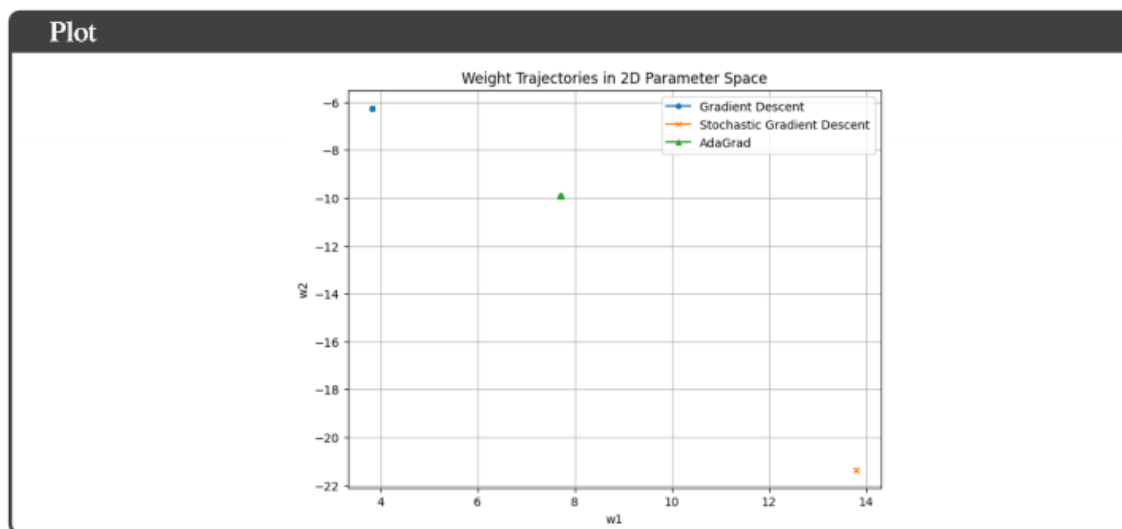


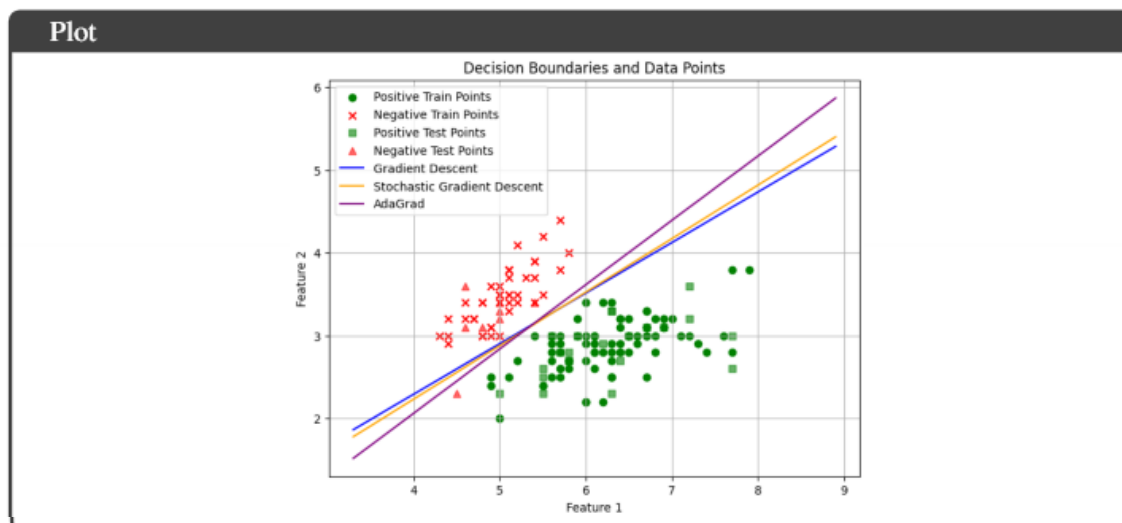
2 Unconstrained Loss Optimization for ML (Programming) (15 points)

In this question, you will implement *Gradient Descent*, *Stochastic Gradient Descent*, and *AdaGrad*. Please code the update rules for each of these optimization techniques in the `loss_optimization.py` file provided to you. After you are done, you will submit your code to Gradescope, where we will run your code on a full suite of tests and the autograder will assign your code points based on whether it passes the tests. After implementing the update rules, please answer the following questions:

