COMPENG 4SL4

Assignment 2

Instructor: Dr. Dumitrescu

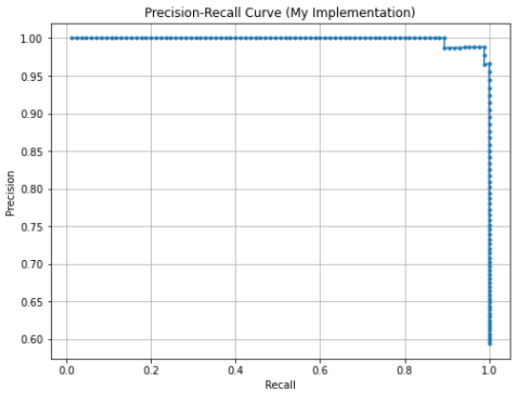
Hritheekka Chinnakonda – chinnakh – 400292782 – C01 – L03

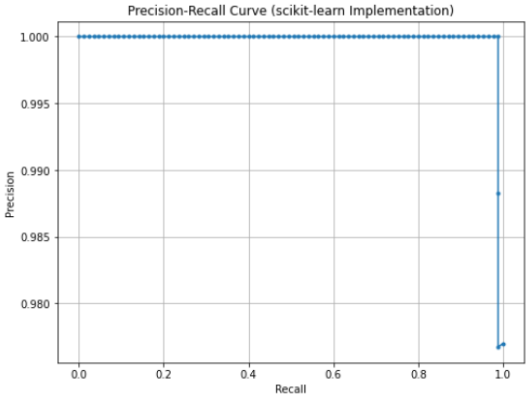
As a future member of the engineering profession, the student is responsible for performing the required work in an honest manner, without plagiarism and cheating. Submitting this work with my name and student number is a statement and understanding that this work is my own and adheres to the Academic Integrity Policy of McMaster University and the Code of Conduct of the Professional Engineers of Ontario. Submitted by [Hritheekka Chinnakonda, chinnakh, 400292782]

Vector of Parameters:

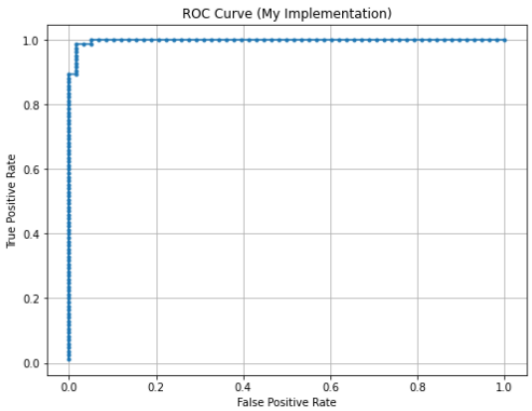
	Vector of Parameters
My Implementation	[-0.23234412 -0.73467356 -0.2984505 -0.48802305 -0.31508491 1.30561492 -1.04024892 -1.21479167 0.57097103 0.01364804 -1.89996626 0.3804992 -1.34372941 -1.53794046 -0.37459183 0.99267396 0.15474089 -0.30189868 0.34411568 1.15740993 -1.61574694 -1.90498077 -1.46089806 -1.68205399 -1.22209231 0.30875204 -1.16726401 -1.10589259 -1.50431249 -0.84907469]
Scikit-learn Implementation	[-0.38636982 -0.59761858 -0.41848032 -0.47761004 -0.12799587 0.50784832 -0.75277025 -0.821396 0.19676161 0.17626671 -0.94489501 0.16768047 -0.68938702 -0.79427393 -0.25791574 0.61806478 0.11145375 -0.12590233 0.2052673 0.6044158 -1.0081727 -1.1954253 -0.93714499 -1.00607813 -0.82247658 0.06621283 -0.74554489 -0.76520825 -0.92037128 -0.40983316]

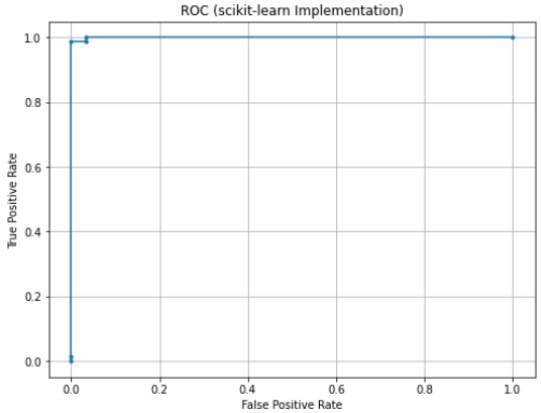
PR Curves:





ROC Curves:





Weight vector w and the train and test misclassification rates and F1 scores:

	Misclassification Error	F1-Score
My Implementation	0.018779342723004744	0.9822485207100591
Scikit-learn Implementation	0.020979020979020935	0.9820359281437125

For your model, you must also plot learning curves. For each method of training (i.e, BGD and SGD) select at least three values of the learning rate: one for which the learning algorithm does not converge (α1), one for which the learning is too slow (α2) and one for which the learning is fast enough (α3). Then plot the three learning curves in the same figure (you should have one figure for BGD and one for SGD).

