LOGISTIC REGRESSION

Group Members:

Sarthak Gaur - 2017A7PS0250H
Smit D Sheth - 2017A7PS1666H
Sourav Sanganeria - 2017A7PS1625H

- ## Introduction

  We are required to analyze the logistic regression model we built based on accuracy,f-score, recall, and precision that is generated by our dataset.
  The model is built using numpy, sklearn and matplotlib.

- ## Dataset

  We have 1372 data points each having 4 features and we are required to classify these points as 1 or 0 ie. notes are forged or unforged. We standardize the given data and randomly split the dataset into a test set and train set in the ratio 2:8.

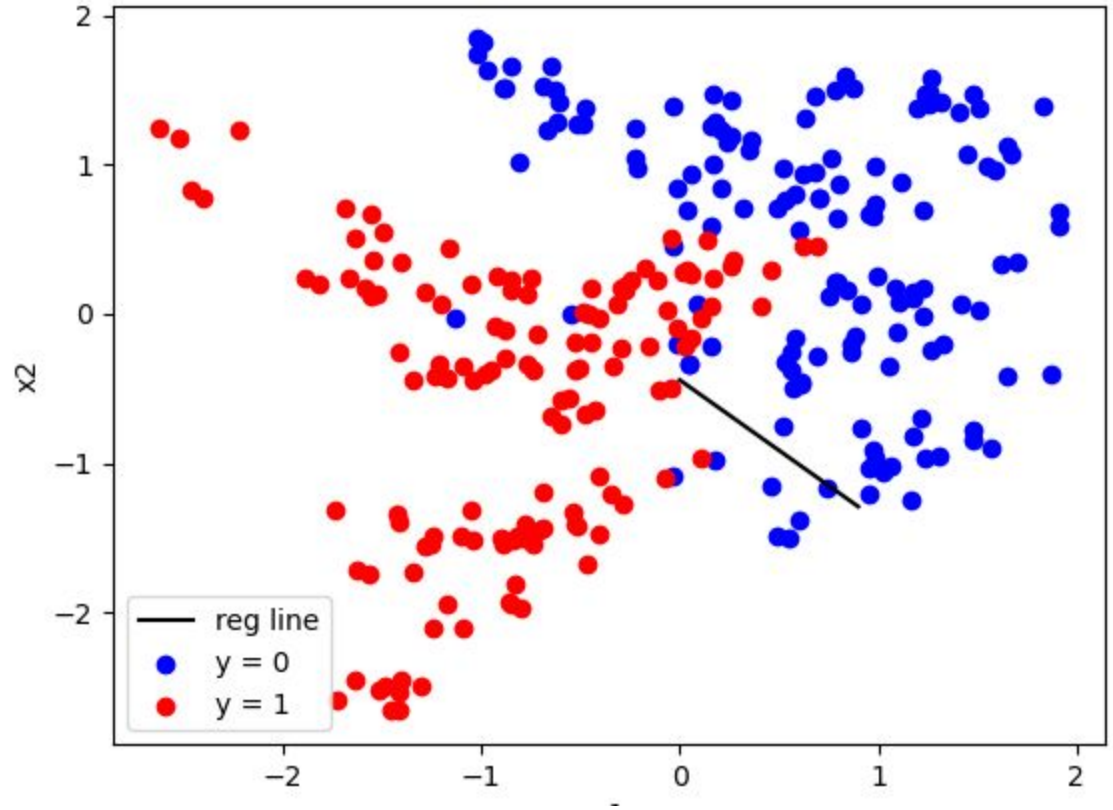
- ## Equations Employed

  Here we employ sigmoid function in order to calculate gradient of parameters.
  Sigmoid Function: S(x)=1/(1+e^-x)
  The Loss function is given as: **L(x)= -t*log(y) - (1-t)*(log(1-y))**

