Questions: (Activities- Week 5)

Part 1

1. Do you think the slope of linear regression is always positive? Explain it.

Slope of the line can be positive or negative, this just shows relationship between the two variables. If Y increases with X, then slope is positive, if Y decreases with increasing X, then the slop is negative.

2. Is a straight line always a good model? In which situations, it is not?

A straight line may not be a good model in those cases where there is non-linear relationship between the input and output variables

Part 2

3. Play with m and b to fit the model best. What are you best values?

You should be able to change values of m and b, and see the new plot. In fact the given values of m=10.5 and b=49 seem to give the best fit but we are not sure if this is the optimal unless we compute the error.

Part 3

4. Is there always one line that fits best?

No, there might be more than one lines which give the same error.

Part 4

5. Is gradient descent only applies to two dimensions?

No, we have used 2-dimensional because this is easy to understand but it can be applied to N-dimensions.

Part 5

6. Why m and b values are very different to the ones we chose in Part 2?

m and b value are very different, in fact they are not optimal because we have done only one iteration. We need to do a number of iterations to get the best answer. You will seethis in Part 6.

7. What is the step here?

Step is 0.01 here as shown in the step_gradient function.

Part 6

8. What happens of there are outliers?

Outliers can happen in reality and are not always avoidable. In addition, unless there is a very good reason to remove them, outliers cannot be easily excluded from real-world data.
As a result, outliers can appear somewhat frequently in datasets. When there are outliers, they can have a fairly strong influence on the slope and intercept of the dataset. As a consequence, the line of best fit for the data can be slightly off, or possibly even completely wrong

9. How to treat outliers, so that it has minimal effect on the slope and intercept?

There are actually some algorithms used to detect and remove outliers.

One common method is to clean sensors or instrument data. Sometimes the instrument catches noise from the environment, the electronic circuits experience random drift, and other factors that makes what is probably just a false reading manifesting as an outlier in the data. So it is always best to visualise data and identify outliers.