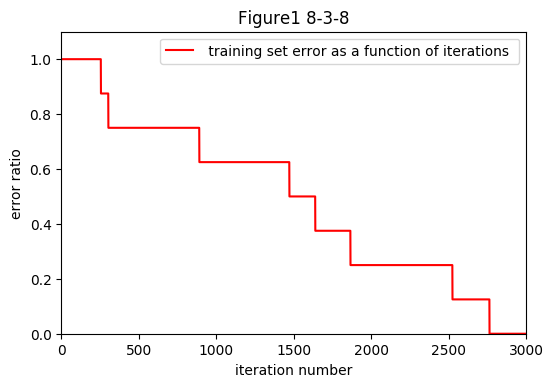
Project 4

Zheyi Yi

一.

￼

What can you observe about the learning curve? and is the “binary representation” discovered by your code?

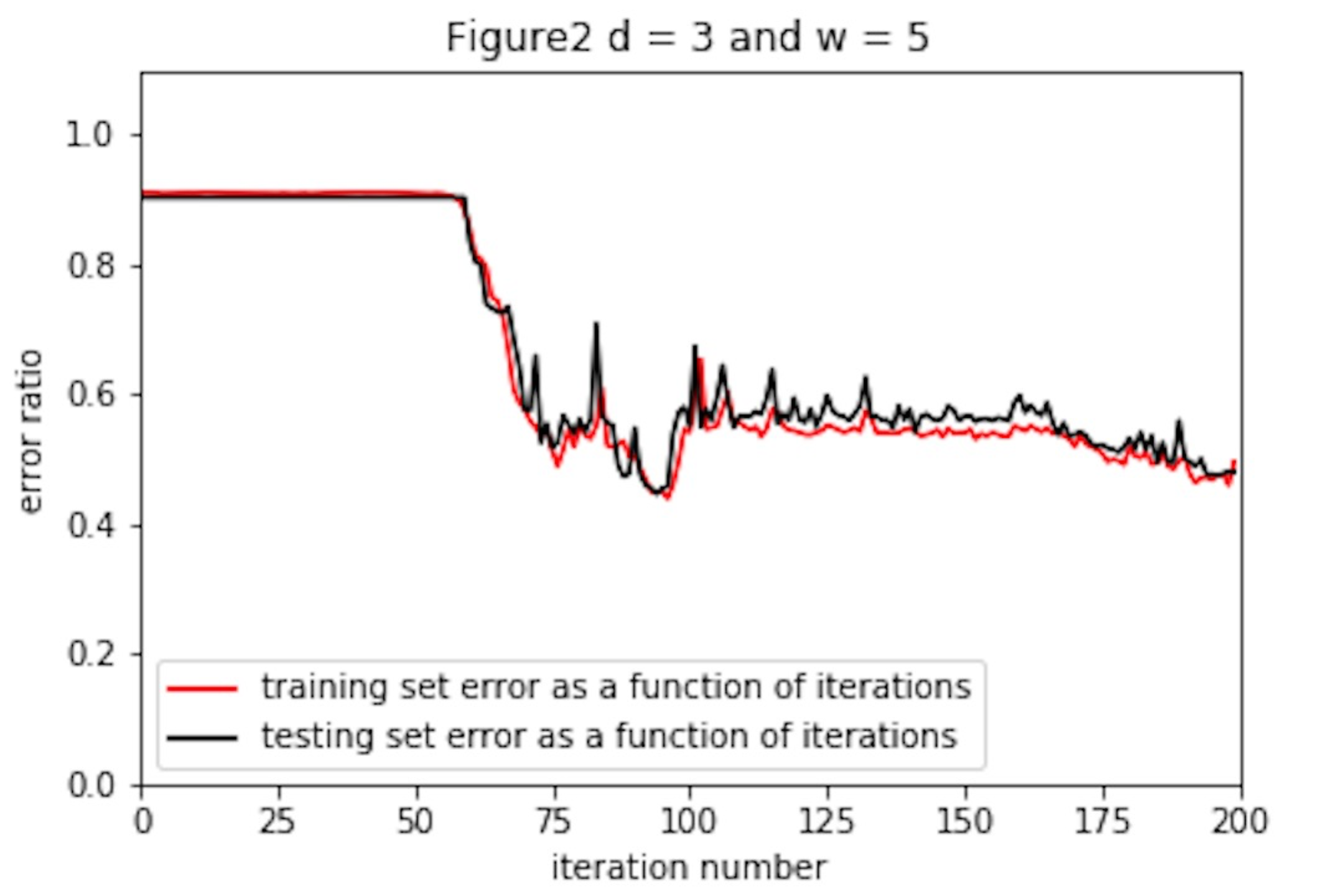
From the figure 1, we can see that as the iteration number increase, our error ratio decreased gradually until to 0.