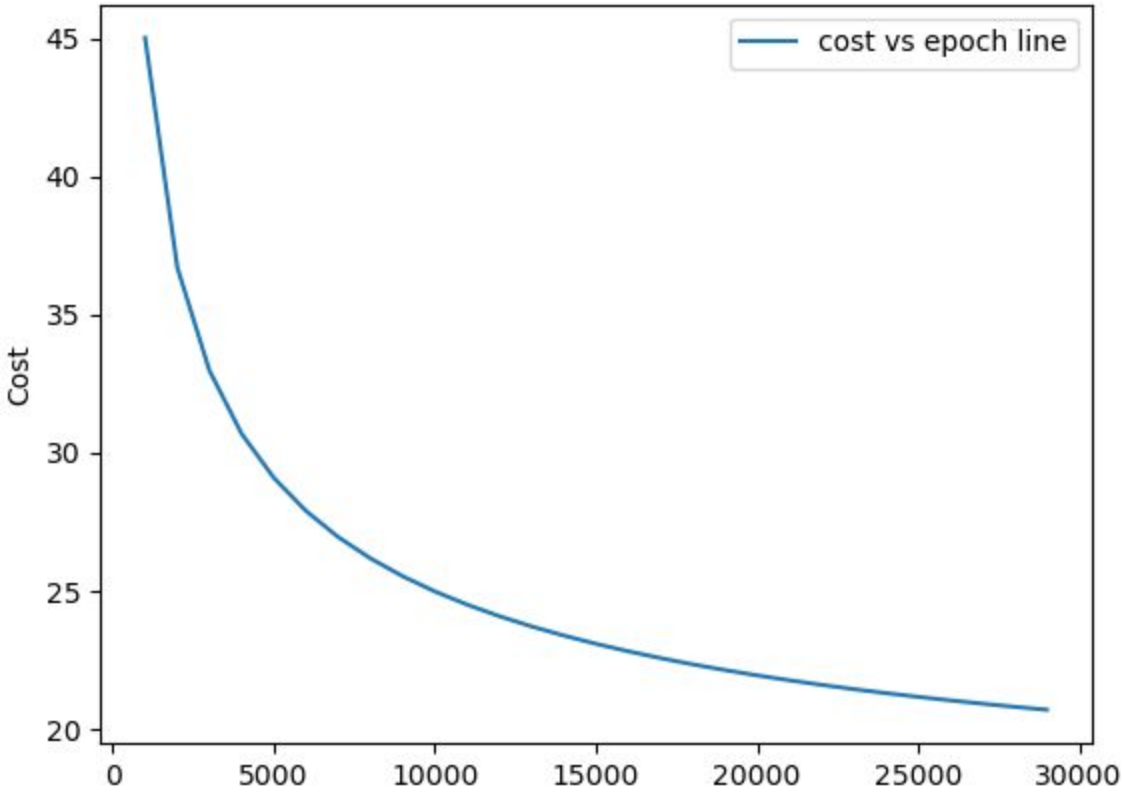
- ## Obserations

  1) **Learning rate=0.0005**
  Convergence rate=0.0001

Classification of points



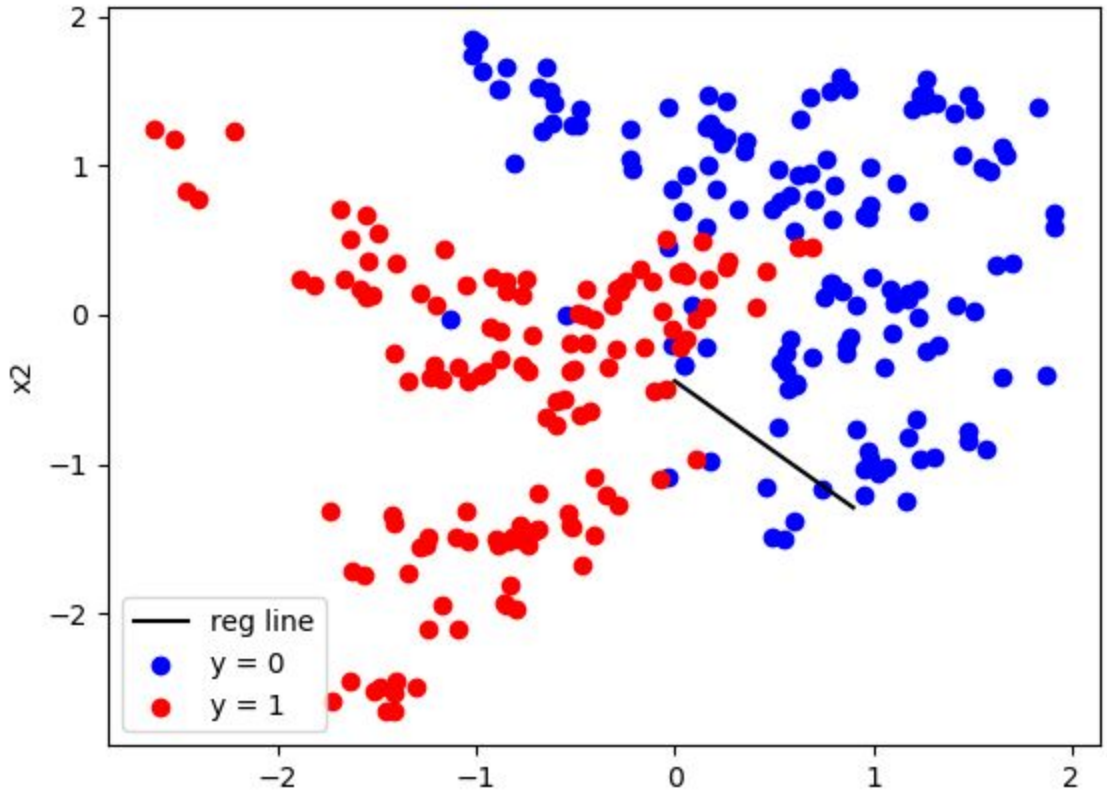
Cost vs Epoch



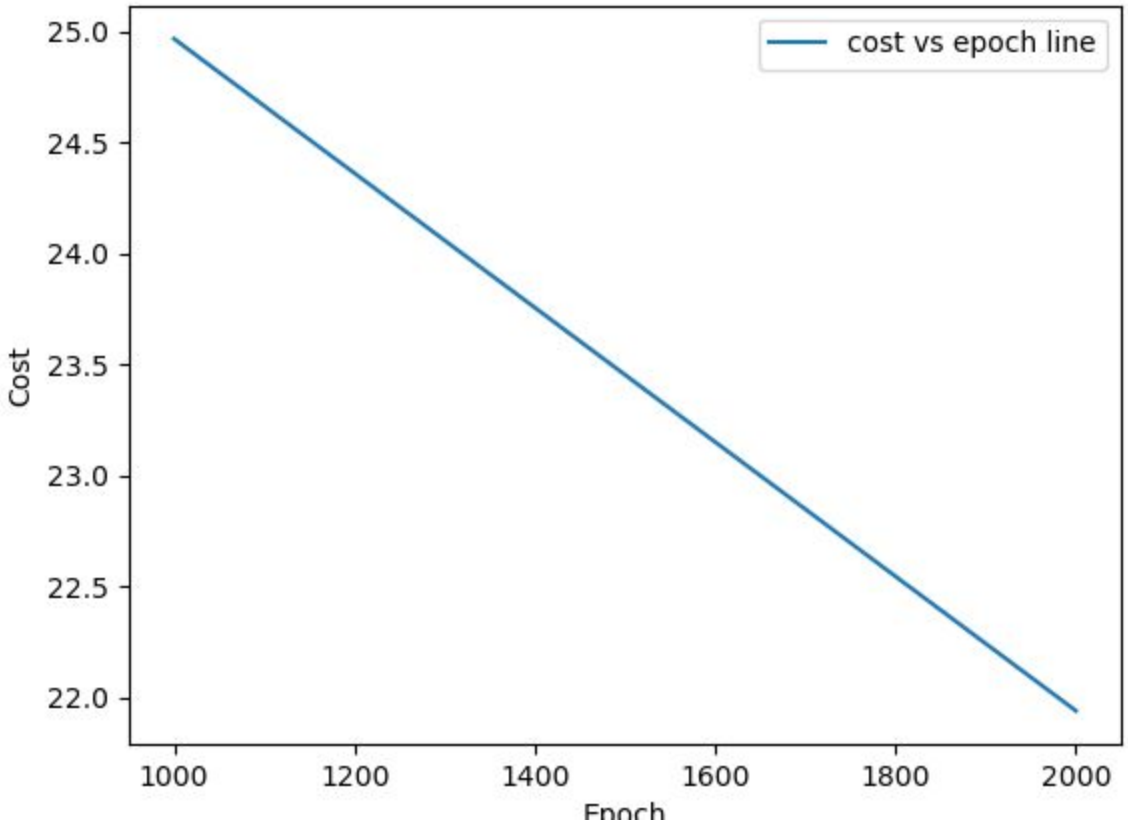
F-Score: 0.9886792452830188
Accuracy: 0.9890909090909091
Estimated regression coefficients: [[-6.21292708 -13.23125955 -14.04399947 -13.03591364 -0.02716142]]

2) **Learning rate=0.005**
Convergence rate=0.001

Classification of points



Cost vs Epoch



F-Score: 0.988593155893536
Accuracy: 0.9890909090909091
Estimated regression coefficients: [[-6.21372505 -13.23280798 -14.0455964 -13.03742507 -0.02723242]]

LOGISTIC REGRESSION(L1)

Introduction

We are required to analyze the logistic regression model with L1 regularization we built based on accuracy,f-score, recall, and precision that is generated by our dataset.
The model is built using numpy, sklearn and matplotlib.

Dataset

We have 1372 data points each having 4 features and we are required to classify these points as 1 or 0 ie. notes are forged or unforged.We standardize the given data and randomly split the dataset into test set and train set in the ratio 2:8.

