Faculty of Engineering
Cairo University

Machine Learning

ML - Task 1

Submitted to:
DR. Inas
Eng. Christin

Submitted by:

Name	Sec	BN
Mohamed Abdelkareem Seyam	2	17

the pre Model class contain the functions that make the preprocessing steps and make predictions

- **read data ()**
  using id2numpy and pd we can read ubyte files and convert it into data frame
- **normalize ()**
  to normalize data by dividing on 255 to make gradient descent faster
- **plot summary of data ()**
  plot summary of data to get intuition about it

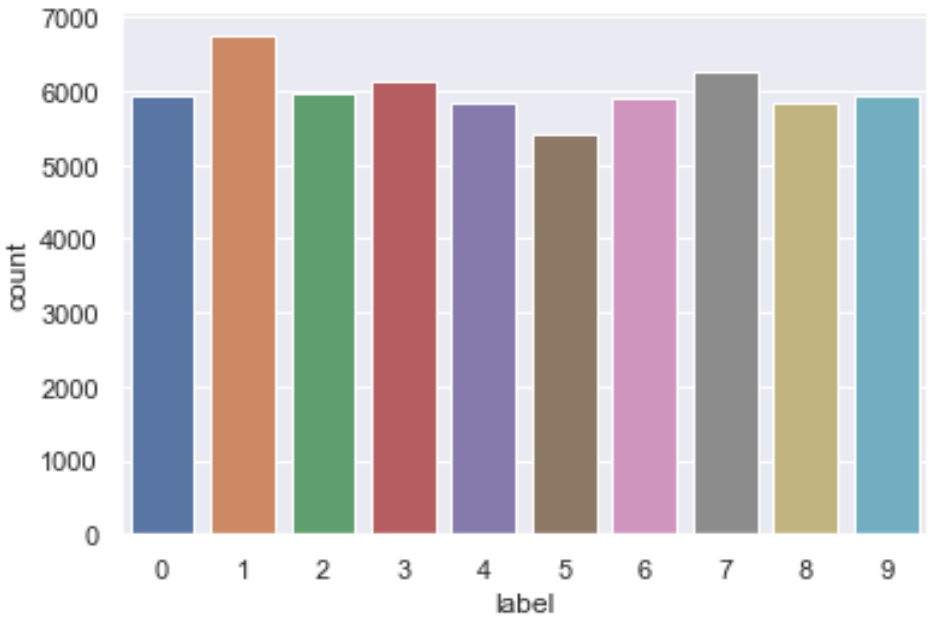

Figure 1: summary of data

- **show examples ()**
  show random number of examples from data set

Figure 2: examples

- **convert MC to BC ()**
  convert the multi classification problem to binary one depending on the target number you wanna classify in our task i used to classify 0
- **make test predictions ()**
  build the model we choose Logistic Regression classifier build in sklearn and train it with the labeled data and using the labeled x data to make predictions and using the labeled y data to make accuracy especsifically test accuracy

  and the resulting for test and train accuracy was
  train accuracy = .99328 and the **test accuracy = 0.9922**

  which is very acceptable

- **make plot for testing data ()**

  pick a random examples of test data and try to classify it

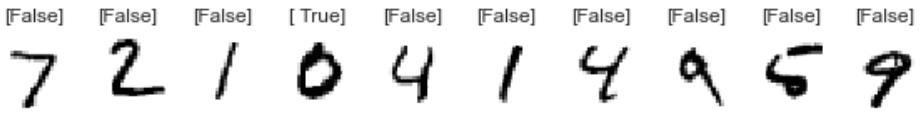

Figure 3: model predictions for binary classification problem for number 0

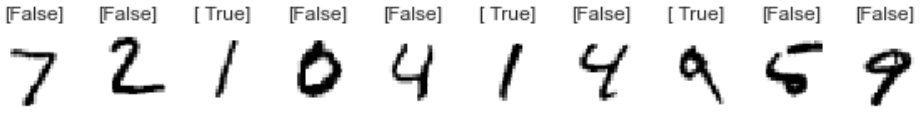

Figure 4: model predictions binary classification prob for number 1

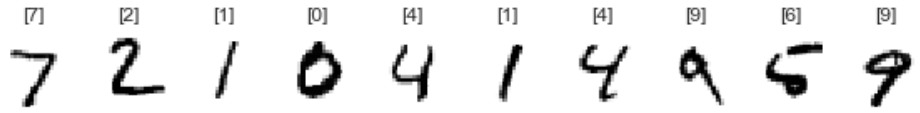

Figure 5: model predictions for number 9 number - multi classification prob

- **confusion matrix ()**

  build the confusion matrix and calc the TP, TN, FP, FN and also calculate F1 score

  it indicate that the TP which is classifying the number correctly is high about 948 digit and classifying the other numbers not equal to the target is 8974 although the alg misclassify small number of digit aroud 46 + 32 digit = 78 digit which is really small

  and the precision is : 0.99490022172949 , recall is = 0.9964468132356207 , f1score = 0.9956729168978143