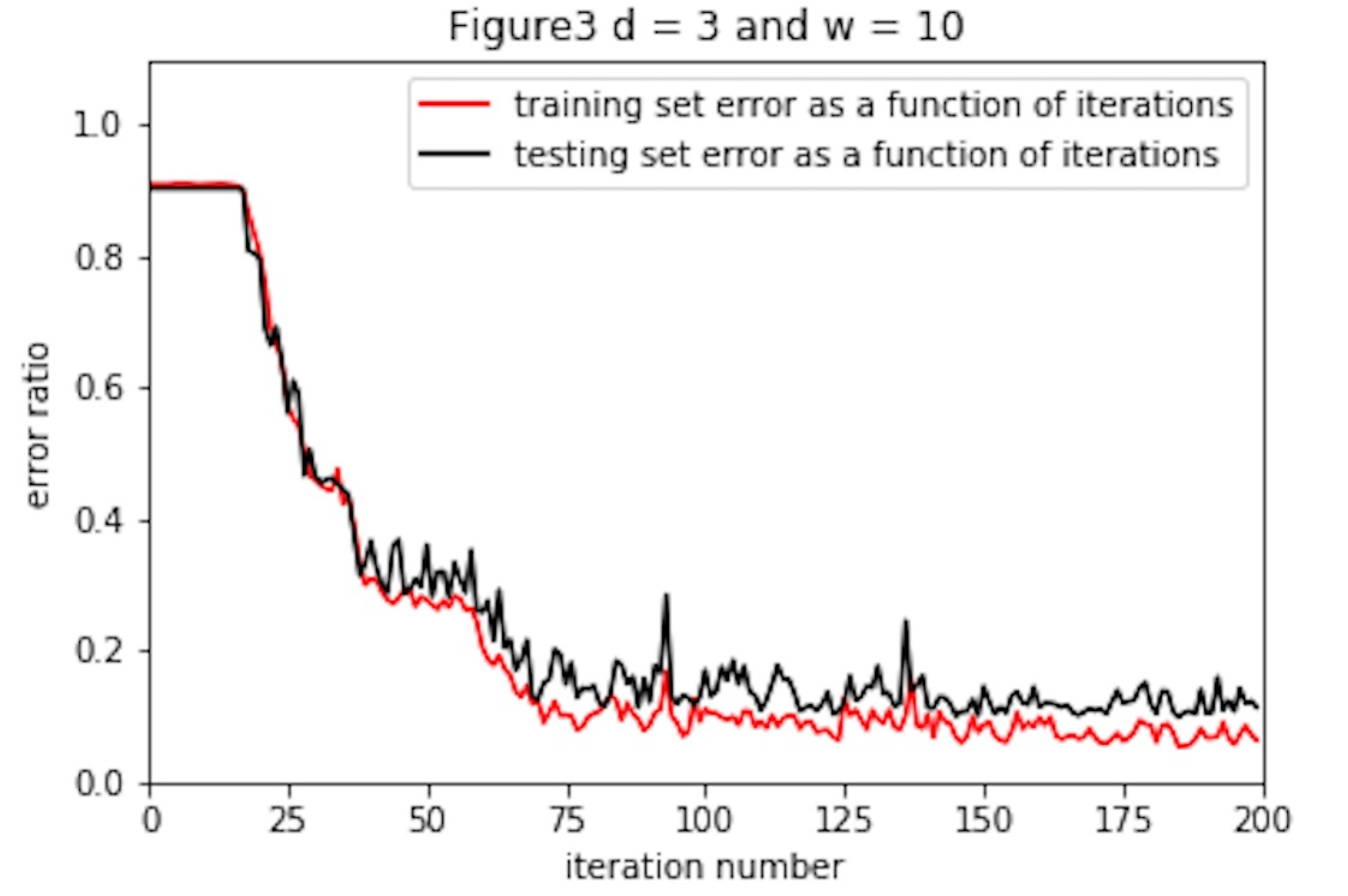
The input is [[1 0 0 0 0 0 0 0]] [[0 1 0 0 0 0 0 0]] [[0 0 1 0 0 0 0 0]] [[0 0 0 1 0 0 0 0]] [[0 0 0 0 1 0 0 0]] [[0 0 0 0 0 1 0 0]] [[0 0 0 0 0 0 1 0]] [[0 0 0 0 0 0 0 1]]

