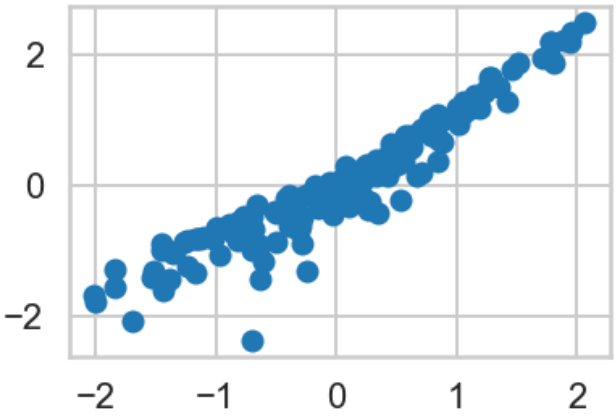
Limuelle Alamil
CMSC 197 Machine Learning
Problem Set

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

$$h_\theta(x) = -\mathbf{0.0058} + \mathbf{0.7476}TV + \mathbf{0.5441}Radio + \mathbf{0.0115}Newspaper$$

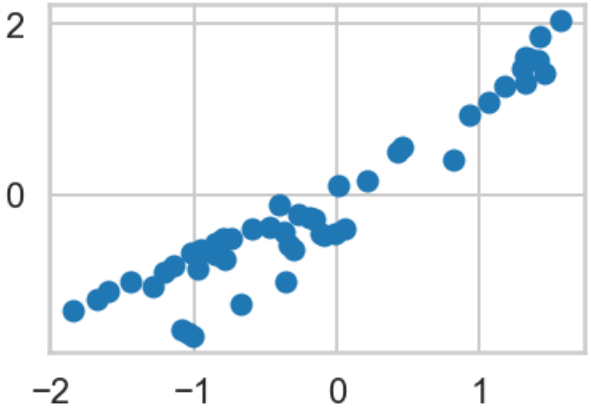
2. Provide a scatter plot of the y_hat$^{(i)}$ and y$^{(i)}$ for both the train and test set. Is there a trend? Provide an r2 score (also available in sklearn).

Train set



r2 score = 0.8846634083526498

Test set

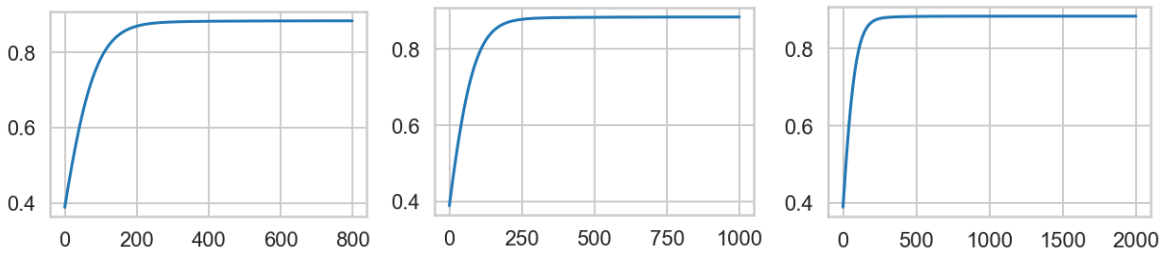


r2 score = 0.8842428941776628

In both training and testing set, there is an upward trend wherein as the predicted sales (y_hat) increases, the actual sales (y) also increase.

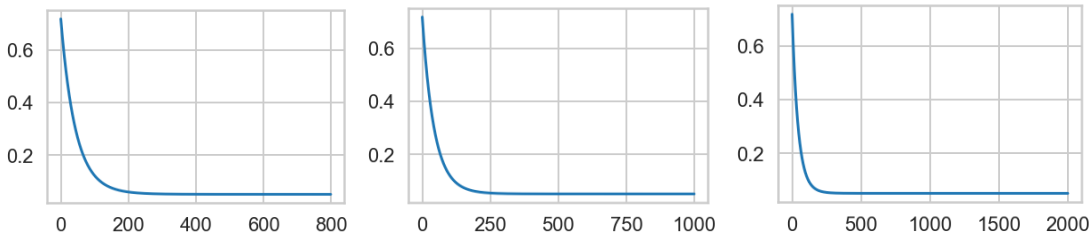
3. What happens to the error, r2, and cost as the number of iterations increase? Show your data and proof. You can alternatively plot your result data for visualization and check until 50000 iterations or more (actually).

