Assignment 2 - 4SL4
Parthav Patel
400251342
patep62

Seed Used: 1342
Batch Gradient Descent

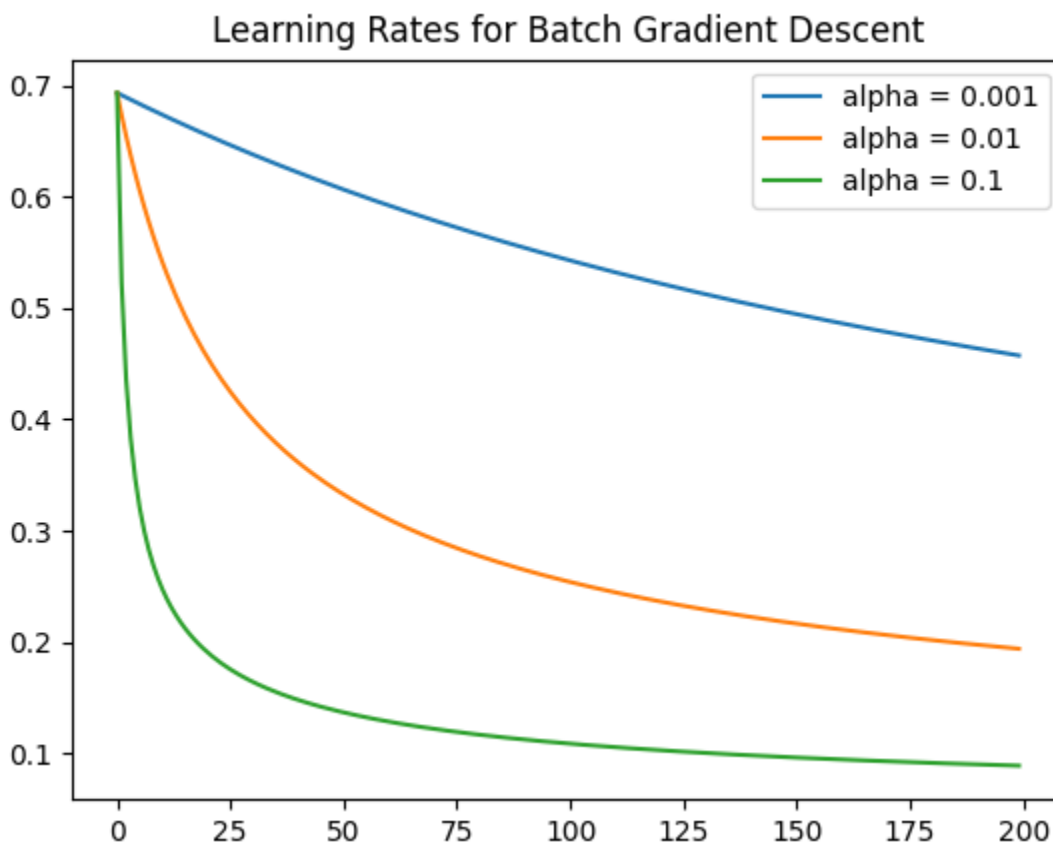### PR Curve for various Beta values

ROC Graph
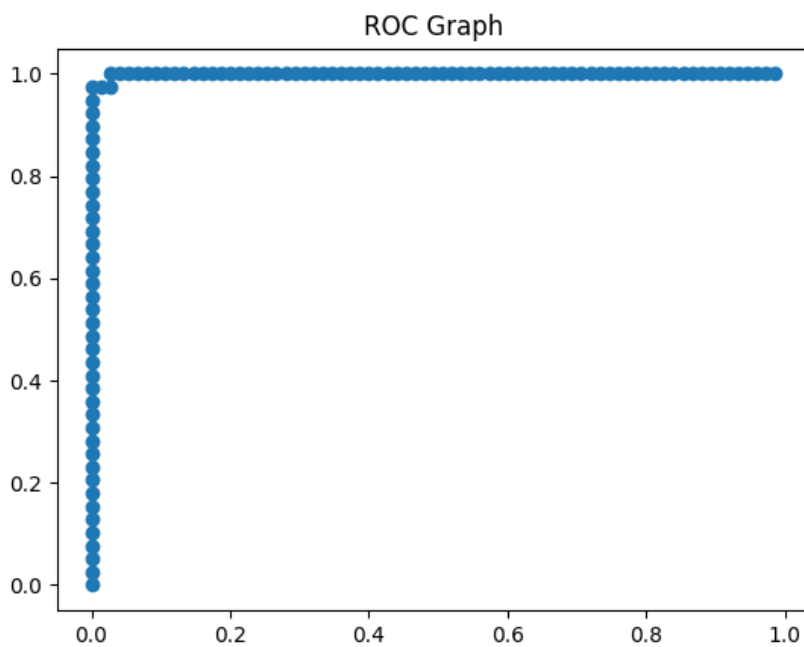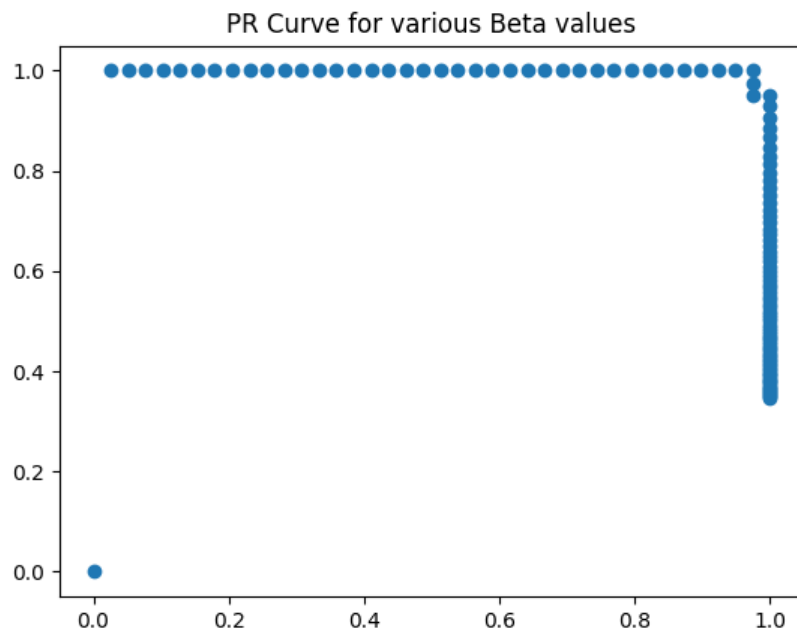
Lowest Misclassification rate: 0.008771929824561403
Highest F1 Score: 0.9870129870129869
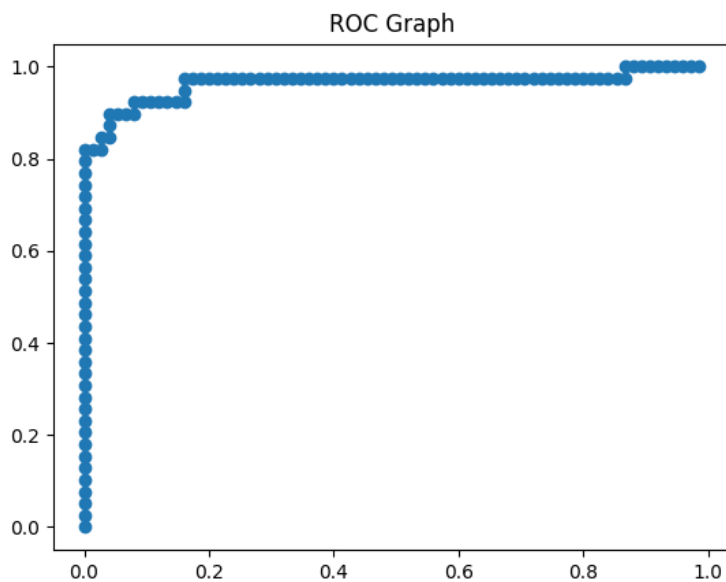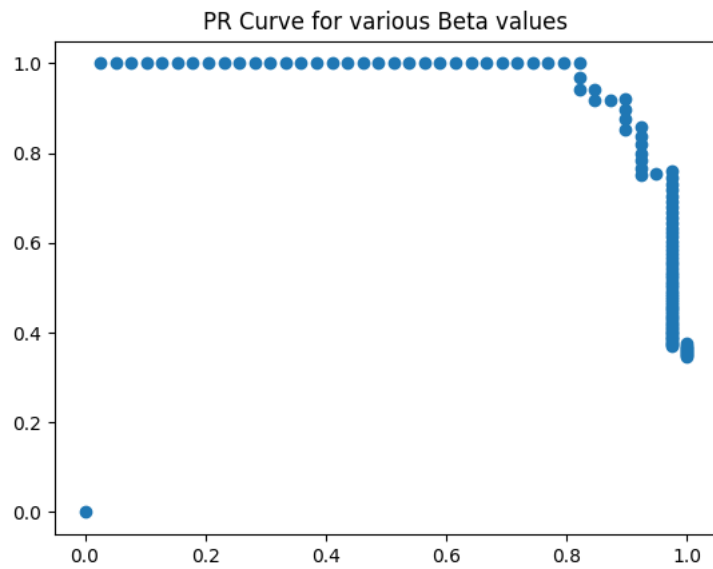
W vector parameters:

```
[ 0.          -0.46429681 -0.46879868 -0.45501313 -0.47807709 -0.131358
 -0.08462026 -0.38585154 -0.46702625 -0.06959066  0.2334463  -0.47560064
 -0.04458231 -0.38200904 -0.42028867 -0.04353836  0.21888215  0.05889141
 -0.07304947  0.18973584  0.26589422 -0.57185071 -0.58306687 -0.54133367
 -0.55941937 -0.40322824 -0.20072842 -0.38914828 -0.49131271 -0.32461387
 -0.11448711]
```