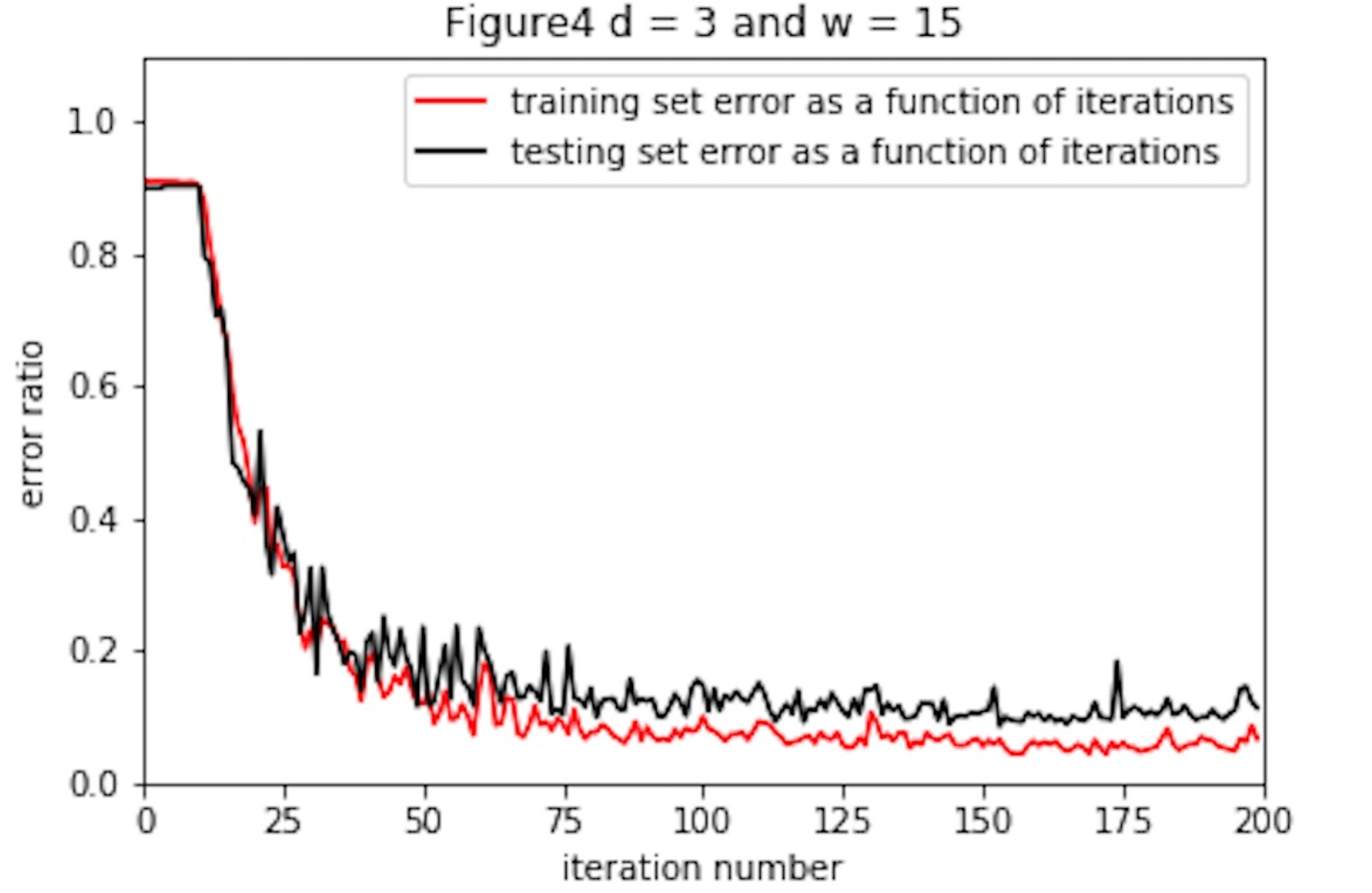
The representation is

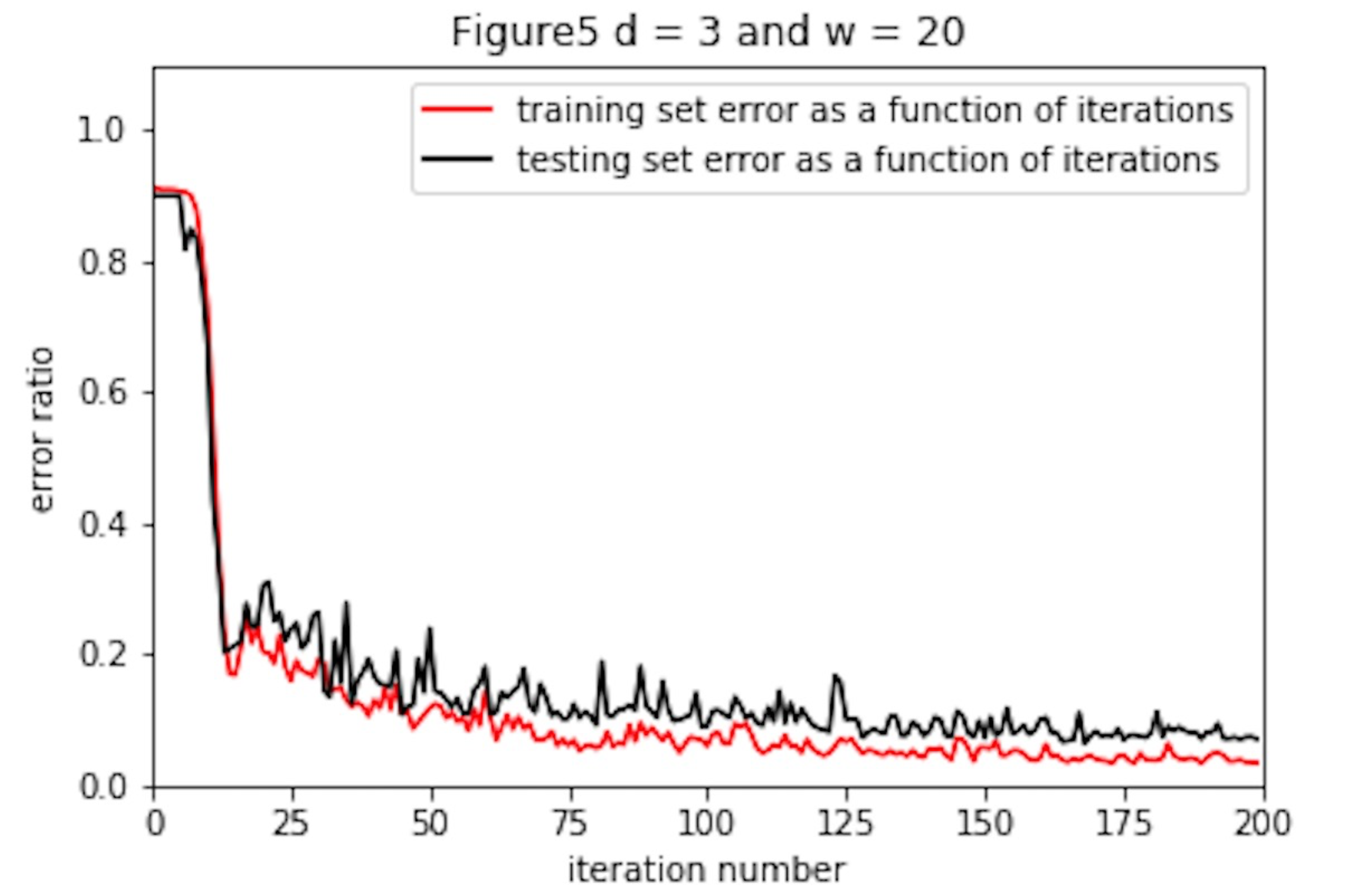
[[1, 0, 1], [1, 1, 0], [1, 0, 0], [0, 0, 0], [0, 1, 0], [0, 