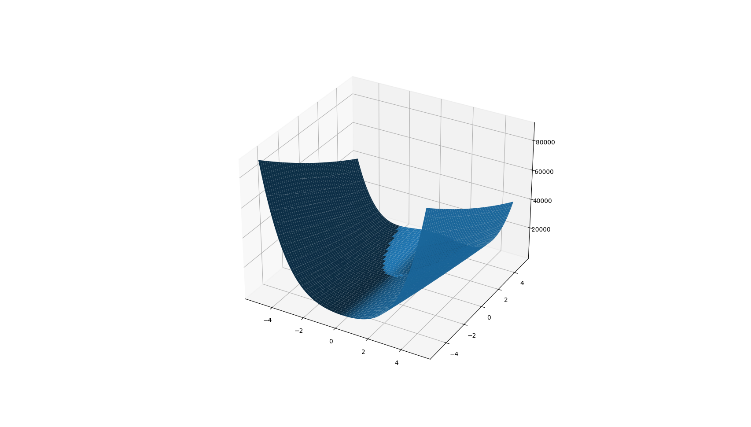
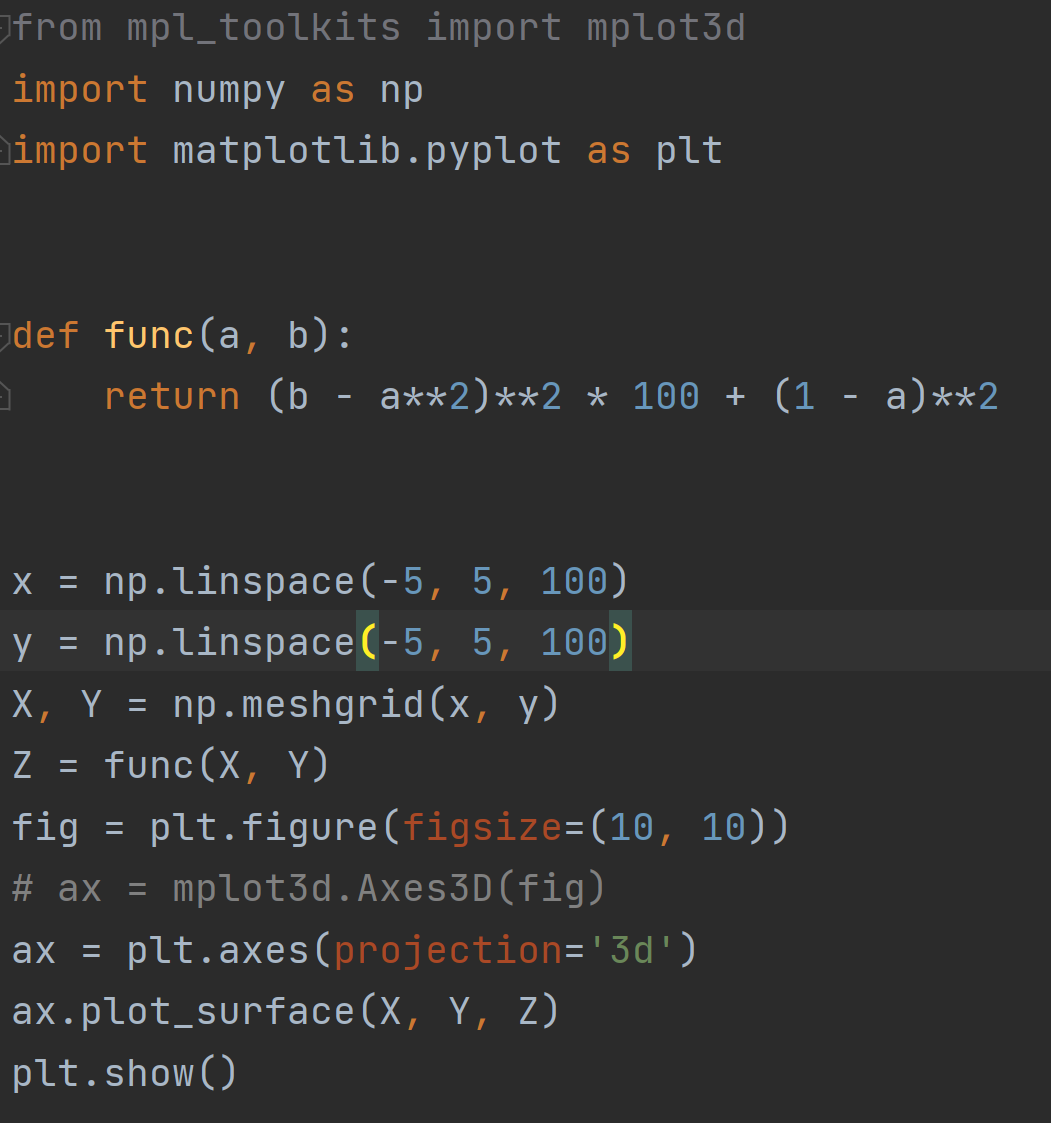
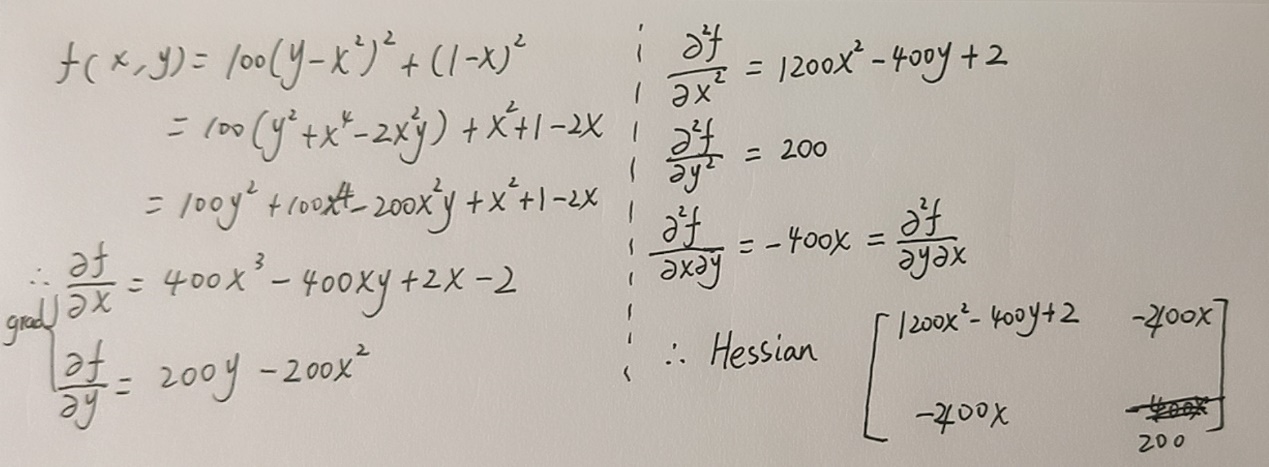
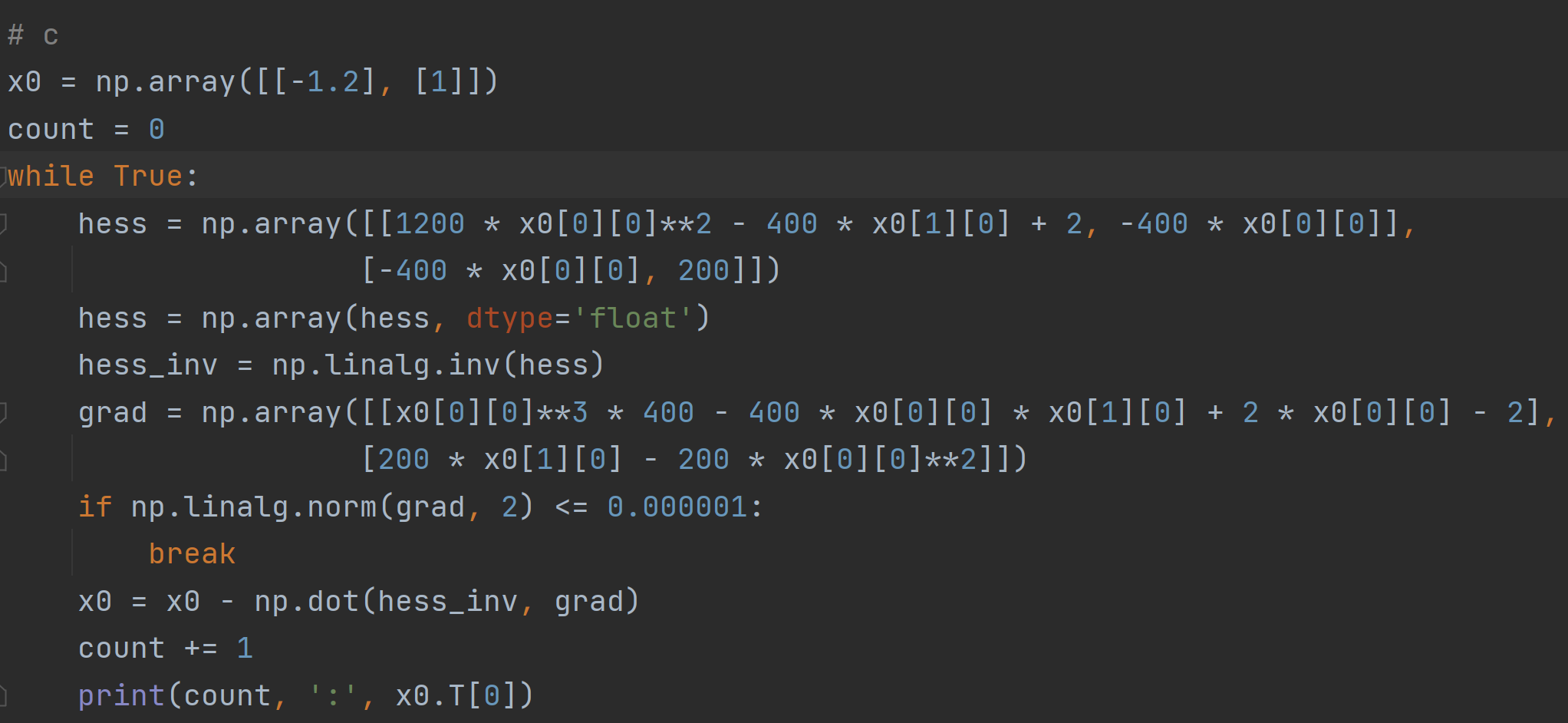
Question 1:

