BITS F464 - Machine Learning

Assignment-2A: Logistic Regression

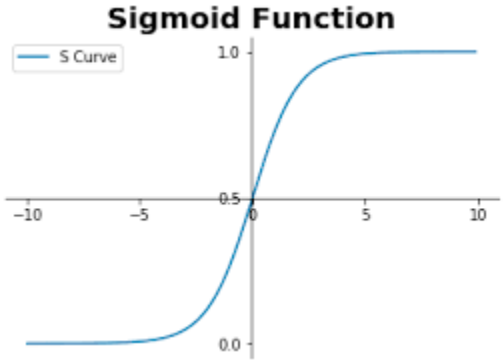
Made by -
Kriti Jethlia (2018A7PS0223H)
Tanay Gupta(2018AAPS0343H)
Anusha Agarwal (2018A3PS0032H)

Model Description

- We have implemented Logistic regression for binary classification of a dataset.

$$y = \frac{1}{1 + e^{-(wx+b)}}$$

- Here w is the weight of every feature and b is the bias. This y is the predicted y and the function is a sigmoid function. It takes values between 0 and 1 therefore can be used in probabilistic models. It is continuous and differentiable at all points.



- We need to find a w that maximizes the likelihood of training examples being predicted correctly to its corresponding class.
- We can find the optimal using 2 methods-
- **Gradient Descent Algorithm-** It is the process of minimizing a function by following the gradients of the cost function. We take eta step size and reach the global minima by reducing the error in every iteration. We walk in the negative direction of gradient. When the error between successive iterations comes close to 0, we stop.The following is the loss function in logistic regression.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$$

$$m = number\ of\ samples$$

- **Stochastic Gradient Descent Algorithm-** this is the same as gradient descent but here we use only 1 single value rather than the whole dataset to find the appropriate w. It is a faster algorithm.
- After we find the w, we predict the class using the sigmoid function. The threshold is set to be 0.5 If the function value is greater than the threshold, the class is 1 or else 0.
- According to the predicted values, we calculate accuracy, precision , recall and f score.

Results

a)Gradient Descent

Weights: [-7.06833619 -3.79093865 -4.78008557 -0.4997776]
Bias: 6.695597508493346
Loss : 16.671909166857183
Training parameters
Accuracy : 0.9907291666666667
Precision : 0.988628889354495
Recall : 0.9906661582497778
F score : 0.9896445566048753
Testing parameters
Accuracy : 0.9895631067961165
Precision : 0.9882641064116828
Recall : 0.9878896500746192
F score : 0.9880333310241675

b)Stochastic Gradient Descent

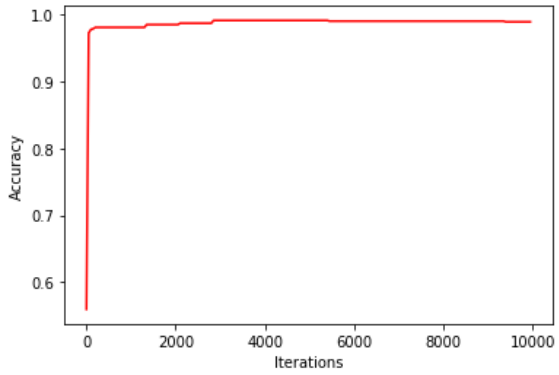
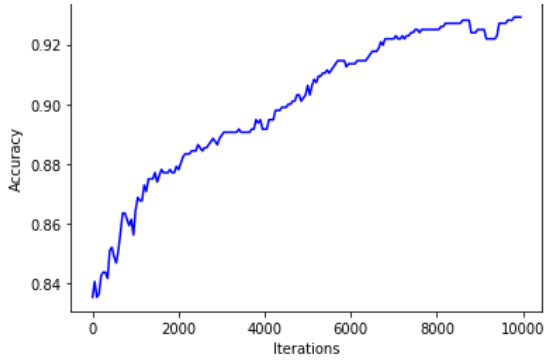
weights: [-1.18574627 -0.62340557 -0.65449352 -0.26792482]
bias: 0.5377976272539726
Loss : 0.14067225490084

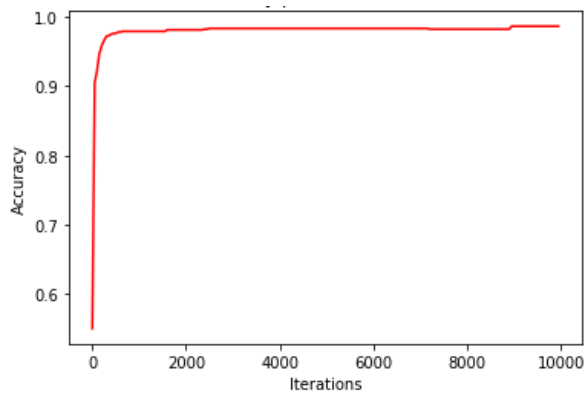
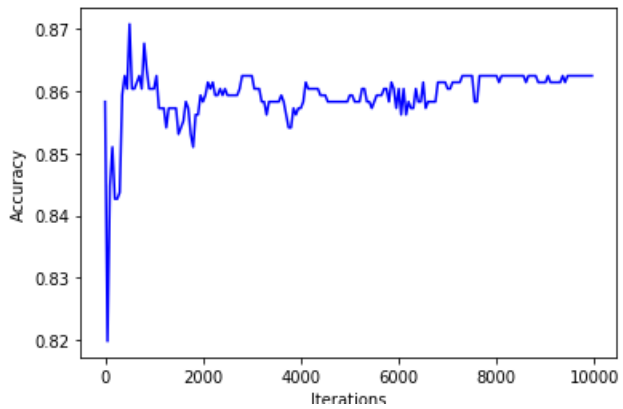
Training parameters
Accuracy : 0.9726041666666667
Precision : 0.9813638167419462
Recall : 0.9568573465068745
F score : 0.96895166607547
Testing parameters
Accuracy : 0.9733009708737864
Precision : 0.977432589122159
Recall : 0.9614048882224535
F score : 0.9693020027189149

Attribute 1 is the most important feature

b)Plots of accuracy for 3 learning rates

Learning Rates = [0.001 , 0.0001 , 0.00001]

lr	Gradient Descent	Stochastic Gradient Descent
0.001		

0.0001		
0.00001