Earl James Q. Rentillo
CMSC 197 – Assignment 3

1. What are the optimal weights found by your implemented gradient descent? Plug it into the linear model:

$$h_\theta(x) = \theta_0 + \theta_1 TV + \theta_2 Radio + \theta_3 Newspaper$$

What are your interpretations regarding the formed linear model?

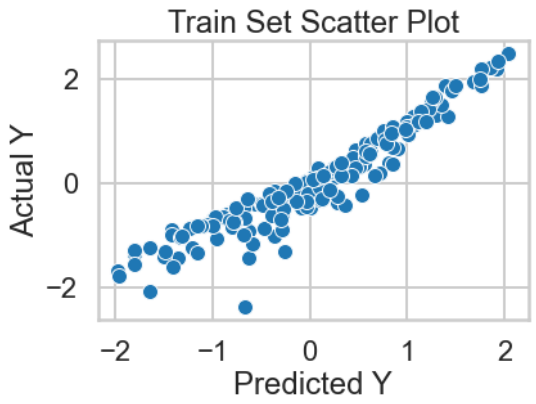
The linear model can be written as:

$h\theta(x)$ = 4.35589486e-04 + 7.37383146e-01 * TV + 5.36307180e-01 * $Radio$ + 3.14254020e-03 * $Newspaper$

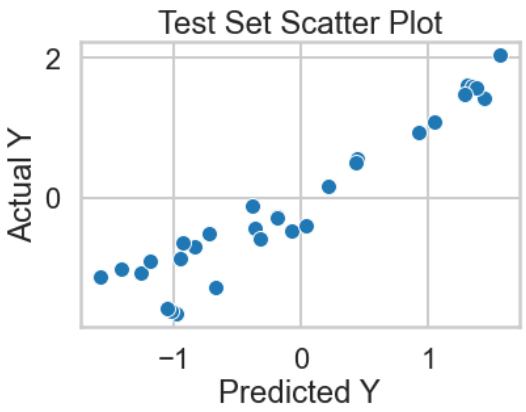
In this linear model, we can say that TV has the highest influence on sales as it bears the highest weight among the three predictors. This is then followed by radio then lastly, newspaper which has the least effect on sales.

2. Provide a scatter plot of the y_hat and y for both the train and test set. Is there a trend? Provide an r2 score (also available in sklearn).

```
R2 score using train set:  0.8936793584593189
```




```
R2 score using test set:  0.911027570209169
```



Both the train and test sets follow a linearly increasing trend between their respective y_hat (predicted) and y (actual) values. Backed by high R2 scores, these imply that the model shows good fit for the data set.

3. What happens to the error, r2, and cost as the number of iterations increase? Show your data and proof. You can alternatively plot your result data for visualization and check until 50000 iterations or more (actually).

```
cost for 1000 iterations:  0.051079831065374885
r2 score for 1000 iterations:  0.8936787899351376
MAE for 1000 iterations:  0.23638702622051314

cost for 10000 iterations:  0.05107955472978693
r2 score for 10000 iterations:  0.8936793584593188
MAE for 10000 iterations:  0.2363335722057068

cost for 50000 iterations:  0.05107955472978693
r2 score for 50000 iterations:  0.8936793584593188
MAE for 50000 iterations:  0.2363335722057068
```

As the number of iterations increases, so do the r2 scores. The mean absolute error and cost, however, decrease as the number of iterations increases, indicating an inverse relationship among these values. Additionally, the changes also become more minute as iterations are increased. In fact, between the 10000th and 50000th iterations, no changes among the r2 score, cost, and the MAE can be seen.

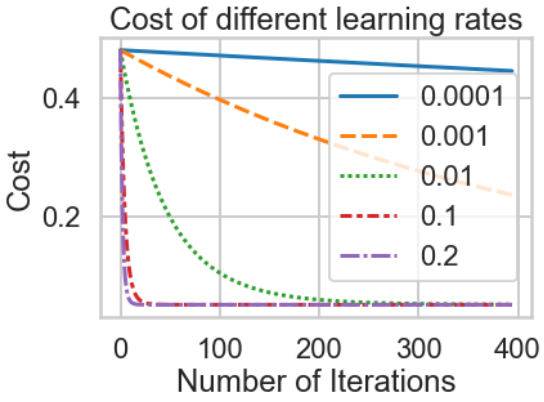
4. Once you determine the optimal number of iterations, check the effect on the cost and error as you change the learning rate. The common learning rates in machine learning include 0.1, 0.01, 0.001, 0.0001, 0.2 but you have the option to include others. Visualize the cost function (vs the optimal number of iterations) of each learning rate in ONLY ONE PLOT. Provide your analysis.

```
For alpha rate, a = 0.0001:
Cost:  0.44536428998765176
R2 score: 0.07316466002826261
MAE: 0.7659393426083329
```

```
For alpha rate, a = 0.1:
Cost:  0.051079554729786925
R2 score: 0.8936793584593189
MAE: 0.2363335722067986
```

```
For alpha rate, a = 0.001:
Cost:  0.23622673852442644
R2 score: 0.5091073142387439
MAE: 0.5448373669343117
```

```
For alpha rate, a = 0.2:
Cost:  0.051079554729786925
R2 score: 0.8936793584593189
MAE: 0.23633357220570614
```



From the plot, we can see that among the different learning rates, a = 0.0001 is the least steep which means that it would take the longest time for it to reach the optimal iteration, as 0.0001 is just too little of a step in obtaining the optimum value. This is followed by a = 0.001 which can still be improved.

The rate starts to get better at a = 0.01 as the rate is not too high or low with the MAE decreasing as well. Increasing it to a = 0.1 however furthers the improvement. Lastly, the best learning rate can be seen at a = 0.2, showing the most drastic decrease in cost in the least possible iterations.

5. Is there a relationship on the learning rate and the number of iterations?

As previously observed in (4), as the learning rate is gradually increased, the number of iterations to get to the optimal number also decreases, which shows an inversely proportional relationship. Higher learning rates in the plot also show that costs lessen more drastically in lesser number of iterations, while lower learning rates would take more iterations.

6. Compare the results with the results of ordinary least squares function

Gathering the results of the gradient descent ($r2 = 0.894$, at $a = 0.2$) and linear regression ($r2 = 0.906$), the difference between the r2 scores is seen to be minimal, with OLS scoring slightly higher. Despite both being good fit to the data, the results also imply that OLS provides a slightly better fit since it has a higher r2 score than a gradient descent with a learning rate of 0.2.