Figure 1, Question 1(a): Moons Training Data
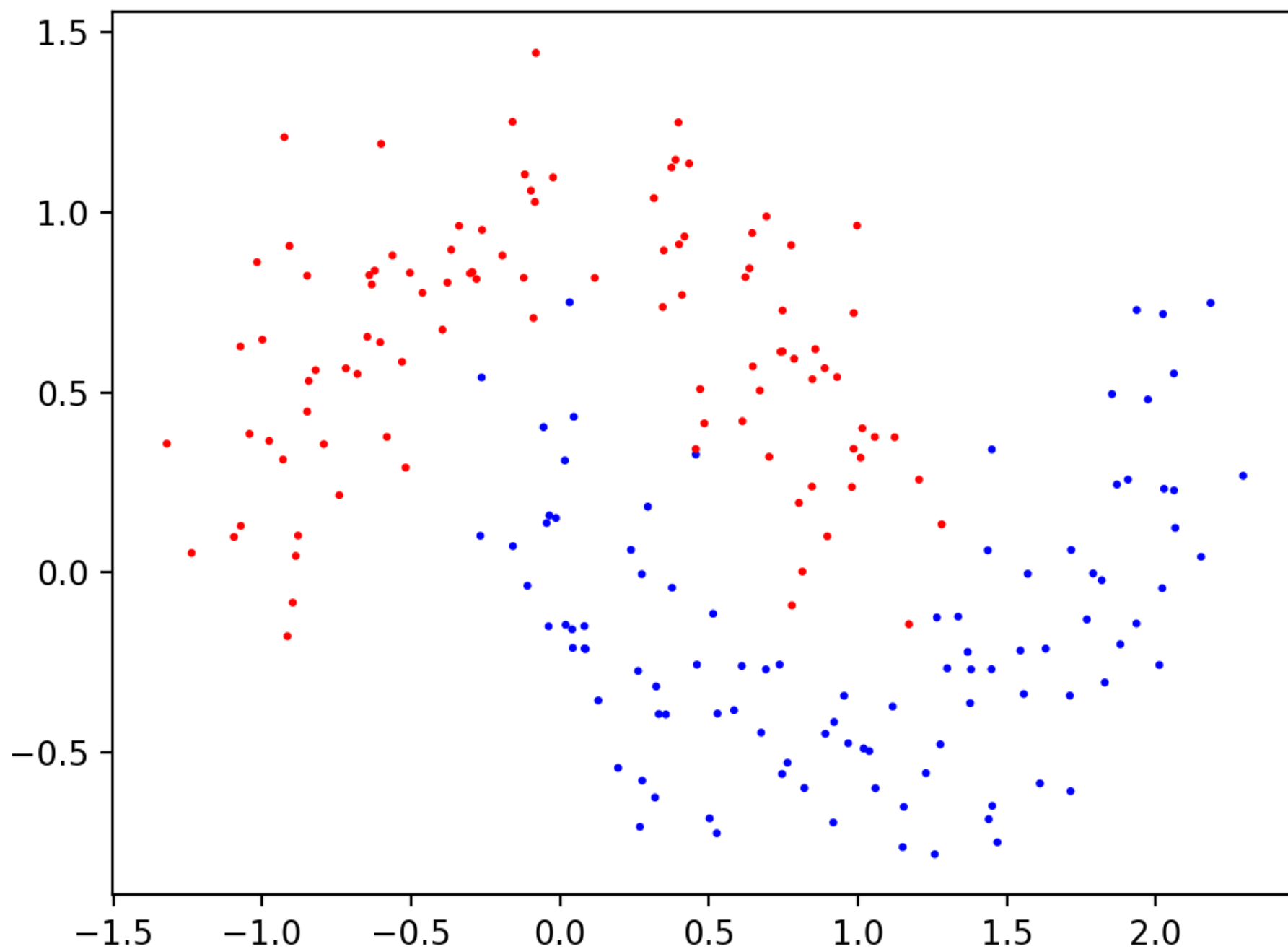
Figure 2, Question 1(a): Moons Test Data
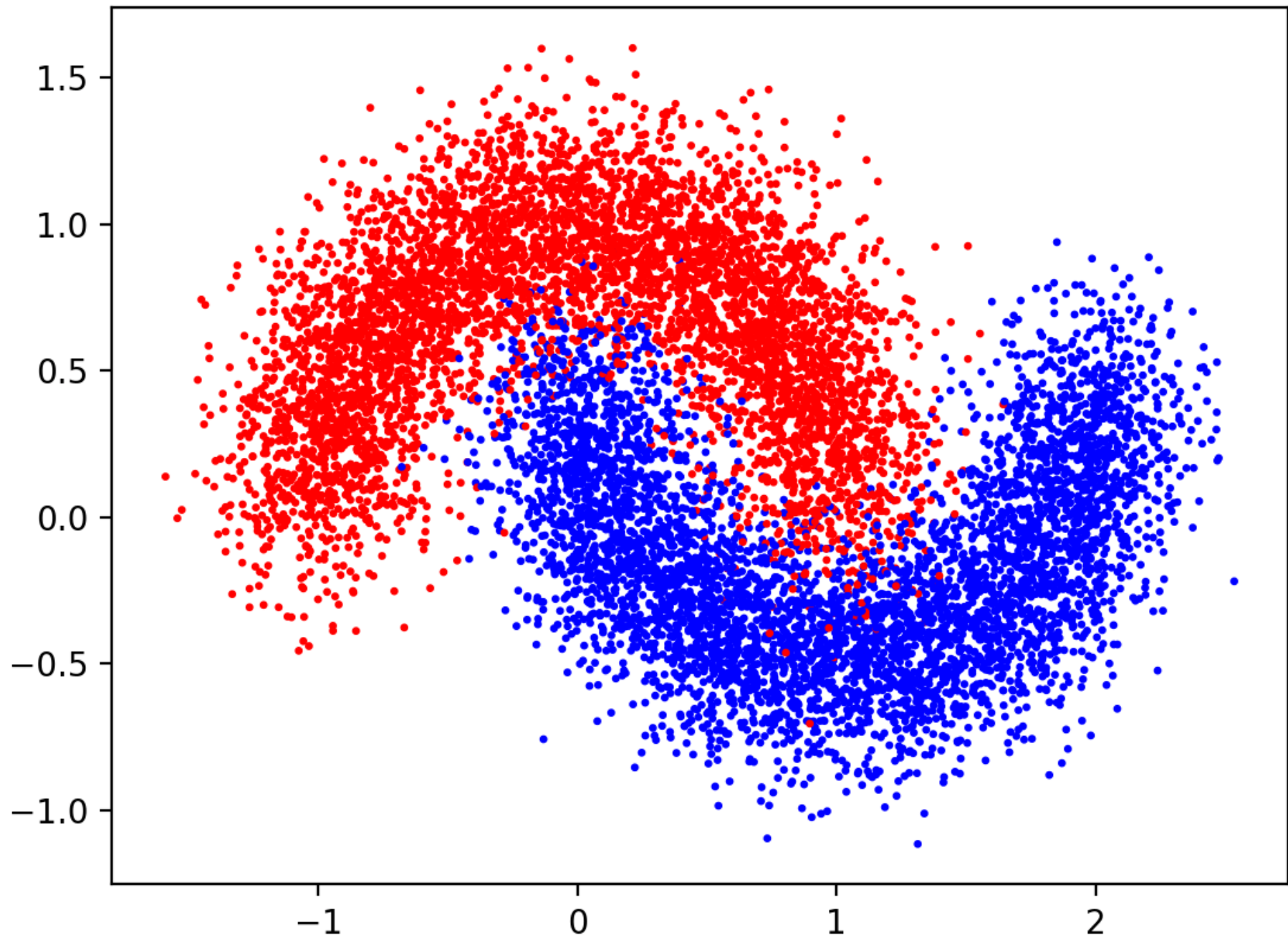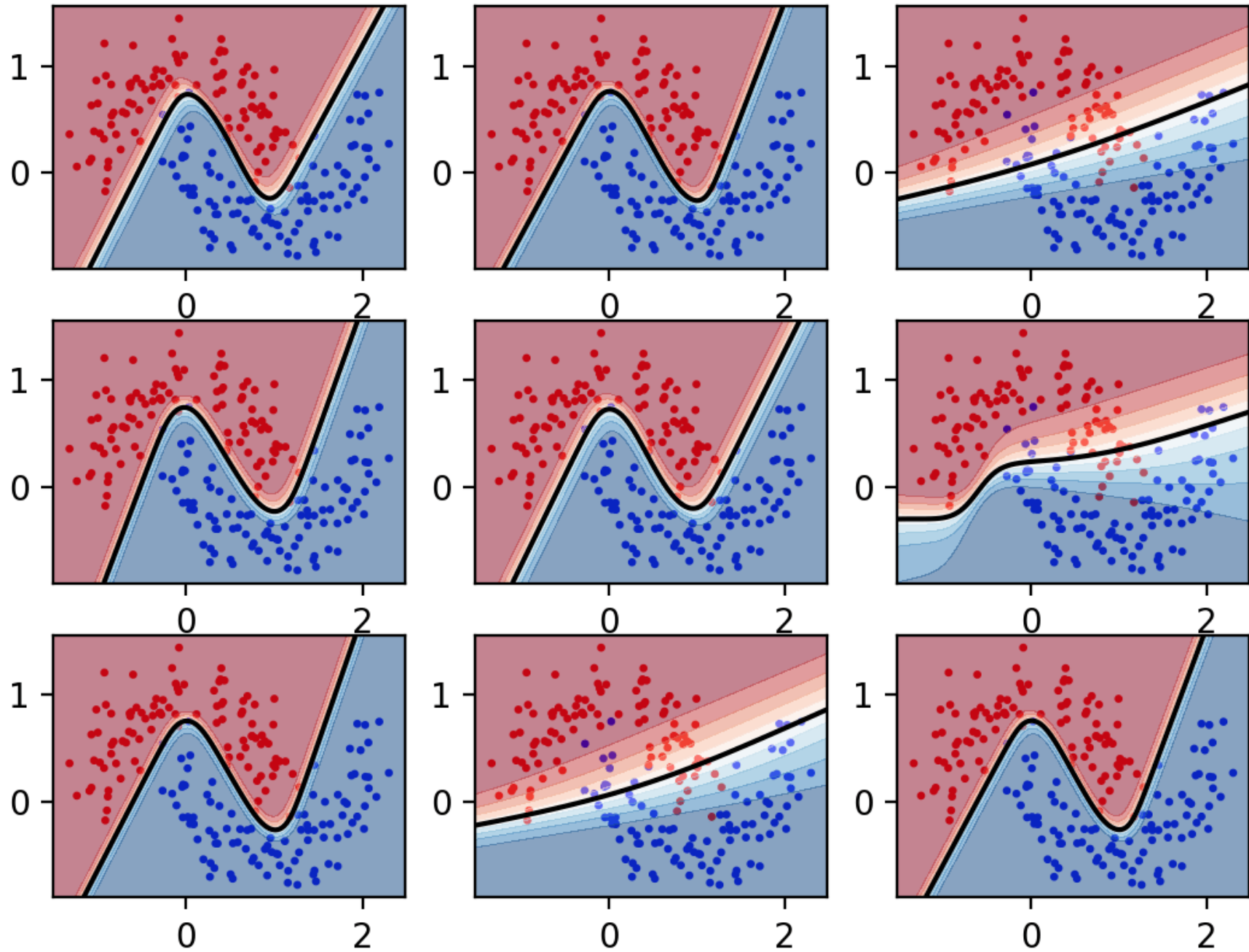
