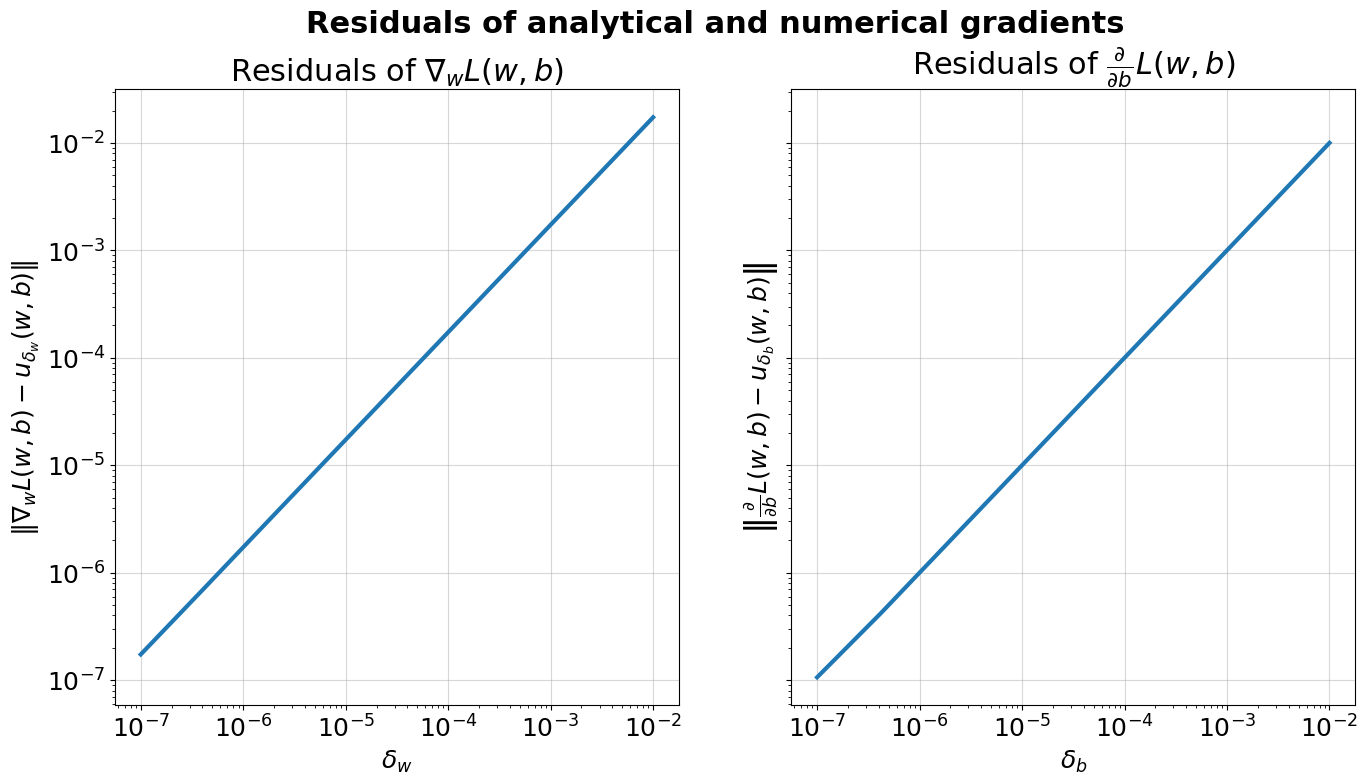
Homework 3

**Introduction to Machine Learning (02360766)**

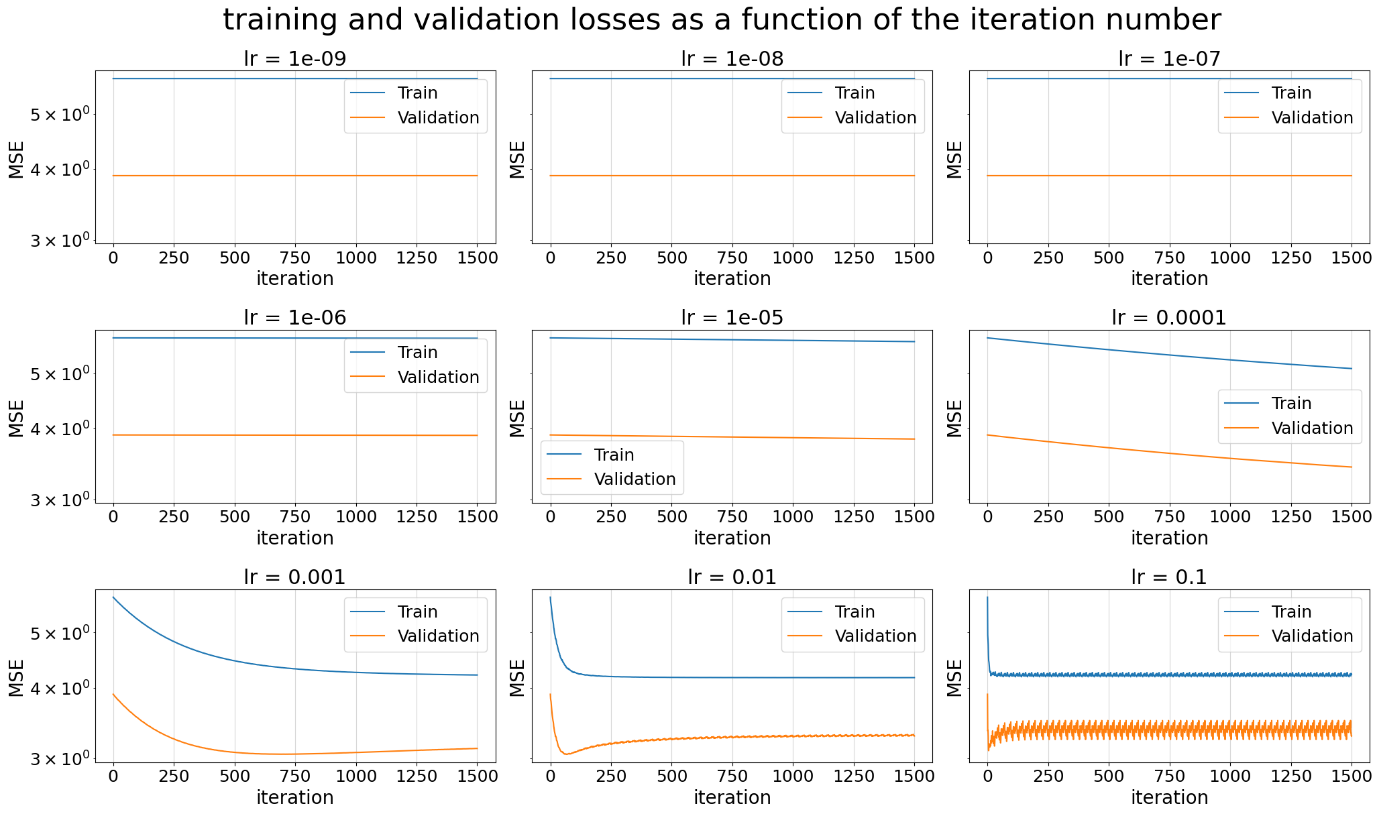
Submitted by: Gal Porat, Eyal Amdur

ID: 318849270, 315116095

**Section 1: Linear regression implementation   
(Q1)** The analytical partial derivative is:

**(Q2)** Our results:

**(Q3)** Here are the graphs of the learning rates:



As we can see, as we increase the learning rate, there is an improvement in the rate of error reduction. We can explain this by noting that for 1500 iterations, small steps are not "sufficient" to improve the model and reduce the error, whereas larger steps lead to faster convergence.  
The “optimal” learning rate is 0.01 lr because it achieves stable convergence in about 100 iterations. There’s no need to add more iterations, as they will not significantly improve the error.

If we want to achieve the minimum error with the minimum number of iterations, we should choose 0.1 lr. Although it is less stable than 0.01, it converges faster to the close error and stays around those values.

* **Note:** the training error is higher than the validation error, which is strange at first but we found out that it depends on the partition of the data, so it can be explained by “lucky” partition.

**Section 2: Evaluation and Baseline**

**Top of Form**