Class1_dataset

December 11, 2024

[43]:
```
from scipy.io import loadmat
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_iris
```

0.0.1 Load iris dataset from sklearn & Retrive chunck of data

[46]:
```
iris_dataset = load_iris()
print("Dataset properties", list(iris_dataset))
print("Dataset shape:", iris_dataset['data'].shape)
print("Dataset label:", set(iris_dataset['target']))
```

```
Dataset properties ['data', 'target', 'frame', 'target_names', 'DESCR',
'feature_names', 'filename', 'data_module']
Dataset shape: (150, 4)
Dataset label: {0, 1, 2}
```

[47]:
```
chunck_data = iris_dataset['data'][20:30, :]
chunck_data.shape
```

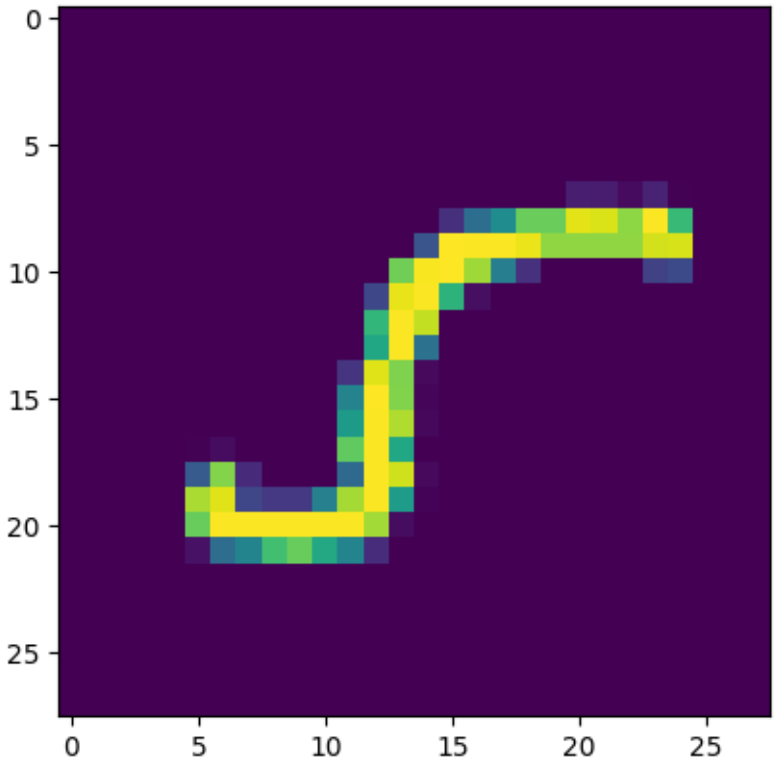
[47]:
```
(10, 4)
```

0.0.2 Load .mat dataset & Display data in the index of 32100

[ ]:
```
mnist_dataset = loadmat(r"./dataset/mnist-original.mat")
print("Dataset properties", list(mnist_dataset))
print("Dataset shape:", mnist_dataset['data'].shape)
print("Dataset label:", set(mnist_dataset['label'][0]))
```

[ ]:
```
import numpy as np
## Show

plt.imshow(np.reshape(mnist_dataset['data'][:, 32100], (28, 28)))
plt.show()
```



0.0.3 Train Test split

[38]:
```
mnist_dataset['label'].shape
```

[38]:
```
(1, 70000)
```

[39]:
```
x_train, y_train, x_test, y_test = train_test_split(np.
↪reshape(mnist_dataset['data'], (70000, 28, 28)), mnist_dataset['label'].T,␣
↪test_size=0.2, shuffle=True)
```

[41]:
```
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)
```

```
(56000, 28, 28)
(14000, 28, 28)
(56000, 1)
(14000, 1)
```