Figure 3, Question 1(b): Contour plots for various training sessions
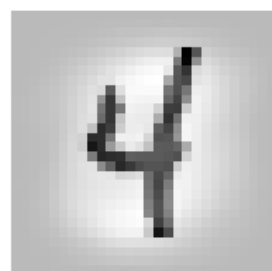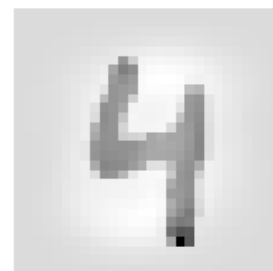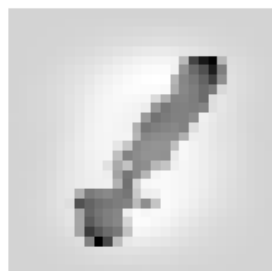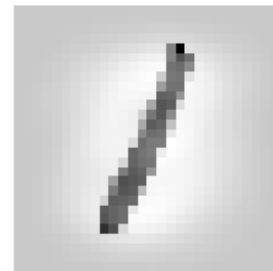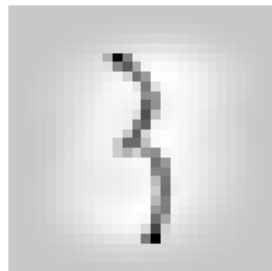
Question 3(c): some normalized MNIST digits

Figure 4, Question 1(b): Contour plot for best training session
