## 第一次作业

- 秦绪博
- 2017000621

## 数据集和模型概述

使用sklearn自带的数据包提供的手写体数字识别数据集load\_digits。输入digits.data.shape可以看到数据规模为(1797,64)。即一共有1797组数据,每幅图片由8X8=64的像素矩阵表示。

使用sklearn自带模块进行交叉验证数据切分,按照训练集:测试集=3:1的规模进行分割。

本次实验使用基于线性假设的支持向量机分类器LinearSVC,随机森林分类器RandomForestClassifier和 XGBOOST分类器XGBClassifier。

## 代码

```
from sklearn.datasets import load digits
# 从通过数据加载器获得手写体数字的数码图像数据并储存在digits变量中。
digits = load digits()
from sklearn.model selection import train test split
# 随机选取75%的数据作为训练样本; 其余25%的数据作为测试样本。
X_train, X_test, y_train, y_test = train_test_split(digits.data,
digits.target, test_size=0.25, random_state=33)
from sklearn.preprocessing import StandardScaler
# 从sklearn.svm里导入基于线性假设的支持向量机分类器LinearSVC。
from sklearn.svm import LinearSVC
# 从仍然需要对训练和测试的特征数据进行标准化。
ss = StandardScaler()
X train = ss.fit transform(X train)
X test = ss.transform(X test)
# 初始化线性假设的支持向量机分类器LinearSVC。
lsvc = LinearSVC()
#进行模型训练
lsvc.fit(X train, y train)
# 利用训练好的模型对测试样本的数字类别进行预测,预测结果储存在变量y predict中。
lsvc_y_predict = lsvc.predict(X_test)
print ('The Accuracy of Linear SVC is', lsvc.score(X test, y test))
from sklearn.metrics import classification report
print (classification_report(y_test, lsvc_y_predict))
from sklearn.ensemble import RandomForestClassifier
```

```
rfc=RandomForestClassifier()
rfc.fit(X_train,y_train)
rfc_y_predict=rfc.predict(X_test)

print ('The Accuracy of Random Forest is', rfc.score(X_test, y_test))
from sklearn.metrics import classification_report
print (classification_report(y_test, rfc_y_predict))

from xgboost import XGBClassifier
xgbc=XGBClassifier()
xgbc.fit(X_train,y_train)
xgb_y_pred = xgbc.predict(X_test)
xgb_predictions = [round(value) for value in xgb_y_pred]
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test, xgb_predictions)
print("The Accuracy of XGBOOST is: %.2f%%" % (accuracy * 100.0))
print (classification_report(y_test,xgb_predictions))
```

## 输出结果

使用sklearn自带的classification\_report模块输出报告:

## Precision recall f1-score support  ## 0	Th	e Accuracy			95333333333	
1 0.96 0.98 0.97 54 2 0.98 1.00 0.99 44 3 0.93 0.93 0.93 46 4 0.97 1.00 0.99 35 5 0.94 0.94 0.94 48 6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556			precision	recall	f1-score	support
1 0.96 0.98 0.97 54 2 0.98 1.00 0.99 44 3 0.93 0.93 0.93 46 4 0.97 1.00 0.99 35 5 0.94 0.94 0.94 48 6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556		0	0.92	1.00	0.96	35
2 0.98 1.00 0.99 44 3 0.93 0.93 0.93 46 4 0.97 1.00 0.99 35 5 0.94 0.94 0.94 48 6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.935555555556 precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35		1				
4 0.97 1.00 0.99 35 5 0.94 0.94 0.94 48 6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556		2				
5 0.94 0.94 0.94 48 6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556 precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 0.95 44 3 0.95 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35						
6 0.96 0.98 0.97 51 7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556		4	0.97	1.00	0.99	35
7 0.92 1.00 0.96 35 8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.93555555556     precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 0.95 44 3 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35		5	0.94	0.94	0.94	48
8 0.98 0.84 0.91 58 9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.935555555556     precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 0.95 44 3 0.95 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35		6	0.96	0.98	0.97	51
9 0.95 0.91 0.93 44  avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.935555555556     precision recall f1-score support  0 0.92 1.00 0.96 35     1 0.93 0.96 0.95 54     2 0.95 0.95 0.95 44     3 0.95 0.85 0.90 46     4 0.94 0.91 0.93 35     5 0.87 0.94 0.90 48     6 1.00 0.98 0.99 51     7 1.00 1.00 1.00 35		7	0.92	1.00	0.96	35
Avg / total 0.95 0.95 0.95 450  The Accuracy of Random Forest is 0.935555555556 precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35		8	0.98	0.84	0.91	58
The Accuracy of Random Forest is 0.93555555556     precision    recall f1-score    support  0     0.92     1.00     0.96     35 1     0.93     0.96     0.95     54 2     0.95     0.95     0.95     44 3     0.95     0.85     0.90     46 4     0.94     0.91     0.93     35 5     0.87     0.94     0.90     48 6     1.00     0.98     0.99     51 7     1.00     1.00     1.00     35		9	0.95	0.91	0.93	44
precision recall f1-score support  0 0.92 1.00 0.96 35 1 0.93 0.96 0.95 54 2 0.95 0.95 0.95 44 3 0.95 0.85 0.90 46 4 0.94 0.91 0.93 35 5 0.87 0.94 0.90 48 6 1.00 0.98 0.99 51 7 1.00 1.00 1.00 35	av	g / total	0.95	0.95	0.95	450
0       0.92       1.00       0.96       35         1       0.93       0.96       0.95       54         2       0.95       0.95       0.95       44         3       0.95       0.85       0.90       46         4       0.94       0.91       0.93       35         5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35	Th	e Accuracy	of Random	Forest is	0.9355555	5556
1       0.93       0.96       0.95       54         2       0.95       0.95       0.95       44         3       0.95       0.85       0.90       46         4       0.94       0.91       0.93       35         5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35			precision	recall	f1-score	support
2       0.95       0.95       0.95       44         3       0.95       0.85       0.90       46         4       0.94       0.91       0.93       35         5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35		0	0.92	1.00	0.96	35
3       0.95       0.85       0.90       46         4       0.94       0.91       0.93       35         5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35		1	0.93	0.96	0.95	54
4       0.94       0.91       0.93       35         5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35		2	0.95	0.95	0.95	44
5       0.87       0.94       0.90       48         6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35		3	0.95	0.85	0.90	46
6       1.00       0.98       0.99       51         7       1.00       1.00       1.00       35		4	0.94	0.91	0.93	35
7 1.00 1.00 1.00 35		5	0.87	0.94	0.90	48
		6	1.00	0.98	0.99	51
8 0.90 0.93 0.92 58		7	1.00	1.00	1.00	35
		8	0.90	0.93	0.92	58

g	9 0	.93	0.84	0.88	44
avg / total	L 0	.94	0.94	0.94	450
The Accurac	cy of XGB	BOOST is:	94.89%		
	precis	ion r	ecall f	1-score	support
(	9 0	.92	1.00	0.96	35
1	L 0	.96	0.98	0.97	54
1	2 0	.98	0.93	0.95	44
	3 0	.98	0.91	0.94	46
4	1 0	.97	0.94	0.96	35
ŗ	5 0	.96	0.90	0.92	48
(	5 0	.96	0.96	0.96	51
<del>,</del>	7 0	.92	0.97	0.94	35
{	3 0	.93	0.95	0.94	58
9	9 0	.91	0.95	0.93	44
avg / total	L 0	.95	0.95	0.95	450