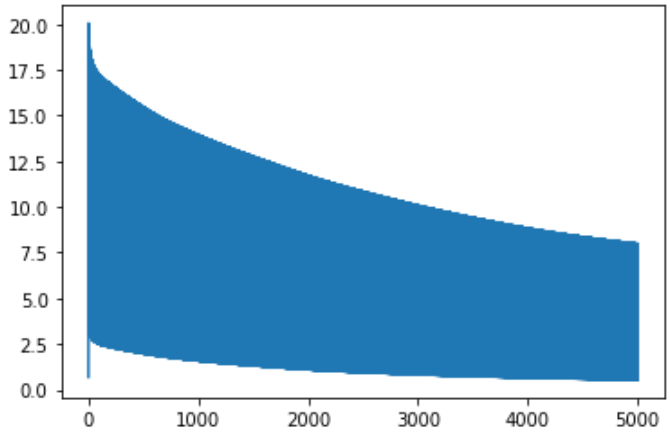
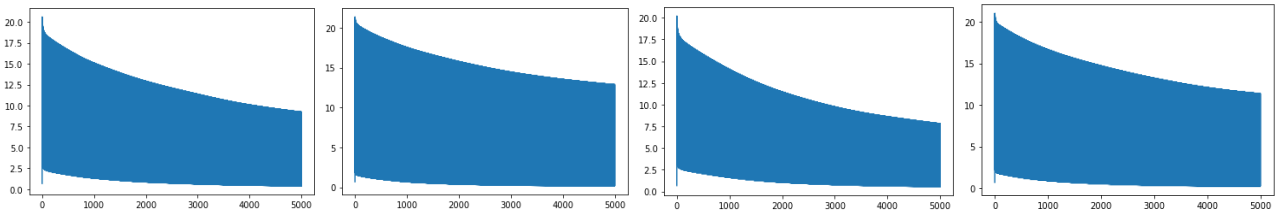
Step1

With step size = 0.1:

Test error: 0.444898 (the average of the errors of five test folds)

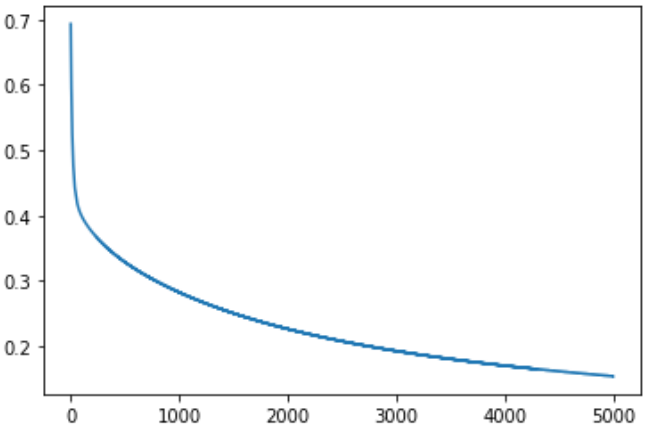


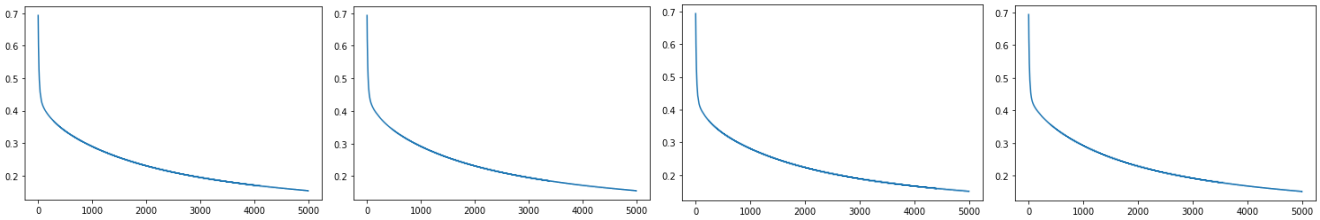


Errors for each fold: [0.621589, 0.257974, 0.394310, 0.491202, 0.459416]

