ECE421: Introduction to Machine Learning — Fall 2024

Assignment 2: Gradient Descent, Multiclass Logistic Regression, and K-Means

Due Date: Friday, October 18, 11:59 PM

General Notes

1. Programming assignments can be done in groups of up to 2 students. Students can be in different sections.
2. Only one submission from a group member is required.
3. Group members will receive the same grade.
4. Please post assignment-related questions on Piazza.

Turning It In

You need to submit your version of the following files:

- `myTorch.py`
- `PA2_qa.pdf` that answer questions related to the implementations.
- The cover file with your name and student ID filled (it can be as the first page of your `PA2_qa.pdf` or as a separate PDF file.)

Please pack them into a single folder, compress into a zip file and name it as PA2.zip. Please submit the zip file to Quercus.

Group Members

Name (and Name on Quercus)	UTORid

1 Gradient Descent

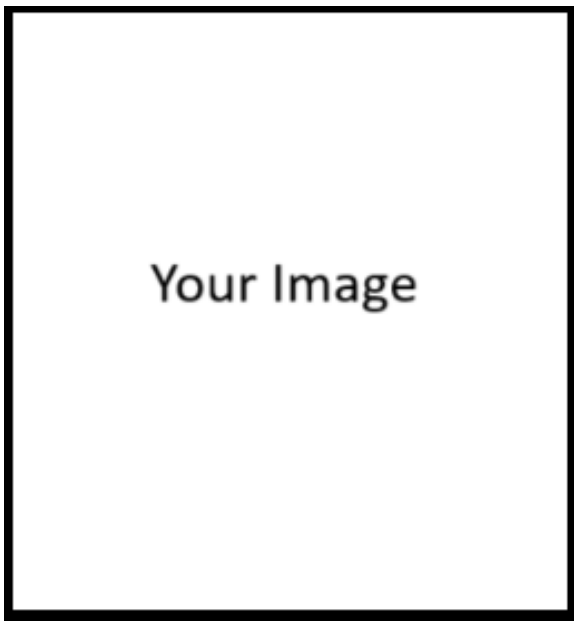
1.1 `Optimizer.sgd` method

1.1.a Test function `q1()`.

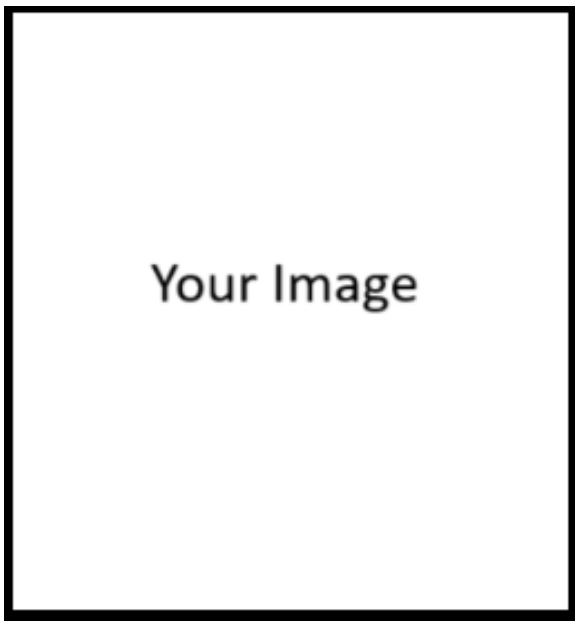
1.1.a.i Describe the termination criteria used in the `test_sgd` function in the `tests_A2.py` file. (1 marks)

Answer. Your answer ...

1.1.a.ii Include the figures generated by `q1()` in your `PA2_qa.pdf` file. (1 marks)



(a) SGD decision parameter trace.



(b) SGD Loss vs. iteration.

