

Hackathon1

Mason Lien

learning rate greatly influences the convergence of gradient descent to the solution. When I increased the lr to 1.0, the squared error increased. This is due to learning too quickly and basically making big adjustments too quickly. I kept the iterations constant when adjusting lr for the sake of understanding. When I reduced learning rate .25 \rightarrow 0.1, the loss value we that we want to minimize was reduced. This means making smaller adjustments improves convergence, meaning we are closer to the solution. When changing iterations, this is essentially how many times we want to update x and optimize the gradient. in the first couple of iterations, there are big changes, but as it gets closer and closer to the solution, the changes become gradually less. The lr and iterations need to be optimized in order for good convergence. Typically, this is when you would plot the loss value over time (iterations) to see where optimal iterations are for that specific learning rate.