Learning Curves with Alpha1 = 0.001, Alpha2 = 0.01, Alpha3 = 0.1

# of Iterations for BGD: 200

Stochastic Gradient Descent:

### PR Curve for various Beta values

ROC Graph

W Vector Parameters

```
[ 0.          0.87739582 -0.92467205  0.33290743  2.14259123 -0.5224064
  3.25193943 -4.43403913 -5.91178054  2.43927076  0.68666608 -7.56438553
  1.56355063 -1.01195319 -1.89630008 -0.20622005  3.6276587  -0.07995155
 -5.6642076   3.90535129  5.22604278 -3.97964086 -4.10586985 -1.87851619
 -0.808991    0.165504    1.09739092 -2.5240855   0.26123953 -3.10492833
 -4.29435494]
```

Lowest Misclassification rate: 0.06140350877192982
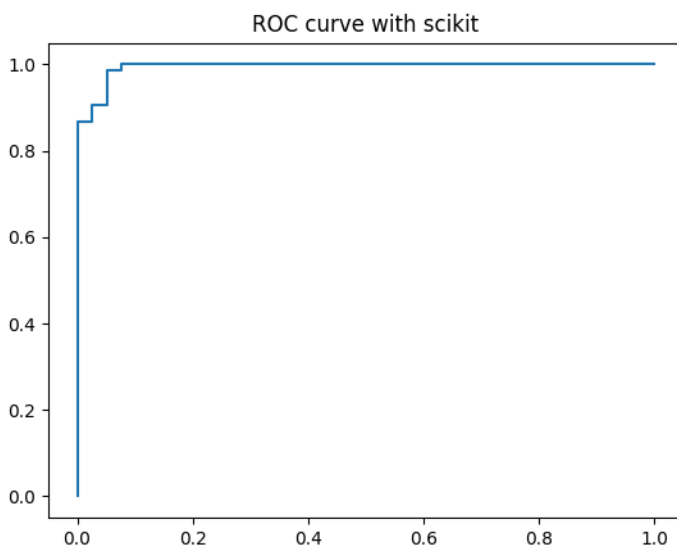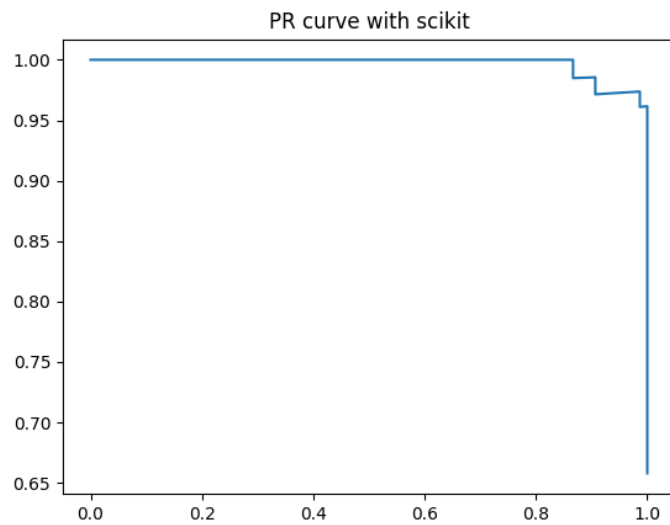Highest F1 Score: 0.9090909090909091