Here are the plots of r2 [of training set] as the iterations increases.



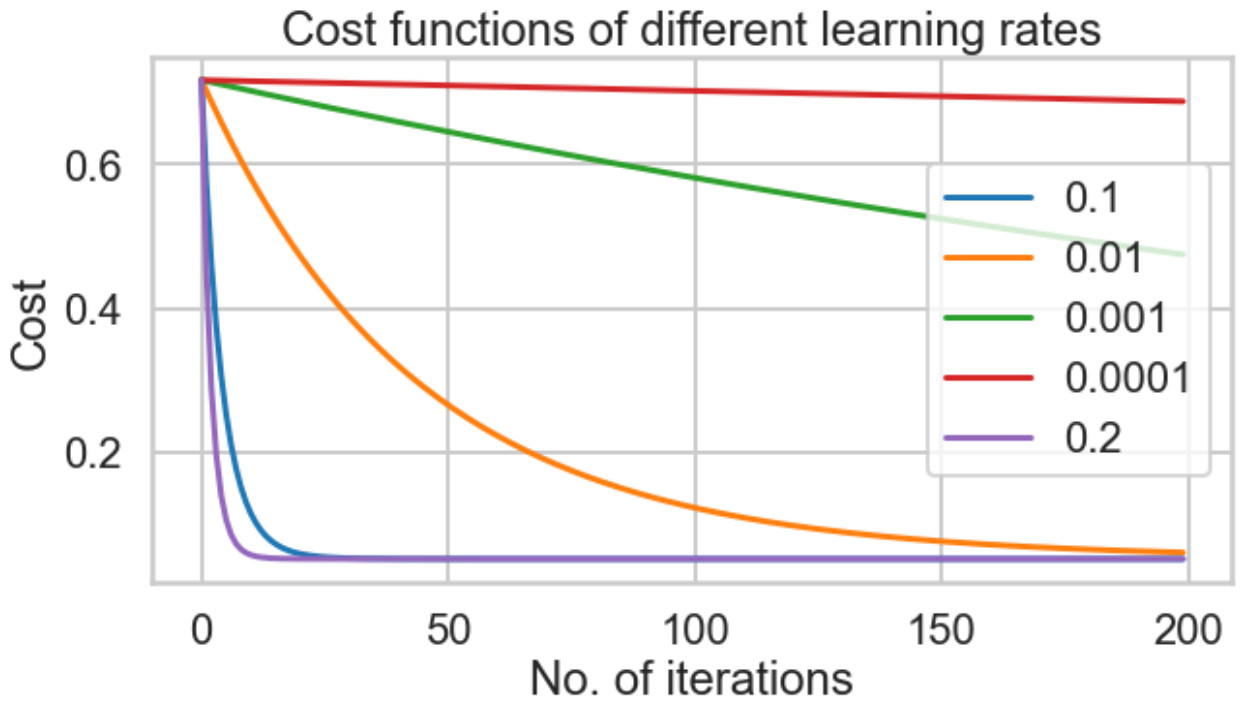
As seen from the plots, the r2 increases as the iteration increases and it increases negligibly when it reached around 200 iterations.

Here are the plots of costs [of training set] as the iterations increases.



As seen from the plots, the costs decrease as the iteration increases and it decreases negligibly when it reached around 200 iterations.

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, and 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.



The plot shows the cost functions versus the optimal number of iterations which is 200, as derived from the previous question. As seen from the plot, the smaller the learning rate becomes, the longer it takes for the cost to decrease negligibly. On the other hand, the larger the learning rate becomes, the faster it arrives at a minimum. In this case, a = 0.2 decreased negligibly quicker than the rest of the learning rates.

By visualizing the cost function of each learning rate, it can be concluded that choosing a good learning rate is critical since if it is too small, it might take a very long time to converge.

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

Based on the plot, there is no significant relationship on the learning rate and the number of iterations.

By comparing the two plots below, where the first one have 200 iterations and the second one have 1000 iterations, it can be seen that the algorithm will always arrive at a minimum at the exact same point no matter how many iterations happen.