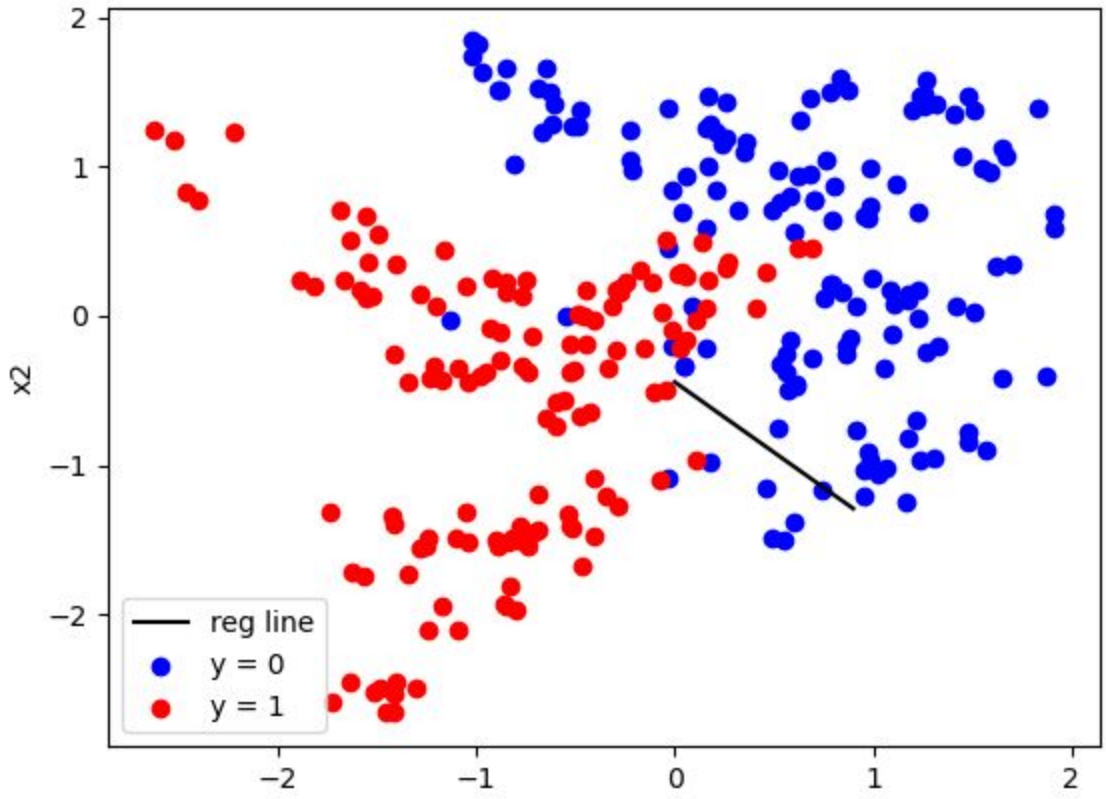
Equations Employed

Here we employ sigmoid function in order to calculate gradient of parameters.
Sigmoid Function: S(x)=1/(1+e^-x)
The Loss function is given as:
L(x)= -t*log(y) - (1-t)*(log(1-y)) + learning rate * (beta[i])
{where beta[i] is value of ith coefficient}

