CSE: 575 PROJECT 1

-Harshith Chittajallu, ASU ID- 1218707243

FEBRUARY 19, 2020

INTRODUCTION

In this project, we have 4 data sets, from the MNIST data images of numbers 7 and 8. The data sets are:

TrX, TrY, TsX and TsY. Each data set is an array, with X describing the features (from 0 to 1) and Y describing the label (either 0 or 1). In the data sets of Y, 0 represents the number/class '7', and 1 represents the number/class '8'.

The number of elements in the data sets are as follows:

TrX and TrY have a total of 12116 rows, with the number of images of '7' are 6265 and the number of images of '8' are 5852. TrX has 784 columns (representing the black and white pixel values of the MNIST images with pixel density of 28x28) and TrY has only 1 column, telling if the corresponding data in X belongs to class '7' or class '8' as 0 or 1 respectively.

These are the arrays which need to be trained.

TsX and TsY have a total of 2002 rows, with the number of rows belonging to class '7' being 1028 and number of rows belonging to class '8' being 974

These are the arrays in which the training model needs to be tested upon

PROCEDURE

1. We extract the mean and the standard deviation of all the columns of each row in the training and testing data sets corresponding to X

   Mean = $\sum_i^N X_i \div N$

   Standard Deviation = $(\sum_i^N (X_i - Mean))^\wedge 2 \times \sqrt{1 \div (N-1)}$

   Variance = Standard Deviation × Standard Deviation

Where N- no. of Data items

Xi – ith row of the training set

The new resultant array has 2 columns and 12116 rows for TrX and 2002 rows for TsX

2. Now we make 2 new arrays which are data sets for class ‘7’ and class ‘8’ individually

Naïve Bayes Classification:

1) First, we calculate the mean and variances of the extracted features each (i.e the mean and the standard deviation columns of the above matrix) individually for the digit classes ‘7’ and 8’
2) Now we use these 8 obtained values to calculate the conditional probabilities and posterior probabilities as follows:

   P(Y=0|X) = P(Y=0) × $\prod_i P(X_i|Y = 0)$

   P(Y=1|X) = P(Y=1) × $\prod_i P(X_i|Y = 1)$

   Where,

   P(Y=0|X) and P(Y=1|X) are posterior probabilities of data belonging to either class ‘7’ or class ‘8’ respectively

   P(Y=0) and P(Y=1) are the prior probabilities of classes ‘7’ and ‘8’

   $\prod_i P(X_i|Y = 0)$ and $\prod_i P(X_i|Y = 1)$ are the Normal distribution PDF’s using the parameters of the corresponding mean and standard deviation for a particular class

3. The normal distribution PDF formula is :

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\begin{aligned} \mu &= \text{ Mean} \\ \sigma &= \text{ Standard Deviation} \\ \pi &\approx 3.14159\cdots \\ e &\approx 2.71828\cdots \end{aligned}$$

Where, x is the 'i'th test data column from TsX
μ is the mean of the 'i'th test data column for a particular class
sigma is the standard deviation of the 'i'th test data column for a particular class

4. We give the label to a data item of a particular class, if its posterior probability is greater than the other class.

LOGISTIC REGRESSION:

1. We can find the class probability using the sigmoid function to calculate the predicted values where,

P = 1/(1+ $e^{W^T \times X}$)

Where W= weight matrix, initially set to [0,0,0]

X- Training Matrix containing the features

2. To calculate the weights, we use gradient ascent algorithm:

W_bias = 0, W =[0,0,0] initially
A = W_bias + $\sum W_i X$
P(Y=1|X,W) = $e^A \div (1 + e^A)$, which is the sigmoid function of –A

W_bias = W_bias + learning rate × ($\sum_j (Y_{j-train} - P(Y = 1|X$,W))

W = W + learning rate × ($\sum_j X_j (Y_{j-train} - P\left(Y = 1\middle|X_j, W\right)$))

Where $Y_{j-train} - P(Y = 1|X$,W)) gives the error value.

3. Now we set a threshold value and classify the test data:

If(P(Y=1|X,W)> threshold value)
Predicted Y = 1
Else
Predicted Y = 0

RESULTS

To calculate accuracy, we use the formula:

Accuracy = number of predicted test values equal to the actual test values (TsY) ÷ total number of predicted values

Accuracy of an individual class = number of predicted test values of a class equal to actual test values of a class ÷ total number of predicted values in a class

```
Accuracy of Naive Bayes:  69.43056943056943
Accuracy of predicting 7 using Naive Bayes:  75.77821011673151
Accuracy of predicting 8 using Naive Bayes:  62.73100616016427
```

Fig.1 Results of Naïve bayes Algorithm

```
weights: [-1.7947380818189749, 2.1716291315467076, 2.2223632068257744]
Accuracy of logistic Regression:  0.6918081918081919
Accuracy of logistic Regression for predicting 7:  0.7626459143968871
Accuracy of logistic Regression for predicting 8:  0.6170431211498973
```

Fig.2 Results of Logistic Regression Algorithm

Methods	Accuracy of predicting ‘7’	Accuracy of Predicting ‘8’	Overall Accuracy
Naïve Bayes	75.77	62.73	69.43
Logistic Regression (epochs = 300, learning rate = 0.005 Threshold value = 0.301)	76.26	61.70	69.18