1, 1], [0, 0, 1], [0, 0, 0]]

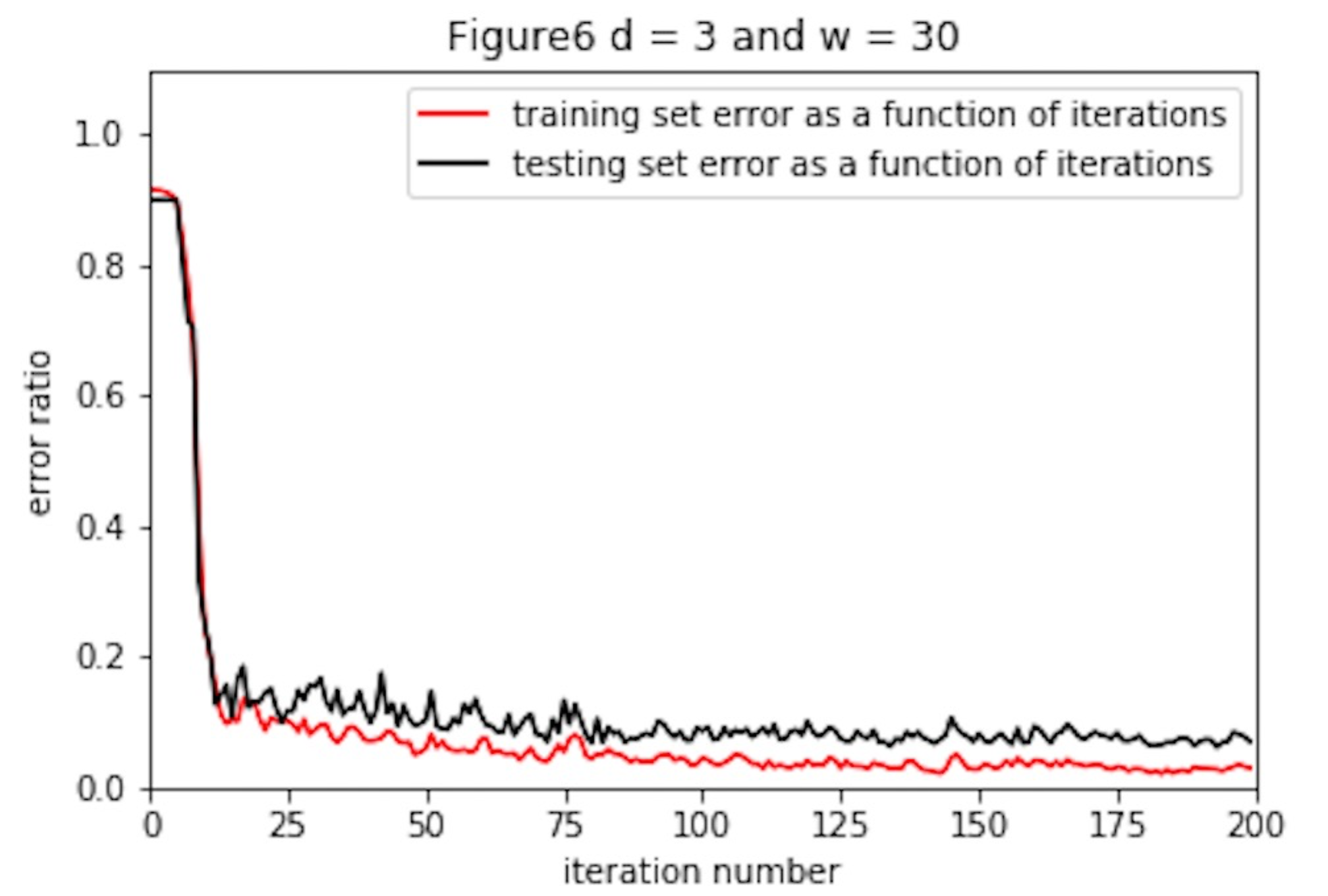
No, from the result, we can see there are the same representations for different input.

