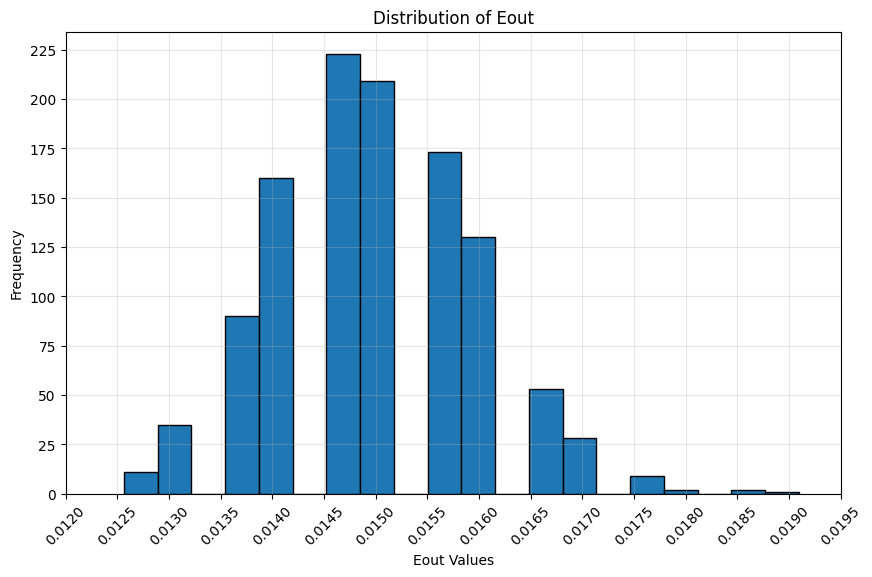
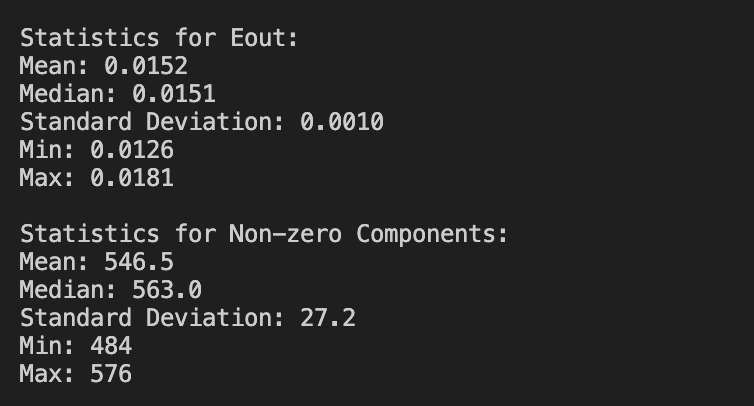
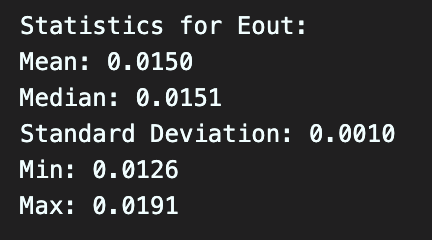
# HTML homework 5: q11 report

The resulting picture is as follows:



In order to compare with the previous question, some statistics are:

[[1]](#footnote-1)

