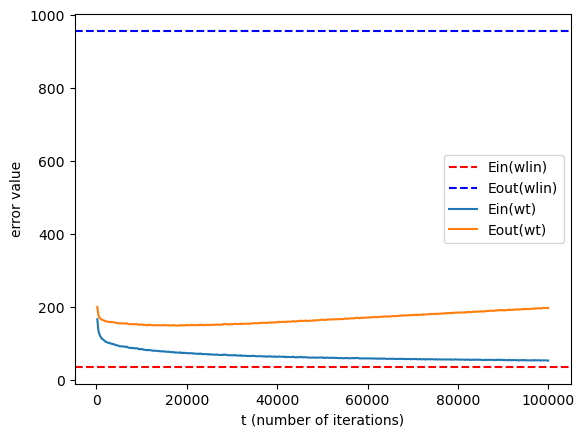
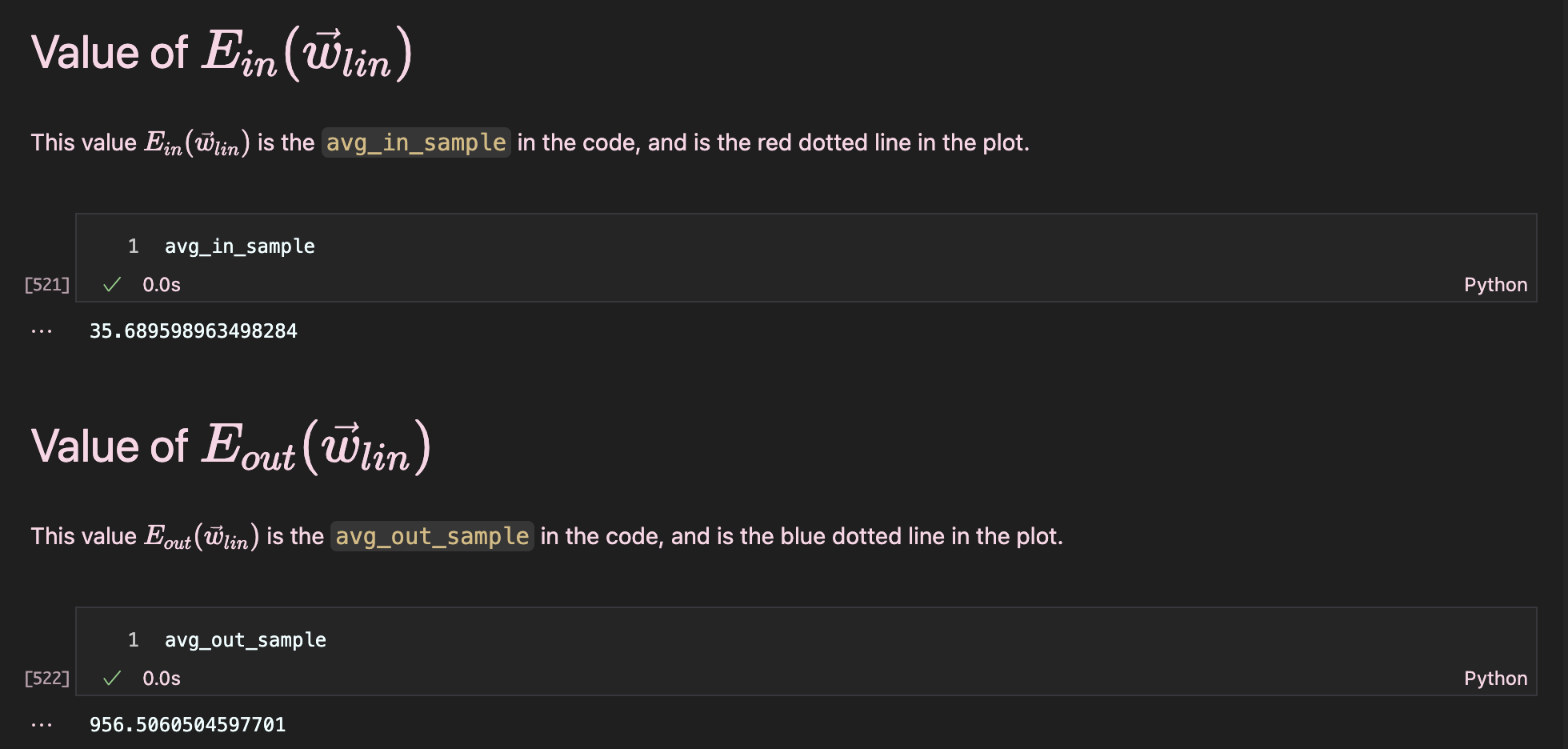
# ML homework 4: question 10

The resulting figure is as below:



Where the values of the in sample error and out of sample error which is not mentioned in the above plot is as below:



## Meaning:

We can see that at the initial values of , both the average in-sample error and out of sample error using SGD decreases rapidly, this dues to the fact that we initialize as a zero vector, which is far from the real weight vector. Therefore, at the first iterations, we make great change on the weight vector, causing great error reduction (because the gradient is large.) However, as the weight vector gets near to the optimal weight vector, the decrease of the error became smaller.

The reason why our resulting SGD has a higher in-sample error on average may due to the reason that we only use a small subset of the dataset (choose from examples) to update the iterations. This would introduce variability, and would make it hard to reach the weight vector that can have same low error as normal linear regression.

Also, another reason is that we have a closed form equation to calculate , but we update randomly at each iteration, so we may not reach the exact solution as the normal regression.

However, we can see that the gap between the in-sample error and out of sample error for SGD is smaller than the normal regression, meaning that SGD provides smaller generalization error. I think it might due to the randomness we introduced in SGD, which is the reason of higher in-sample error as we mentioned. This also meet our expectation that smaller in-sample error is not better, we should not choose weight vectors that simply generates small in-sample error.

## Code:

## 

## 

## 

## 

## 