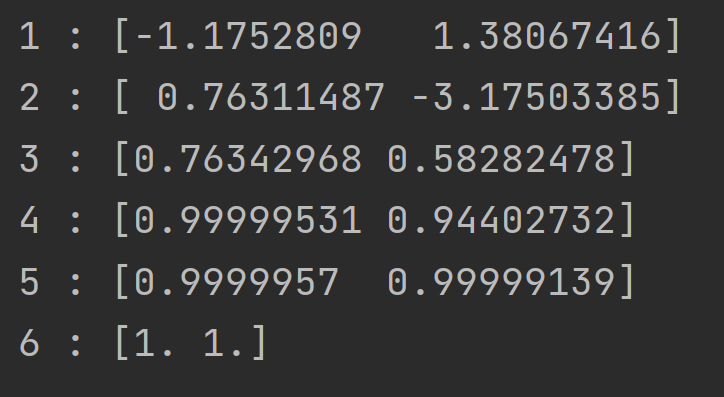
1. Given the initial iteration point x(0), iterate with Eq.(1) repeatedly until a point with derivative 0 is reached or the maximum number of iterations is reached.



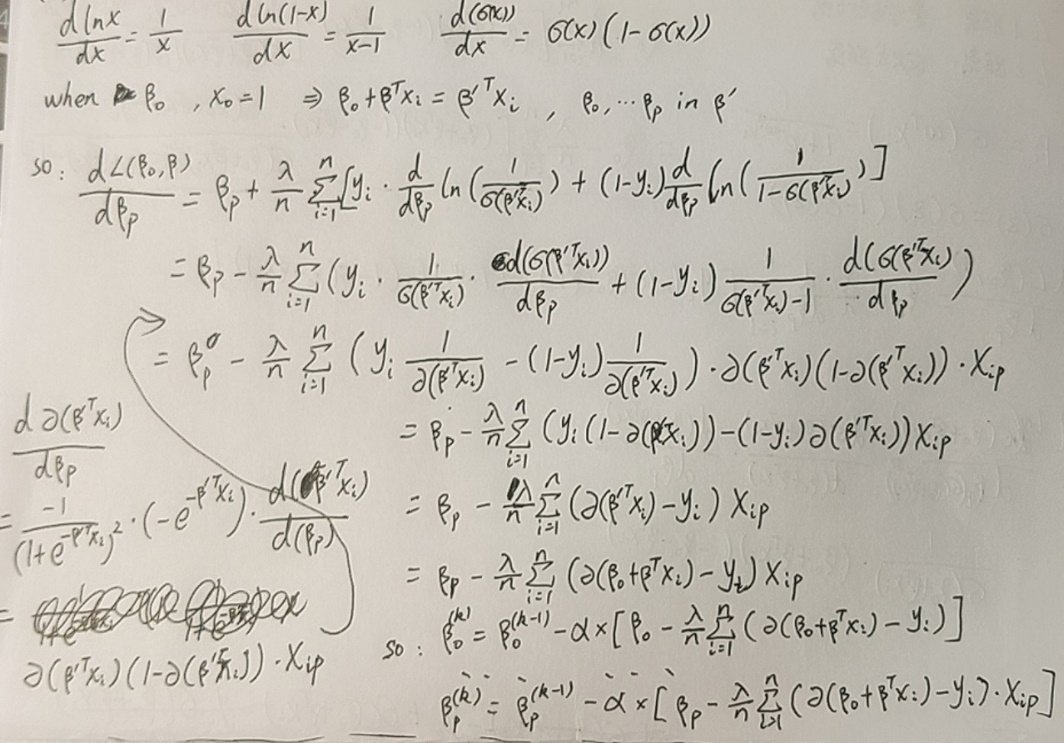


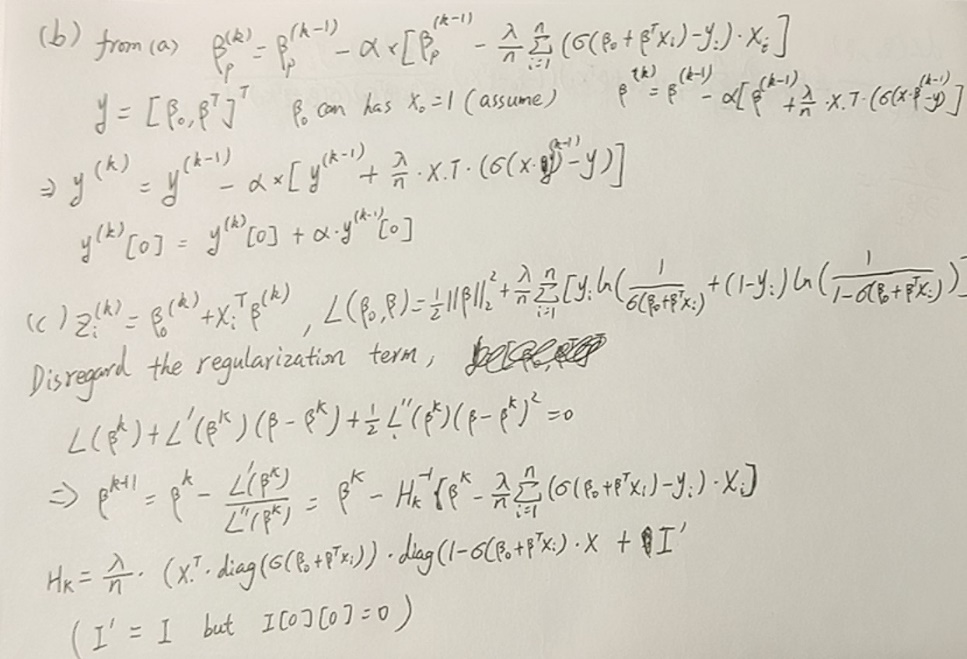
(c）



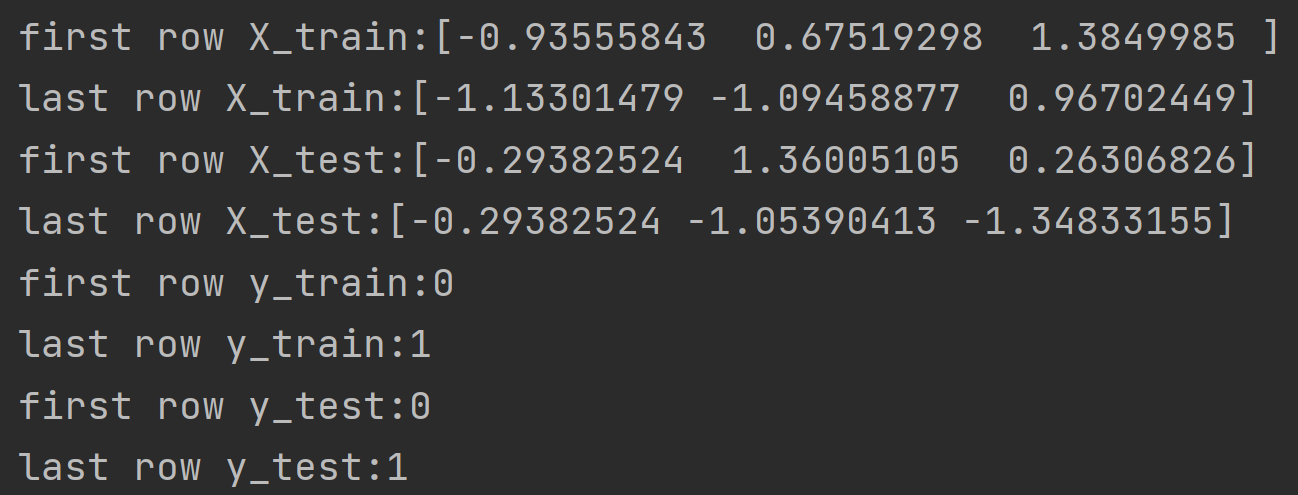


Question 2:

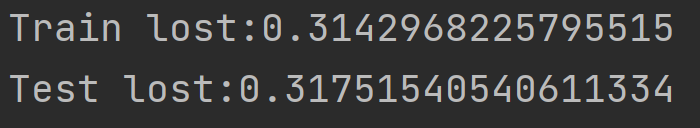
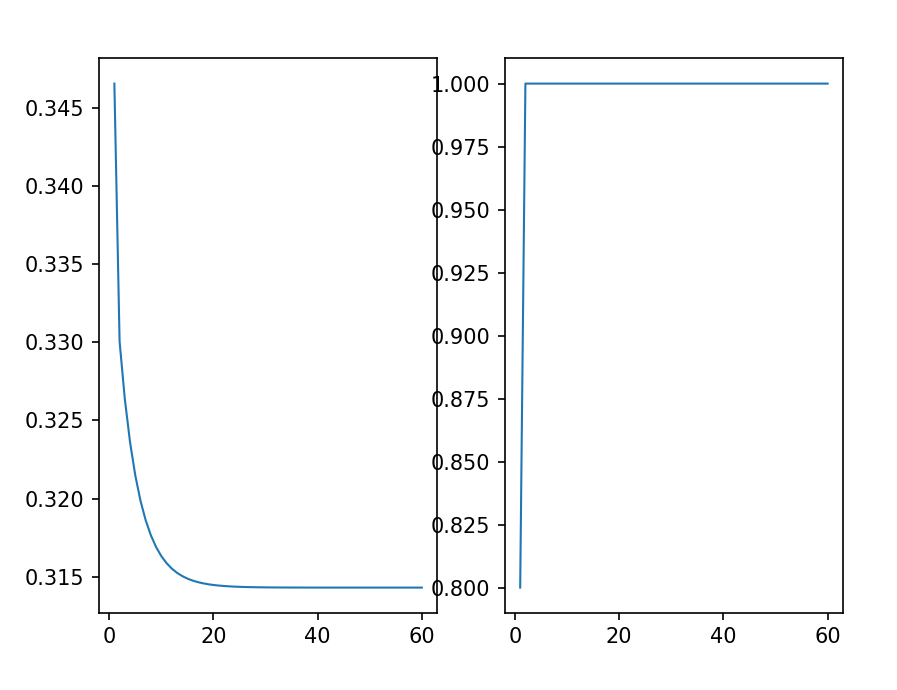


