

机器学习作业第六章

161910126 赵安

课后习题6.2

基于西瓜数据集，使用LIBSVM分别用线性核与高斯核训练一个SVM

```
1  from libsvm.svmutil import *
2
3  y, x = svm_read_problem('data.txt')
4  porblem = svm_problem(y, x)
5
6  print("线性核:")
7  param1 = svm_parameter('-t 0 -c 50 -b 1')
8  model1 = svm_train(porblem, param1)
9  p_label1, p_acc1, p_val1 = svm_predict(y, x, model1)
10 print(p_label1)
11 print(p_acc1)
12 print(p_val1)
13
14
15 print("高斯核:")
16 param2 = svm_parameter('-t 2 -c 50 -b 1')
17 model2 = svm_train(porblem, param2)
18 p_label2, p_acc2, p_val2 = svm_predict(y, x, model2)
19 print(p_label2)
20 print(p_acc2)
21 print(p_val2)
22
```

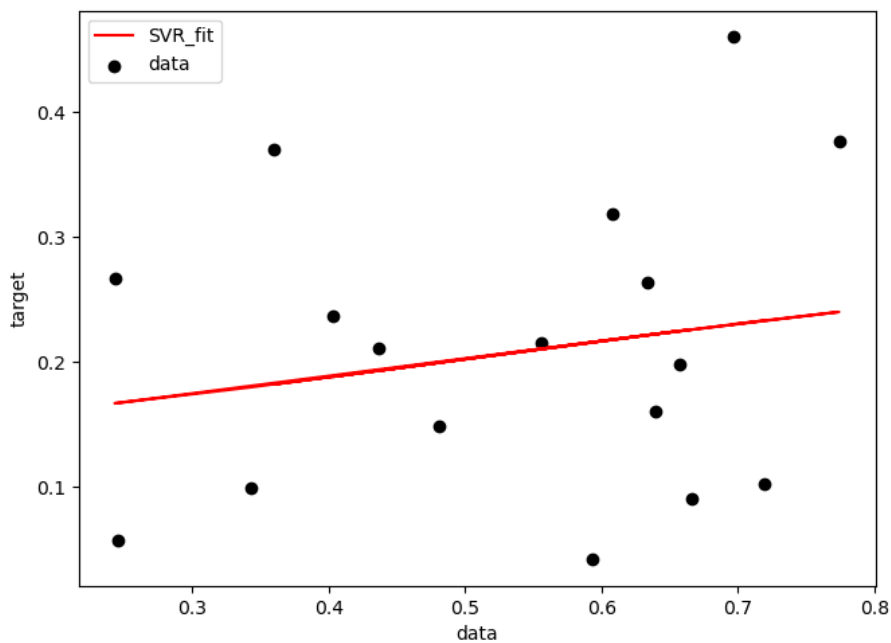
	线性核	高斯核
迭代次数	20	28
参数	0.694176	0.687546
判决函数的偏置项b	1.952743	4.589314
支持向量个数	13	13
准确率	82.3529%	88.2353%

由于数据集太小，二者在支持向量个数与准确率差别不大

课后习题6.8

```
1 from sklearn.svm import SVR
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # 数据准备
6 x = np.loadtxt('Input.txt').reshape(-1, 1)
7 y = np.loadtxt('Output.txt')
8
9 # 自动选择合适的参数
10 svr = SVR(kernel='rbf', degree=3, gamma='auto', coef0=0, C=0.5)
11 svr.fit(x, y)
12
13 gauss_svr = svr.fit(x, y)
14 y_pred = gauss_svr.predict(x)
15
16 plt.scatter(x, y, c='k', label='data', zorder=1)
17 plt.plot(x, y_pred, c='r', label='SVR_fit')
18 plt.xlabel('data')
19 plt.ylabel('target')
20 plt.legend()
21 plt.show()
22
```