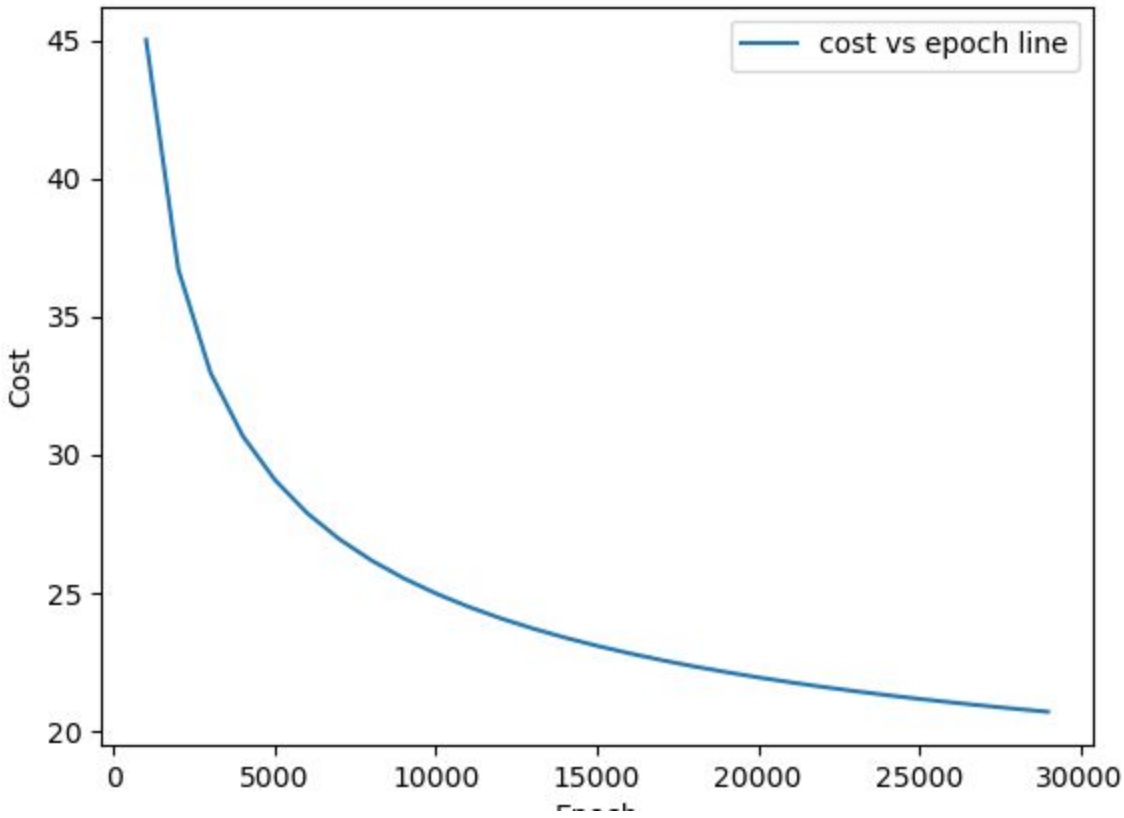
Observations

1) **Learning rate=0.0005**
Convergence rate=0.0001

Classification of points



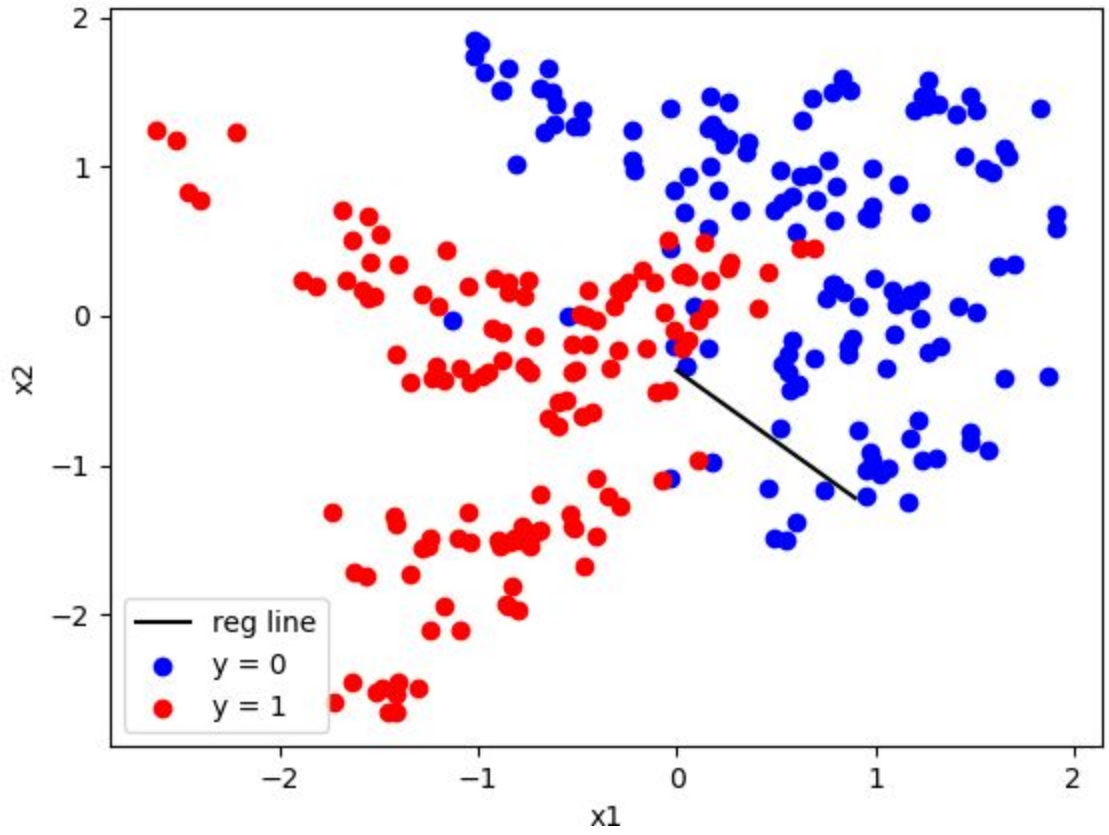
Cost vs Epoch



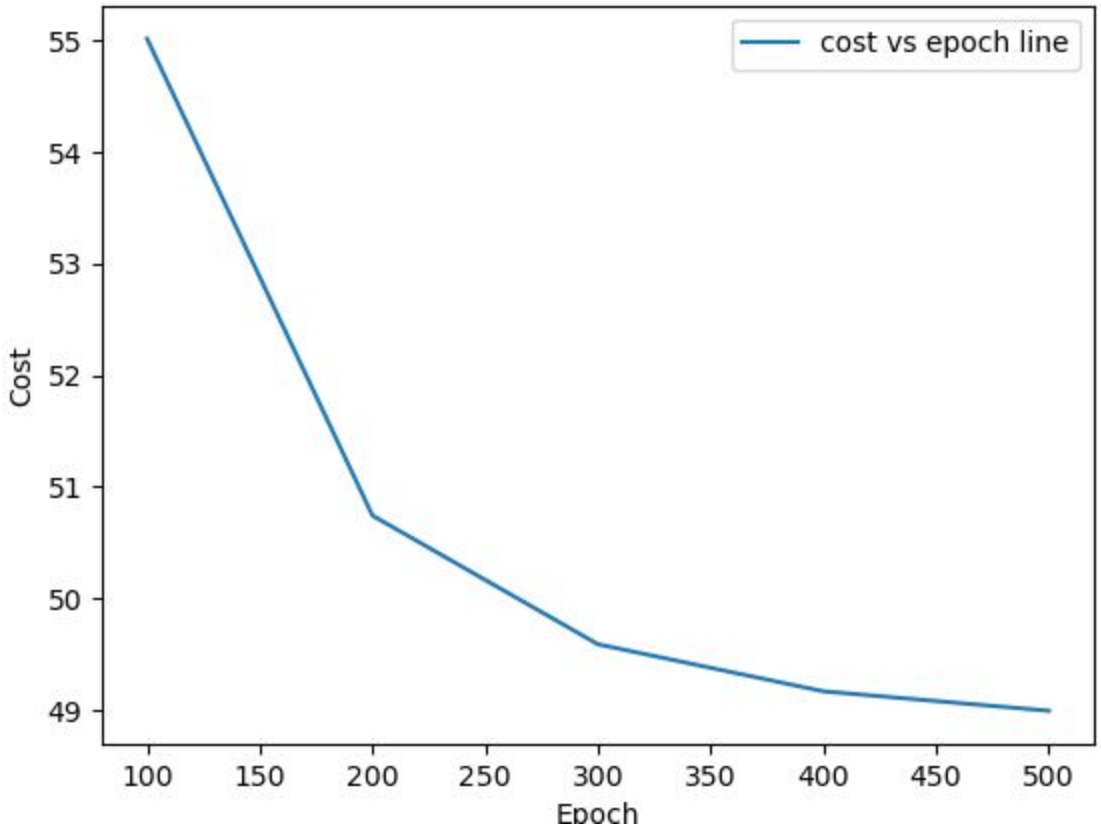
F-Score: 0.98792452830188
Accuracy: 0.98890909090909091
Estimated regression coefficients: [[-2.54176394 -6.6079528 -6.95584991 -6.41641277 0.26455088]]