With step size = 0.001

Test error: 0.163421

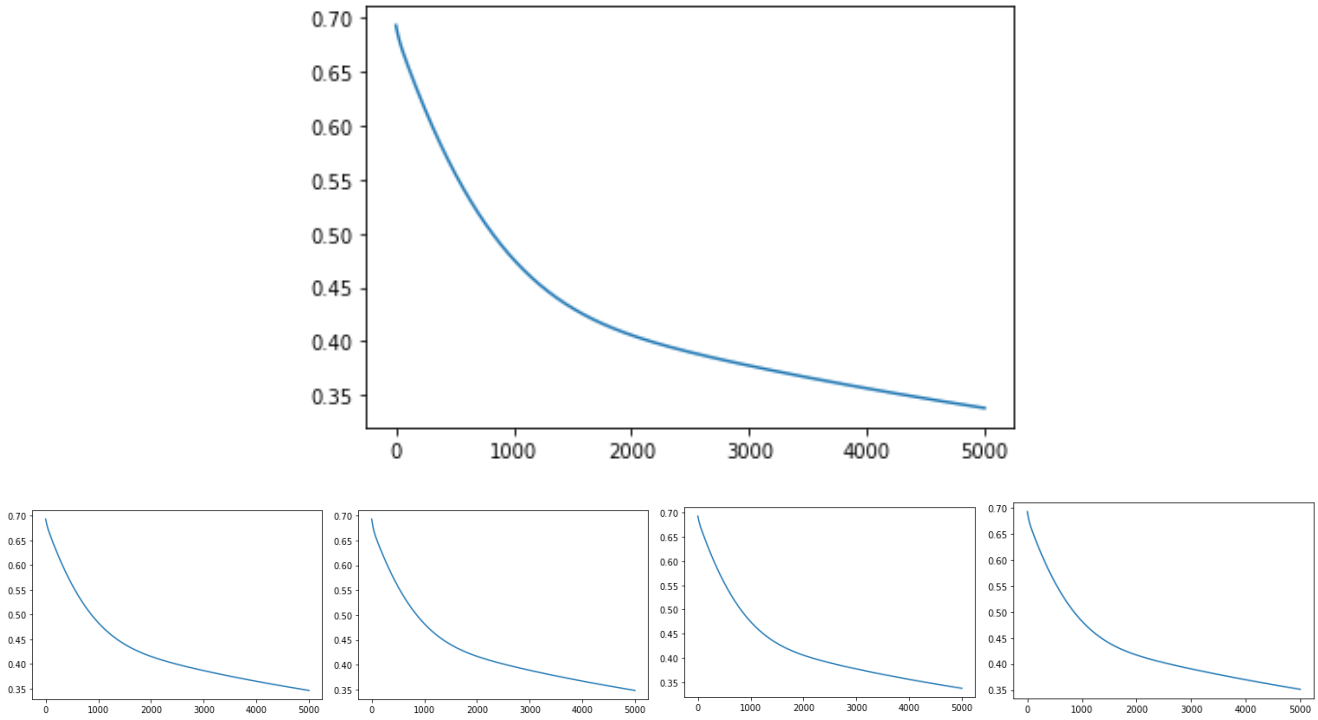




Errors for each fold: [0.166125, 0.150709, 0.163170, 0.176681, 0.160418]

