

For this question you will change some parameters in the “Deep dive: gradient descent” notebook, re-run the notebook with the new parameters, and answer questions about the results. You do not have to write any new code, and you should not submit any code. (Copy the relevant output images to a regular document, answer the questions there, and submit that document - don’t submit a Colab notebook.)

A. Re-run the “Descent path” section with three different learning rates:  $lr = 0.0002$ ,  $lr = 0.002$ , and  $lr = 0.02$  (and leave other parameters at their default settings). For each learning rate,

- Show the plot of coefficient value vs. iteration, and the plot of the descent path on the MSE contour.
- What is the estimate of  $w$  after 50 iterations?
- Describe whether the gradient descent diverges, converges within 50 iterations, or starts to converge but does not get close enough to the optimum value within 50 iterations.

B. Re-run the “Stochastic gradient descent” section with  $lr=0.1$  and  $n=1$ , then with  $lr=0.01$  and  $n=10$ , and finally with  $lr = 0.001$  and  $n = 100$  (and leave the other parameters at their default settings). For each,

- Show the plot of coefficient value vs. iteration, and the plot of the descent path on the MSE contour.
- Comment on the descent path. Does it converge smoothly to the optimal solution?