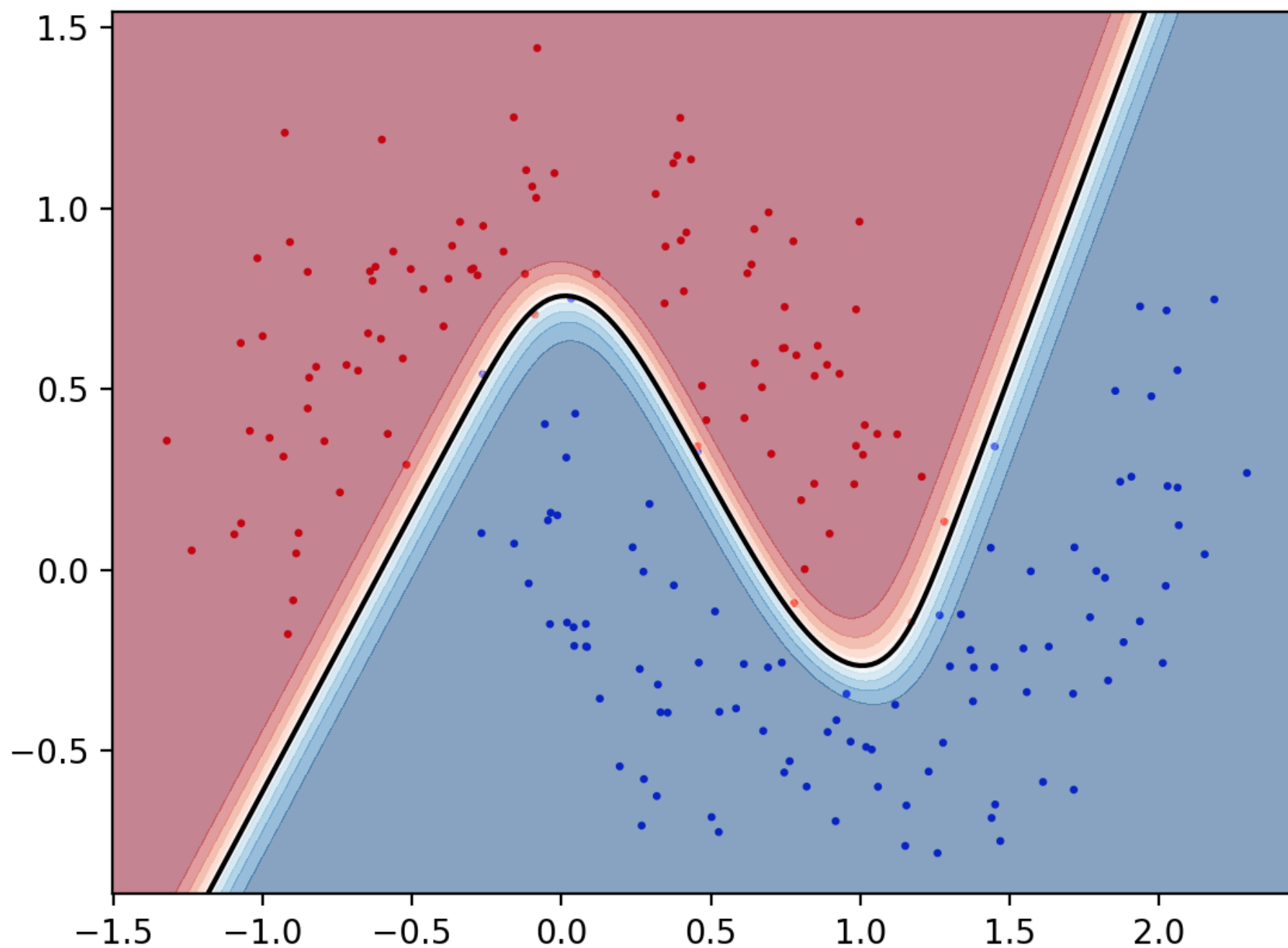
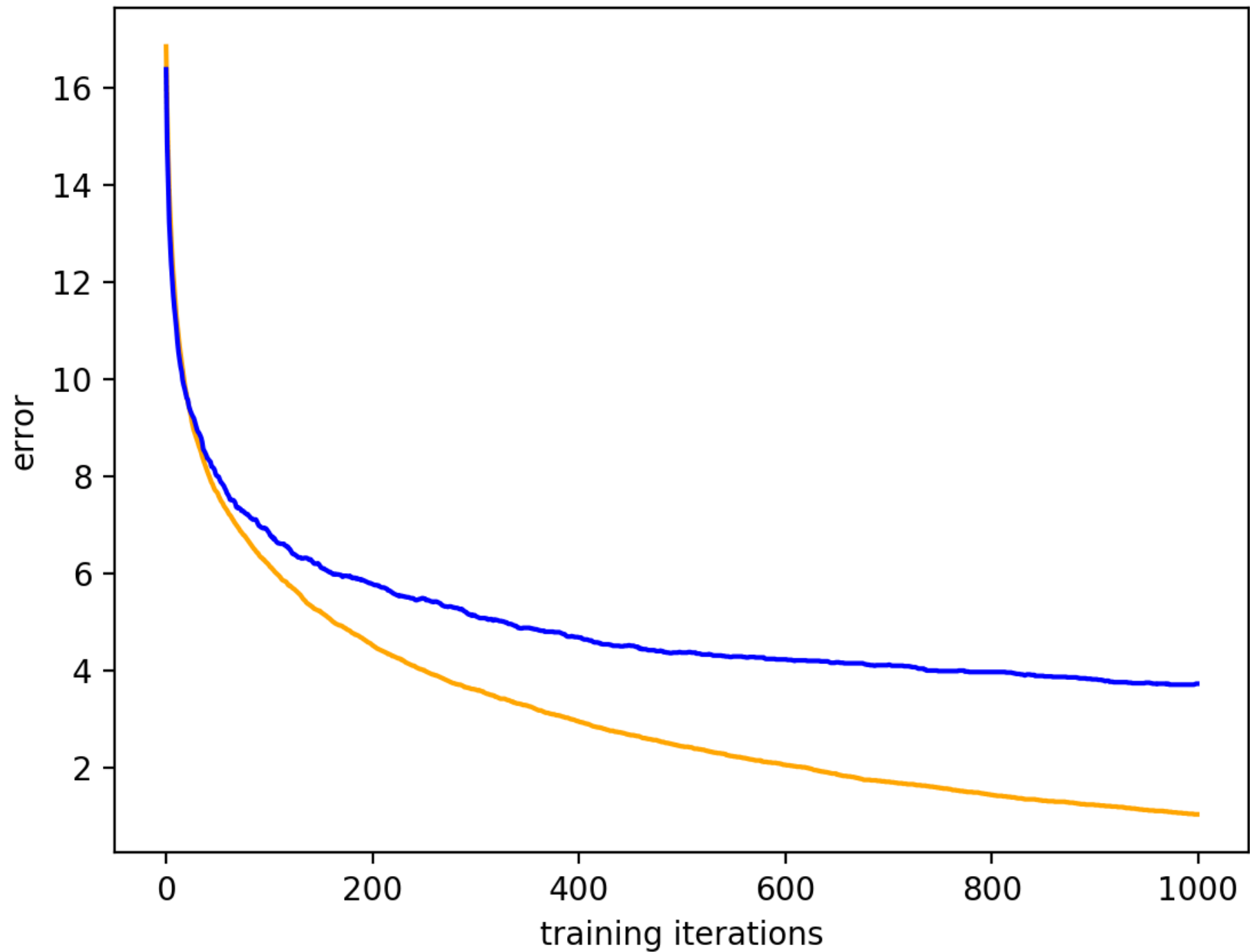Figure 5, Question 3: training and test error in batch mode