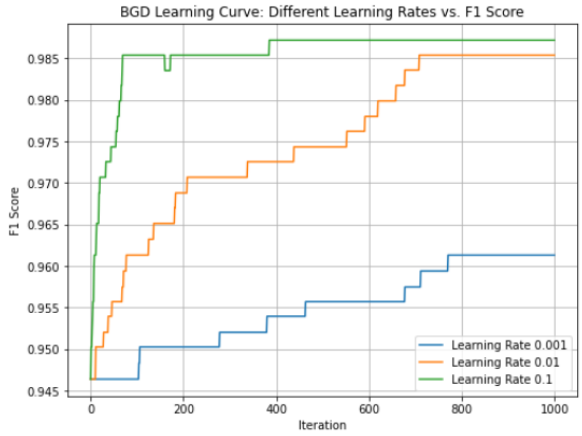
Values of the learning rate [α1, α2, α3]:

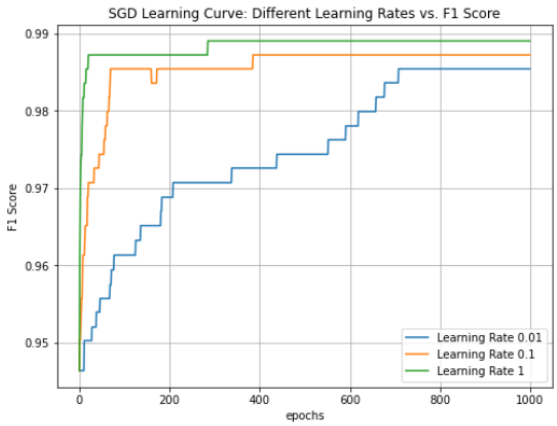
BGD learning rates = [0.001, 0.01, 0.1]

SGD learning rates = [0.01, 0.1, 1]

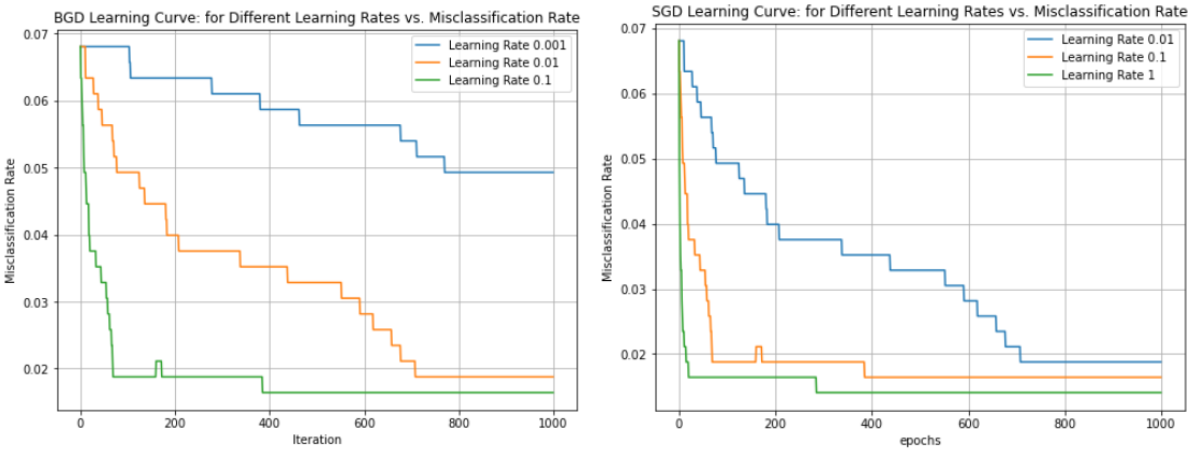
I chose these values because BGD typically uses smaller learning rates, while SGD can benefit from using a larger learning rate.

Metric: F1 Score





Metric: Misclassification Rate



Number of iterations to converge to α3: 780

Number of epochs to converge to α3: 700

How do the results obtained with your implementation compare with the results obtained using scikit-learn? Which of the two models is the best overall? To decide this, use first the misclassification rate as the metric, next use the F1 score as the metric. Is the winner the same in the two cases?

Overall, the scikit-learn implementation is better than my implementation. My misclassification rate is significantly lower than the scikit-learn implementation while the F1 score is very similar, but the scikit-learn implementation is slightly lower. The winner is not the same when looking at each metric separately.

Then compare the two models using the PR curves, then the ROC curves. Is the PR curve of one model better than for the other model? Is the ROC curve of one model better than for the other model?

The scikit-learn implementation is better in both the PR and ROC curves compared to my implementation is more precise.