**Module 2 Predictive Analytics**

**Author’s Name:** Shivani Sharma

**Course Number:** ALY6020

**Course Title:** Predictive Analytics

**Academic Term:** Fall 2019 CPS Analytics

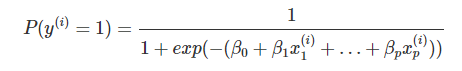
**Instructor’s Name:** Marco Montes de Oca

**Assignment Completion Date:** 09-29-2020



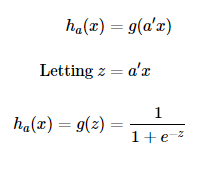
**Introduction**

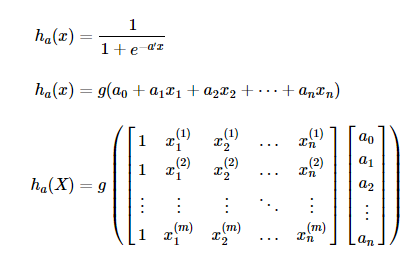
This week’s assignment is based on “Logistic Regression” algorithm on the same data set we used in the last week i.e “MNIST” Fashion dataset. This is basically a classification problem set where we have to use logistic regression technique. I would like to give a brief about logistic regression. It is one of the supervised machine learning algorithms used for classification problems. It is considered as a linear classifier, also known as “logit function”. “Logistic Function” is mathematically a sigmoid function and provides the output close to zero or one and provides the weights b0,b1,b2 …….. bn and this is known as model fitting. Here is the generalized formula for the same:



We label the training data into 0 and 1 and assign labels to each target variable separately.

Model hypothesis function is as follows:





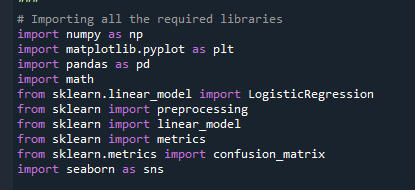
And the above form is a mathematical representation of logistic model which is nothing but a sigmoid function/logistic function.

I am going to explain the implemented algorithm in the analysis section.

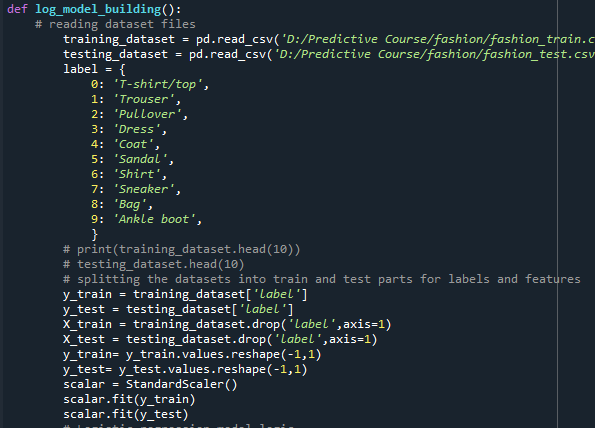
**Analysis**

We have to apply logistic regression model on the same MNIST fashion data set that was provided to us in the last week assignment.

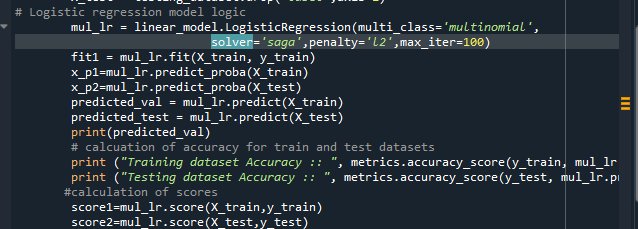
Firstly, I have imported all the required libraries:



Then I have worked on the data set processing and divided the provided training and testing datasets into features and labels.



Now I have applied multinomial logistic regression model by putting the parameter “multi\_class” in the LogisticRegression(). I have used the solver ‘saga’ (Based on my results with different type of solvers this is the best one) and charged the penalty as “l2”. Then performed the fitting of model and predicted the values for each type of product i.e the model is applied on 10 different models altogether and provided output for all type of fashion items. Then I calculated the accuracy and scores for each item.

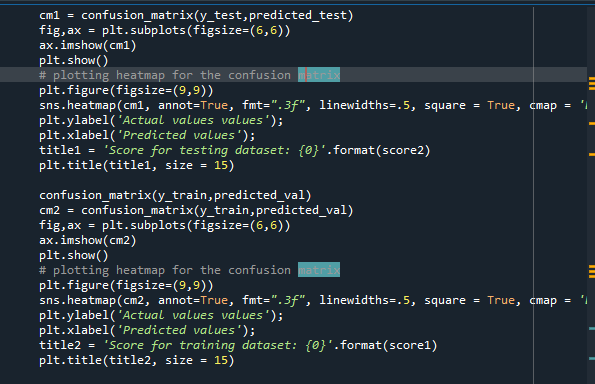


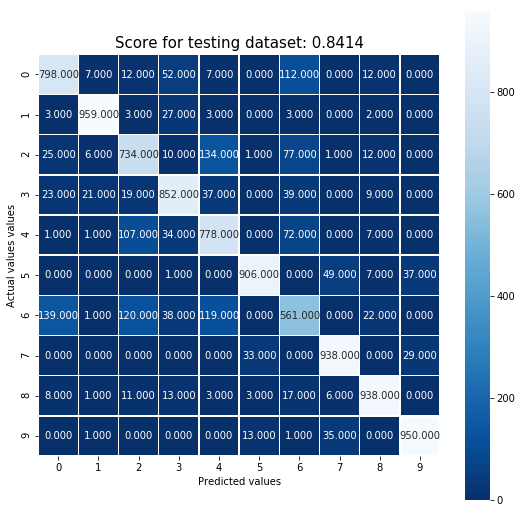
Here is the output for the value for the training and testing datasets accuracy:

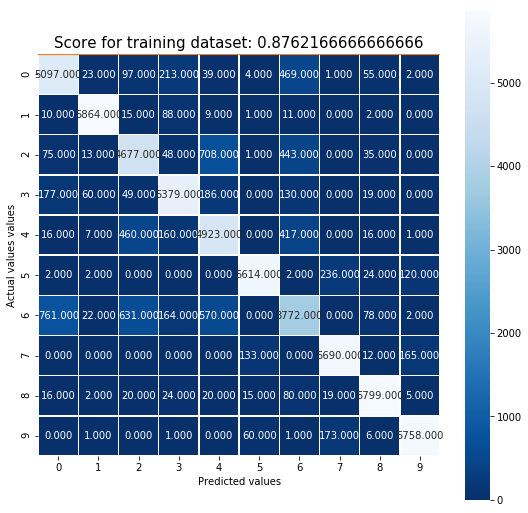


Finally I plotted the confusion matrix for training and testing dataset predicted values. In order to show the detailed probabilities I created heatmaps as well for showing the predicted outputs.

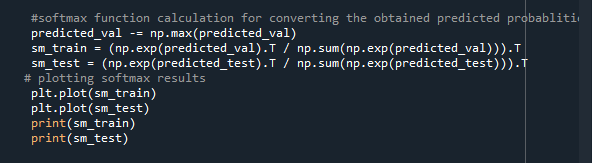
**NOTE:** **Reason behind calculating every parameter for both training and testing dataset is that I wanted to apply “softmax function” on both and transform the probabilities into a probability distribution separately on both datasets provided.**

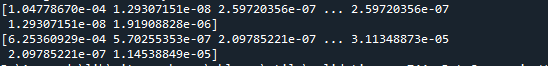






Now as mentioned in the question we have to convert the predicted probabilities into probability distribution using “softmax” function so here is the logic:





**Conclusion**

The accuracy for the logistic model is coming as 84.14 percent on testing dataset and for training dataset it came out 87.62% which is pretty high value. And now if we compare the accuracy to the results we obtained with KNN which was 84.97%,84.95% and 84.07% for k=1,11 and 21 respectively (which was calculated for testing dataset.) So, we can say that there is slight decrease in the accuracy in case of logistic regression model as compared to KNN.

# References

1. Sklearn.metrics.confusion\_matrix¶. (n.d.). Retrieved September 29, 2020, from <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html>
2. 1.1. Linear Models¶. (n.d.). Retrieved September 30, 2020, from https://scikit-learn.org/stable/modules/linear\_model.html
3. 6.3. Preprocessing data¶. (n.d.). Retrieved September 30, 2020, from https://scikit-learn.org/stable/modules/preprocessing.html
4. Multiclass logistic regression from scratch¶. (n.d.). Retrieved September 30, 2020, from https://gluon.mxnet.io/chapter02\_supervised-learning/softmax-regression-scratch.html
5. Raschka, S. (n.d.). Softmax Regression. Retrieved September 30, 2020, from http://rasbt.github.io/mlxtend/user\_guide/classifier/SoftmaxRegression/
6. Juliani, A. (2016, February 22). Simple Softmax Regression in Python - Tutorial. Retrieved September 30, 2020, from https://medium.com/@awjuliani/simple-softmax-in-python-tutorial-d6b4c4ed5c16
7. What is Softmax Regression and How is it Related to Logistic Regression? (n.d.). Retrieved September 30, 2020, from https://www.kdnuggets.com/2016/07/softmax-regression-related-logistic-regression.html
8. Kinghorn, D. (n.d.). Machine Learning and Data Science: Logistic Regression Theory. Retrieved September 30, 2020, from https://www.pugetsystems.com/labs/hpc/Machine-Learning-and-Data-Science-Logistic-Regression-Theory-988/
9. Real Python. (2020, September 21). Logistic Regression in Python. Retrieved September 30, 2020, from https://realpython.com/logistic-regression-python/
10. MNIST classification using multinomial logistic + L1¶. (n.d.). Retrieved September 30, 2020, from https://scikit-learn.org/stable/auto\_examples/linear\_model/plot\_sparse\_logistic\_regression\_mnist.html
11. User3697903user3697903 1766 bronze badges, JerilJeril 5, & Arad HaseliradArad Haselirad 12666 bronze badges. (1968, September 01). Maximum number of iterations must be positive ERROR when using Logistic Regression (python). Retrieved September 30, 2020, from https://stackoverflow.com/questions/56022400/maximum-number-of-iterations-must-be-positive-error-when-using-logistic-regressi/56023028
12. Galarnyk, M. (2020, April 29). Logistic Regression using Python (scikit-learn). Retrieved September 30, 2020, from https://towardsdatascience.com/logistic-regression-using-python-sklearn-numpy-mnist-handwriting-recognition-matplotlib-a6b31e2b166a
13. 2 Ways to Implement Multinomial Logistic Regression In Python. (2017, May 15). Retrieved September 30, 2020, from https://dataaspirant.com/implement-multinomial-logistic-regression-python/
14. Kinghorn, D. (n.d.). Machine Learning and Data Science: Multinomial (Multiclass) Logistic Regression. Retrieved September 30, 2020, from https://www.pugetsystems.com/labs/hpc/Machine-Learning-and-Data-Science-Multinomial-Multiclass-Logistic-Regression-1007/