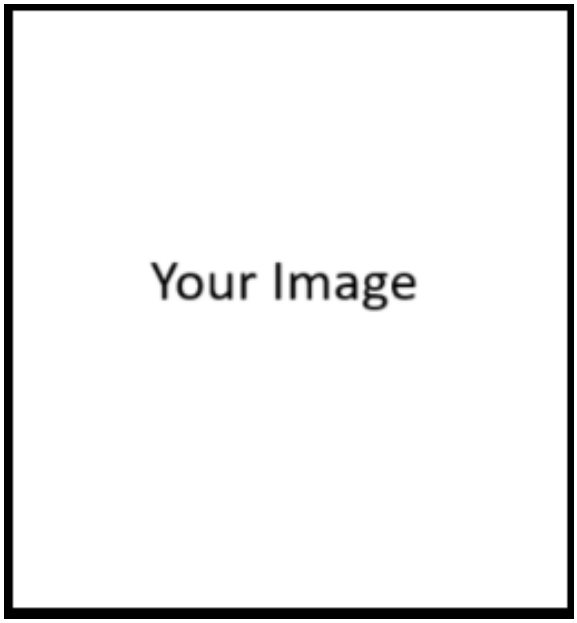
Figure 1: Figures generated by `q1()`.

1.1.a.iii With learning rate $\eta = 0.05$, what would be the value of w_1, *i.e.*, after one iteration of SGD update. Show your mathematical process. If you implemented SGD correctly, the figures generated by `q1()` should verify your w_1. (1 marks)

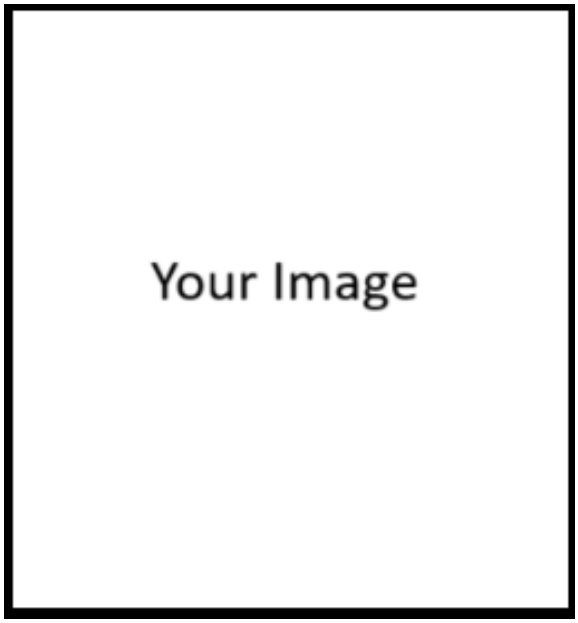
Answer. Your answer ...

1.1.b Test function `q2()`.

1.2.b.i Include the figures generated by `q2()` in your `PA2_qa.pdf` file. (1 marks)



(a) SGD decision parameter trace.



(b) SGD Loss vs. iteration.

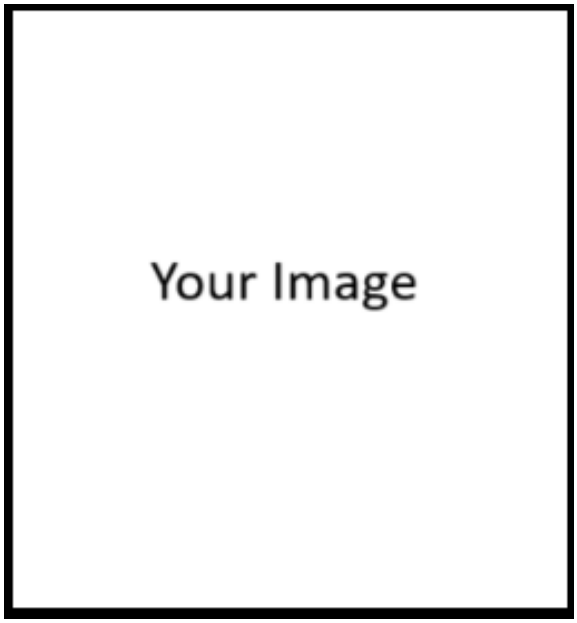
Figure 2: Figures generated by `q2()`.

