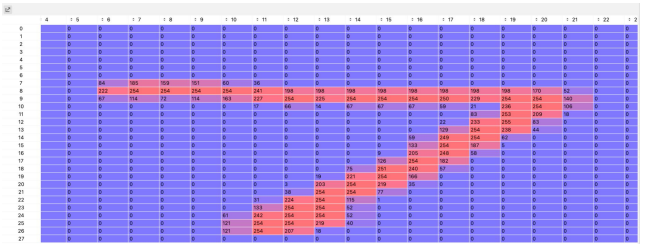
COEN 140L Lab 6 Intro

Xuyang Wu

Dataset



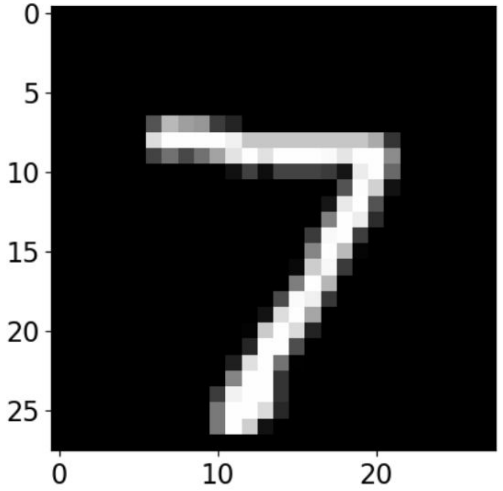
```
import tensorflow as tf

# import the data
mnist = tf.keras.datasets.mnist
from sklearn.model selection import train test split
# load the data, train = (60000,28,28), test = (10000,28,28)
(x traino, y train), (x testo, y test) = mnist.load data()
# convert the 28x28 images to vectors train = (60000,784), test = (10000,784)
x train = np.reshape(x traino, (60000, 28 * 28))
x test = np.reshape(x testo, (10000, 28 * 28))
# normalize the pixel values to be real numbers in [0,1], It's fine if you don't normalize them
# x_train, x_test = x_train / 255.0, x_test / 255.0
```

Plot grayscale image

```
import matplotlib.pyplot as plt

plt.imshow(x_testo[0, :, :], cmap='gray', vmin=0, vmax=255)
plt.show()
```



Explanation of example code

logistic regression documentation link:
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Based on the example code(IrisDatasetLogReg_MulticlassClassification.py):

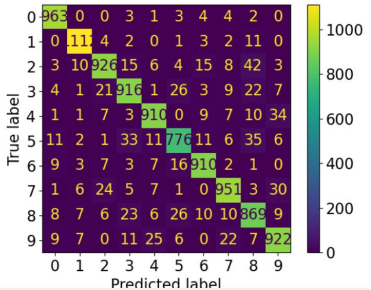
a. fit the logistic regression model with training data

b. do prediction

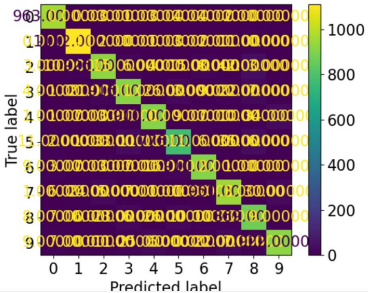
c. plot confusion matrix

plot_confusion_matrix(logreg, x_test, y_test, values_format = '.5g')

d. calculate classification accuracy rate



values_format = '.5g'



values_format = '.5f'

Include in the report

1. Display 10 selected images from the test set, as gray-scale images, each with a different class label.
2. Give the recognition accuracy rate of the whole test set, and show the confusion matrix.
3. Analyze the experimental results you obtain (that is, explain the accuracy rate and confusion matrix, and comment on the performance of your classification model, such as why it makes correct classifications and why some errors occur).

Lab Tasks

- **Need demo** for week 5 assignment(10% points).
- Submit to Camino a pdf report with answers(60% points), the report contains some **results** which required by lab document, you also need to add some **observations** for the questions.
- Submit all the source code needed to generate these answers to Camino(30% points).