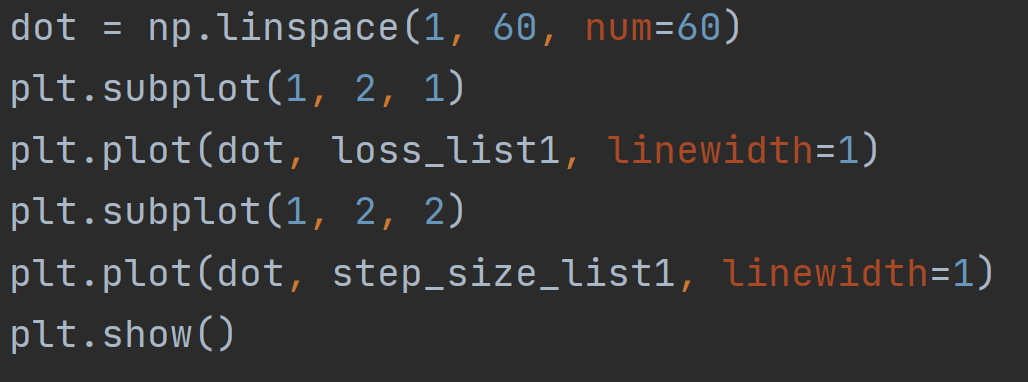
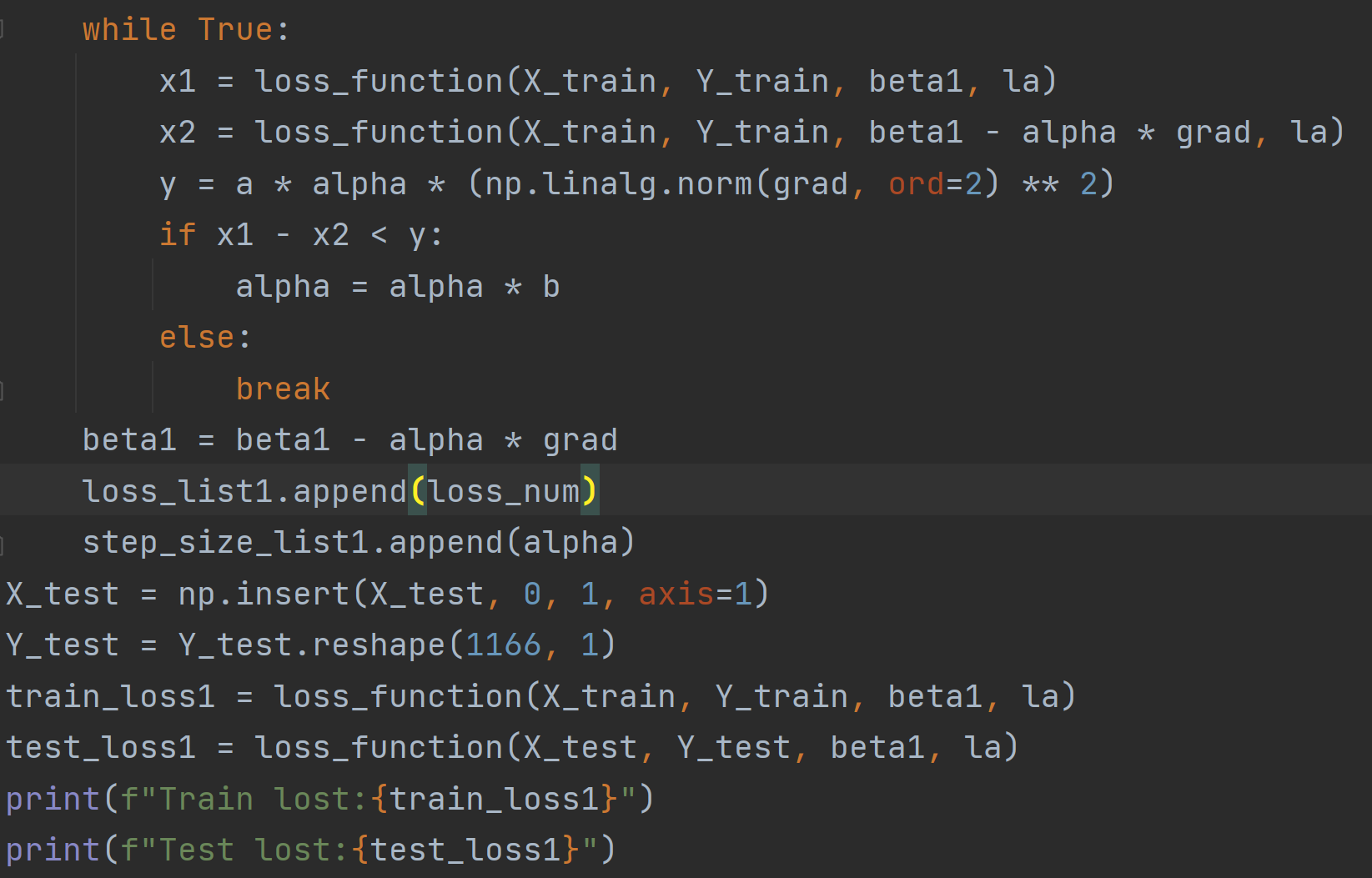
