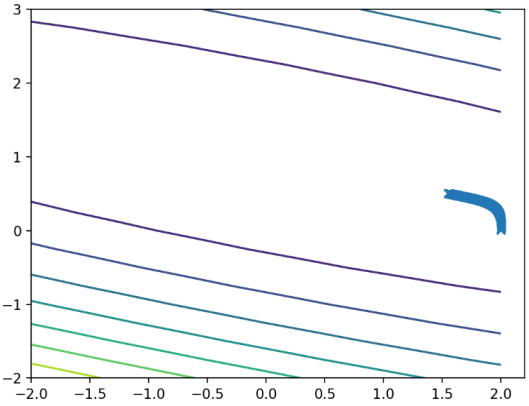
Report - LAB 5: Linear Regression

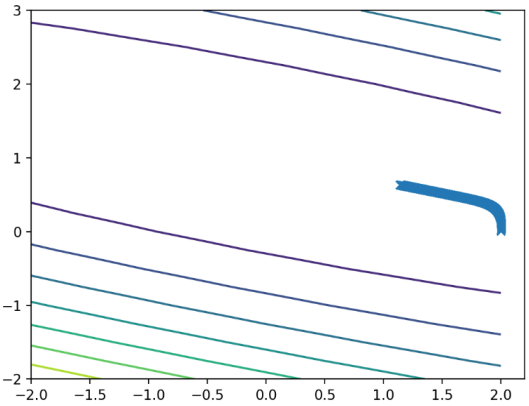
1. **What happens if the learning rate is too low?**

Hint: check different values for the learning grade and analyse their impact

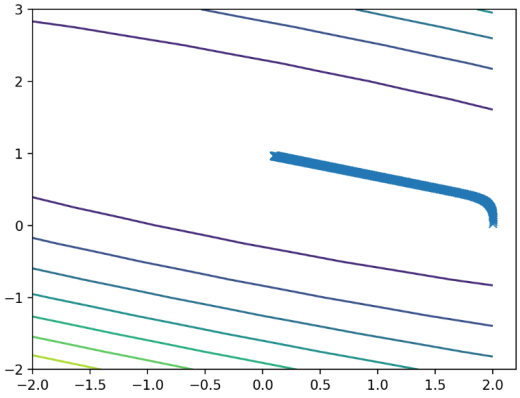
Answer: When the learning rate is too low, the convergence may slow down, however, it is usually more stable, as it takes more steps. Since the convergence takes longer, it is also possible that the maximum steps are not enough to reach to the point of convergence. The examples from below might help us demonstrate this better.



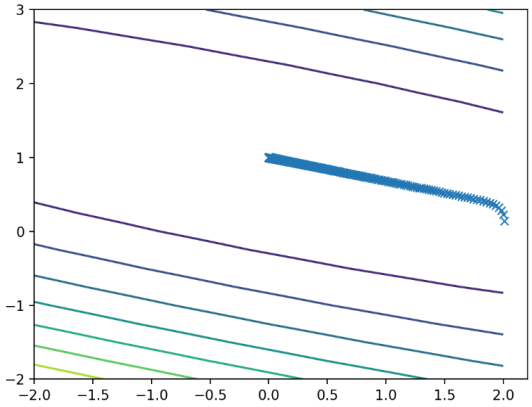
Learning rate = 0.001



Learning rate = 0.002



Learning rate = 0.01



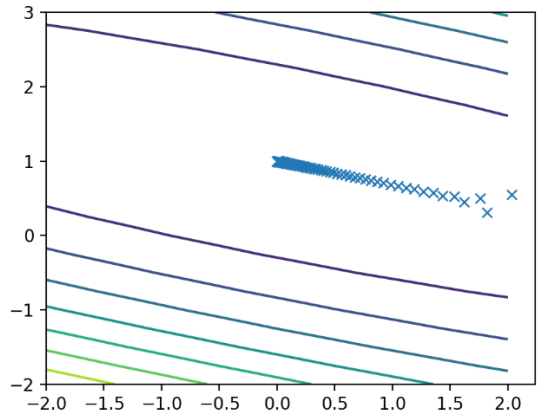
Learning rate = 0.05 (the one given to us)

These plots demonstrate the impact of low learning rates. While they are stable and appear to reach the same point, some were unable to reach within the steps that were allocated to them.

