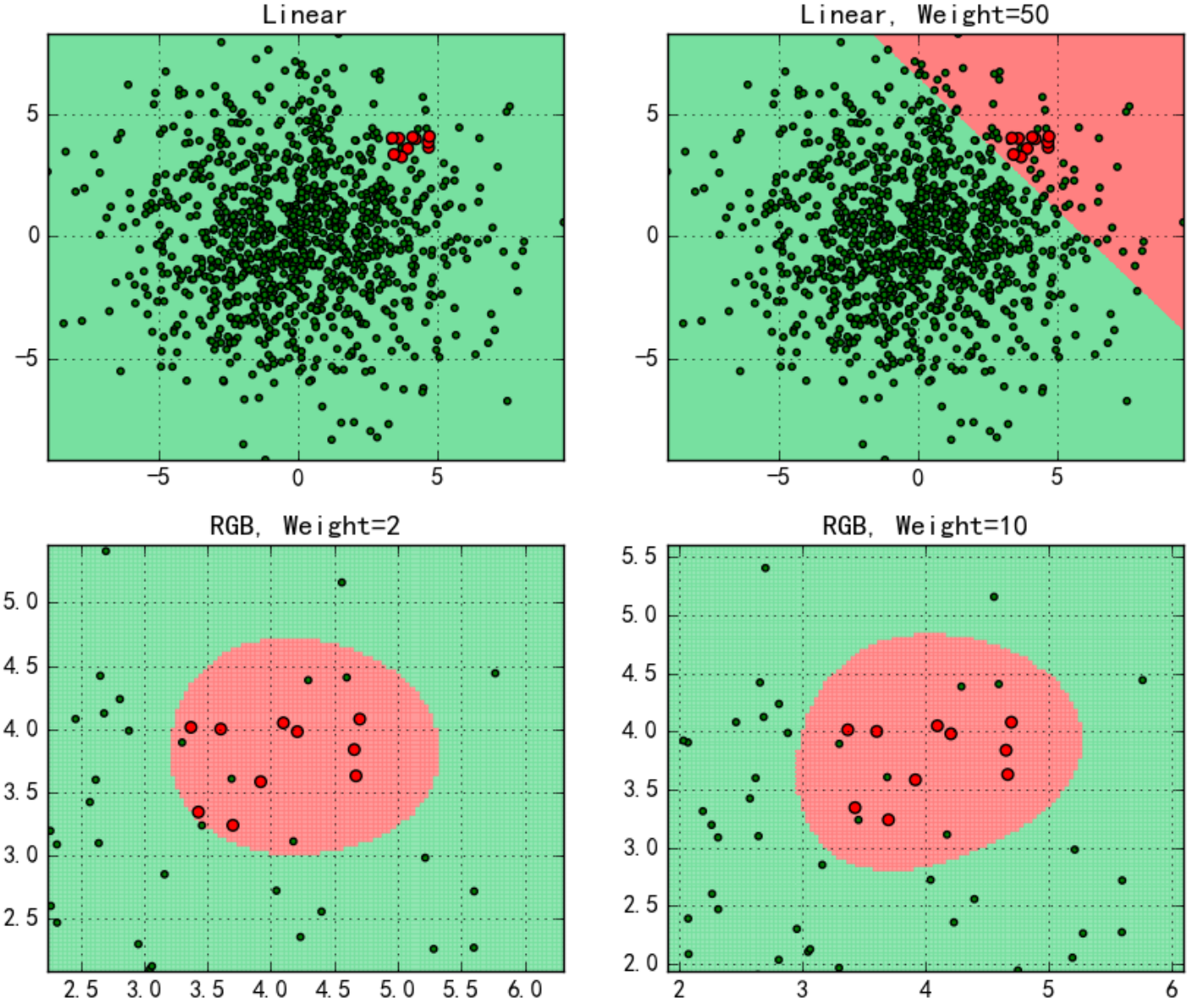
调参



Unbalanced


不平衡数据的处理
Linear
Linear, Weight=50
RGB, Weight=2
RGB, Weight=10
5
0
−5
−5
0
5

放大

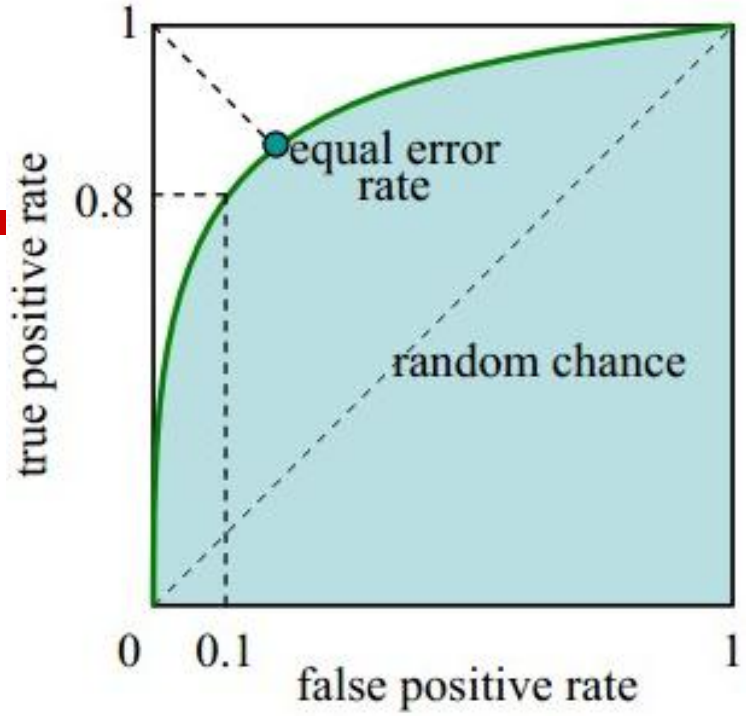


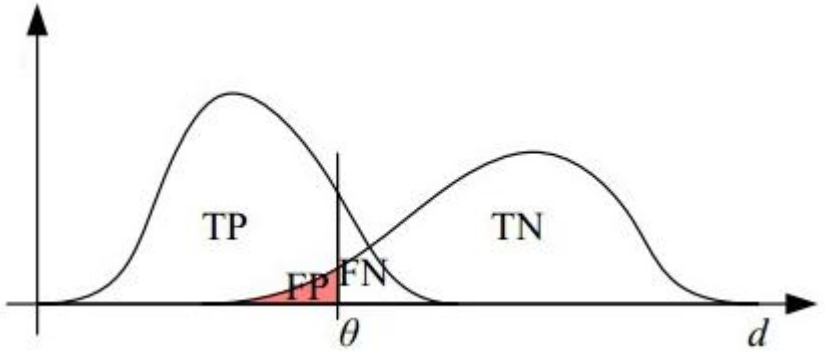
复习：AUC

实际值 \ 预测值	Positive	Negtive
正	TP	FN
负	FP	TN

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$





Receiver Operating Characteristic

Area Under Curve

分类器指标