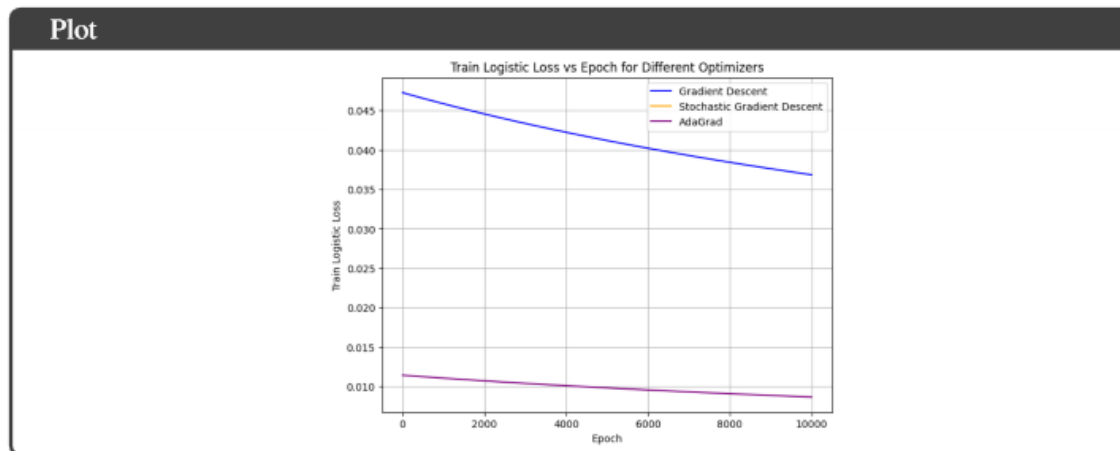
- (3 points) Create a plot for the 2D parameter space w_1, w_2 corresponding to the non-bias weight parameters. These are the second and third elements in the weight vector since the first element w_0 corresponds to the bias term. In this plot, show the trajectory of the weight parameters captured at the end of each epoch as they are updated during the training process. The plot should include three trajectories, one for each of the three parameter update rules. Include a legend indicating which trajectory corresponds to which optimizer.



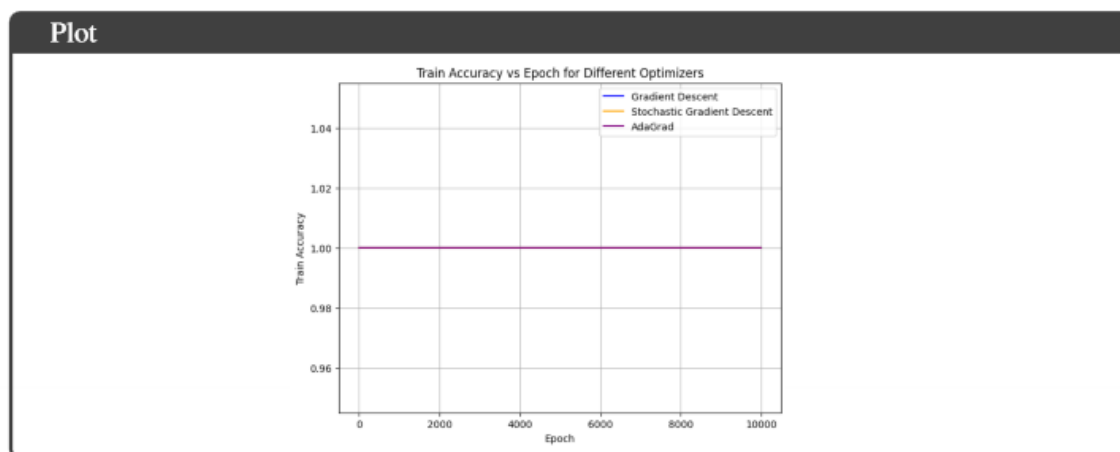
- (3 points) Plot the two-dimensional data with the positive datapoints (label=1) in green and negative datapoints (label=0) in red. Display the train and test datapoints using different market shapes. Also include in the plot the classifier decision boundary learned by the three different optimizers. Include a legend indicating which decision boundary corresponds to which optimizer.



3. (3 points) Plot the curve for train logistic loss versus train epoch for each of the three optimizers in a single plot. Include a legend indicating which curve corresponds to which optimizer.



4. (3 points) Plot the curve for train accuracy versus train epoch for each of the three optimizers in a single plot. Include a legend indicating which curve corresponds to which optimizer.



5. (1 point) What is the accuracy obtained by the gradient descent optimizer on the test data?

Solution

0.967

Test Accuracy (Gradient Descent): 0.966666666666667

Test Accuracy (Stochastic Gradient Descent): 0.966666666666667

Test Accuracy (AdaGrad): 0.966666666666667

6. (1 point) What is the accuracy obtained by the stochastic gradient descent optimizer on the test data?

Solution

0.967

Test Accuracy (Gradient Descent): 0.966666666666667

Test Accuracy (Stochastic Gradient Descent): 0.966666666666667

Test Accuracy (AdaGrad): 0.966666666666667

7. (1 point) What is the accuracy obtained by the Adagrad optimizer on the test data?

Solution

0.967