Programming Assignment #1

Logistic Regression

February 17th, 2023

Gautam Sapre (*gss170001*)

Sanjeev Penupala (*sxp170022*)

a) Reporting Error Rates

```
Iterations: 10, Learning Rate: 0.01, Train Error: 0.49180327868852464, Test
Error: 0.5277777777777778
Iterations: 10, Learning Rate: 0.1, Train Error: 0.319672131147541, Test
Error: 0.3611111111111111
Iterations: 10, Learning Rate: 0.33, Train Error: 0.2213114754098361, Test
Error: 0.2546296296296296
Iterations: 100, Learning Rate: 0.01, Train Error: 0.3114754098360656, Test
Error: 0.3587962962962963
Iterations: 100, Learning Rate: 0.1, Train Error: 0.20491803278688525, Test
Error: 0.23148148148148145
Iterations: 100, Learning Rate: 0.33, Train Error: 0.18852459016393444,
Test Error: 0.21990740740740738
Iterations: 1000, Learning Rate: 0.01, Train Error: 0.20491803278688525,
Test Error: 0.23148148148148145
Iterations: 1000, Learning Rate: 0.1, Train Error: 0.18032786885245902,
Test Error: 0.2037037037037037
Iterations: 1000, Learning Rate: 0.33, Train Error: 0.20491803278688525,
Test Error: 0.19907407407407407
Iterations: 10000, Learning Rate: 0.01, Train Error: 0.18032786885245902,
Test Error: 0.2037037037037037
Iterations: 10000, Learning Rate: 0.1, Train Error: 0.18852459016393444,
Test Error: 0.18518518518518517
Iterations: 10000, Learning Rate: 0.33, Train Error: 0.18032786885245902,
Test Error: 0.18287037037037035

Best Parameters: Iter=10000, LR=0.33
Our Best Model Train Error: 0.18032786885245902
Our Best Model Test Error: 0.18287037037037035
```

c) scikit-learn

```
Our Best Model Train Error: 0.18032786885245902
Our Best Model Test Error: 0.18287037037037035

Sklearn Train Error: 0.16393442622950816
Sklearn Test Error: 0.1782407407407407
```

There are a couple of reasons why sklearn model is performing a bit better than our custom model.

- Implementation: The sklearn one could be using different functions than we did in our custom one, like a different loss function, which could be better for the dataset.
- Regularization: Logistic regression models use regularization techniques such as L1 or L2 regularization to prevent overfitting. Sklearn's default logistic regression model uses L2 regularization, and it could be improving its performance on the test set.
- Hyperparameters: Scikit-learn's logistic regression model has many hyperparameters that can be tuned to improve performance. Although we did not change the default parameters, it's possible that the default parameters are better suited for the dataset.
- Data Preprocessing: Sklearn may be performing some additional data preprocessing like scaling the data or performing feature selection to improve model performance.

d) Plotting Curves