Exercise 1A:

Using gradient descent, train a linear regressor to solve the function Ax + By + C with 150 random samples (0-1) for x and y using the **mean** **absolute error** [1] objective/loss function. Then, perform the following task and questions below. Feel free to use the Gradient Descent example notebook as reference. Submit the notebook and this document containing the answers.

* 1. A is based on your first name first character (A=1, B=2, … Z=26) while C is based on the last name first character. B is C – A.
  2. Submit the notebook that produced the correct coefficients (within 1e-3 difference). Use the same parameters in the following items.
  3. What happens if the learning rate is multiplied by 300? What are the learned parameters for A, B, and C?
* Having a very high learning rate of 3 does not allow the gradient to converge and loop finished quickly. It caused an OverflowError. The learned parameters are as follows:

(-1.6780838825362156e+154, -3.2500436928549913e+154, -4.768518127920449e+154)

* 1. What happens if the learning rate is divided by 150? What are the learned parameters (A, B, and C)?
* Having a learning rate of 6.666666666666667e-05 makes the loop go continuously. The learned parameters are as follows:

(11.126659420828128, 8.421468097508921, 18.332508793481615)

* 1. What happens if inputs X and Y are multiplied by 1000? What are the learned parameters (A, B, and C)?
* Multiplying the inputs by 1000 makes the loop finish quickly and causes an error. The learned parameters are as follows:
* (3.826735902939471e+153, 1.9510721348892192e+154, 2.004511217826663e+151)
  1. What happens if the inputs are reduced to just 15 samples? What are the learned parameters of (A, B, and C)?
* Reducing the inputs to 15 samples allows the loop to go continuously. The error becomes significantly small but still the coefficients are not the desired value. The learned parameters are as follows:
* (12.769872686511782, 6.949813720734405, 18.622981550575048)
  1. What happens if the inputs are increased to 1500? What are the learned parameters of (A, B, and C)?
* Increasing the inputs to 1500 samples allows the loop to go continuously for a long time. The loop went on continuously however my device could not handle processing and keeps crashing before I can get the learned parameters.

Note:  
[1]: *This means that the gradient must be computed from the derivative of the absolute value function. The n in the mean absolute error also refers to the number of data you used in an update. In the example’s case, 1*.