HW3: Logistic Regression

CSE6242 - Data and Visual Analytics - Spring 2017

Ebeid ElSayed - Ebeid@gatech.edu

March 14, 2017

```
setwd("C:/Users/eelsayed/Google Drive/CSE 6242/2017 Spring")
```

0. Data Preprocessing

0.a & 0.b

```
rawDatalLoaded <- TRUE

if(file.exists("mnist_train.csv")){
  train <- read.csv(file="mnist_train.csv", header = FALSE)
}else{
  rawDatalLoaded <- FALSE
}

if(file.exists("mnist_test.csv")){
  test <- read.csv(file="mnist_test.csv", header = FALSE)
}else{
  rawDatalLoaded <- FALSE
}

if(!rawDatalLoaded){
  print("Data wasn't loaded correctly.")
}

train <- as.data.frame(t(train))
names(train)[785] <- "Label"

test <- as.data.frame(t(test))
names(test)[785] <- "Label"
```

0.c Partition the training set for classification of 0, 1 and 3, 5 classes based on the class label (last row 785): train_0_1, train_3_5.

```
train_0_1 <- train[(train$Label == 0) | (train$Label == 1),]
train_3_5 <- train[(train$Label == 3) | (train$Label == 5),]
```

0.d Do the same for the test set: test_0_1, test_3_5.

```
test_0_1 <- test[(test$Label == 0) | (test$Label == 1),]
test_3_5 <- test[(test$Label == 3) | (test$Label == 5),]
```

0.e & 0.f | Separate the class label from all the partitions created (remove row 785 from the actual data and store it as a separate vector).

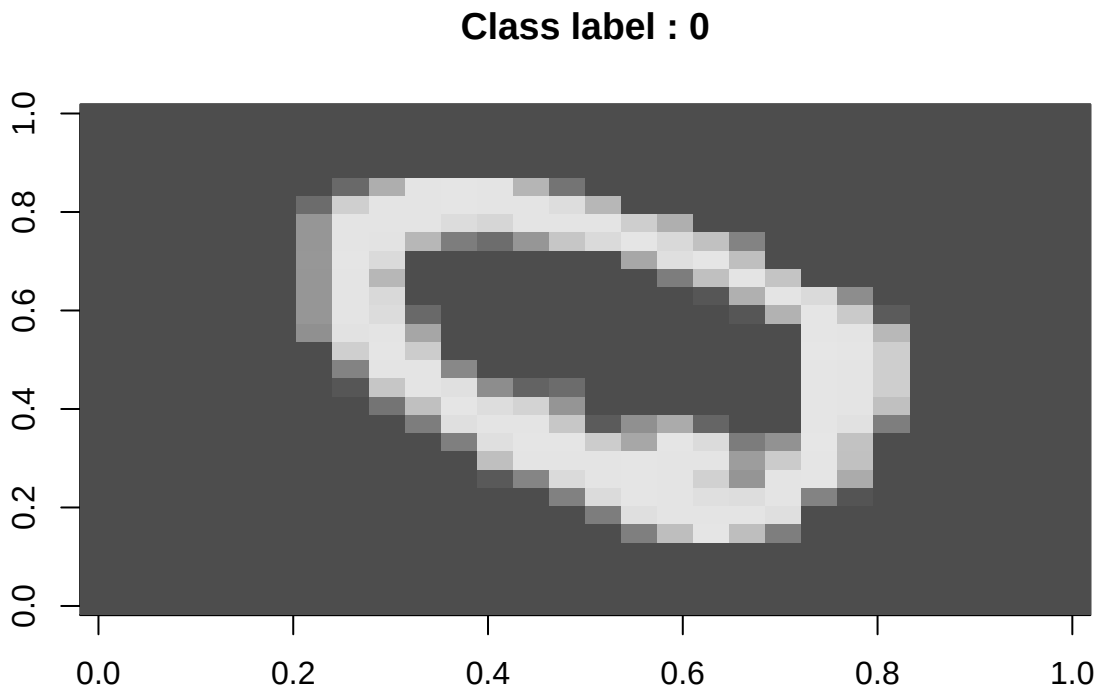
```
true_label_train_0_1 <- train_0_1$Label
train_0_1 <- subset(train_0_1, select = names(train_0_1) != "Label" )
```

```
true_label_train_3_5 <- train_3_5$Label
train_3_5 <- subset(train_3_5, select = names(train_3_5) != "Label" )

true_label_test_0_1 <- test_0_1$Label
test_0_1 <- subset(test_0_1, select = names(test_0_1) != "Label" )

true_label_test_3_5 <- test_3_5$Label
test_3_5 <- subset(test_3_5, select = names(test_3_5) != "Label" )
```

