CSE 6363 Machine Learning
Assignment 02
Abhijit Deshpande
UTA ID 1001677938
11/04/2020

The logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1, with a sum of one.

Problem 1

Logistic regression has been used for classification of data. Data has been generated using gaussian random samples. Each data labelled with class 0 and 1 combing with both has 1000 data points for the classification.

In logistic regression for hypothesis, sigmoid function used to get probabilities for labels between 0 and 1. Threshold can be set to define class. Here, 0.5 threshold has been set. Threshold depends on data set and labels we have. Logistic Regression has fit method to select the best hypothesis with minimum loss that fits model well in all hypothesis spaces. It uses gradient decent to converge the model and find the minima for the model. Learning rate used to define the speed of convergence, if learning rate is too big or too small then model may not converge, so learning should be selected properly. Also having too small learning rate may take longer time to converge the model so stopping criteria has been added to avoid the long-time delay.

Learning rate set to different values and checked the accuracy and time required to fit the model using python magic function. Here are the results with different learning rate:

```
Learning Rates = [ 0.0001, 0.001, 0.01, 0.1, 1]
Accuracies = [ 81.8, 81.5, 81.3, 81.2, 81.2]
```
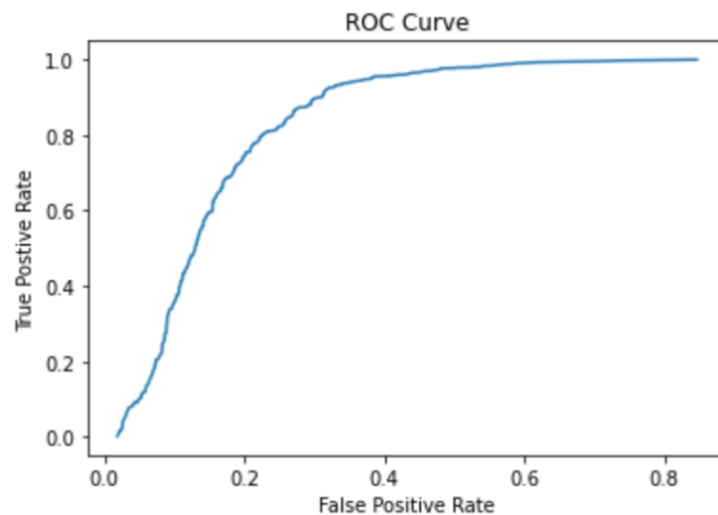
It clearly shows that as learning rate decreases accuracy tends to decrease as it tries to minimize model at faster rate.
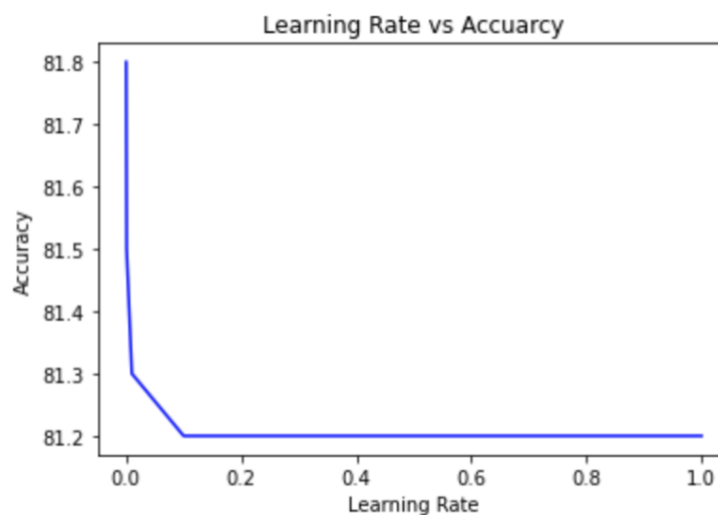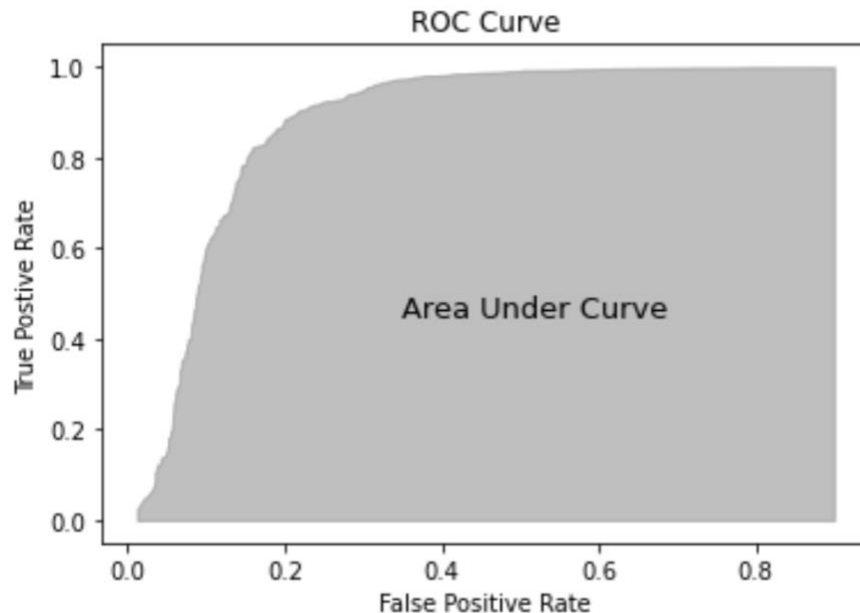
ROC Curve

