

Why is the before value of the fourth point nonzero?

The fourth point has similarities with the second point so it remembers the weights from the second point when it calculates the $f(x,y)$.

Why does the MSE not decrease further towards zero?

There is still random noise in the data so the function cannot be exactly perfect.