1.2.b.ii When $\eta = 0.05$, SGD would fail to converge to the optimal solution. What causes such behavior? (1 marks)

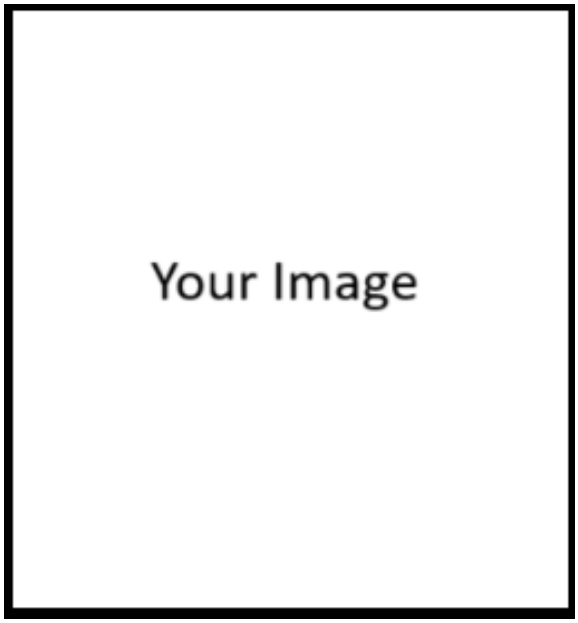
Answer. Your answer ...

1.1.c Test function `q3()`.

1.2.c.i Include the figures generated by `q3()` in your `PA2_qa.pdf` file. (1 marks)



(a) SGD decision parameter trace.



(b) SGD Loss vs. iteration.

Figure 3: Figures generated by `q3()`.

1.2.c.ii In 1-2 sentences describe the behavior of SGD in `q3()` when $\eta = 0.001, 0.005$, and 0.01. Explain why SGD fails to find the global optimum point? (1 marks)

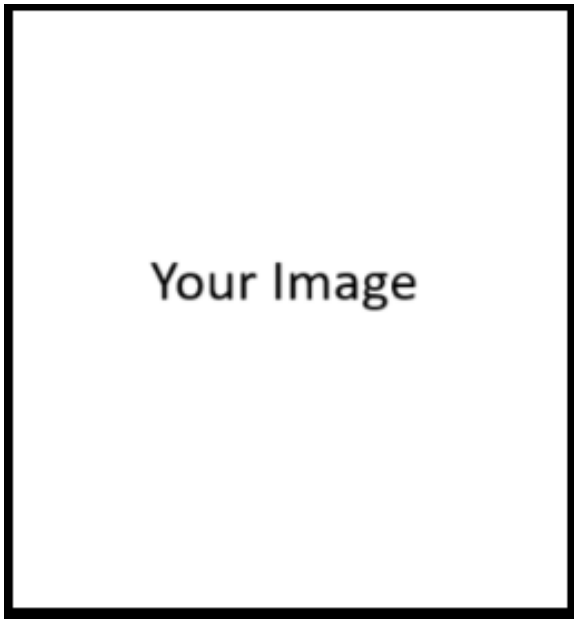
Answer. Your answer ...

1.2.c.iii In 1-2 sentences describe the behavior of SGD in `q3()` when $\eta = 0.05$. (1 marks)

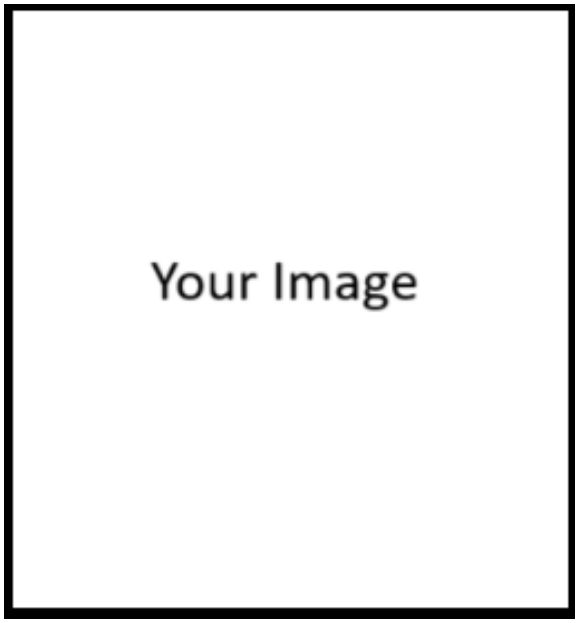
Answer. Your answer ...

1.1.d Test function `q4()`.

1.2.d.i Include the figures generated by `q4()` in your `PA2_qa.pdf` file. (1 marks)



(a) SGD decision parameter trace.



(b) SGD Loss vs. iteration.

Figure 4: Figures generated by `q4()`.

1.2.d.ii In 1-2 sentences describe the behavior of SGD in `q3()` when $\eta = 0.005$ and 0.01. How is this behavior related to the stretched nature of the function $f(\underline{w})$? (1 marks)

Answer. Your answer ...