Learning Curves with Alpha1 = 0.001, Alpha2 = 0.01, Alpha3 = 0.1:
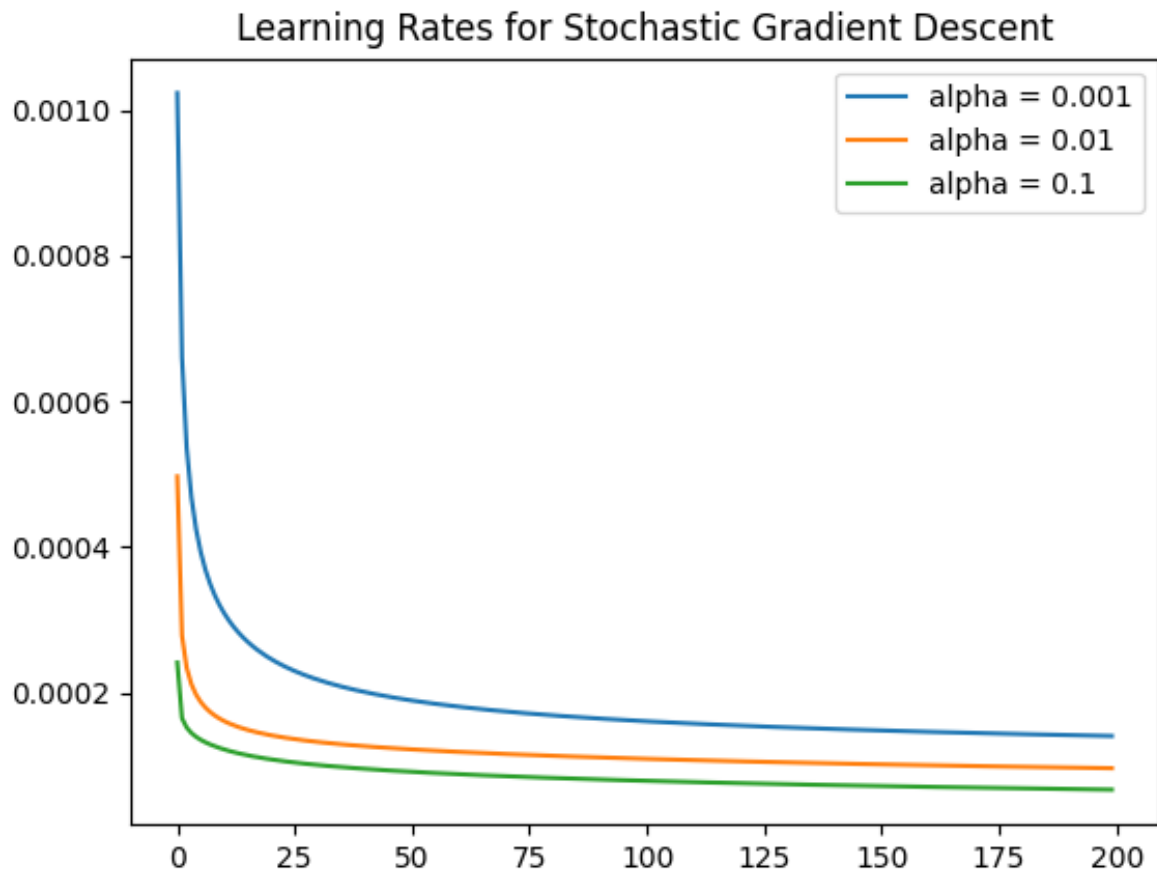
# of Epochs for SGD: 200

SCI-KIT Implementation:

Batch Gradient Descent:

PR curve with scikit

ROC curve with scikit

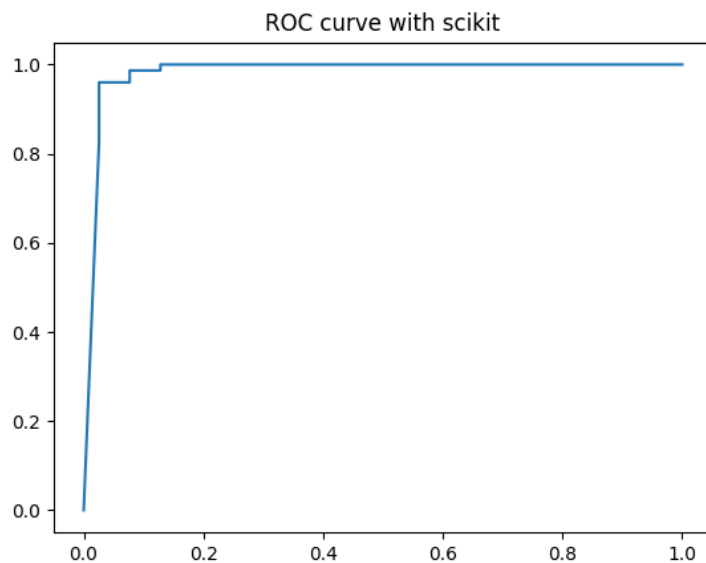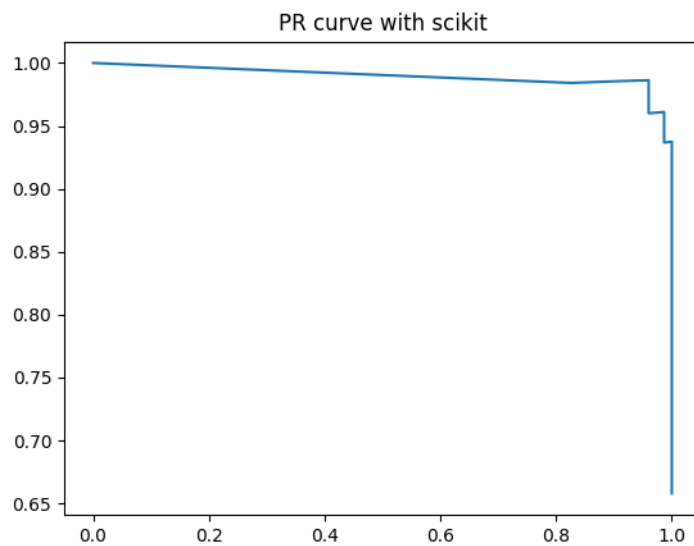W Vector Parameters

```
[[ 0.          -0.32680678 -0.50137818 -0.33026224 -0.40150338 -0.07433856
   0.41313898 -0.91588199 -0.91923994  0.2077316   0.20280371 -1.31559135
   0.15035841 -0.64887818 -0.8900624  -0.26249558  1.07982211 -0.19859072
  -0.49007728  0.73109224  0.8252055  -0.96168378 -1.19262385 -0.77933185
  -0.91855582 -0.50494724  0.05791358 -0.87414446 -0.77115855 -0.73936934
  -0.5082076 ]]
```

Lowest Misclassification rate: 0.03508771929824561
Highest F1 Score: 0.9733333333333334
# of Iterations for BGD: 200

Stochastic Gradient Descent:

PR curve with scikit

ROC curve with scikit

W Vector Parameters:

```
[[  0.          -1.17939679  -1.70931745  -1.13592159  -2.63800539
   -3.68197759   6.23155085 -12.09684314 -11.00917012   9.597362
   -0.43842244 -23.55552811   6.16817325  -8.84625954 -13.71361078
   -7.55604995  16.17811188   4.68401305  -5.96674456   6.52864085
   12.92289483 -13.05110237 -18.86933061  -8.62479392 -10.78464512
   -1.92201834   0.45115953 -10.50300451  -8.57864479 -13.91423917
   -7.40438557]]
```

Lowest Misclassification rate: 0.052631578947368474
Highest F1 Score: 0.96
# of Epochs for SGD: 200

Comparison:

Misclassification Rate:
With an equal # of iterations for both versions of BGD, my implementation performed significantly better with a misclassification rate of 0.008771929824561403 compared to scikit's 0.03508771929824561. However, scikit's implementation for SGD performed slightly better at 0.052631578947368474 compared to my 0.06140350877192982.

F1 Score:
All Scores for F1 rate fall between the range of 0.9 to 1, with my implementation of BGD performing the best at 0.9870129870129869. While scikit's implementation of SGD performed better than my version. These results align with the conclusions drawn from the misclassification rates.

PR and ROC Curves:
There are no visible differences between scikit's and my implementation of BGD and SGD.