

Q2 – Brief Discussion

Generally, with a smaller step-value ' ϵ ', the algorithm requires more steps to adequately converge to within a given distance of the 'true value' of the modeled loss function and thus results in slower processing times. However, results, especially at each step, are also generally more accurate w/ smaller steps. The gradient descent algorithm developed here tends to fail when, for a fixed number of iterations, the steps are 'too small' and the initial guess is too far away from the true value of the underlying loss function. Notably, larger step sizes also tend to result in two failures: (a) they can trigger overflow errors due to size of intermediate values involved in calculating gradient, or (b) they occasionally cause the gradient descent approximator to diverge, especially if initial guess is too far off.