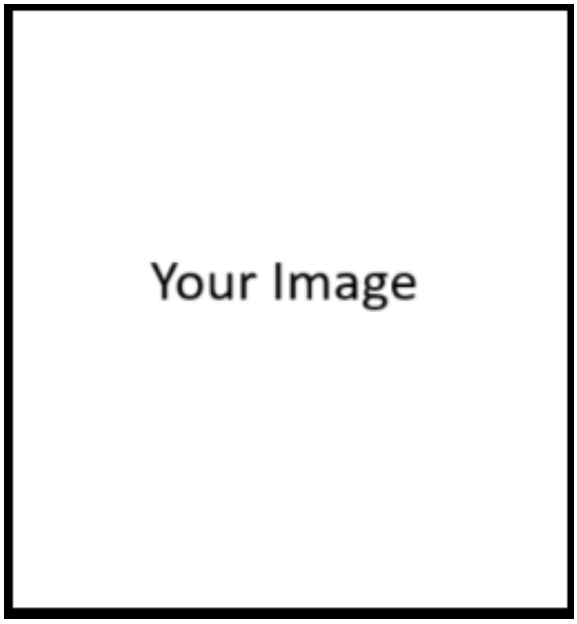
1.2.d.iii In 1-2 sentences describe the behavior of SGD in `q3()` when $\eta = 0.03$. (1 marks)

Answer. Your answer ...

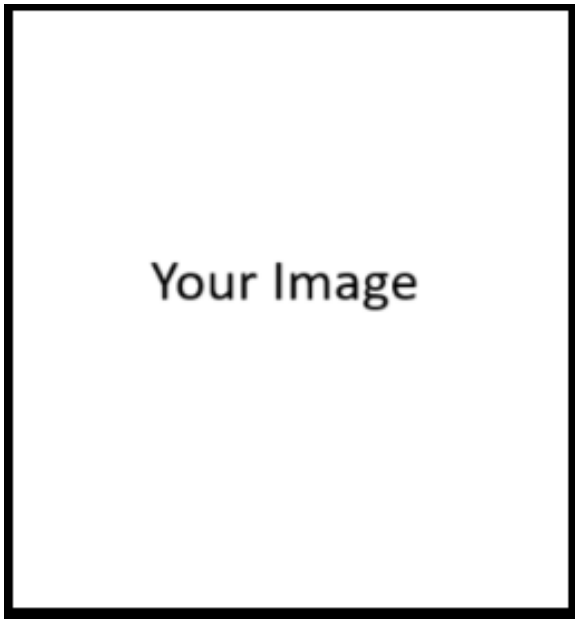
1.2 `Optimizer.heavyball_momentum` and `Optimizer.nestrov_momentum` methods

1.2.a Test function `q5()`.

1.2.a.i Include the figures generated by `q5()` in your `PA2_qa.pdf` file. (1 marks) use proper address to your png files



(a) Heavy-ball momentum decision parameter trace.



(b) Heavy-ball momentum Loss vs. iteration.

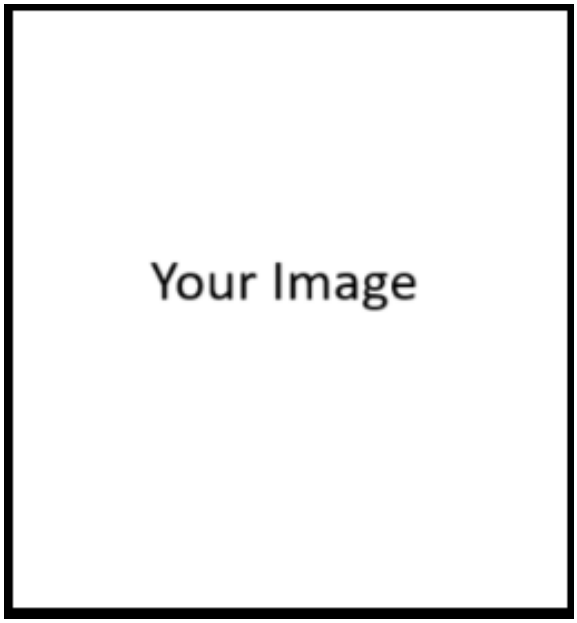
Figure 5: Figures generated by q5().