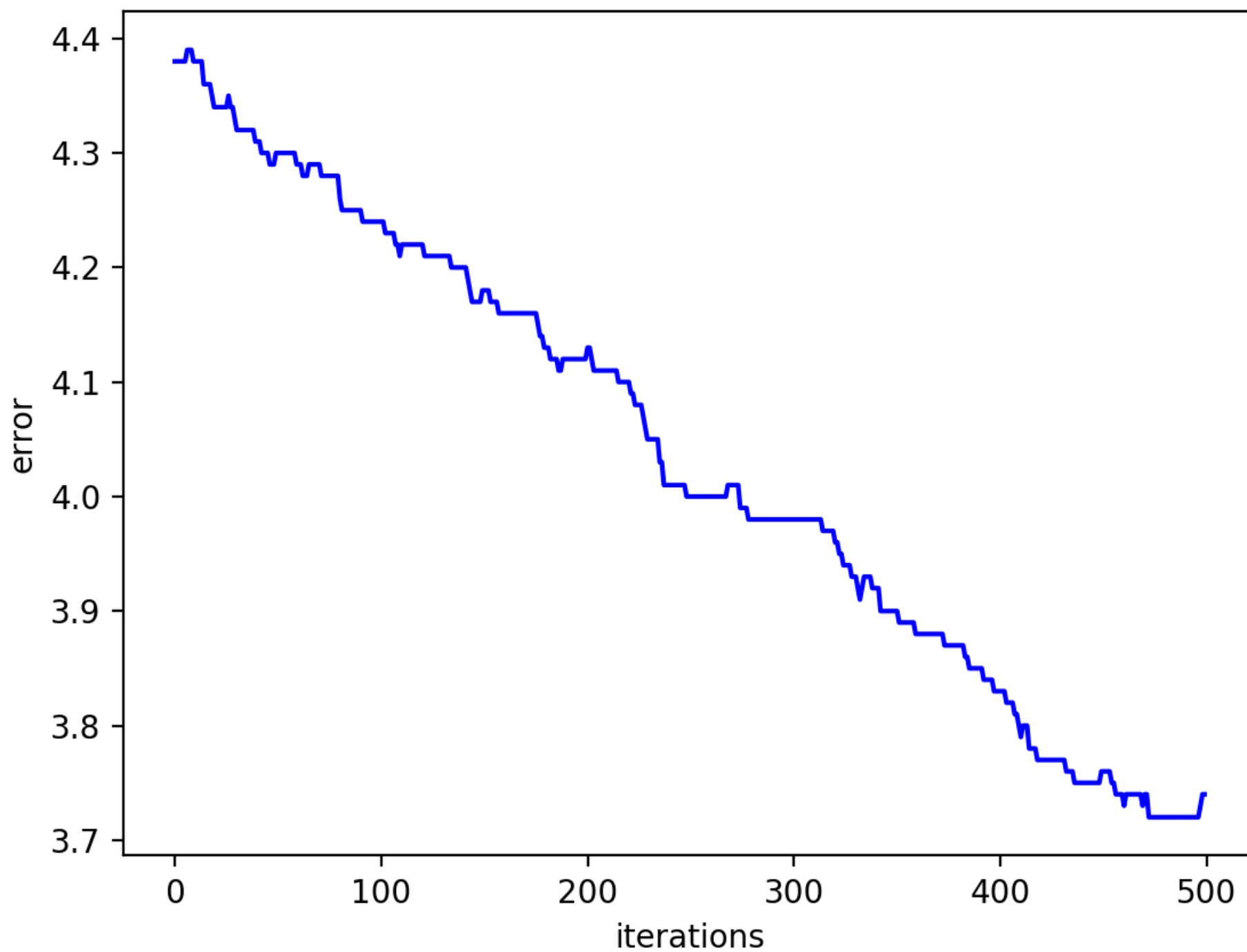Figure 6: Question 3: test error during last 500 iterations of batch training

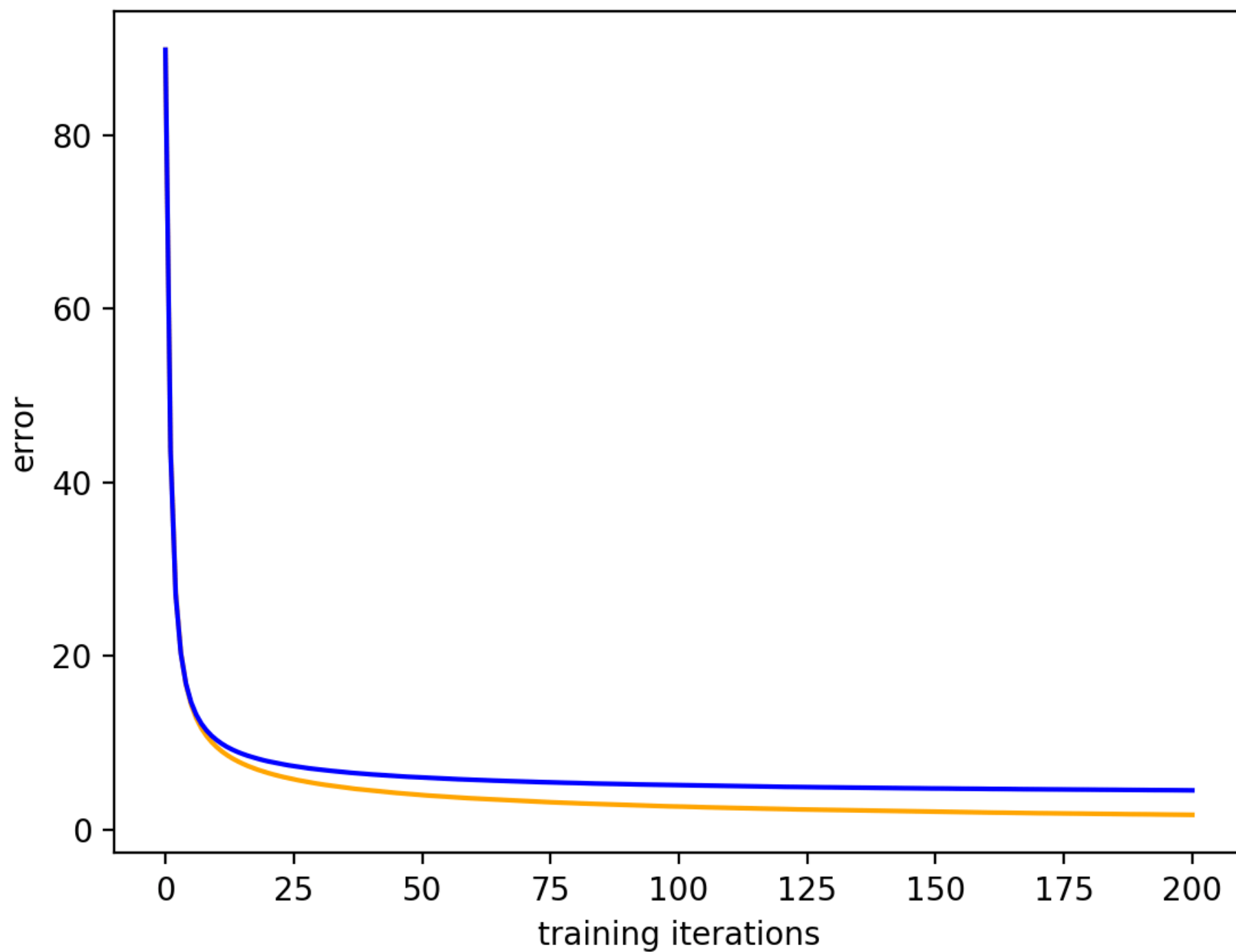Figure 7, Question 4(d): training and test error for batch gradient descent

