

Random Forest

68_Adnan Shaikh

```
In [1]: import pandas as pd  
        from sklearn.datasets import load_digits  
        digits = load_digits()
```

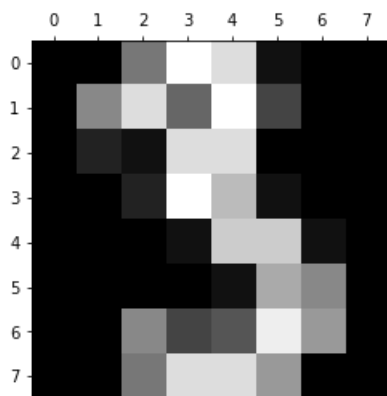
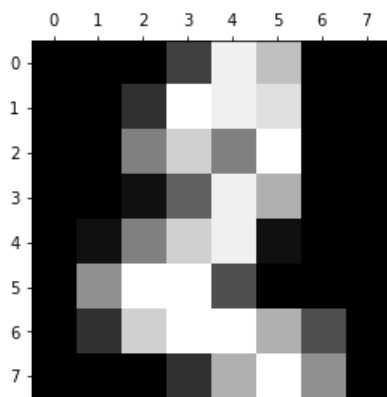
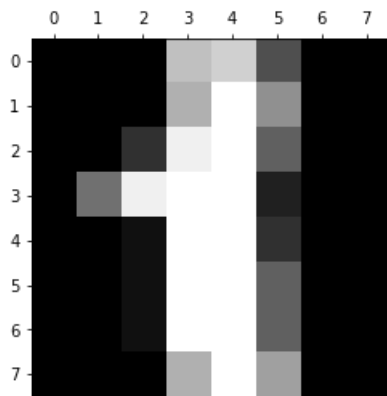
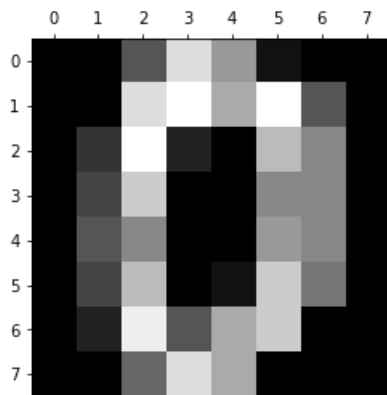
```
In [2]: dir(digits)
```

```
Out[2]: ['DESCR', 'data', 'feature_names', 'frame', 'images', 'target', 'target_names']
```

```
In [3]: import matplotlib.pyplot as plt
```

```
In [4]: plt.gray()
for i in range(4):
    plt.matshow(digits.images[i])
```

<Figure size 432x288 with 0 Axes>



```
In [5]: df = pd.DataFrame(digits.data)
df.head()
```

Out[5]:

	0	1	2	3	4	5	6	7	8	9	...	54	55	56	57	58	59	60	61	62	63
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	6.0	13.0	10.0	0.0	0.0	0.0
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	11.0	16.0	10.0	0.0	0.0
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	0.0	0.0	0.0	...	5.0	0.0	0.0	0.0	0.0	3.0	11.0	16.0	9.0	0.0
3	0.0	0.0	7.0	15.0	13.0	1.0	0.0	0.0	0.0	8.0	...	9.0	0.0	0.0	0.0	7.0	13.0	13.0	9.0	0.0	0.0
4	0.0	0.0	0.0	1.0	11.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	2.0	16.0	4.0	0.0	0.0

5 rows × 64 columns

```
In [6]: df['target'] = digits.target
```

```
In [7]: df[0:12]
```

Out[7]:

	0	1	2	3	4	5	6	7	8	9	...	55	56	57	58	59	60	61	62	63	target
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	6.0	13.0	10.0	0.0	0.0	0.0	0
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	11.0	16.0	10.0	0.0	0.0	1
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	3.0	11.0	16.0	9.0	0.0	2
3	0.0	0.0	7.0	15.0	13.0	1.0	0.0	0.0	0.0	8.0	...	0.0	0.0	0.0	7.0	13.0	13.0	9.0	0.0	0.0	3
4	0.0	0.0	0.0	1.0	11.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	2.0	16.0	4.0	0.0	0.0	4
5	0.0	0.0	12.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	9.0	16.0	16.0	10.0	0.0	0.0	5
6	0.0	0.0	0.0	12.0	13.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	9.0	15.0	11.0	3.0	0.0	6
7	0.0	0.0	7.0	8.0	13.0	16.0	15.0	1.0	0.0	0.0	...	0.0	0.0	0.0	13.0	5.0	0.0	0.0	0.0	0.0	7
8	0.0	0.0	9.0	14.0	8.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	11.0	16.0	15.0	11.0	1.0	0.0	8
9	0.0	0.0	11.0	12.0	0.0	0.0	0.0	0.0	0.0	2.0	...	0.0	0.0	0.0	9.0	12.0	13.0	3.0	0.0	0.0	9
10	0.0	0.0	1.0	9.0	15.0	11.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	10.0	13.0	3.0	0.0	0.0	0
11	0.0	0.0	0.0	0.0	14.0	13.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	13.0	16.0	1.0	0.0	1

12 rows × 65 columns

Train and the model and prediction

```
In [8]: X = df.drop('target',axis='columns')
y = df.target
```

```
In [9]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2)
```

```
In [10]: from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier(n_estimators=20)
model.fit(X_train, y_train)
```

Out[10]: RandomForestClassifier(n_estimators=20)

```
In [11]: model.score(X_test, y_test)
```

Out[11]: 0.975

```
In [12]: y_predicted = model.predict(X_test)
```

Confusion Matrix

```
In [13]: from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_predicted)
cm
```

```
Out[13]: array([[34,  0,  0,  0,  0,  0,  0,  0,  0,  0],
 [ 0, 38,  0,  0,  0,  0,  0,  0,  0,  0],
 [ 0,  0, 37,  0,  0,  0,  0,  0,  1,  0],
 [ 0,  0,  0, 42,  0,  1,  0,  0,  0,  0],
 [ 0,  0,  0,  0, 35,  0,  0,  0,  0,  0],
 [ 0,  0,  0,  0,  0, 36,  0,  0,  0,  0],
 [ 0,  0,  0,  0,  0,  0, 36,  0,  0,  0],
 [ 0,  0,  0,  0,  0,  0,  0, 35,  0,  0],
 [ 0,  2,  0,  1,  0,  0,  0,  1, 30,  0],
 [ 0,  1,  0,  1,  0,  0,  0,  0,  1, 28]], dtype=int64)
```

```
In [14]: import matplotlib.pyplot as plt
import seaborn as sn
plt.figure(figsize=(10,7))
sn.heatmap(cm, annot=True)
plt.xlabel('Truth')
plt.ylabel('Predicted')
```

```
Out[14]: Text(69.0, 0.5, 'Predicted')
```