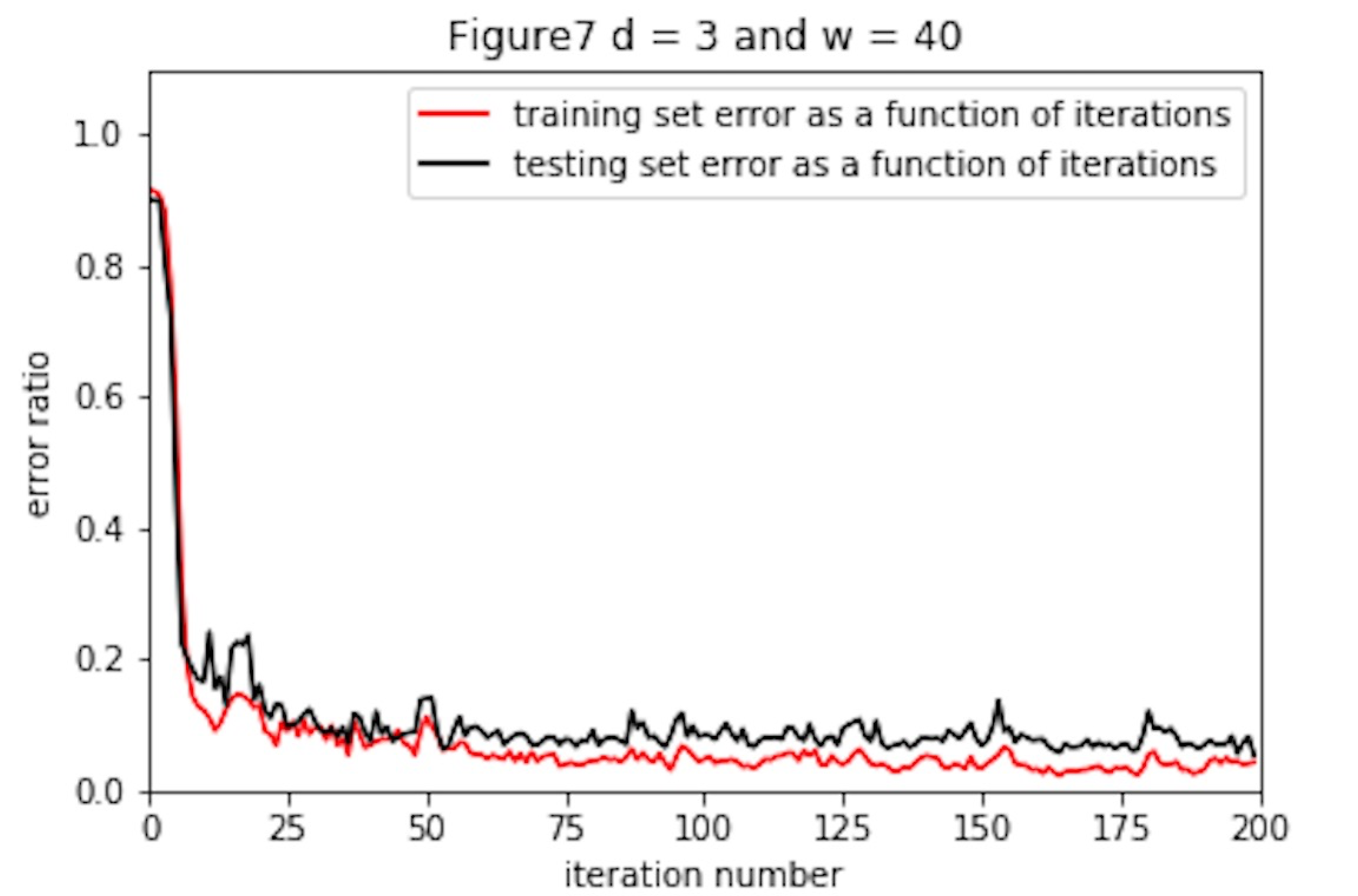


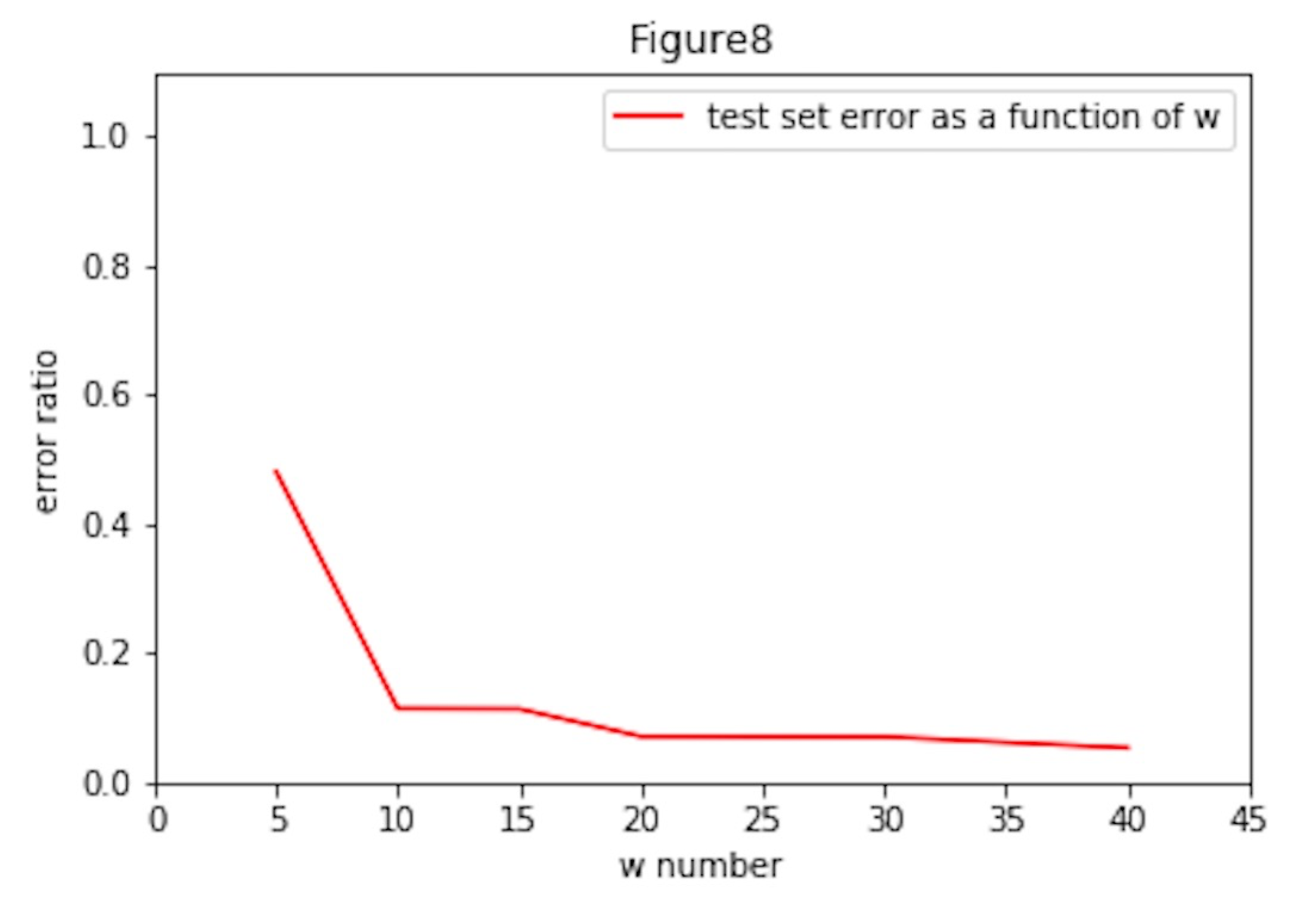


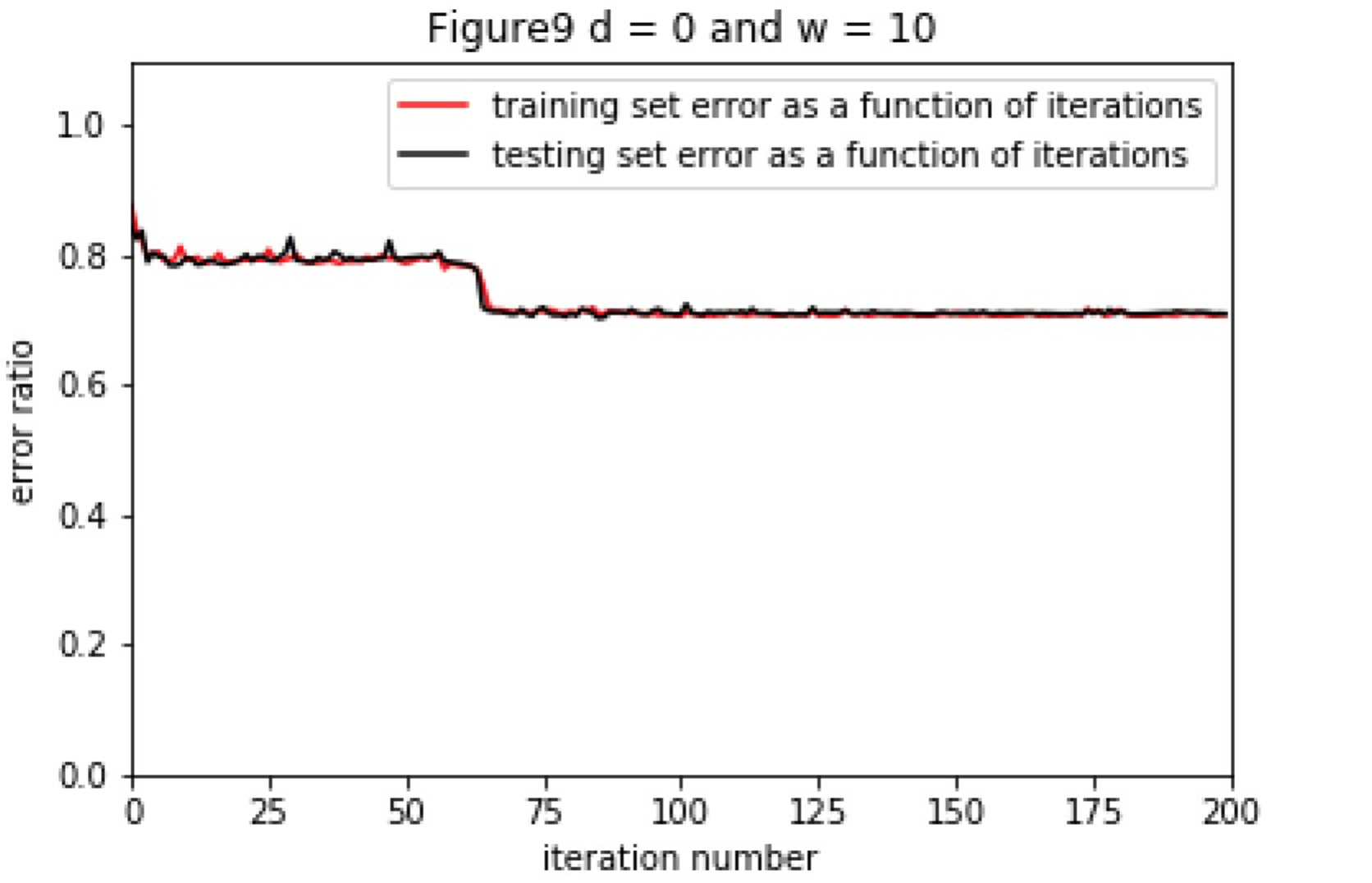


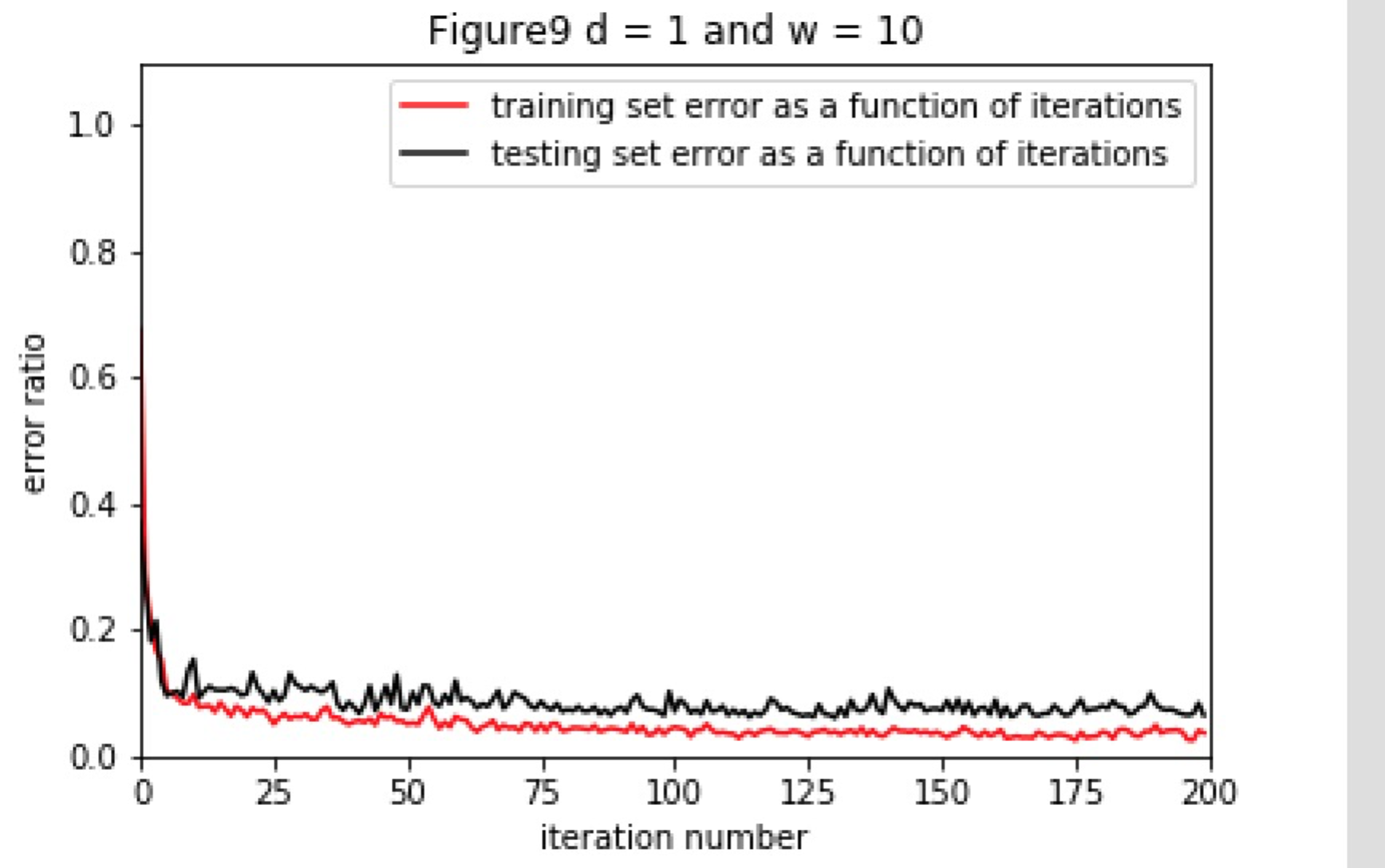


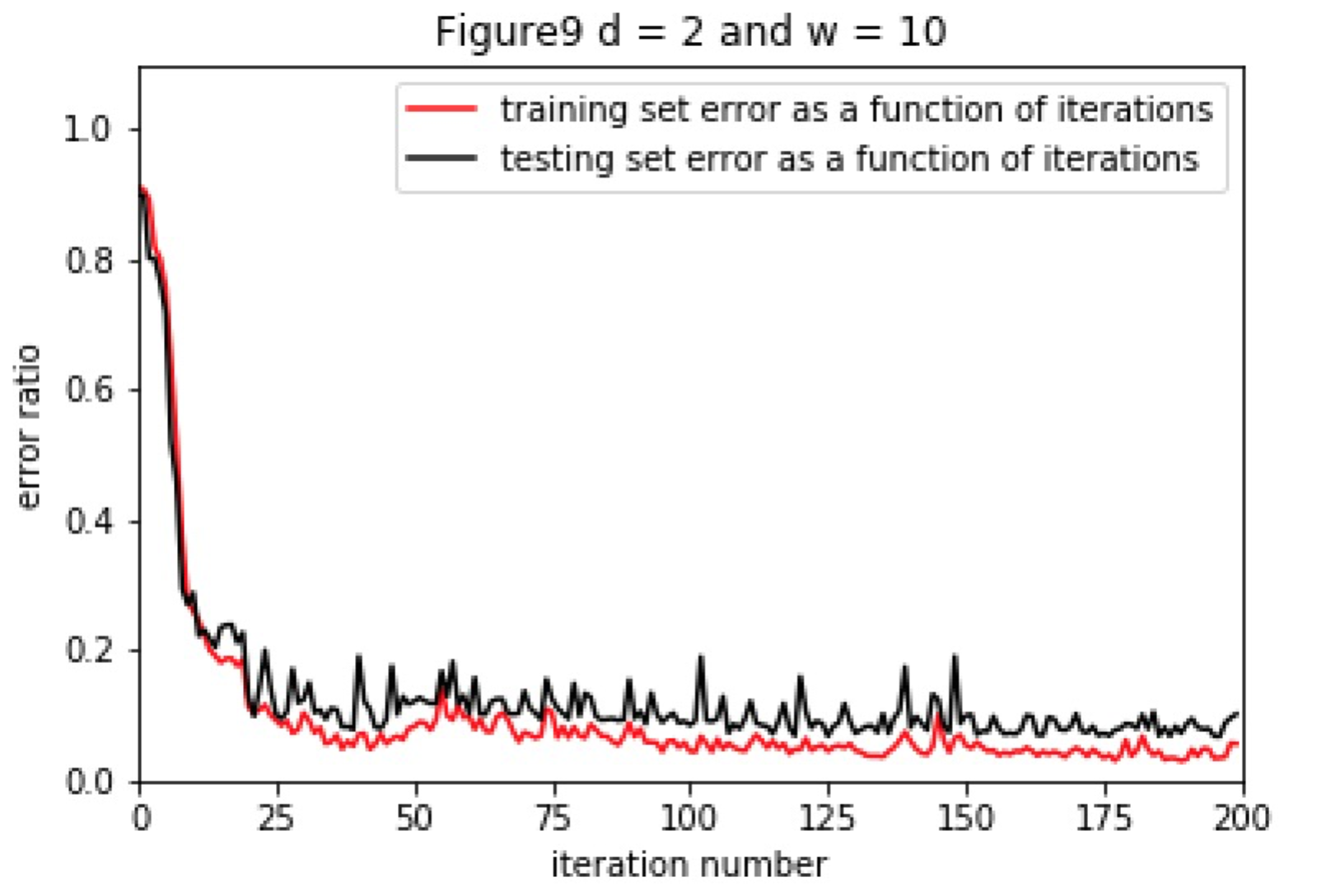


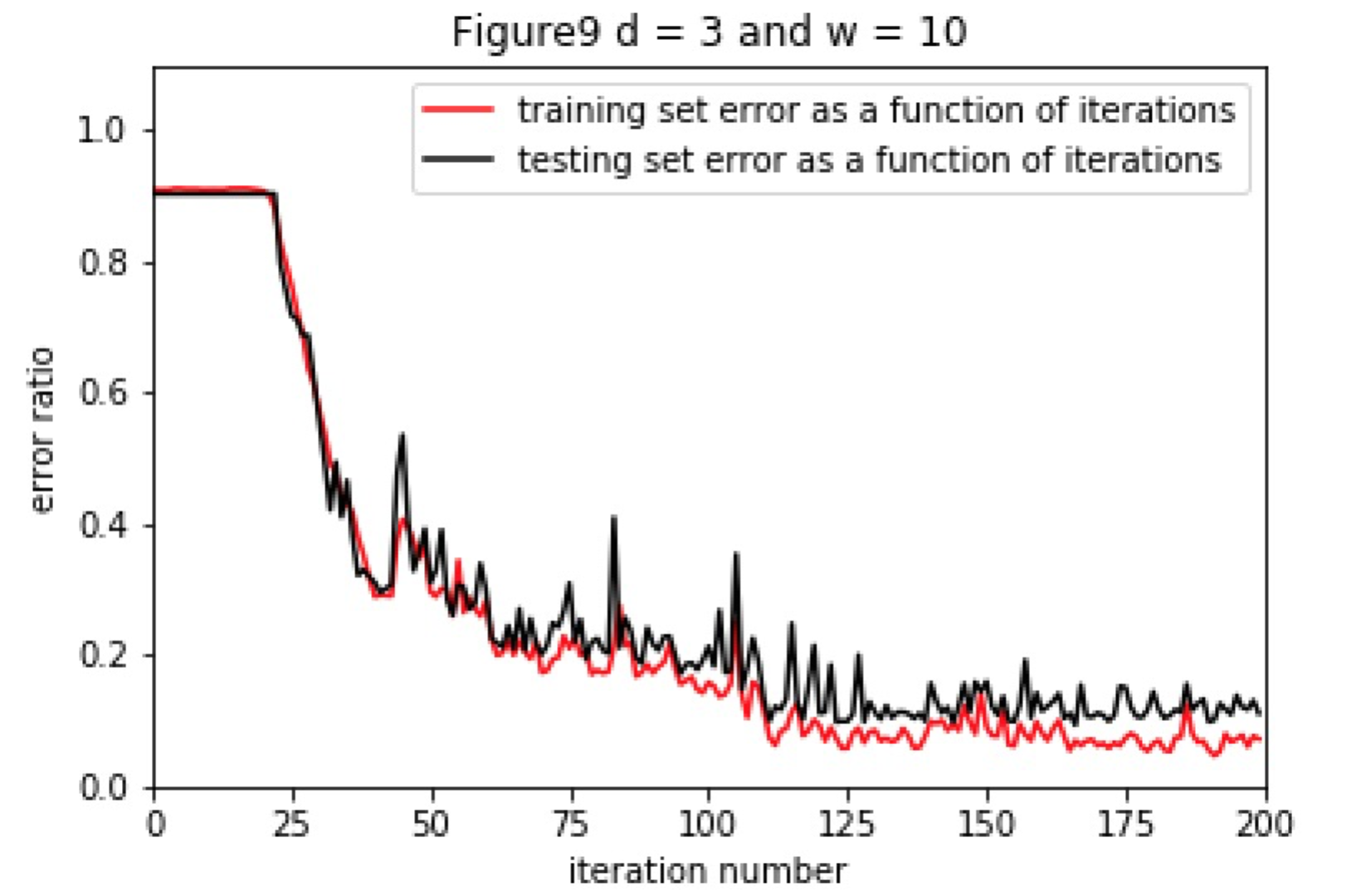


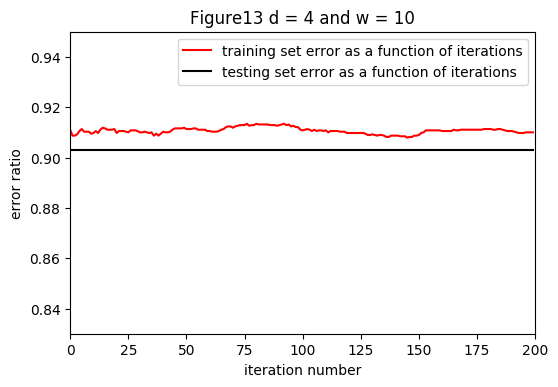


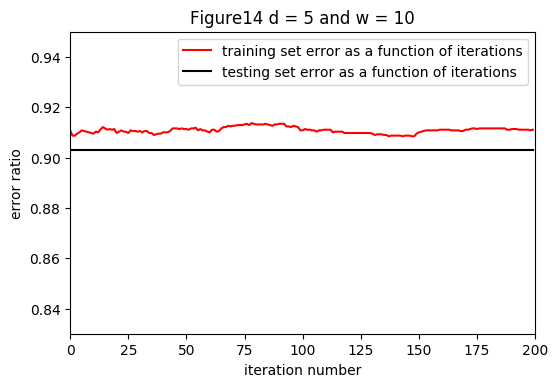




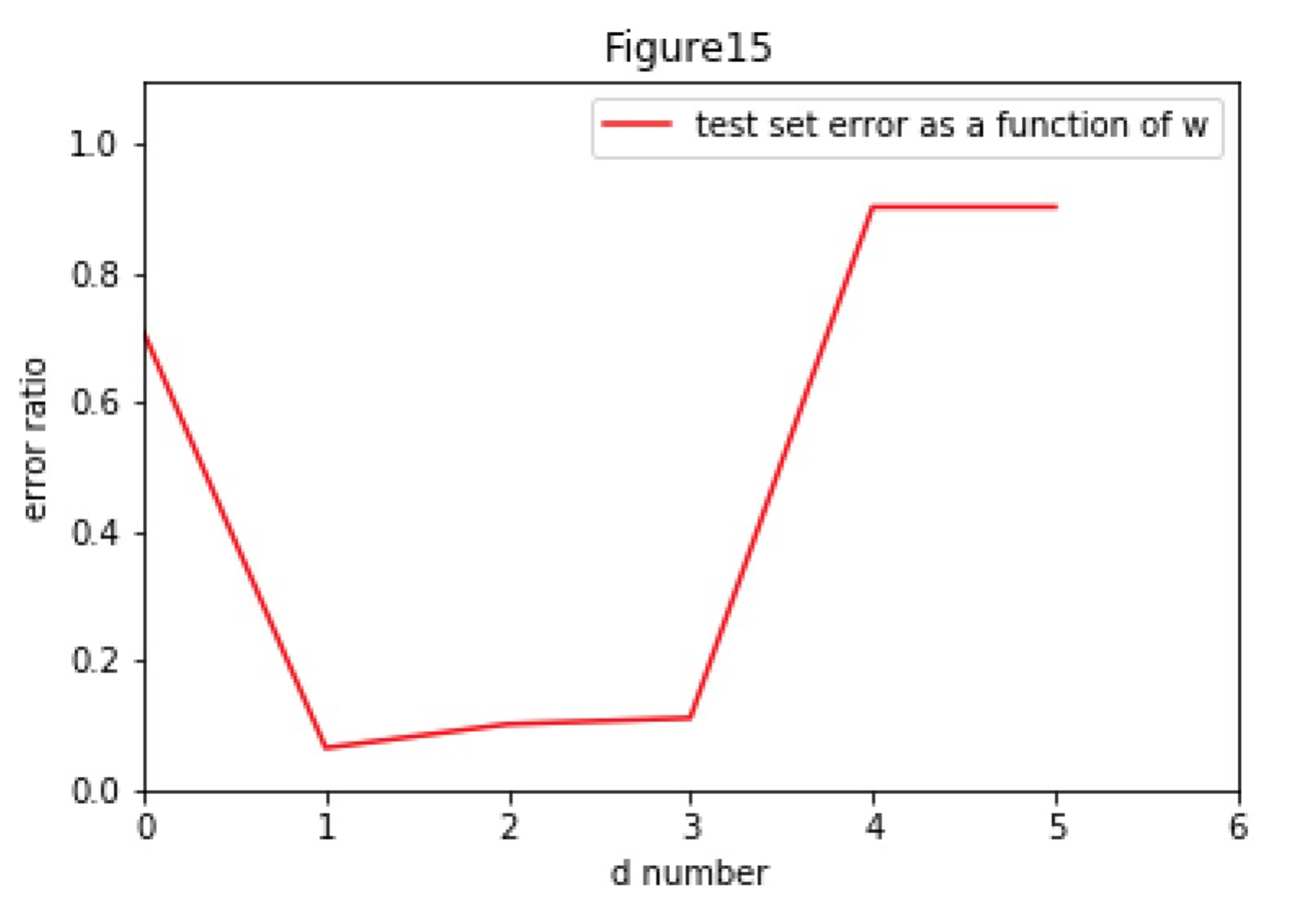








￼



二 What can you conclude from the learning curves and additional plots? How does the performance vary as a