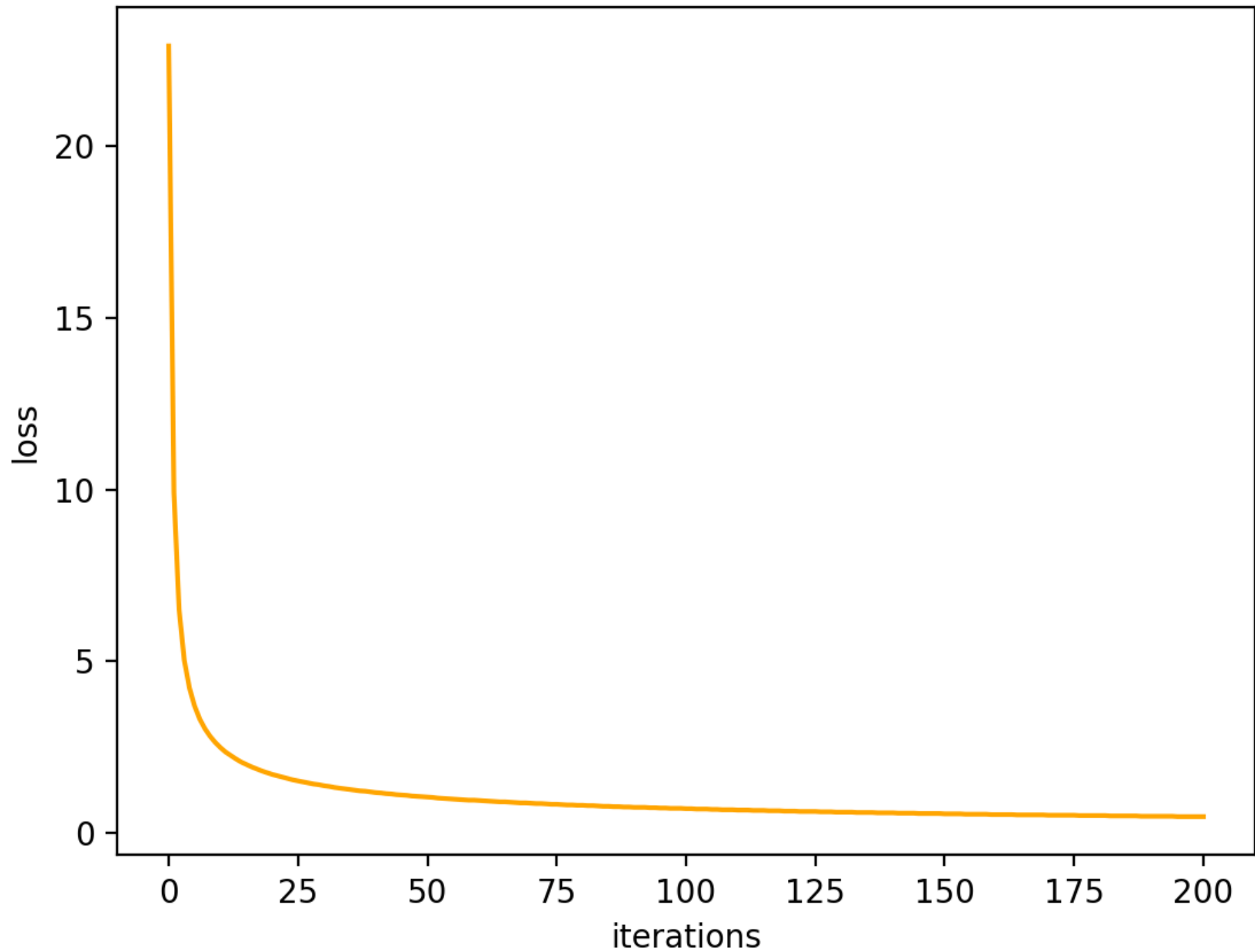Figure 8: Question 4(d): mean training loss for batch gradient descent

Figure 9: Question 4(d): training and test error for last 500 epochs of bgd