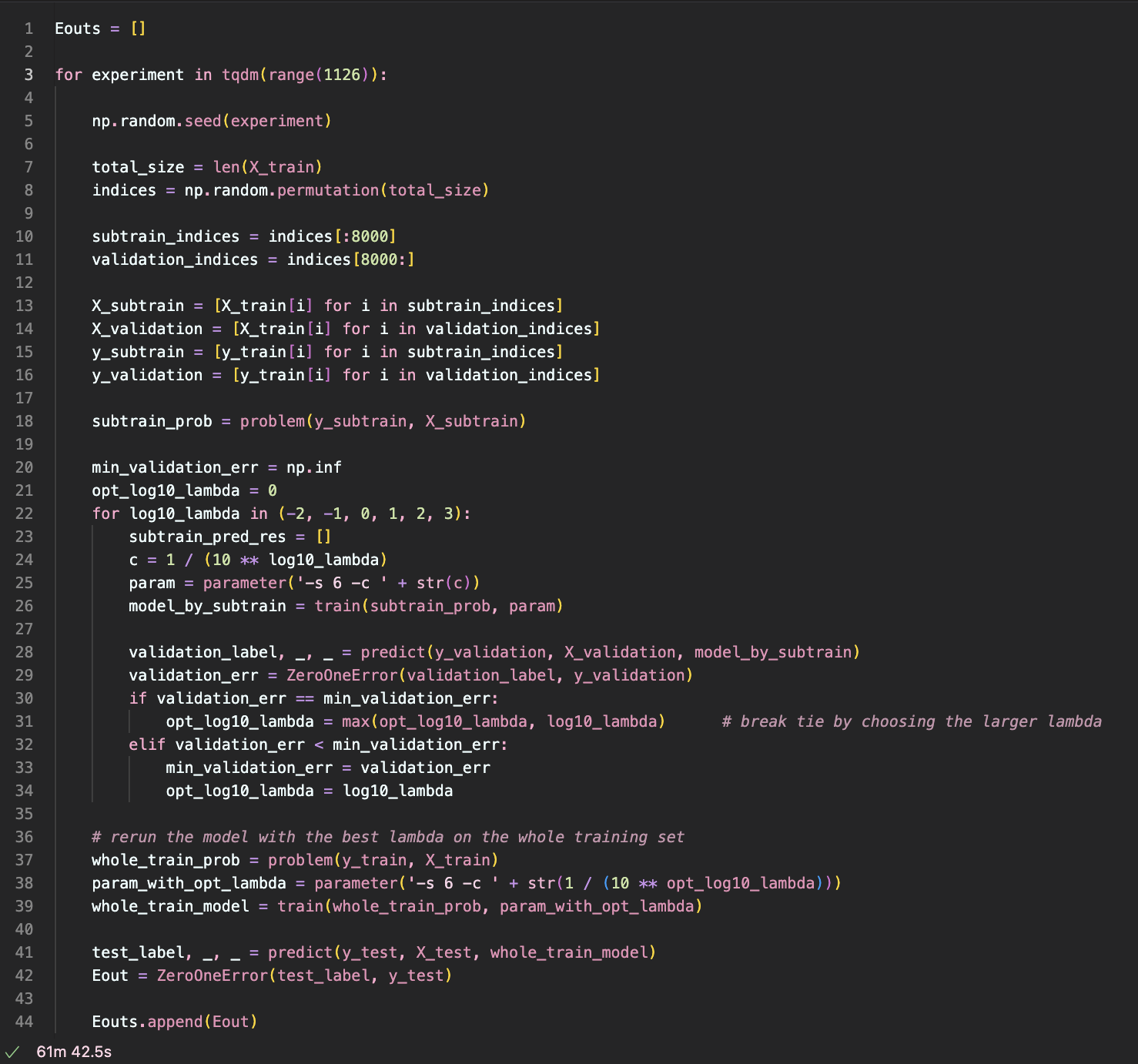
In this problem, the difference from the previous is that we further divide the training set into a subtrain set (with 8000 exampls), and a validation set (3876 examples).

From the statistics, we can see that most of them are quite similar, I think part of the reason is that, even though 3876 examples are taken out from the original training set, so one may expect that as we train on a smaller subtrain set, the out of sample error should be bigger than in q10 while we have the whole training set used, we got an expected value slightly smaller because we retrain the whole training set after the optimal is chosed. We did not use to predict the test set, but instead.

Apart from this reason, I think this result may also due to the precision, if more digits are taken into consideration, maybe we can observe the difference.

Another cause may be the limited number of possible , this constraints the variability of our model, even though there are infinitely many possible models, most of them are alike, while only slightly different by the value chosen.

## Code:



[[2]](#footnote-2)

1. Left: statistics of q11 result; Right: statistics of q10 result [↑](#footnote-ref-1)
2. Other part of codes that are similar to problem 10 (like reading in the data, split the data to get proper form of X and y, the ZeroOneError function…) are omitted in this report, can check for problem 10 for details. [↑](#footnote-ref-2)