该SVR基于高斯核训练，效果如下，使用其他核函数训练结果大致无异



附加题1

```
1 from sklearn.svm import SVC
2 import numpy as np
3 import pandas as pd
4 import matplotlib.pyplot as plt
5
6 # 数据准备
7 x_train = np.genfromtxt('E:\Machine
  Learning\XIGUA\C5_NN\train_feature.csv', delimiter=',')
8 y_train = np.genfromtxt('E:\Machine Learning\XIGUA\C5_NN\train_target.csv',
  delimiter=',')
9 x_test = np.genfromtxt('E:\Machine Learning\XIGUA\C5_NN\test_feature.csv',
  delimiter=',')
10
11
12 # 使用高斯核训练
13 model1 = SVC(kernel='rbf', degree=3, gamma='auto', coef0=0, C=0.5)
14 model1.fit(x_train, y_train)
15
16 gauss_svc = model1.fit(x_train, y_train)
17 y_pred1 = gauss_svc.predict(x_test)
18
19 y_predict1 = []
20 threshold = 0.5
21 for i in y_pred1:
22     if float(i) <= threshold:
23         y_predict1.append(0)
24     else:
25         y_predict1.append(1)
26 test_target = pd.DataFrame(data=y_predict1)
27 print(y_predict1)
28 #保存在test_target_rbf_SVM.csv文件中
29 test_target.to_csv('test_target_rbf_SVM.csv', index=False, encoding='gbk')
30
31 # 使用线性核训练
32 model2 = SVC(kernel='linear', degree=3, gamma='auto', coef0=0, C=0.5)
33 model2.fit(x_train, y_train)
34
35 linear_svc = model2.fit(x_train, y_train)
36 y_pred2 = linear_svc.predict(x_test)
37
38 y_predict2 = []
39 threshold = 0.5
40 for i in y_pred2:
41     if float(i) <= threshold:
42         y_predict2.append(0)
43     else:
44         y_predict2.append(1)
45 test_target = pd.DataFrame(data=y_predict2)
46 print(y_predict2)
47 #保存在test_target_linear_SVM.csv文件中
48 test_target.to_csv('test_target_linear_SVM.csv', index=False,
  encoding='gbk')
```

同样选取0.5为阈值

```
1 #神经网络训练结果
2 [0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1,
3 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1,
4 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1,
5 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1]
6
7 #高斯核SVM训练结果
8 [0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1,
9 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0,
10 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1,
11 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1]
12
13 #线性核SVM训练结果
14 [0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1,
1 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0,
2 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1,
3 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1]
```

附加题2

题目叙述，线性支持向量机还可以定义为以下形式：

$$\begin{aligned} \min_{w,b,\zeta} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \zeta_i^2 \\ \text{s.t. } & y_i(w \cdot x_i + b) \geq 1 - \zeta_i, \forall i = 1, 2, \dots, N \\ & \zeta_i \geq 0, \forall i = 1, 2, \dots, N \end{aligned}$$

试求其对偶形式

解：

加入拉格朗日算子

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \epsilon_i^2 - \sum_{i=1}^N \lambda_{i1} (1 - \epsilon_i - y_i(w x_i + b)) - \sum_{i=1}^N \lambda_{i2} \epsilon_i$$

对 w, b, ϵ_i 求偏导

得

$$w = \sum_{i=1}^N \lambda_{i1} y_i x_i$$

$$0=\sum_{i=1}^N\lambda_{i1}y_i$$

$$2C\epsilon_i=\lambda_{i1}+\lambda_{i2}$$

$$max_{\lambda_{i1},\lambda_{i2}}\sum_{i=1}^N\lambda_{i1}-\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\lambda_{i1}\lambda_{j1}\lambda_{i2}\lambda_{j2}x_i^Tx_j-\frac{1}{4C}\sum_{i=1}^N(\lambda_{i1}+\lambda_{i2})^2$$

$$s.t.\sum_{i=1}^N\lambda_{i1}\lambda_{i2}=0$$

$$\lambda_{i1}\geq 0,\lambda_{i2}\geq 0,i=1,2,3,\ldots,N$$