1.2.a.ii In 1-2 sentences, compare the performance of SGD with and without heavy-ball momentum by comparing the outcome of tests `q3()` and `q5()` (2 marks)

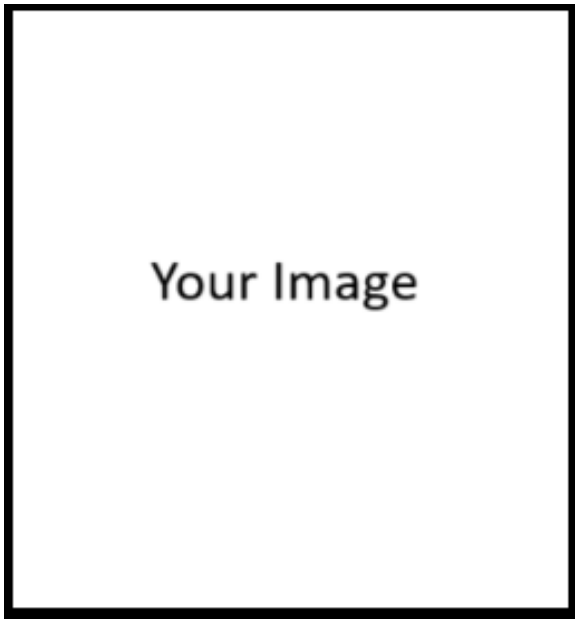
Answer. Your answer ...

1.2.b Test function `q6()`.

1.2.b.i Include the figures generated by `q4()` in your `PA2_qa.pdf` file. (1 marks)



(a) Heavy-ball momentum decision parameter trace.

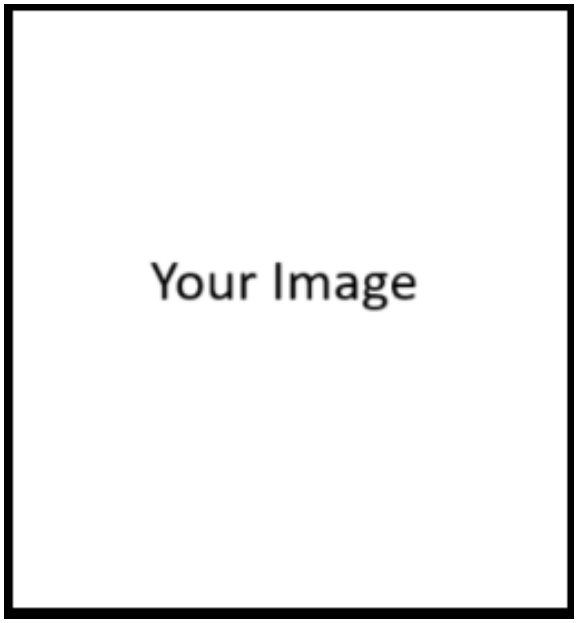


(b) Heavy-ball momentum Loss vs. iteration.

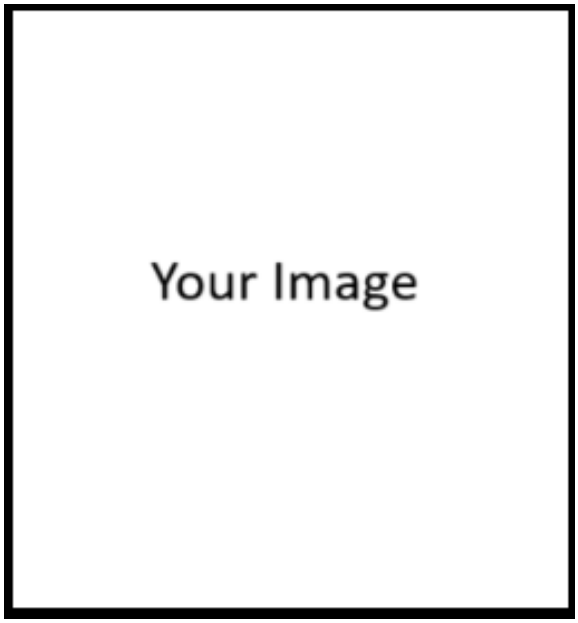
Figure 6: Figures generated by q6().

1.2.c Test function `q7()`.

1.2.c.i Include the figures generated by `q5()` in your `PA2_qa.pdf` file. (1 marks)



(a) Nestrov momentum decision parameter trace.

