Differences in estimated coefficients using gradient decent and ordinary least squares.

OLS and Gradient decent are different methods for determining optimal coefficients in linear regression models. OLS uses a direct formula for the solution of the problem, while Gradient decent computes them by starting with a set of random coefficients and adapts them with the purpose of minimizing a cost function. The cost function in gradient decent, as in OLS, is measured by the mean of squared errors. Differences in the results between the two methods arise due to the nature of each of the methods. Gradient decent keeps trying to improve its solution bit by bit, it can be affected by how it starts, since the initial parameters(thetas) are usually set at random, this might lead to it getting stuck in not-so-great (local minima) solutions, especially in complex situations. In comparison, OLS is quicker for smaller to medium-sized sets of data. Yet on the other hand, Gradient Descent is better for handling bigger sets of data and lots of features.

In our case when looking at the gradient decent cost function plotted against the number of iterations, we see that it showcases a smooth decent, meaning that in the first iterations the cost function decreases rapidly, since the algorithm makes significant improvements to reduce the error. As the iterations progress, the rate of decrease in the cost function remains relatively constant. This indicates a well-balanced learning rate that allows the algorithm to converge efficiently. The overall shape of the plot resembles a smooth curve without erratic jumps or oscillations. This indicates that the learning rate is not too large, causing overshooting, or too small, causing slow convergence. Towards the end of the plot, as the algorithm approaches convergence, the rate of improvement in the cost function starts slowing down and the cost function starts to flatten indicating that the algorithm has found, or at least has come close to, finding the minimum.