2) **Learning rate=0.005**
Convergence rate=0.001

Classification of points



Cost vs Epoch



F-Score: 0.9812734082397004
Accuracy: 0.9818181818181818
Estimated regression coefficients: [[-2.54264496 -6.60933935 -6.95754007 -6.4179259 0.26446348]]

LOGISTIC REGRESSION(L2)

- ## Introduction

  We are required to analyze the logistic regression model with L2 Regularization we built based on accuracy,f-score, recall, and precision that is generated by our dataset.
  The model is built using numpy, sklearn and matplotlib.

- ## Dataset

  We have 1372 data points each having 4 features and we are required to classify these points as 1 or 0 ie. notes are forged or unforged. We standardize the given data and randomly split the dataset into a test set and train set in the ratio 2:8.

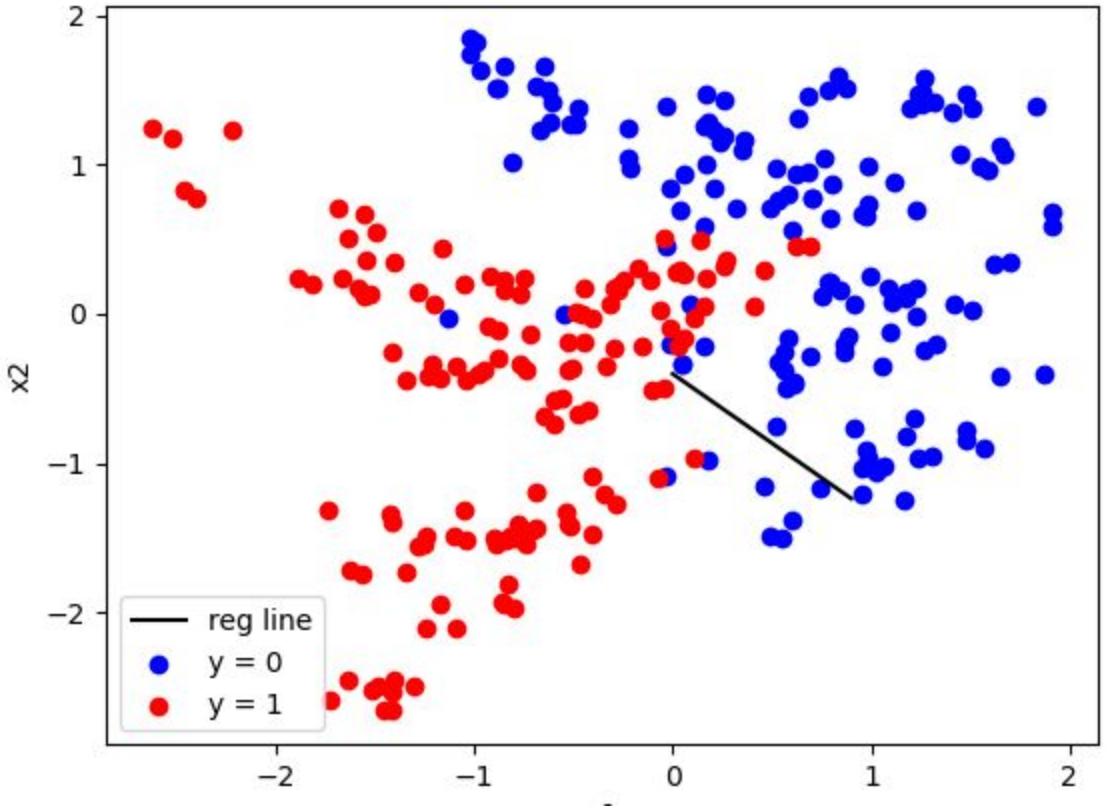
- ## Equations Employed

  Here we employ sigmoid function in order to calculate the gradient of parameters.
  Sigmoid Function: S(x)=1/(1+e^-x)
  The Loss function is given as:
  L(x)= -t*log(y) - (1-t)*(log(1-y)) + sigma(sign(beta[i])) * learning rate
  {where beta[i] is value of ith coefficient}