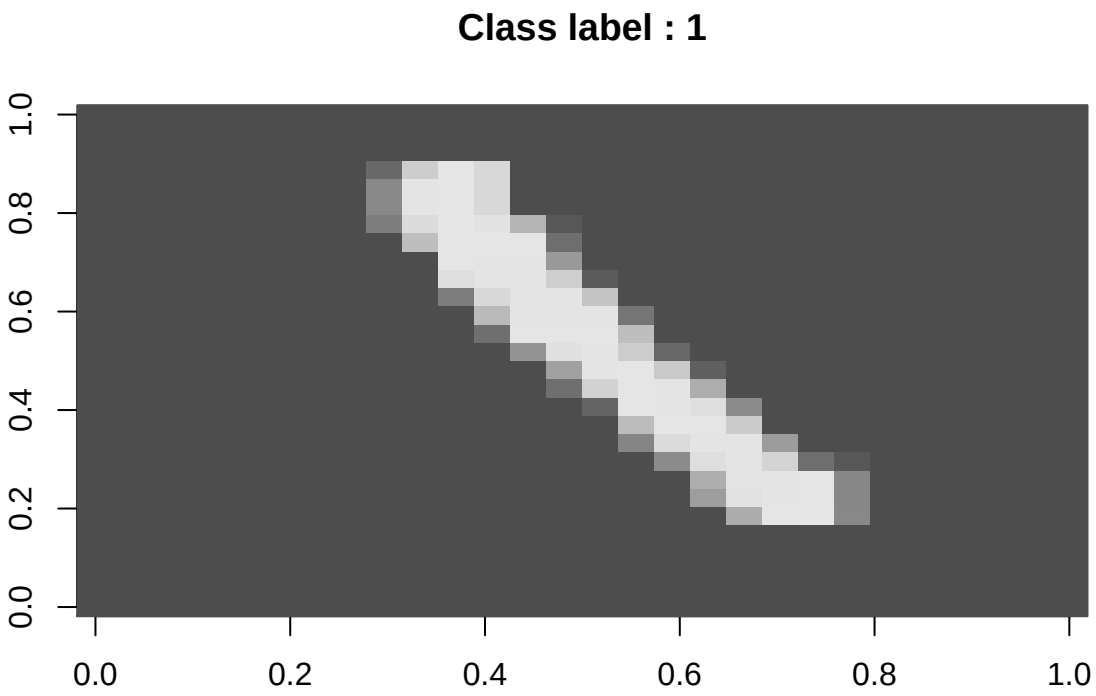
0.g Visualize 1 image from each class to ensure you have read in the data correctly. You will have 4 images corresponding to 0, 1, 3 and 5. You need to convert the 1D image data into 2D for visualisation.

```
show_digit_image <- function(df, digitClass, imageTitle) {
  tmp <- df[df$Label == digitClass,]
  m <- matrix(unlist(tmp[1,1:784]), ncol = 28, byrow = TRUE)
  image(z = m, col = gray.colors(256))
  title(main = imageTitle)
}
```

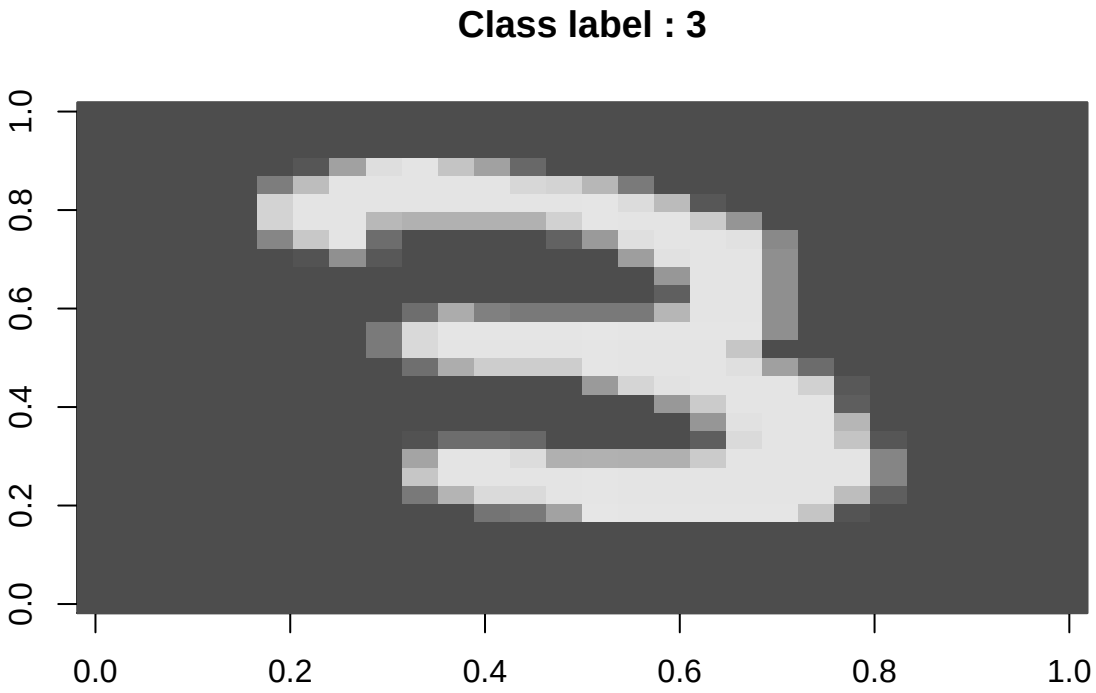
```
show_digit_image(train, 0, "Class label : 0")
```