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

$$F_\beta = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}$$

预测值 / 实际值	Positive	Negtive
正	TP	FN
负	FP	TN

计算

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

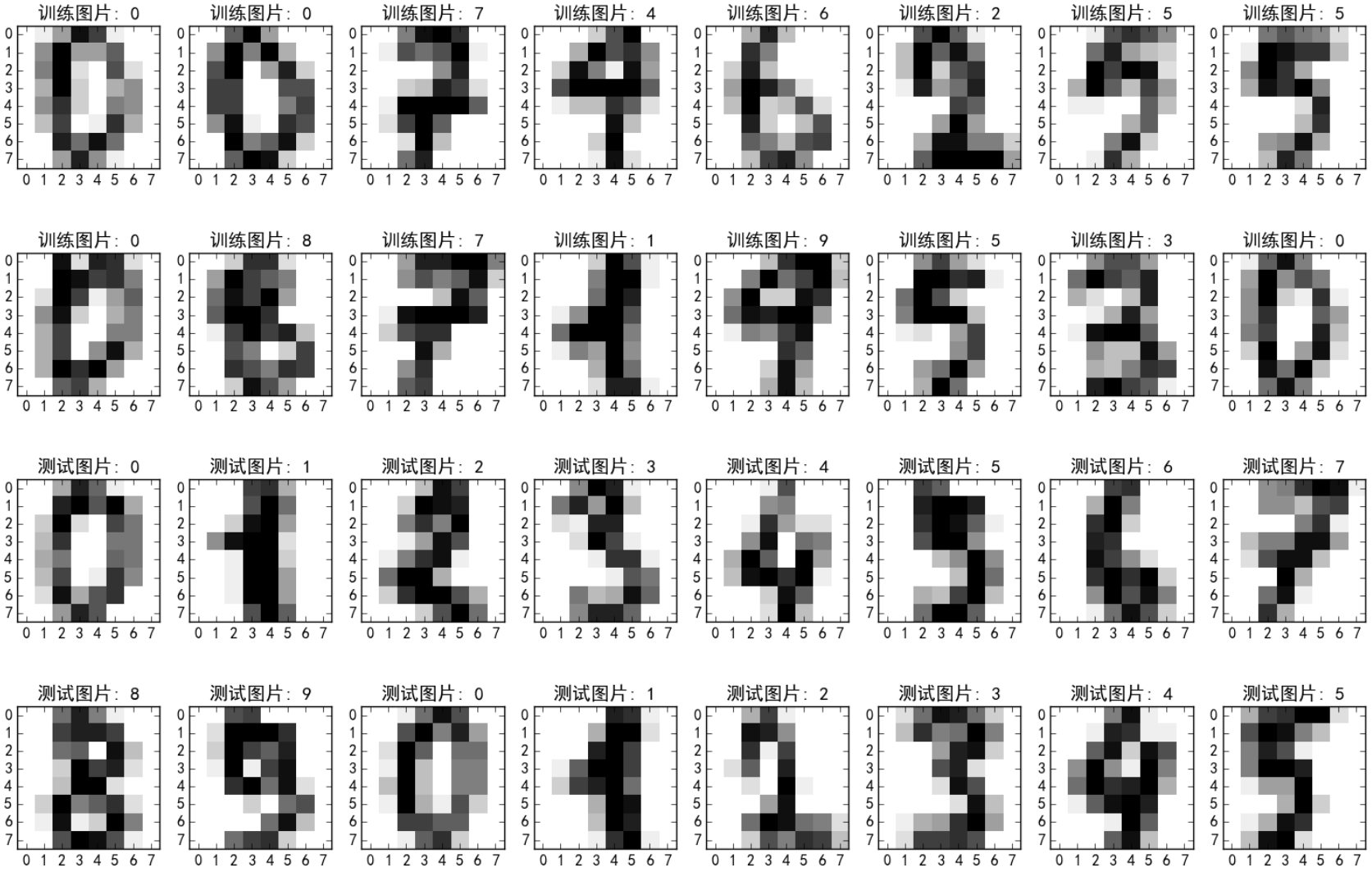
$$F_\beta = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}$$

```
y_true = np.array([1, 1, 1, 1, 0, 0])
y_hat  = np.array([1, 0, 1, 1, 1, 1])
```

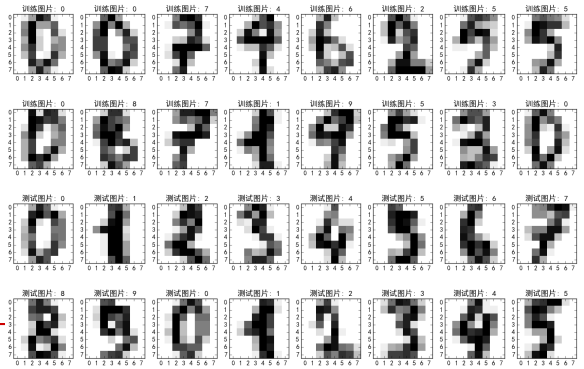
ClassifierIndex

```
Accuracy:   0.5
Precision:  0.6
Recall:     0.75
f1 score:   0.666666666667
F-beta:
    beta=    0.001  F-beta=0.60000
    beta=    0.010  F-beta=0.60001
    beta=    0.100  F-beta=0.60119
    beta=    1.000  F-beta=0.66667
    beta=   10.000  F-beta=0.74815
    beta=  100.000  F-beta=0.74998
    beta= 1000.000  F-beta=0.75000
```

