COEN 140 Machine Learning and Data Mining

Lab Assignment #6: Logistic Regression

Guideline: Submit a pdf report to Camino. Also submit all the source code needed to generate the results as a separate zip file to Camino.

Problem: Use multi-class logistic regression for the hand-written digits recognition task with the MNIST data set. The dataset has 60,000 training images, and 10,000 test images. Each image is a matrix of size 28×28, representing a hand-written digit from 0 to 9. The images are gray-scale, that is, each pixel value is an integer in [0,255]. The 10 class labels are digits 0, 1, 2, …, 9.

The following code snippet is for your reference:

```
import tensorflow as tf
mnist = tf.keras.datasets.mnist
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split

(x_traino, y_train),(x_testo, y_test) = mnist.load_data()
# convert the 28x28 images to vectors
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
logreg = LogisticRegression(multi_class='multinomial',max_iter = 100,verbose=2)
```

Include in the report (60%):

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).

Demo and explain to TA (10%):

1. How did you train the model?

2. How did you get the predicted labels?

3. How did you calculate the classification accuracy rate?

4. How did you plot the confusion matrix?