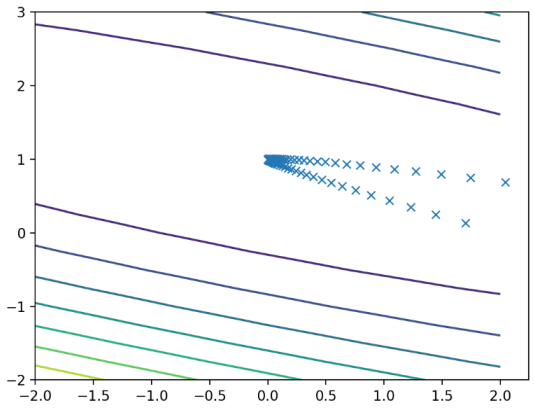
2. What happens if the learning rate is too high?

Hint: check different values for the learning grade and analyse their impact

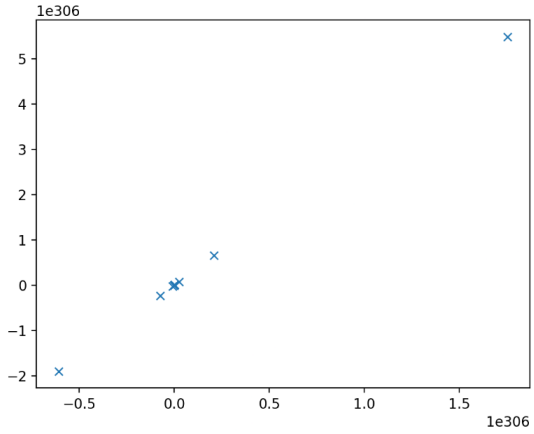
Answer: If the learning rate is too high, the convergence might speed up but the likelihood of overshooting and diverging also increases.



Learning rate = 0.2



Learning rate = 0.25



Learning rate = 0.5

The plots above show us how the algorithm fails as it does not reach convergence. This is can be attributed to overshooting, since the step sizes are too big for this data set.

3. Can Linear Regression really find the absolute global minimum? Explain why/why not.

Hint: you can explain the theoretical approach and prove/disprove it empirically

Answer: Linear regression can find the absolute global minimum but it is more likely that it converges at a local minimum. Since the cost function is convex, there is only one minima, the global minima which would yield an optimal result. The gradient uses the convexity to get closer to said minimum, if maximum iterations are enough with the step size to reach there. The figures from the first two questions can demonstrate how with the correct learning rate, the global minima is achieved, however, with the large step sizes we fail to achieve them. This shows that linear regression CAN but does not necessarily find the absolute global minimum.

4. What effect does it have if you change the initial guess for thet0 and thet1 for the gradient descent to something completely off?

Hint: check different values for the theta values and analyse their impact

Answer: The example “completely off” values that I took to analyse their impact were, with the learning rate of 0.2, the convergence seems to be to the same point within 1000 iterations:

```
theta0 = np.arange(-10, 20, 1)
theta1 = np.arange(-10, 30, 1)
```

```
Optimized Theta0 is  2.2192605568817342e-16
Optimized Theta1 is  0.9999999999999999
```

The example “completely off” values that I took to analyse their impact were, with the learning rate of 0.05, the convergence seems to be to the same point within 1000 iterations:

```
theta0 = np.arange(-10, 20, 1)
theta1 = np.arange(-10, 30, 1)
```

```
Optimized Theta0 is  3.9852928381292103e-07
Optimized Theta1 is  0.9999998725702243
```

Despite guesses being significantly different from the expected optimal values (0 and 1), the gradient descent algorithm consistently converges to values extremely close to these expected optimal parameters within 1000 iterations. This resilience against initializations 'completely off' showcases the algorithm's stability and efficiency in finding the global minimum. Notably, different learning rates (0.2 and 0.05) affect convergence speed, with a higher rate leading to faster convergence. But a very fast learning rate might still lead to instability.

5. What happens if you are not updating thet0 and thet1 “simultaneously” as you should but you are updating both parameters in separate for-loops (see code)?
Hint: evaluate it empirically by modifying the code accordingly
Answer:

6. How many iterations of the gradient descent algorithm do you have to perform to reach the correct exact values of theta0 and theta1?
Hint: explain how you calculated it and demonstrate it with experimental results
Answer: Code crashed when checking results :/