SVM用于手写图片识别

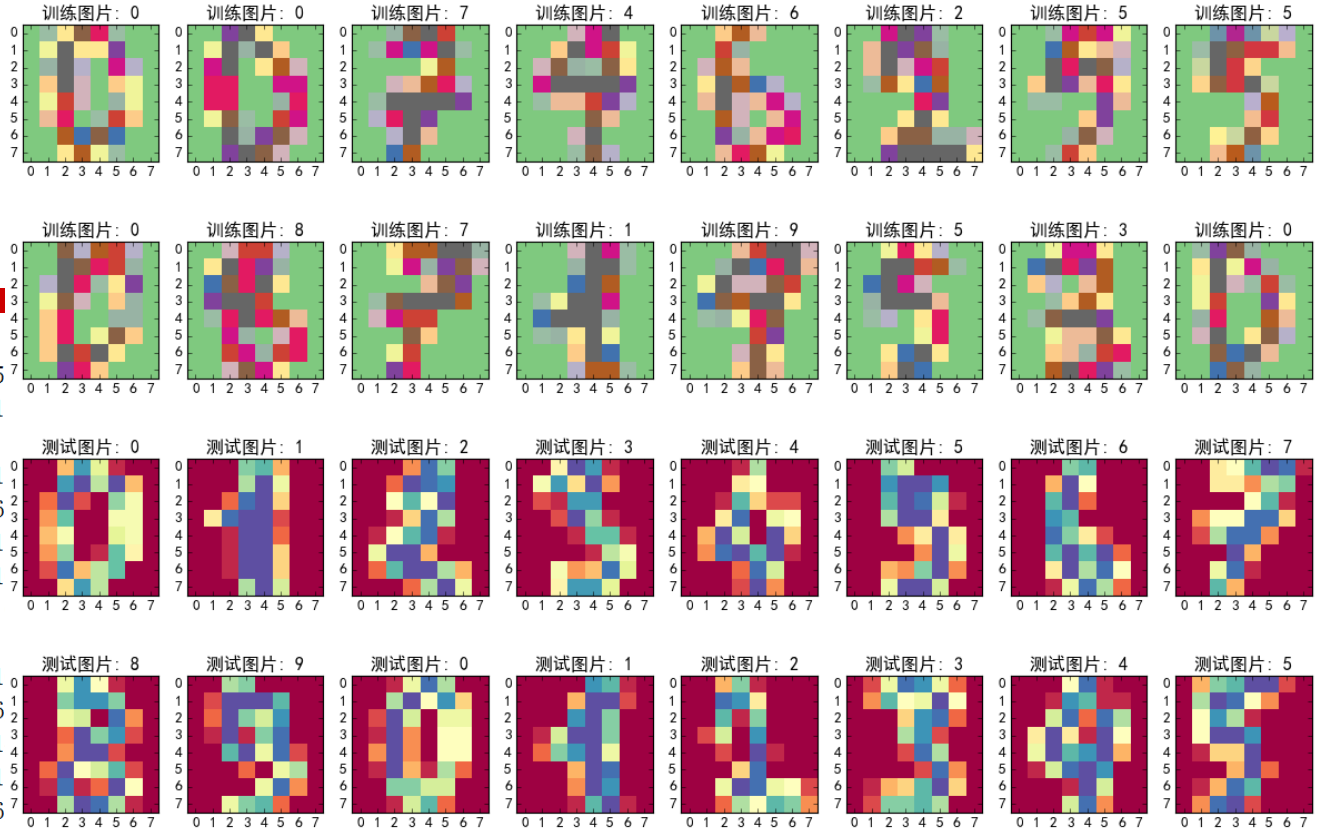


数据描述



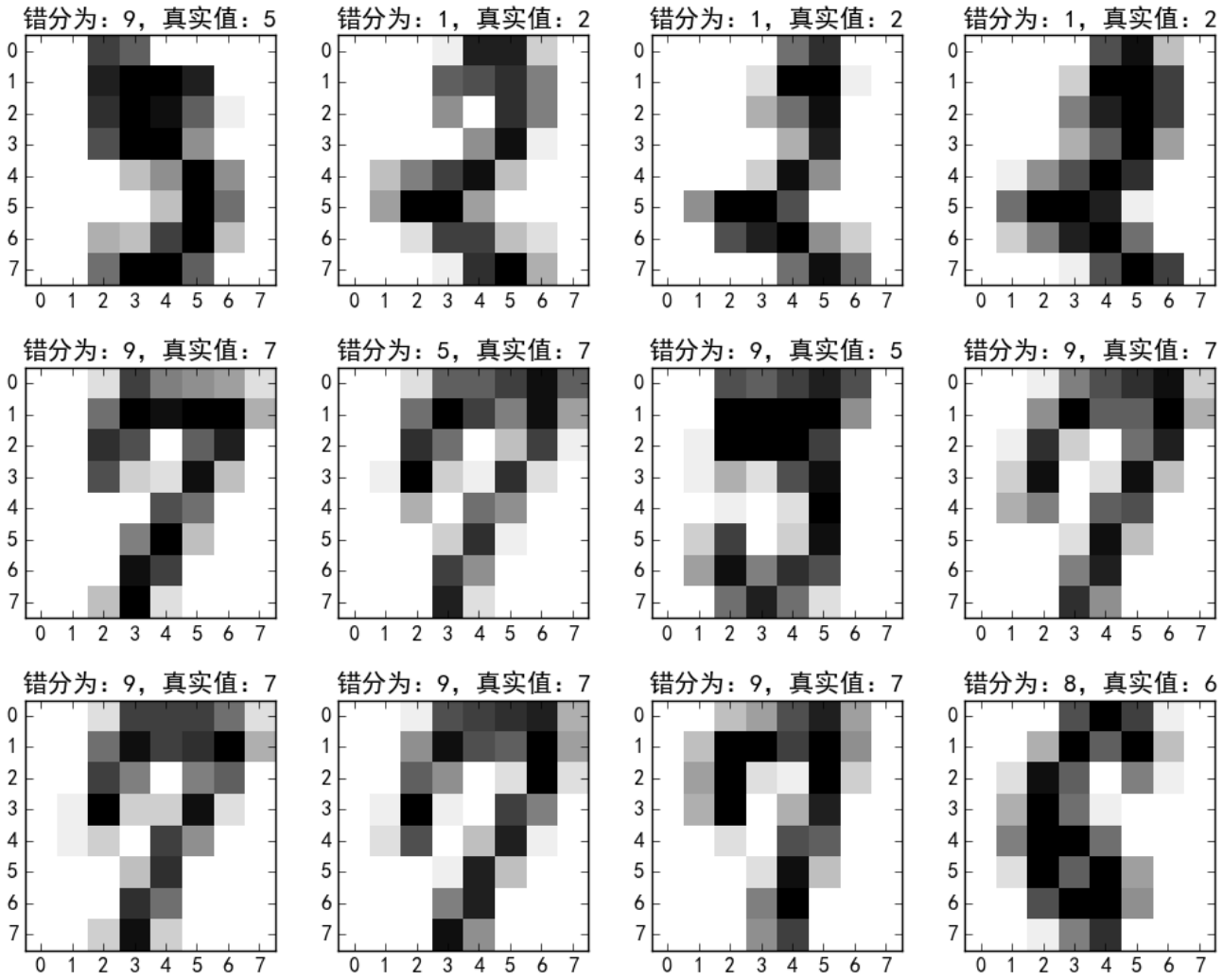
- 该数据来自于43人的手写数字，其中30人用于训练，13人用于测试，训练集共3823个图片，测试集共1797个图片，每个图片为8×8的灰度图像，像素值从0到16，其中，16代表全黑，0代表全亮(与通常的像素亮度习惯正好相反)
- 该数据的下载地址为：
  - http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits

数据存取

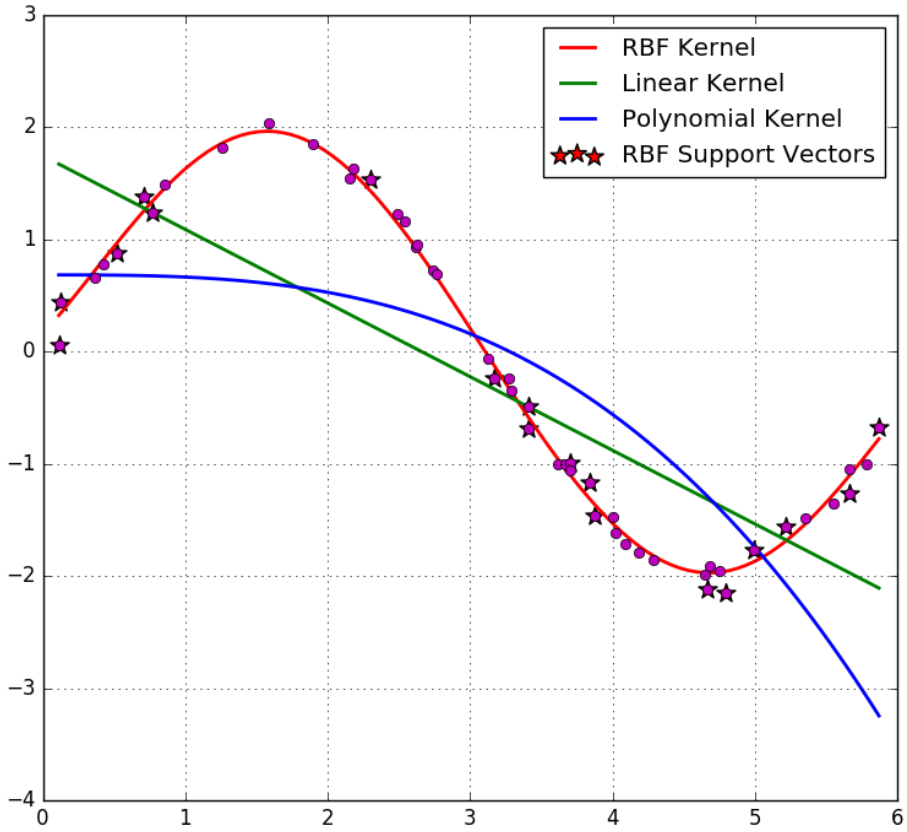


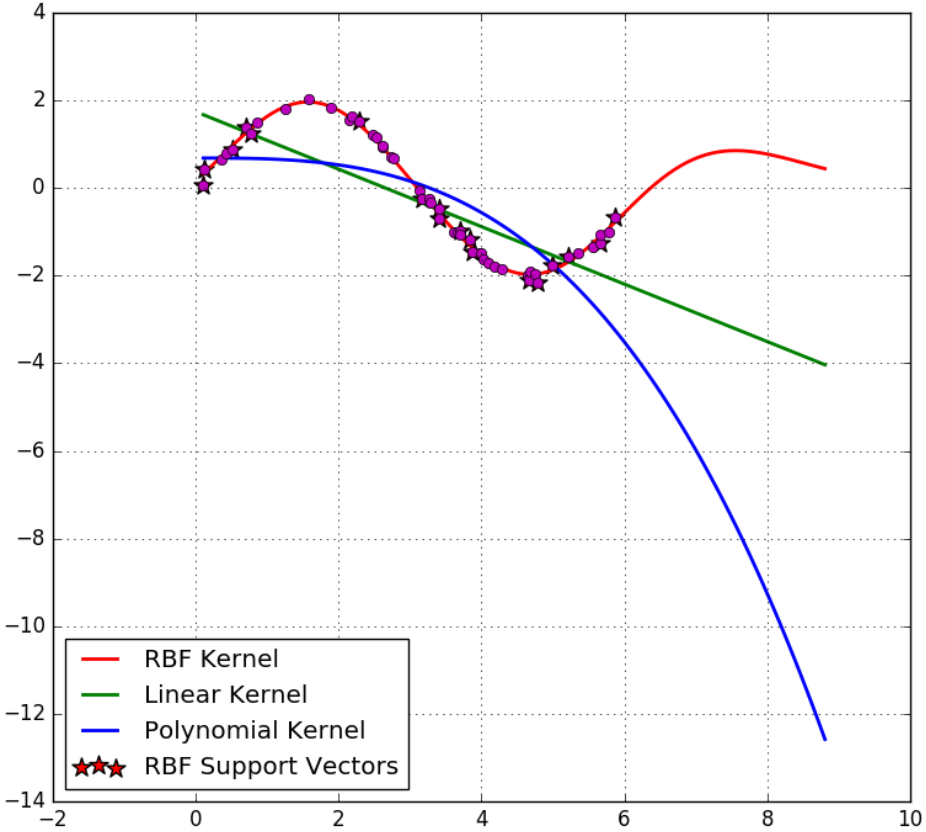
0, 0, 5, 13, 9, 1, 0, 0, 0, 0, 13, 15, 10, 15, 5, 0, 0, 3, 15
0, 0, 0, 12, 13, 5, 0, 0, 0, 0, 0, 11, 16, 9, 0, 0, 0, 0, 3, 1
0, 0, 0, 4, 15, 12, 0, 0, 0, 0, 3, 16, 15, 14, 0, 0, 0, 0, 8,
0, 0, 7, 15, 13, 1, 0, 0, 0, 8, 13, 6, 15, 4, 0, 0, 0, 2, 1, 1
0, 0, 0, 1, 11, 0, 0, 0, 0, 0, 0, 7, 8, 0, 0, 0, 0, 0, 1, 13, 6
0, 0, 12, 10, 0, 0, 0, 0, 0, 0, 14, 16, 16, 14, 0, 0, 0, 0, 1
0, 0, 0, 12, 13, 0, 0, 0, 0, 0, 5, 16, 8, 0, 0, 0, 0, 0, 13, 1
0, 0, 7, 8, 13, 16, 15, 1, 0, 0, 7, 7, 4, 11, 12, 0, 0, 0, 0,
0, 0, 9, 14, 8, 1, 0, 0, 0, 0, 12, 14, 14, 12, 0, 0, 0, 0, 9,
0, 0, 11, 12, 0, 0, 0, 0, 0, 2, 16, 16, 16, 13, 0, 0, 0, 3, 1
0, 0, 1, 9, 15, 11, 0, 0, 0, 0, 11, 16, 8, 14, 6, 0, 0, 2, 16
0, 0, 0, 0, 14, 13, 1, 0, 0, 0, 0, 5, 16, 16, 2, 0, 0, 0, 0, 1
0, 0, 5, 12, 1, 0, 0, 0, 0, 0, 15, 14, 7, 0, 0, 0, 0, 0, 13, 1
0, 2, 9, 15, 14, 9, 3, 0, 0, 4, 13, 8, 9, 16, 8, 0, 0, 0, 0, 6
0, 0, 0, 8, 15, 1, 0, 0, 0, 0, 1, 14, 13, 1, 1, 0, 0, 0, 10, 15, 3, 15, 11, 0, 0, 7, 16, 7, 1, 16, 8, 0, 0, 9, 16, 13, 14, 16, 5, 0, 0, 1, 10, 15, 16, 14, 0, 0, 0, 0, 0, 1, 16, 10, 0, 0, 0, 0, 0, 0, 10, 15, 4, 0, 0, 4