With step size = 0.00001

Test error: 0.351341

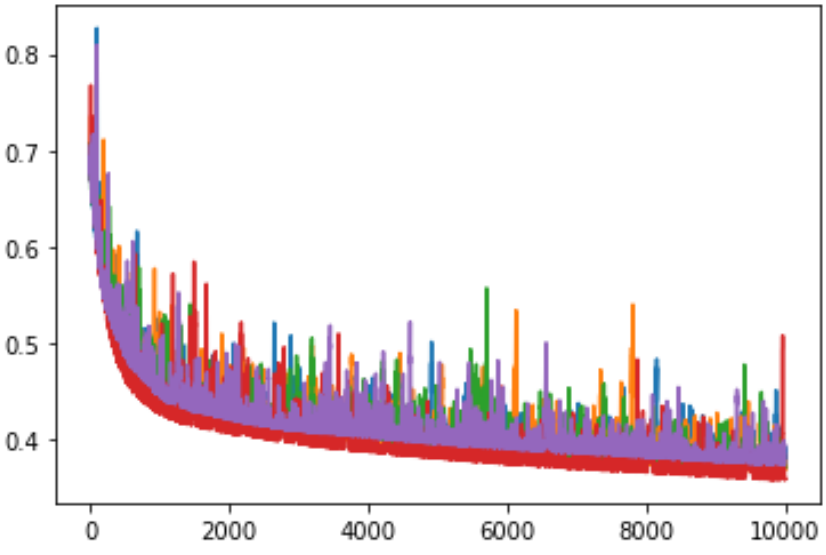


Errors for each fold: [: 0.377450, 0.341822, 0.331629, 0.383972, 0.321834]

Step2

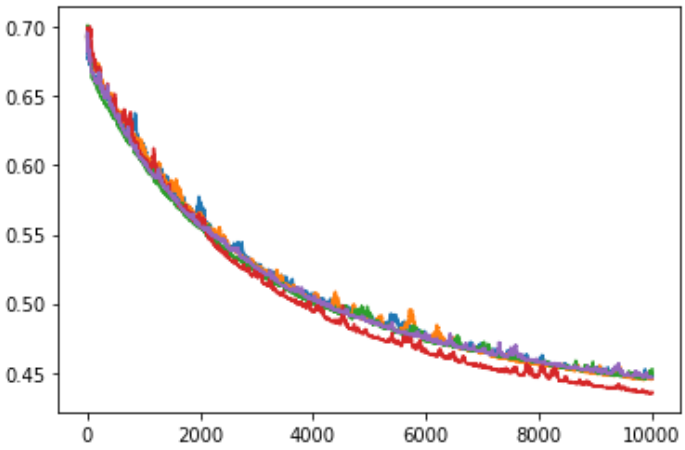
With n = 0.000001

Test error: 0.323816



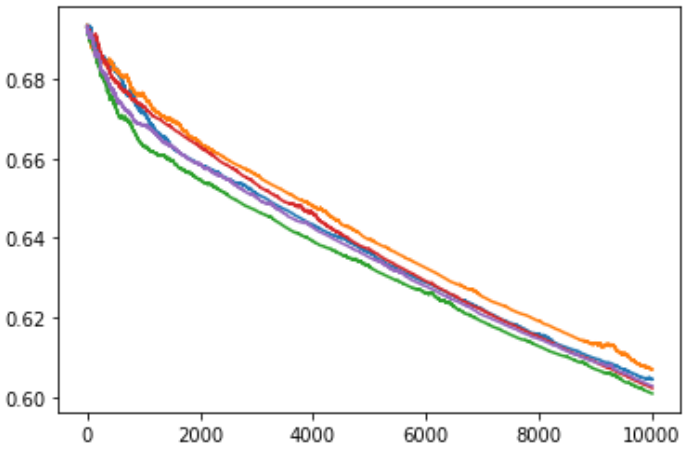
With n = 0.0000001

Test error: 0.474380



With n = 0.00000001

Test error: 0.553690



In these models, stopping criteria does not depend on error difference between iterations, the iteration number is given explicitly by observing the error values with different iteration number.

Batch gradient gets better results in less iteration number than stochastic gradient descent, but the model is more complicated and costly. Since required iteration number is less than stochastic gradient descent, it also takes less time.

While step size is decreasing, the error decreases smoothly. However, it needs more iterations to reach its minimum value. When step size = 0.00001, error reaches the values around 0.35 after 5000 iterations. When step size = 0.001, error reaches the same values after ~50 iterations. Therefore, we can choose the step size 0.001 as the best from our step sizes since its error is the minimum one. We could get the same error also with another step sizes lower than 0.001 but we would need more iterations therefore more time.

In reported test results, the test error of batch gradient descent is minimum with step size = 0.001, when number of iterations fixed to 10000 but it can be a different number with different stopping criteria or different number of iterations. The optimal step sizes differ according to other parameters for each problem.