A **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:
- True Positive Rate
- False Positive Rate

A ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. The following figure shows a typical ROC curve.

AUC: Area Under the ROC Curve

AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1). AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.

For training data, the precision, recall and f1_score is:
```
Precision= 0.9140625
Recall= 0.702
F1-score= 0.7941176470588236
```

For testing data,
```
Precision:  0.9116945107398569
Recall:  0.764
F1-Score:  0.8313384113166485
```

It shows that model is 91% precise in predicting positive values and it captures 70% on training and 76% on testing actual positives.

Problem 2

Multiclass logistic regression is a classification method that generalizes logistic regression to multiclass problems, i.e. with more than two possible discrete outcomes. It is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables.

MNSIT data set has been used to perform on multiclass logistic regression. Here, the change is in hypothesis function in which instead of sigmoid, softmax faction has been used to calculate probabilities. The data set has five classes from 0 to 4. Multinomial Logistic regression used one vs all method to predict the classes. One hot coding method used to get dummies of classes. With 0.001 learning rate model giving **84.32%** accuracy on training data set where as on testing data set it is giving **85.46%** accuracy which shows model performing good on unseen data and predicting pretty accurately.

Precision and recall for testing set:

0	0.75	0.99	0.86
1	0.99	0.86	0.92
2	0.83	0.76	0.79
3	0.79	0.90	0.84
4	0.98	0.76	0.86

Model is precise in predicting 1 and 4 class labels with 99 and 98% precision. Model captures better details for class 0 and 3.

Problem 3

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data becomes available.

The same data set has been used for ANN for training and testing which was used in problem 2. It has 5 classes to predict. TensorFlow and Keras library is used for this problem to train and test the data. Sigmoid and ReLU, the two different activation functions used to evaluate the accuracies with both activation functions. Here, used with 1 hidden layer.

1) With Sigmoid:

   Sigmoid function is used for activation. Results has been calculated with 10 neurons and 1 hidden layer.

   Epoch indicates the number of passes of the entire training dataset. Here, 5 epochs are used. With increasing epochs, accuracy of model increases with more epochs. Model giving 99% accuracy with the sigmoid function.

2) With ReLU and Softmax:

   ReLU is used for activation and softmax is used to activate the function at output. It gives better results than single ReLU function. In most of the cases both combinations are used to get better accuracies. Model is 98% accuracy.

3) With ReLU:

   Looking at results using only ReLU function, it is giving worst results compared to others. Comparing only ReLU with ReLU with Softmax, ReLU with Softmax performs better and gives better results than the single ReLU. It rectifies the output and gives the probabilities ranging between 0 to 1. Thus, softmax is combined with ReLU which gives output with probabilities.

Comparing with Multiclass Logistic Regression

Evaluating the accuracies of both Multiclass logistic and ANN, ANN performs better than multiclass logistic regression. It gives almost 14% more accuracy than the multiclass logistic regression. ANN is more complex than multiclass logistic regression, logistic regression is one-part neural network. Using multiple hidden, more complex model can be prepared to get better results.

Reference:

https://medium.com/@himanshuxd/activation-functions-sigmoid-relu-leaky-relu-and-softmax-basics-for-neural-networks-and-deep-8d9c70eed91e

https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

http://rasbt.github.io/mlxtend/user_guide/classifier/SoftmaxRegression/