function of d and w? What is the best w, d settings among the ones you tested?

For all Figures (2-7), we see as iteration number increase, the error ratio decreased, and error ratio of training data is lower than that of testing data because our network is trained by training data. The plateau period become smaller before the error ratio decreased.

For Figure 8, we can see the error rate of test data is gradually decrease with the increase of w and the lowest value is d = 3 and w = 40.

For d = 0,1,2,3,4,5 and w = 10, the error ratio decreased, and error ratio of training data is lower than that of testing data because our network is trained by training data.

For d = 0 and w = 10, without hidden layer, we can see the error ratio decreased from approximately 80% to 50% after 200 iteration,

When d = 1 and w = 10, we can see the error rate is decreased a lot compared with d =0 and w = 10.

But with d increased to 2 and tyu3, we can see the plateau period become longer before the error ratio decreased.

Especially, d = 4 and 5(Figure 13 and 14), the error rate keep constant to 90% approximately during 200 iteration. The reason is that d = 4 and 5, we don’t have enough iterations to update w because the update of w will weaken as the increase of hidden layers. If we increased iterations, the error ratio when d = 4 and 5, will also decreased a lot.

From the Figure 15, we can see when w = 10, d = 1 it has the lowest error rate compared w = 10, d = 2, 3, 4, 5. The error ratio decreased and then increased.