Assignment 4
Kristen Kaiser

Dataset:

Amazon Data Science Books :
Datasethttps://www.kaggle.com/datasets/die9origephit/amazon-data-science-books

Preprocessing steps

removed colums

- title
- author
- dimensions
- weight
- publisher
- isbn_13
- link
- complete_link

remaining columns

- price (output)
- pages (input)
- ave_review (input)
- n_reviews (input)
- star 5 (input)
- star 4 (input)
- star 3 (input)
- star 2 (input)
- star 1 (input)

OLS Data

parameter vector:

(pages * 0.09) + (avg_reviews * -8.01) + (n_reviews * -0.00) + (star5 * 19.65) + (star4 * -49.31) + (star3 * -33.72) + (star2 * 39.83) + (star1 * 164.04) +(-343.92)

test set data

MSE: 809.9586312975856
MAE: 2.2291413947199263
R^2 28.459772158216335

training set data

MSE: 224.75181419224143

MAE: 0.5884766762134361

R^2 14.991724857141737

Linear regression with gradient descent Data:

parameter vector:

(pages * 0.09) + (avg_reviews * -18.64) + (n_reviews * -0.01) + (star5 * 20.63) + (star4 * -28.33) + (star3 * -30.22) + (star2 * -13.64) + (star1 * 42.94) +(88.10)

test set data:

MSE: 1088.9261107964242

MAE: 21.568972040263848

R^2: 0.2325199858330539

training set data

MSE: 866.3257169682452

MAE: 19.031924206726817

R^2: 0.30524669668400306

Note: exact data will change each time program is run as the program randomizes the distribution included in test and data sets.