  Precision tells us how many of the correctly predicted numbers actually turned out to be true about **99.4**

  Recall tells us how many of the target number we were able to predict correctly with our model. about **99.6**

  f1 score is just a way for averaging them. about **99.5**

  which good for our model

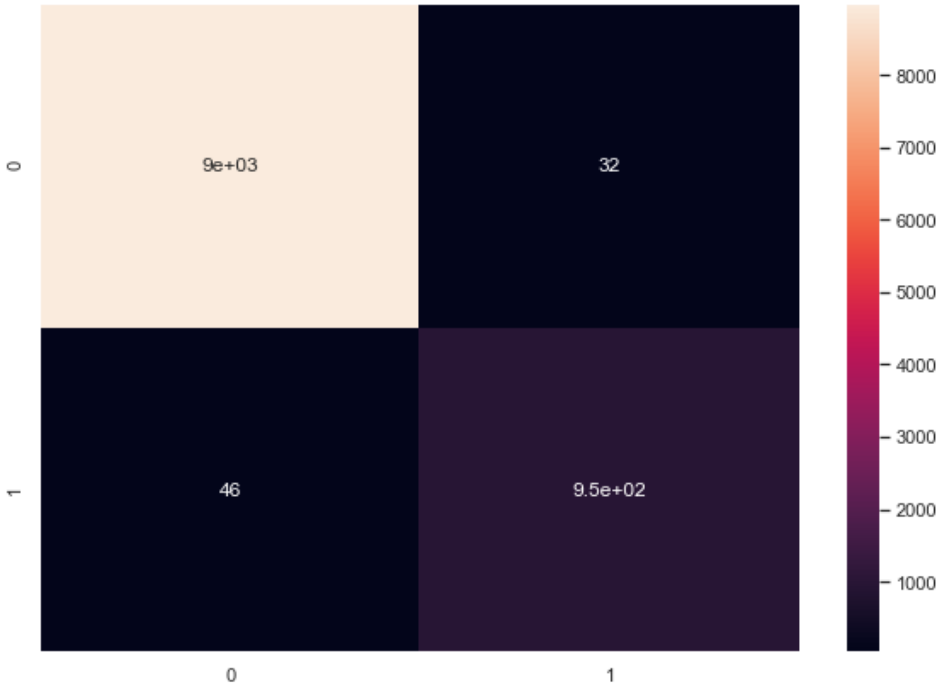

Figure 6: confusion matrix and confusion report

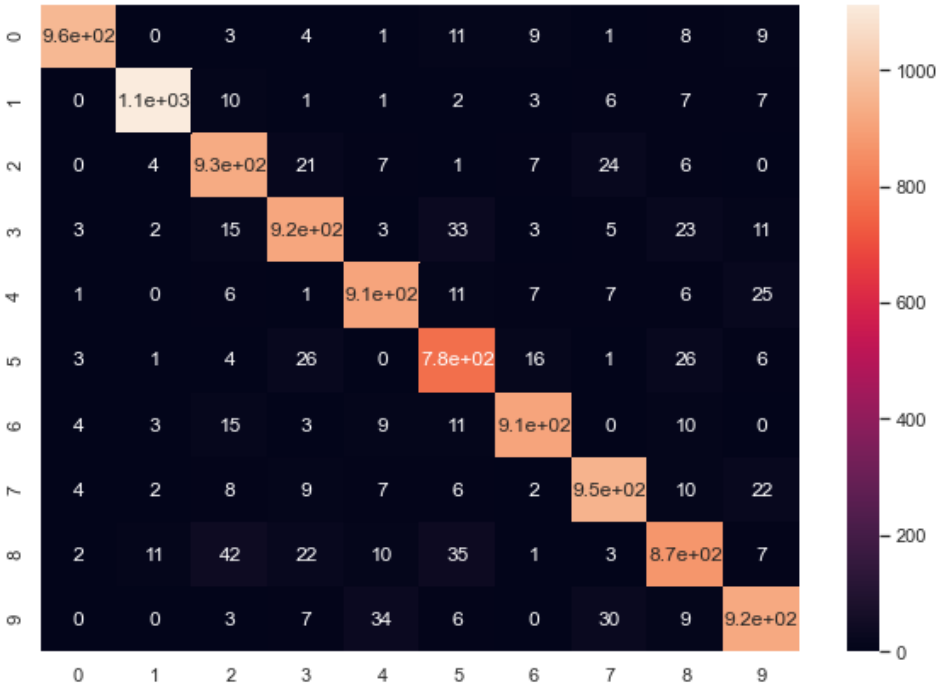

Figure 7: confusion matrix and confusion report