0, 5, 12, 13, 16, 16, 2, 0, 0, 11, 16, 15, 8, 4, 0, 0, 0, 8, 14, 11, 1, 0, 0, 0, 0, 8, 16, 16, 14, 0, 0, 0, 0, 1, 6, 6, 16, 0, 0, 0, 0, 0, 0, 5, 16, 3, 0, 0, 0, 1, 5, 15, 13, 0, 0, 0, 0, 4, 15, 16, 2, 0, 0, 0, 5
0, 0, 0, 8, 15, 1, 0, 0, 0, 0, 0, 12, 14, 0, 0, 0, 0, 0, 3, 16, 7, 0, 0, 0, 0, 0, 6, 16, 2, 0, 0, 0, 0, 0, 7, 16, 16, 13, 5, 0, 0, 0, 15, 16, 9, 9, 14, 0, 0, 0, 3, 14, 9, 2, 16, 2, 0, 0, 0, 7, 15, 16, 11, 0, 6
0, 0, 1, 8, 15, 10, 0, 0, 0, 3, 13, 15, 14, 14, 0, 0, 0, 5, 10, 0, 10, 12, 0, 0, 0, 0, 3, 5, 15, 10, 2, 0, 0, 0, 16, 16, 16, 16, 12, 0, 0, 1, 8, 12, 14, 8, 3, 0, 0, 0, 0, 10, 13, 0, 0, 0, 0, 0, 0, 11, 9, 0, 0, 0, 7
0, 0, 10, 7, 13, 9, 0, 0, 0, 0, 9, 10, 12, 15, 2, 0, 0, 0, 4, 11, 10, 11, 0, 0, 0, 0, 1, 16, 10, 1, 0, 0, 0, 0, 12, 13, 4, 0, 0, 0, 0, 0, 12, 1, 12, 0, 0, 0, 0, 1, 10, 2, 14, 0, 0, 0, 0, 0, 11, 14, 5, 0, 0, 0, 8
0, 0, 6, 14, 4, 0, 0, 0, 0, 0, 11, 16, 10, 0, 0, 0, 0, 0, 8, 14, 16, 2, 0, 0, 0, 0, 1, 12, 12, 11, 0, 0, 0, 0, 0, 0, 0, 11, 3, 0, 0, 0, 0, 0, 5, 11, 0, 0, 0, 1, 4, 4, 7, 16, 2, 0, 0, 7, 16, 16, 13, 11, 1, 9
0, 0, 3, 13, 11, 7, 0, 0, 0, 0, 11, 16, 16, 16, 2, 0, 0, 4, 16, 9, 1, 14, 2, 0, 0, 4, 16, 0, 0, 16, 2, 0, 0, 0, 16, 1, 0, 12, 8, 0, 0, 0, 15, 9, 0, 13, 6, 0, 0, 0, 9, 14, 9, 14, 1, 0, 0, 0, 2, 12, 13, 4, 0, 0, 0