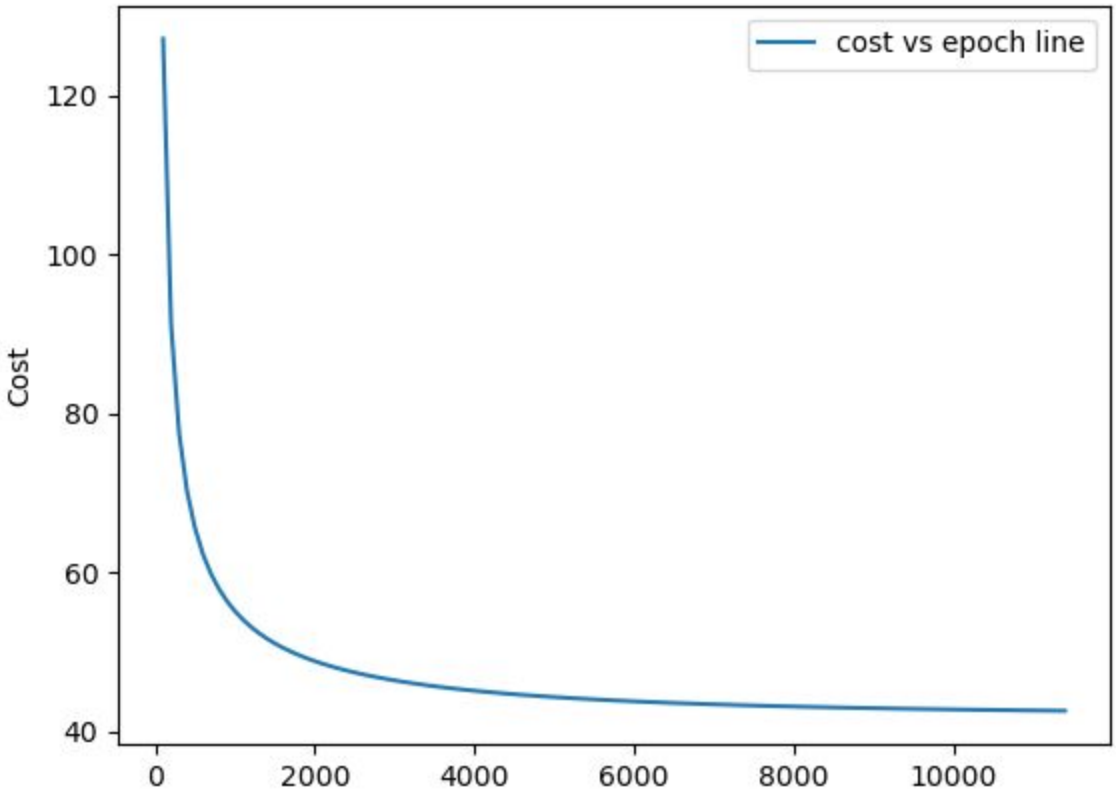
- ## Obserations

  1) Learning rate=0.0005
  Convergence rate=0.0001

Classification of points



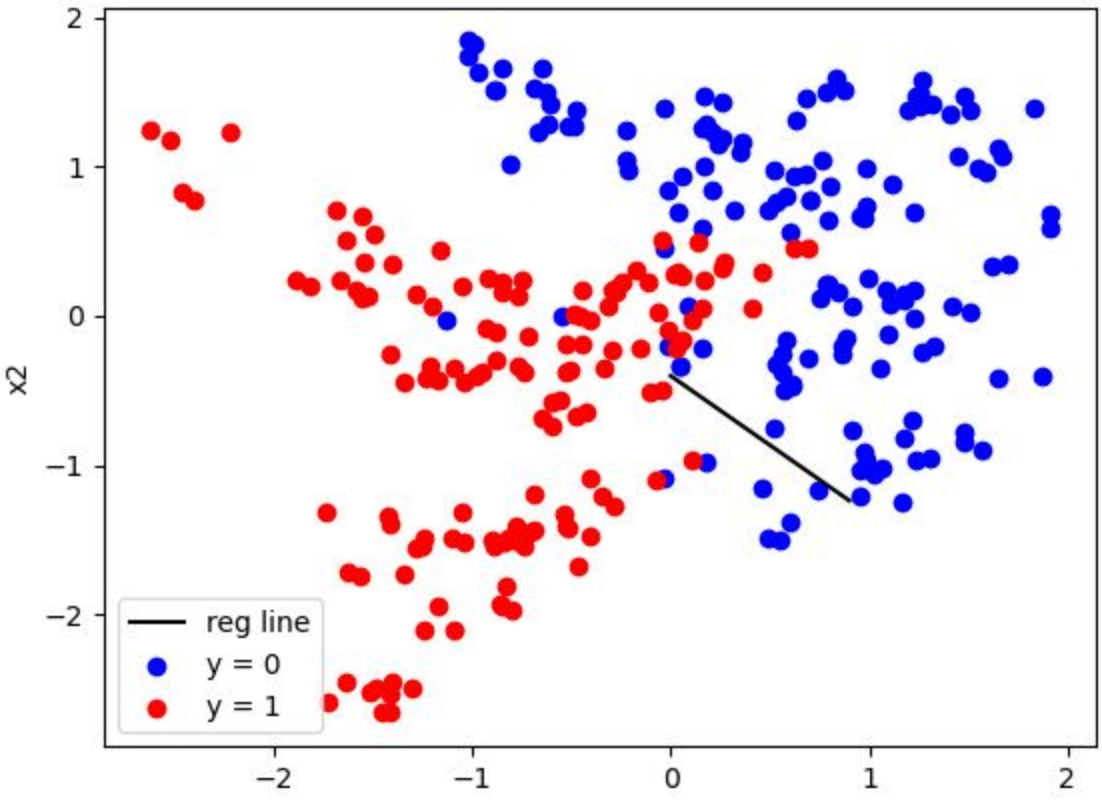
Cost vs Epoch



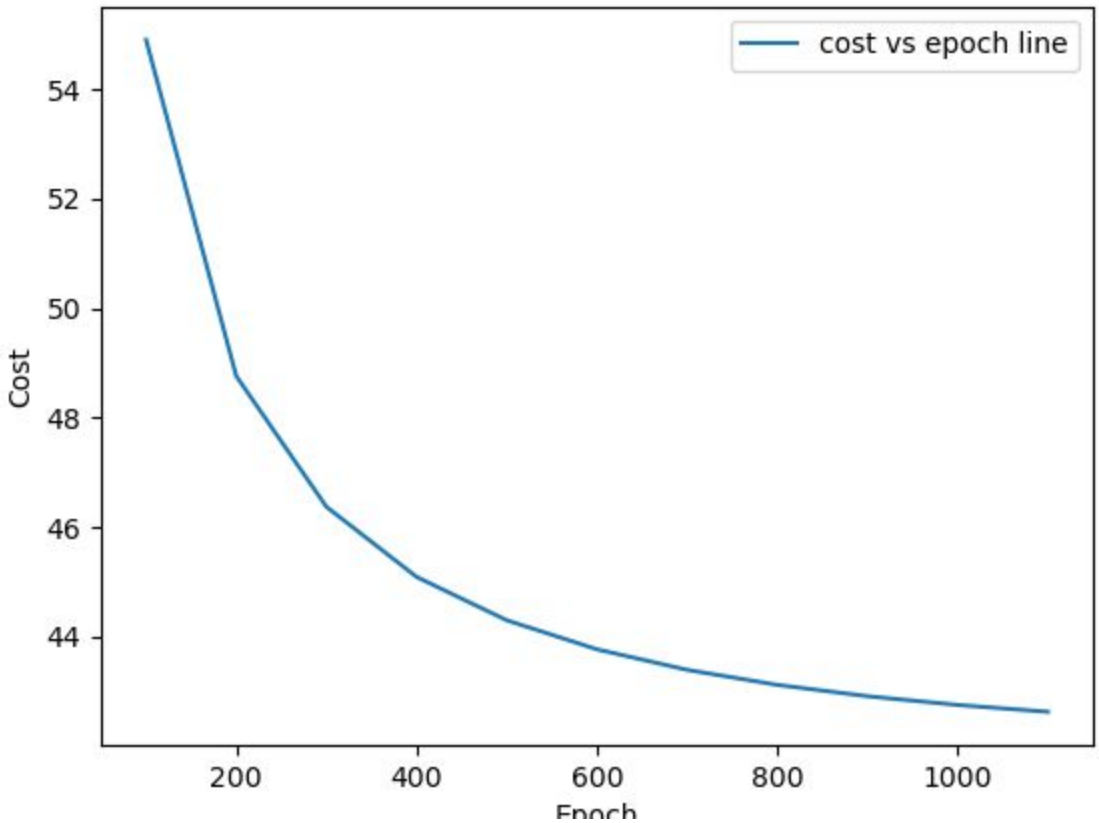
F-Score: 0.9886792452830188
Accuracy: 0.9890909090909091
Estimated regression coefficients: [[-3.67717078 -8.50117925 -9.16239591 -8.41381562 0.11568869]]

2) **Learning rate=0.005**
Convergence rate=0.001

Classification of points



Cost vs Epoch



F-Score: 0.9856792452830188
Accuracy: 0.9860909090909091
Estimated regression coefficients: [[-3.67769225 -8.50207365 -9.16338864 -8.41473007 0.11563314]]

Conclusion

The accuracy and F-score obtained with L2 regularization is more than that obtained in simple logistic regression as well as that obtained in L1 regularization. Thus putting constraints on parameters improves the performance of the model. We can clearly see that the absolute value of the third parameter is highest in all the cases ie. weightage of 3rd feature is highest among all the given features. Thus the decision of whether the note is forged or not is mostly dependent on 3rd feature. Similarly, the 1st feature is the least important deciding factor.