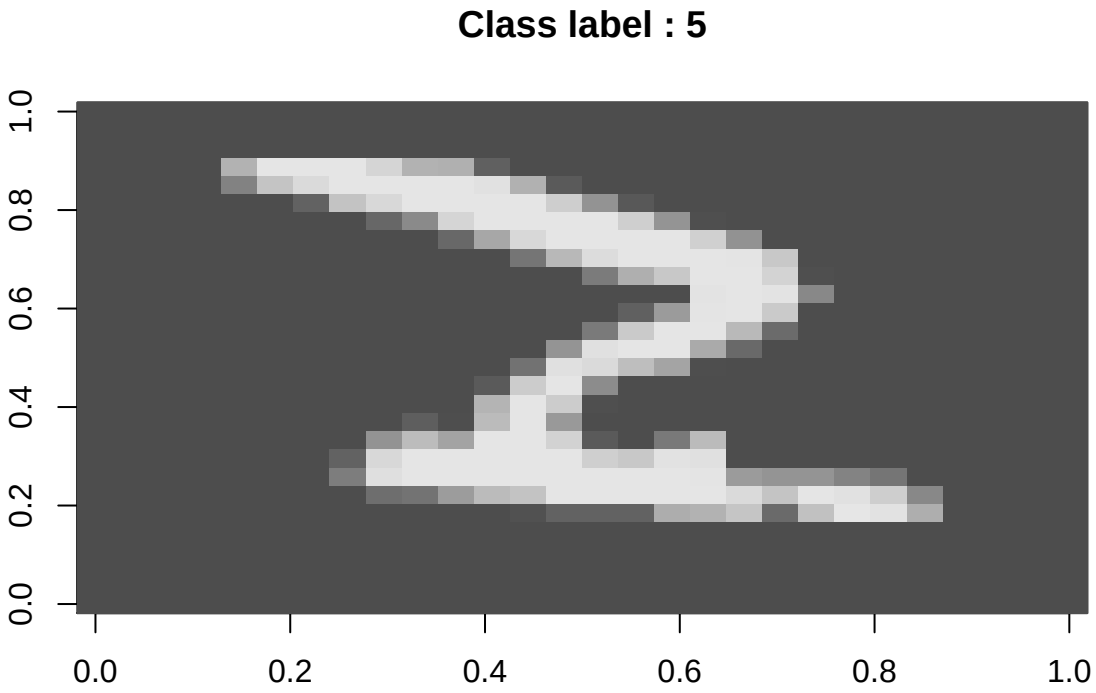

```
show_digit_image(train, 1, "Class label : 1")
```




```
show_digit_image(train, 3, "Class label : 3")
```




```
show_digit_image(train, 5, "Class label : 5")
```



1. Theory

1.a Write down the formula for computing the gradient of the loss function used in Logistic Regression. Specify what each variable represents in the equation.

The formula for the gradient descent is:

$$\theta_j \leftarrow \theta_j - \alpha \sum_{i=1}^{n} \frac{1}{1 + \exp(-y^{(i)} < \theta, x^{(i)} >)}$$

Logistic regression model is built on a likelihood function that gives each data point represented in a vector of features (first part of the data point) a propability of being in class y^i (the second part of the data point) like this