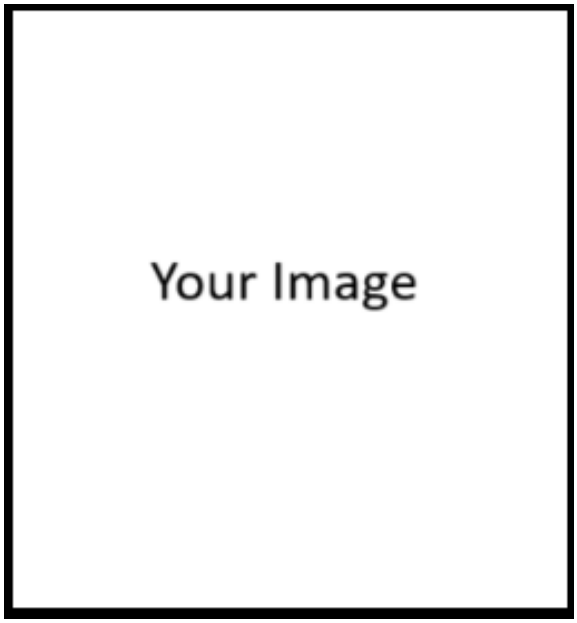


(b) Nestrov momentum Loss vs. iteration.

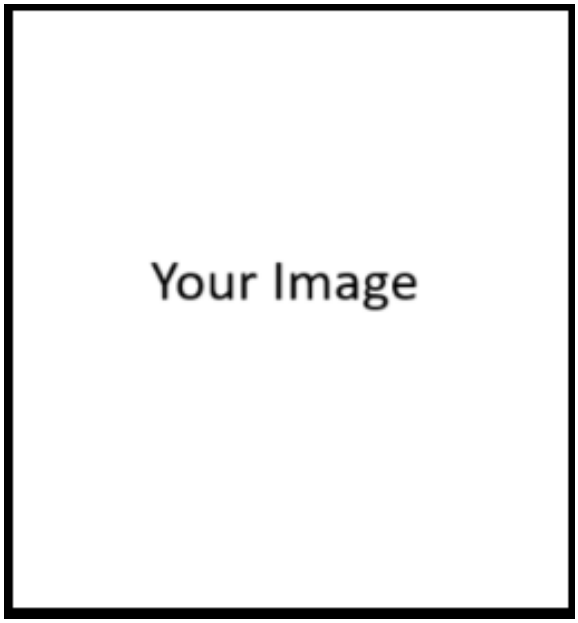
Figure 7: Figures generated by q7().

1.2.d Test function `q8()`.

1.2.d.i Include the figures generated by `q4()` in your `PA2_qa.pdf` file. (1 marks)



(a) Nestrov momentum decision parameter trace.



(b) Nestrov momentum Loss vs. iteration.

Figure 8: Figures generated by q8().

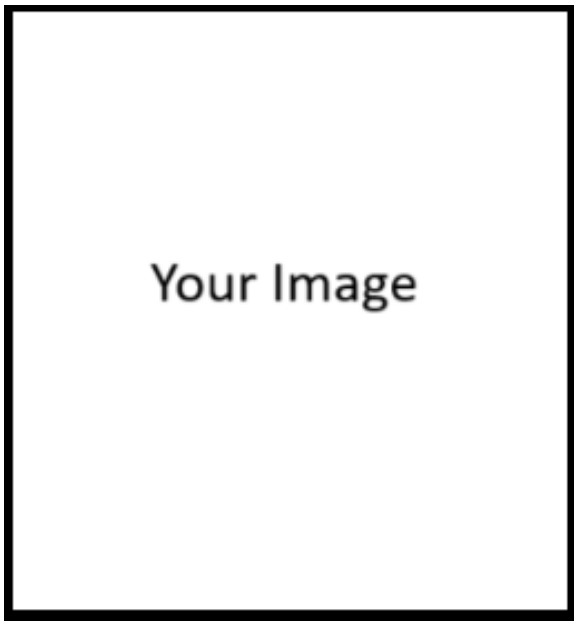
1.2.d.ii In 1-2 sentences, compare the performance of Nestrov Momentum with the heavy-ball momentum by comparing the outcome of tests `q5()` and `q6()` with that of `q7()` and `q8()`. (1 marks)

Answer. Your answer ...