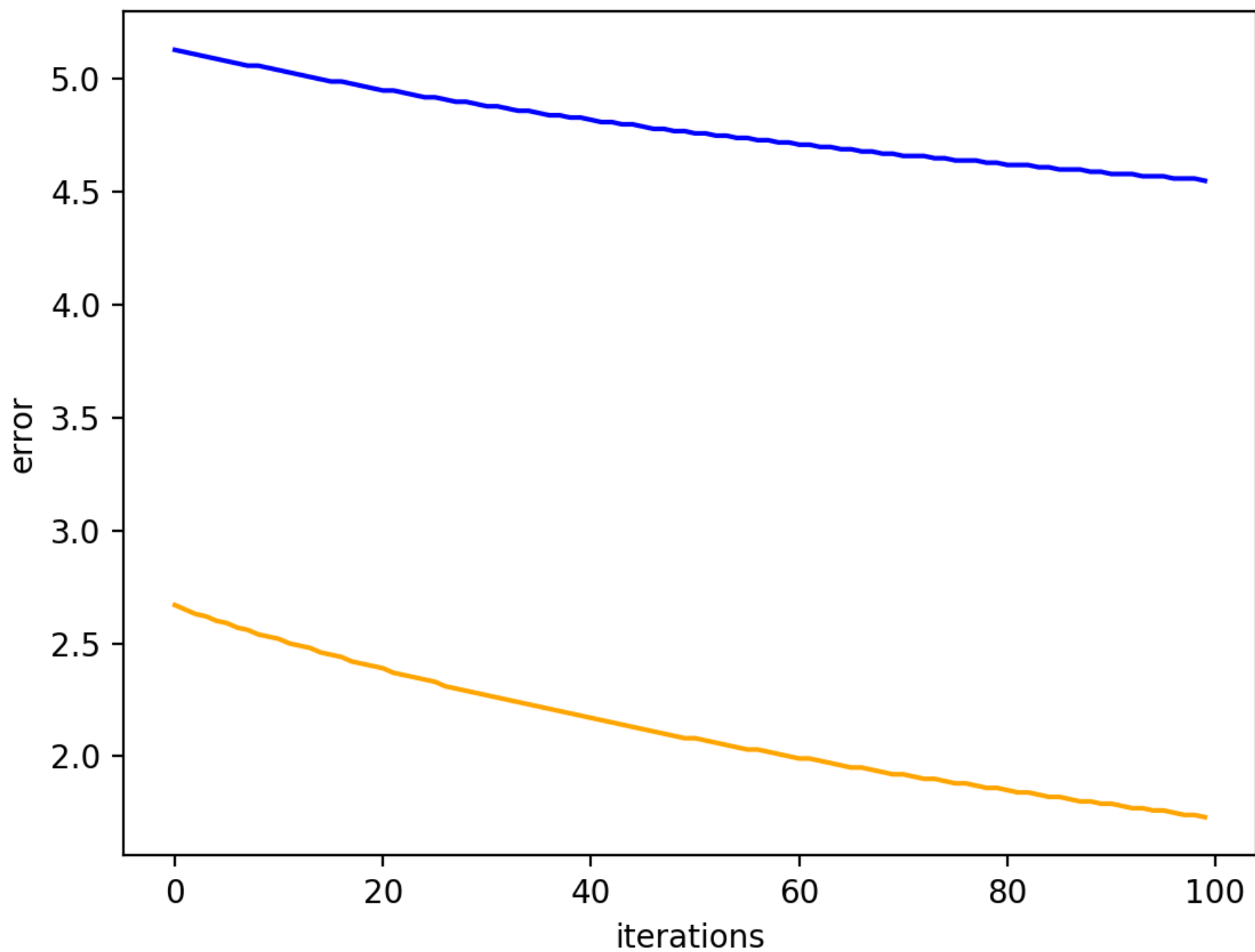
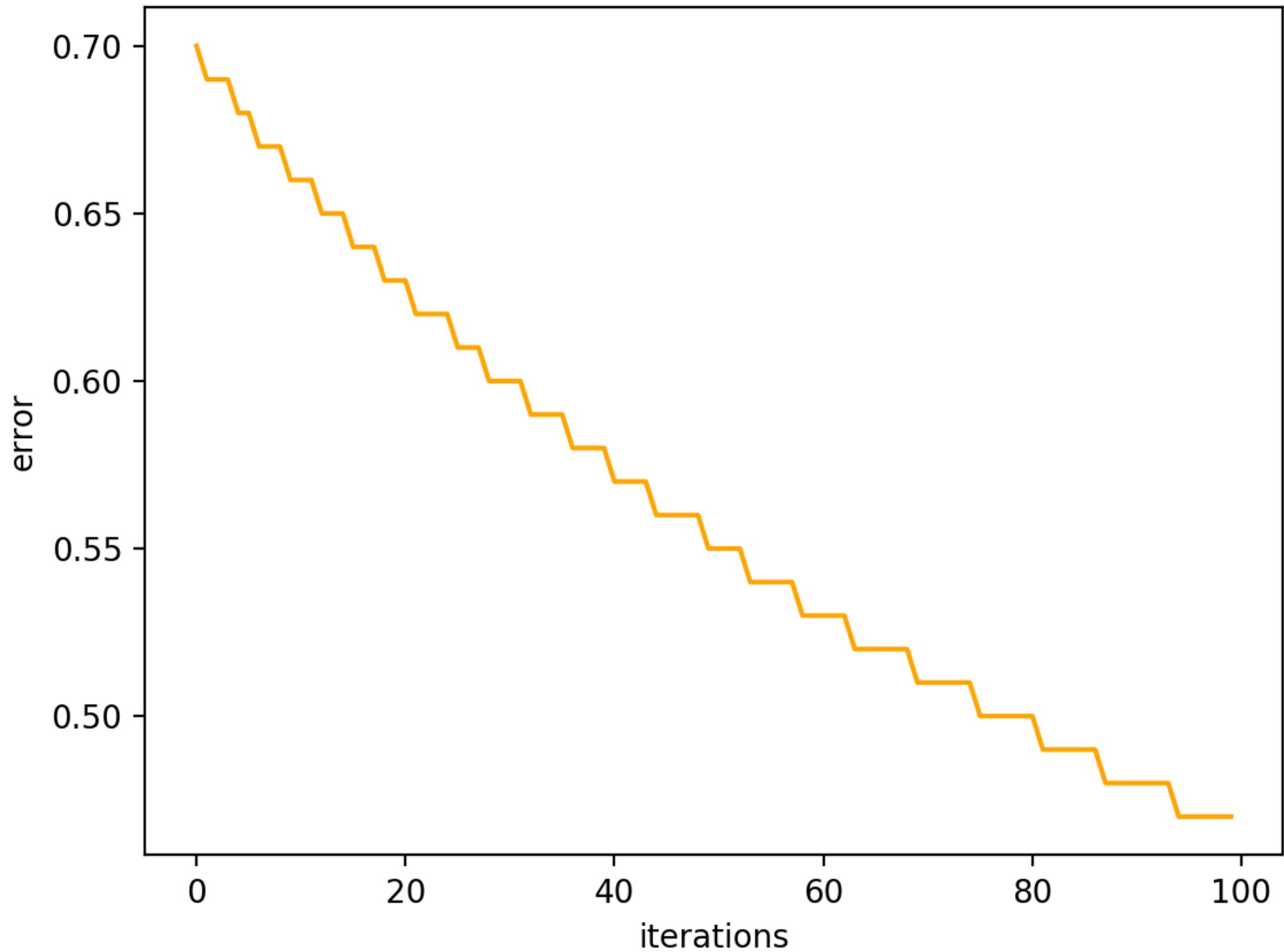Figure 10: Question 4(d): mean training loss for last 500 epochs of bgd