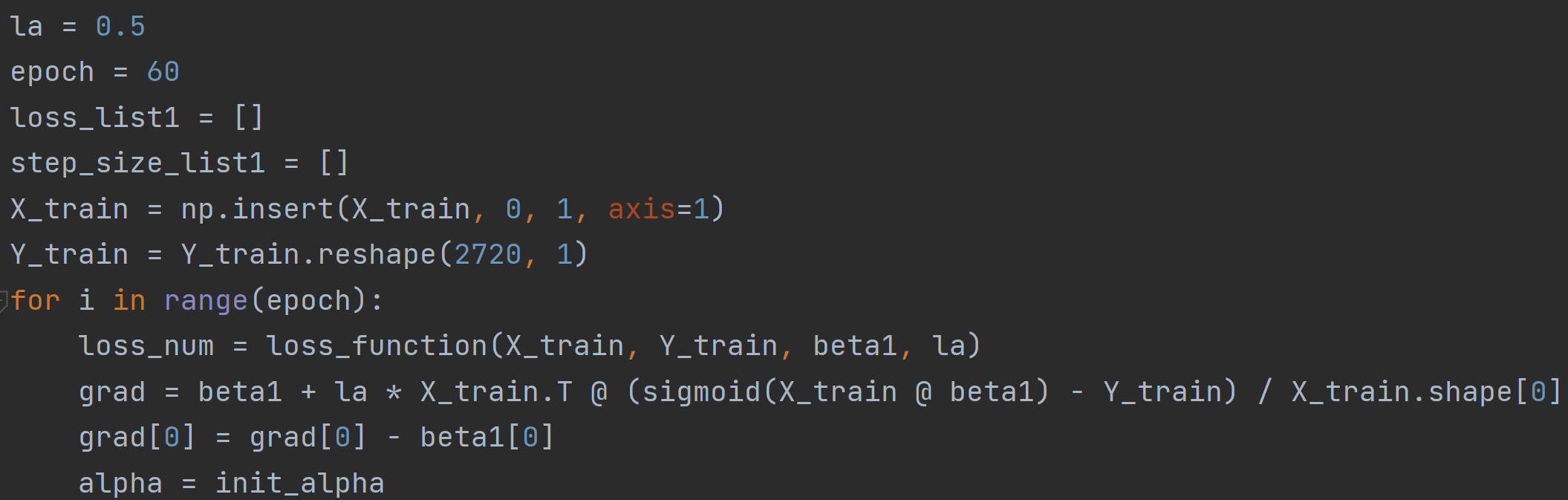
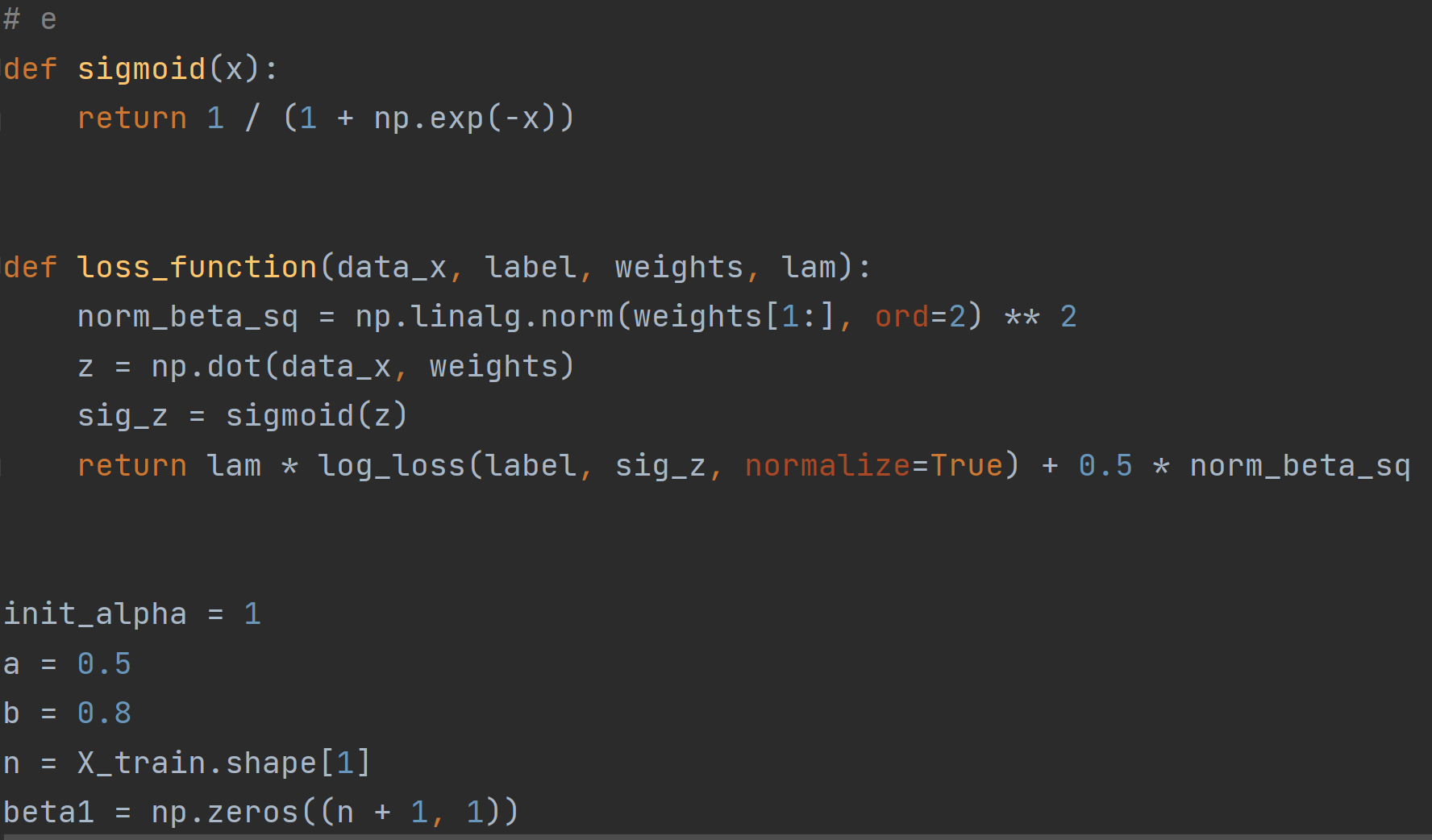