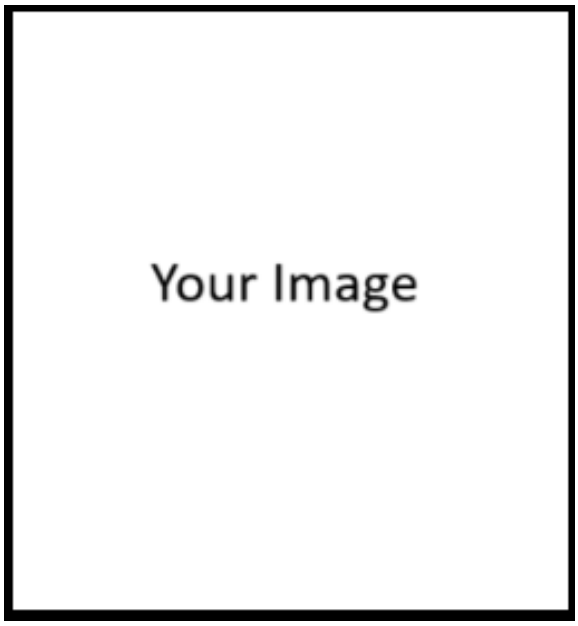
1.3 `Optimizer.adam` method

1.3.a Test function `q9()`

1.3.a.i Include the figures generated by `q9()` in your `PA2_qa.pdf` file. (1 marks)



(a) Adam momentum decision parameter trace.



(b) Adam momentum Loss vs. iteration.

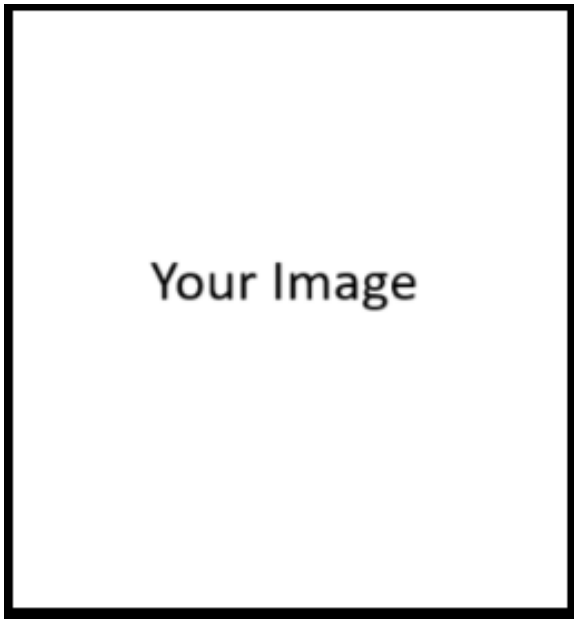
Figure 9: Figures generated by q9().

1.3.a.ii In 1-2 sentences, compare the performance of adam with momentum method (heavy-ball or Nestrov) (2 marks)

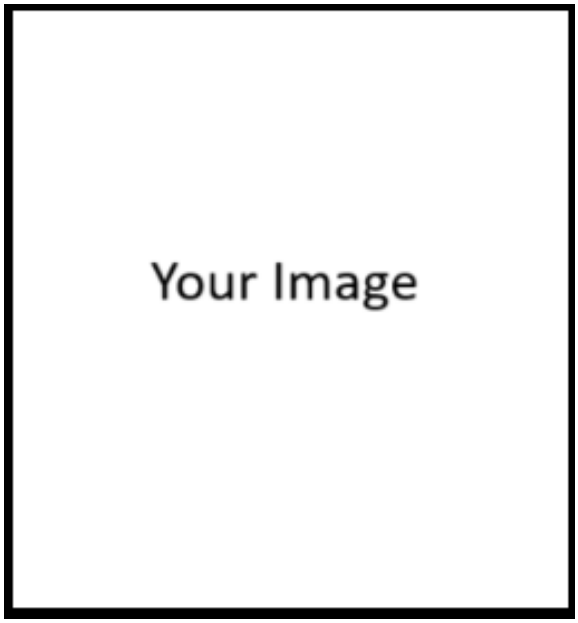
Answer. Your answer ...

1.3.b Test function `q10()`.