Figure 11, Question 4(g): training and test error for stochastic gradient descent
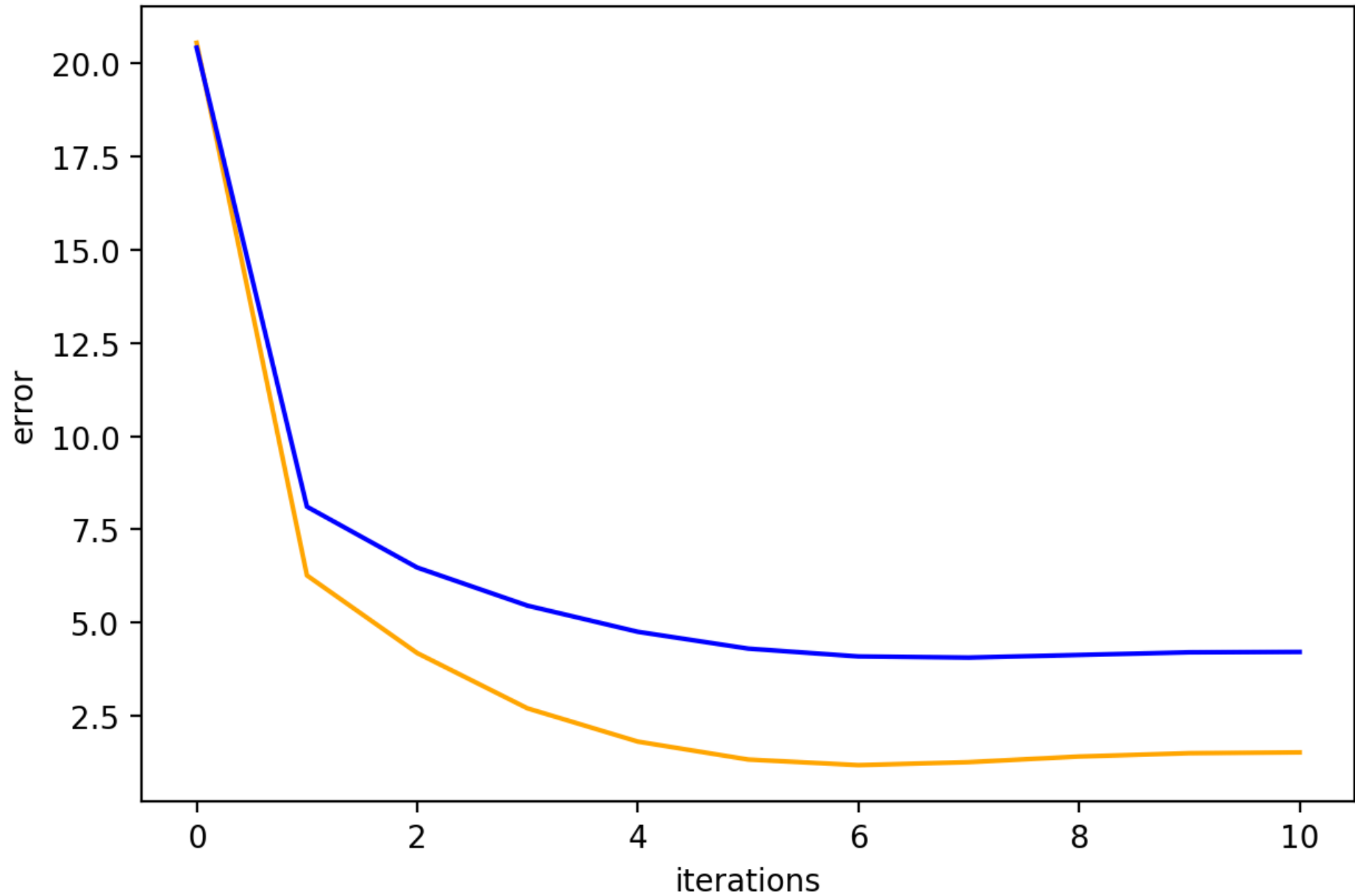
Figure 12: Question 4(g): mean training loss for stochastic gradient descent