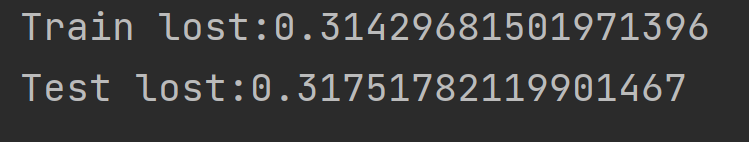
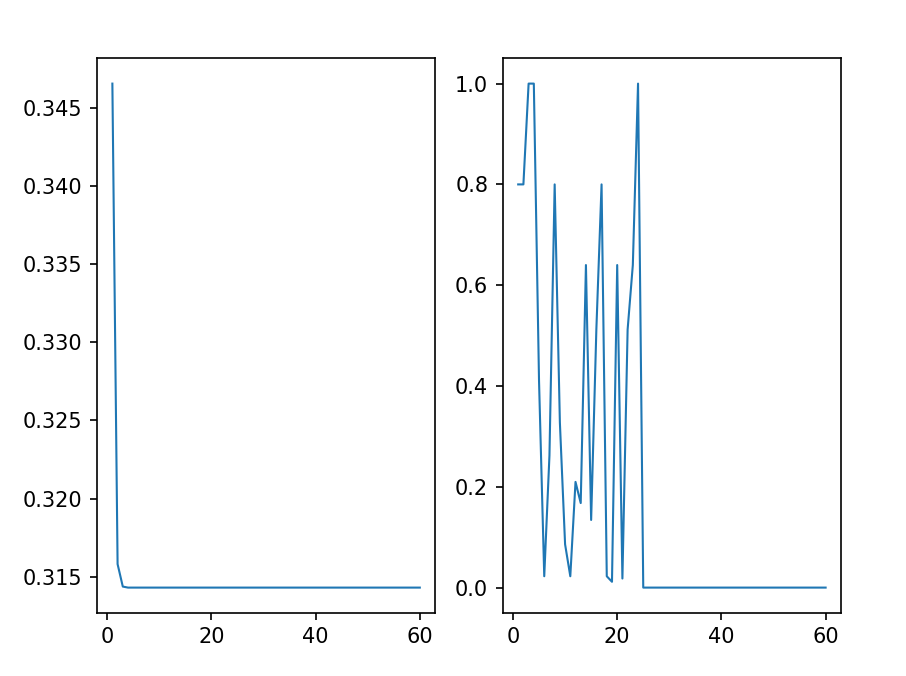
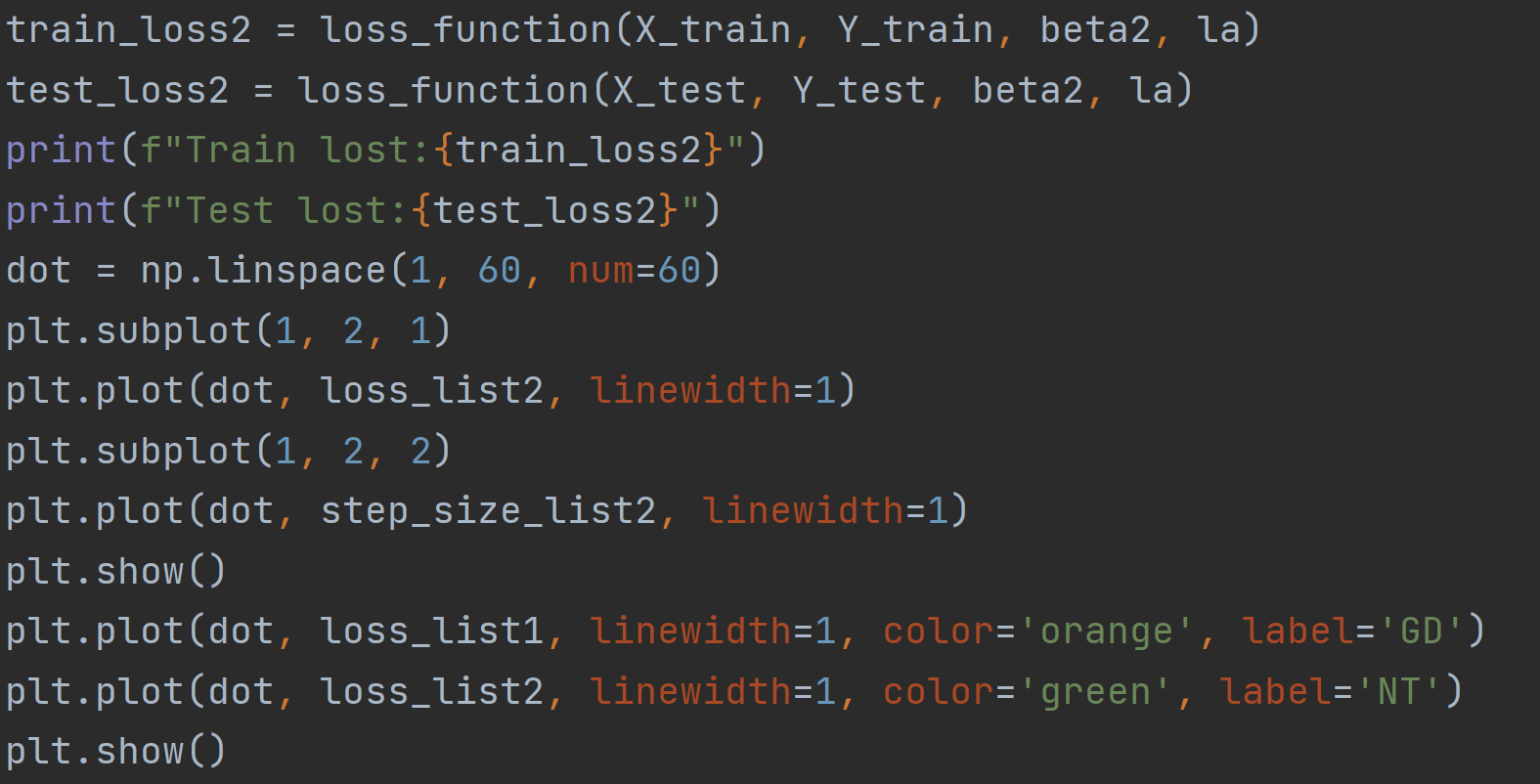
