

Mã sinh viên : 20020688

Họ và tên : Đỗ Đức Mạnh

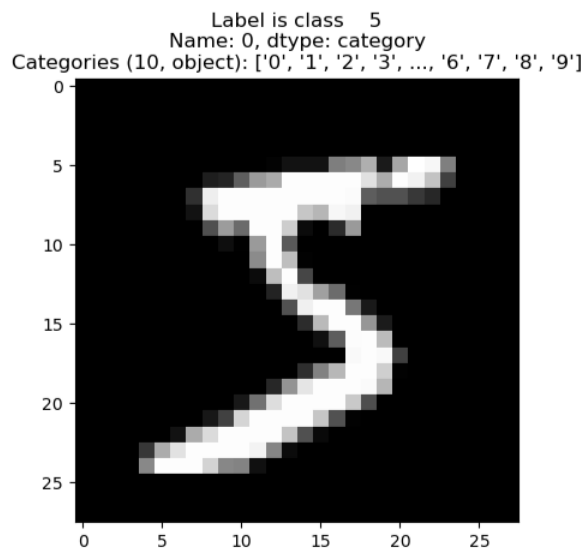
Bài thực hành số 5

Môn : Xử lý ảnh và thị giác robot

Đề bài: Thực hiện bài toán nhận dạng chữ số bằng thuật toán KNN với dữ liệu cho trước.

B1: Sử dụng bộ dữ liệu “mnist_784” từ thư viện sklearn.

In ra phần tử đầu tiên trong ‘mnist_84’ và nhãn của nó:



B2: Sử dụng Model ‘train_test_split’ để chia bộ train và test từ ‘mnist_784’.

Bộ train có kích thước 52500 phần tử và bộ test có 17500 phần tử.

```
In [82]: from sklearn.model_selection import train_test_split
X_train,X_test, y_train,y_test = train_test_split(mnist.data,mnist.target)

In [107]: X_train.shape
Out[107]: (52500, 784)

In [108]: X_test.shape
Out[108]: (17500, 784)
```

B3: Sử dụng Model ‘RandomForestClassifier’ cho bộ dữ liệu train.

B4: Sử dụng hàm ‘predict’ trong ‘RandomForestClassifier’ để dự đoán dữ liệu của bộ test và so sánh kết quả với nhãn đã biết.

Kết quả độ chính xác giữa dự đoán và nhãn bộ test là 96.8%.

```
In [51]: from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
model.fit(X_train, y_train)

Out[51]:
RandomForestClassifier

In [40]: y_pred = model.predict(X_test)

In [41]: y_pred
Out[41]: array(['8', '4', '8', ..., '3', '4', '3'], dtype=object)

In [42]: model.score(X_test, y_test)
Out[42]: 0.968

In [43]: from sklearn.metrics import accuracy_score
accuracy_score(y_pred, y_test)
Out[43]: 0.968
```

Một kết quả trong bộ dữ liệu test:

```
In [115]: y_test[17500]
Out[115]: '2'

In [116]: test = X_test.loc[17500].values
test_label = y_test[17500]

In [110]: test_arr = np.array(test, dtype='uint8')
test_arr = test_arr.reshape((28,28))

In [120]: plt.imshow(test_arr, cmap='gray')
plt.show()
```