0, 0, 0, 2, 16, 16, 2, 0, 0, 0, 0, 4, 16, 16, 2, 0, 0, 1, 4, 12, 16, 12, 0, 0, 0, 7, 16, 16, 16, 12, 0, 0, 0, 0, 3, 10, 16, 14, 0, 0, 0, 0, 0, 8, 16, 12, 0, 0, 0, 0, 0, 6, 16, 16, 2, 0, 0, 0, 0, 2, 12, 15, 4, 0, 1
0, 0, 8, 16, 5, 0, 0, 0, 0, 1, 13, 11, 16, 0, 0, 0, 0, 0, 10, 0, 13, 3, 0, 0, 0, 0, 3, 1, 16, 1, 0, 0, 0, 0, 0, 9, 12, 0, 0, 0, 0, 0, 3, 15, 5, 0, 0, 0, 0, 0, 14, 15, 8, 8, 3, 0, 0, 0, 7, 12, 12, 12, 13, 1, 2
0, 1, 8, 12, 15, 14, 4, 0, 0, 3, 11, 8, 8, 12, 12, 0, 0, 0, 0, 0, 2, 13, 7, 0, 0, 0, 0, 2, 15, 12, 1, 0, 0, 0, 0, 0, 13, 5, 0, 0, 0, 0, 0, 0, 9, 13, 0, 0, 0, 0, 7, 8, 14, 15, 0, 0, 0, 0, 14, 15, 11, 2, 0, 0, 3
0, 0, 0, 0, 12, 2, 0, 0, 0, 0, 0, 6, 14, 1, 0, 0, 0, 0, 4, 16, 7, 8, 0, 0, 0, 0, 13, 9, 0, 16, 6, 0, 0, 6, 16, 10, 11, 16, 0, 0, 0, 0, 5, 10, 13, 16, 0, 0, 0, 0, 0, 0, 6, 16, 0, 0, 0, 0, 0, 0, 12, 8, 0, 0, 4

训练测试正确率：99.82% - 98.27%



SVR – 预测


RBF Kernel
Linear Kernel
Polynomial Kernel
RBF Support Vectors


RBF Kernel
Linear Kernel
Polynomial Kernel
RBF Support Vectors

Demo

我们在这里

- http://wenda.ChinaHadoop.cn
  - 视频/课程/社区
- 微博
  - @ChinaHadoop
  - @邹博_机器学习
- 微信公众号
  - 小象
  - 大数据分析挖掘

感谢大家！

恳请大家批评指正！