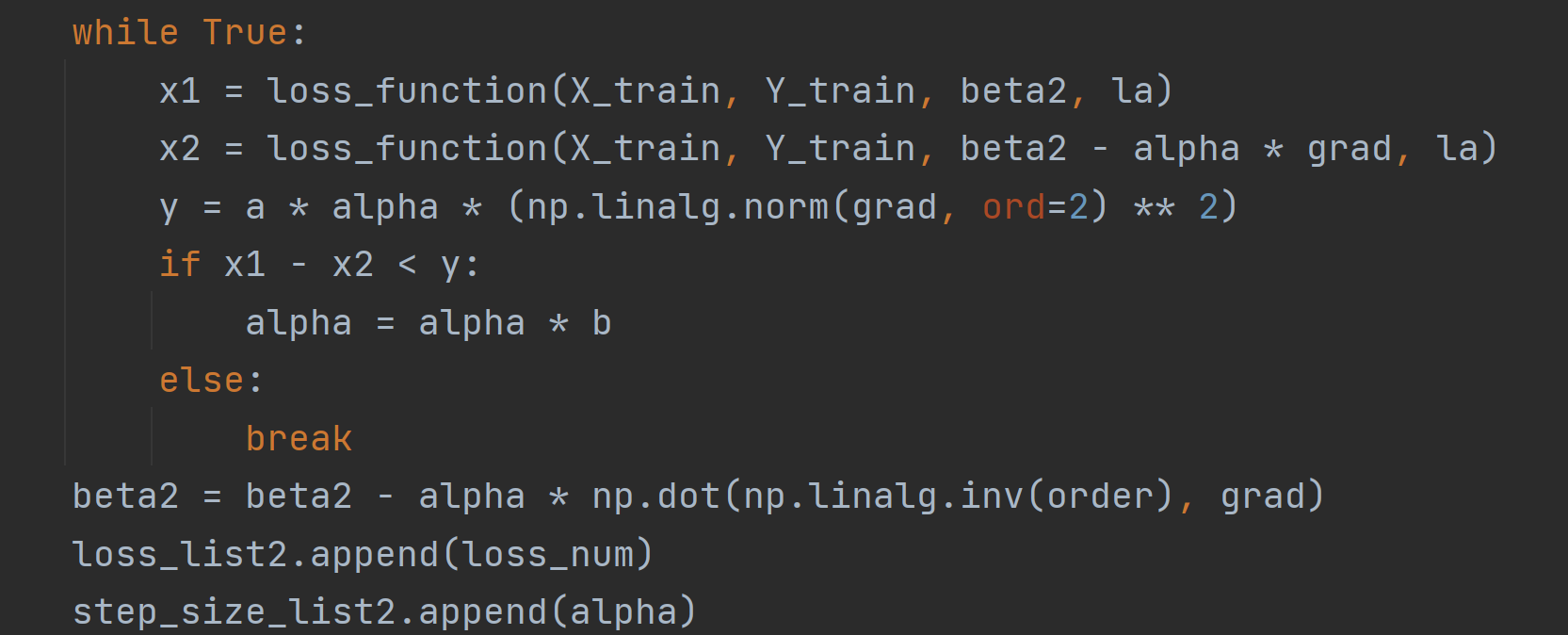
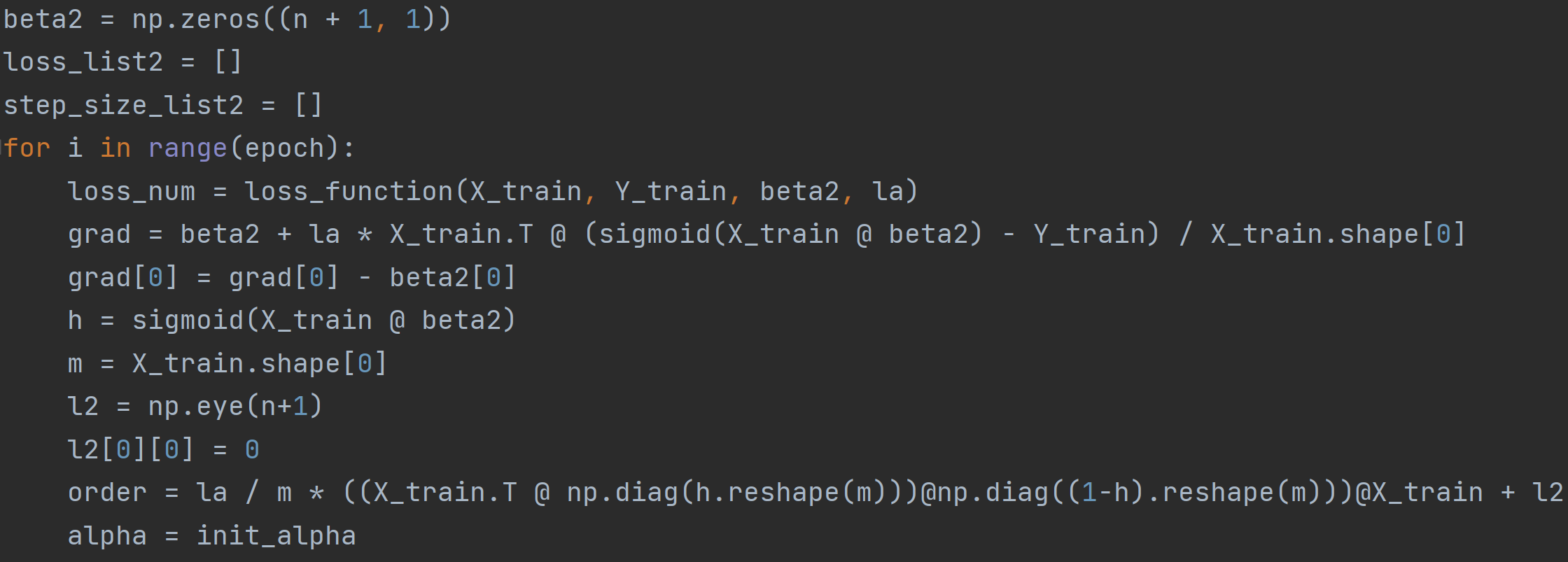
(d)