1.3.b.i Include the figures generated by `q10()` in your `PA2_qa.pdf` file. (1 marks)



(a) Adam momentum decision parameter trace.



(b) Adam momentum Loss vs. iteration.

Figure 10: Figures generated by q10().

1.3.b.ii Based on the outcome of `q9()` and `q10()`, describe the advantage of Adam in 1-2 sentence. (2 marks) [**HINT:** run `q11()` to see what could be the impact of scaling the function (or gradients) on the other optimization method such as gradient descent with Nestrov Momentum. You don't need to report the output of `q11()` in your report. Also, note that `q11()` would most often result in error. Don't worry. That is intentional. Try to understand why this happens.]

Answer. Your answer ...

2 Multiclass Logistic Regression

2.1 Implementing the Learning Model

No written part.

2.2 Implementing the Learning Algorithm

2.2.a The test function `q22()` runs your implementation on the Iris dataset.

2.2.a.i Include the figures generated by `q22()` in your `PA2_qa.pdf` file. (2 marks)

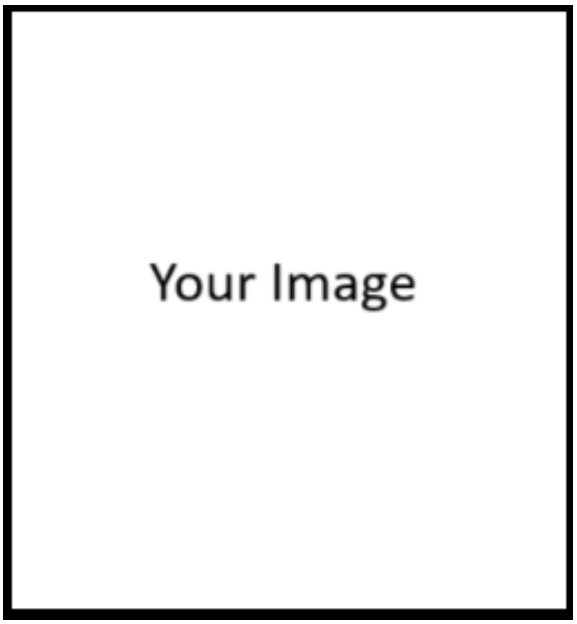


Figure 11: Figures generated by q22().

2.2.a.ii In 1-2 sentences, compare the performance of the four variants of gradient descent on this dataset (2 marks)

Answer. Your answer ...

2.2.a.iii In 1-2 sentences, explain how is it possible that the loss derived by the Adam optimizer is smaller than that of Heavy-ball Momentum, but the evaluation score of Adam is equal to the evaluation score of the heavy-ball momentum. (2 marks)

Answer. Your answer ...

2.2.b The test function `q23()` runs your implementation on the digits dataset.

2.2.b.i Include the figures generated by `q23()` in your `PA2_qa.pdf` file. (2 marks)

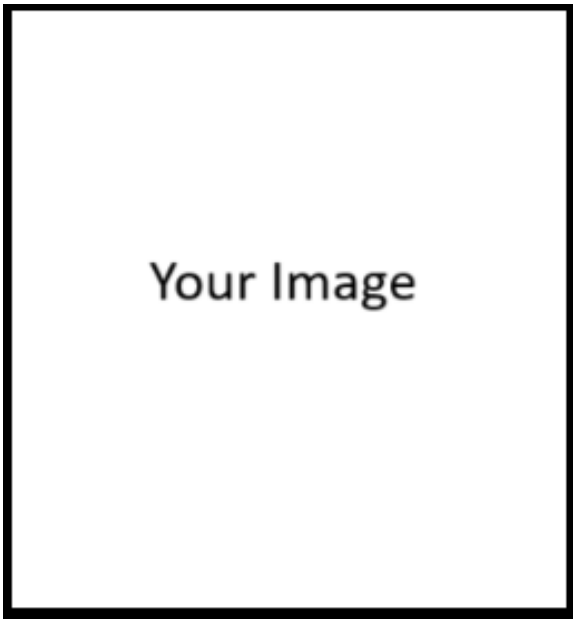


Figure 12: Figures generated by q23().

3 K-Means Clustering (Bonus)

No Written part.

4 Discussion

4.a How much time did you spend on each part of this assignment? (1 mark)

Answer. Your answer ...

4.b Any additional feedback? (optional)

Answer. Your answer ...