(e)



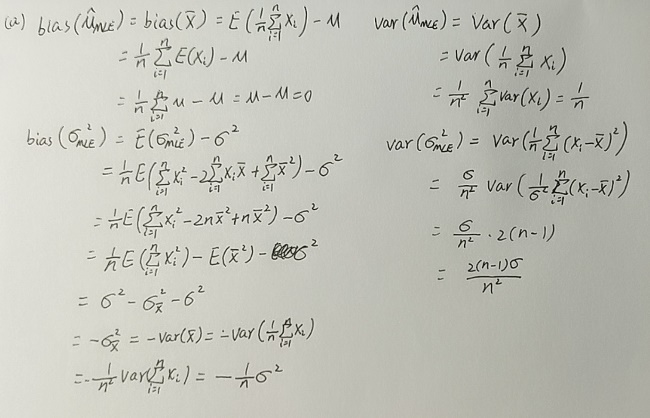


(f) 

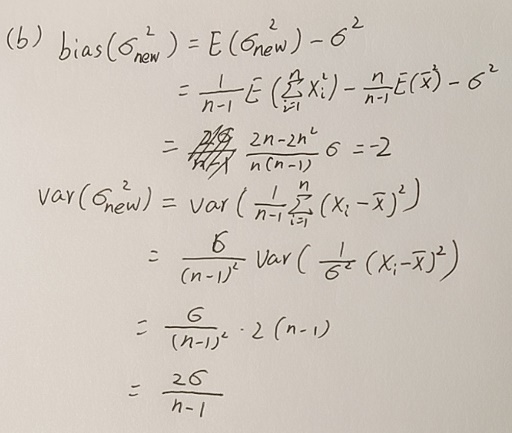
(g) The gradient descent method only requires solving the gradient, while methods such as Newton's method require Hessian matrices or calculating analytic solutions, etc. Since machine learning often requires the use of a large sample size, the gradient descent method takes much less time per iteration.

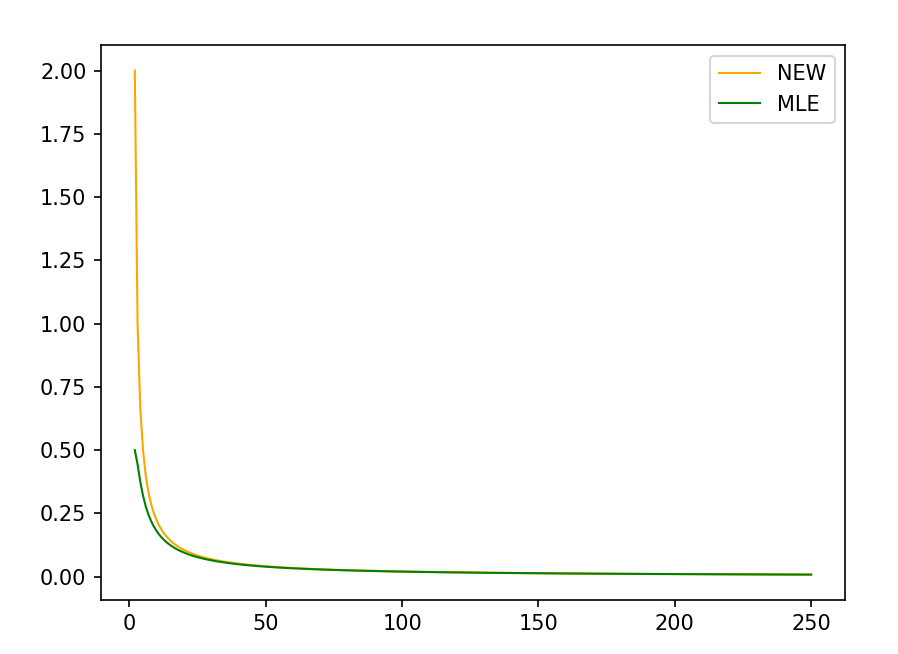
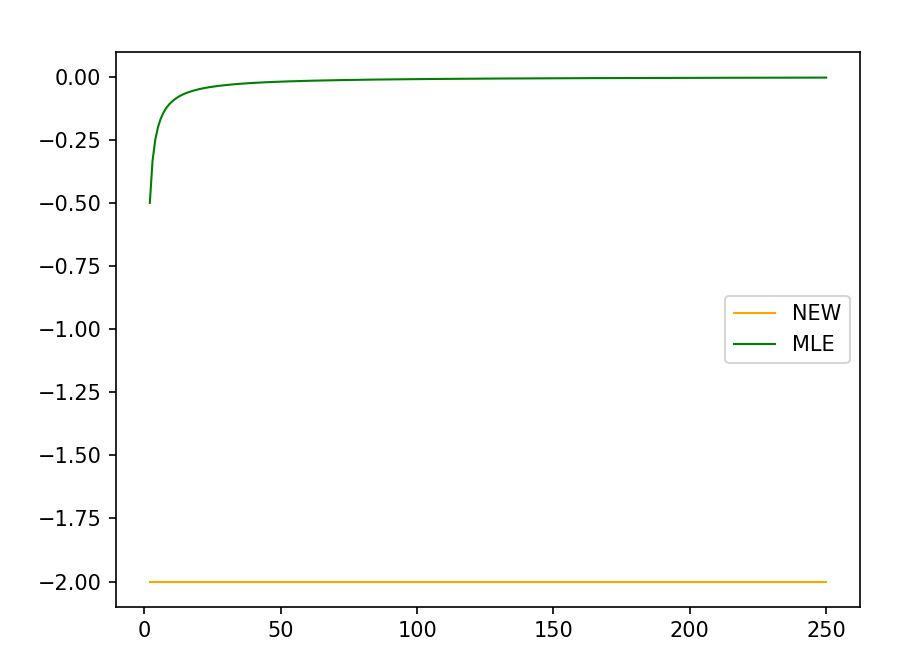
Question3:

(a)



(b)





1. The calculated MSE curve shows that the MLE estimator is better.

When the sample size n increases, the MSE of the two estimators decreases first and then stabilizes at a certain value.

