Camden Alpert, Qasim Bhutta

Anurag Nagar

CS 4375.001

**Assignment 1 Report**

**Dataset**

* Features: dteday, season, yr, mnth, hr, holiday, weekday, workingday, weathersit, temp, atemp, hum, windspeed
* Target: cnt

**Training Log:**

Training Run 1:

Specifications:

All Features Used

Adaptive Learning Rate with Initial Learning Rate of 0.001

1000 Max Iterations

90/10 Train/Test split

Scores:

R Squared Score (Training Set): -1.71e18

R Squared Score (Test Set): -2.14e18

Mean Squared Error (Training Set): 1.71e18

Mean Squared Error (Test Set): 2.19e18

**Training Run 2: (Tied for Best Run)**

Specifications:

Features Used: season, yr, mnth, hr, holiday, weekday, workingday, weathersit, temp, atemp, hum, windspeed

Adaptive Learning Rate with Initial Learning Rate of 0.001

1000 Max Iterations

90/10 Train/Test split

R Squared Score (Training Set): 0.39

R Squared Score (Test Set): 0.39

Mean Squared Error (Training Set): 0.61

Mean Squared Error (Test Set): 0.6

**Training Run 3: (Tied for Best Run)**

Specifications:

Features Used: season, yr, mnth, hr, temp, atemp, hum

Adaptive Learning Rate with Initial Learning Rate of 0.001

1000 Max Iterations

90/10 Train/Test split

R Squared Score (Training Set): 0.39

R Squared Score (Test Set): 0.39

Mean Squared Error (Training Set): 0.61

Mean Squared Error (Test Set): 0.6

Training Run 4:

Specifications:

Features Used: season, yr, mnth, hr, temp, atemp, hum

Optimal Learning Rate with Initial Learning Rate of 0.001

1000 Max Iterations

80/20 Train/Test split

R Squared Score (Training Set): 0.34

R Squared Score (Test Set): 0.31

Mean Squared Error (Training Set): 0.66

Mean Squared Error (Test Set): 0.68

Training Run 5:

Specifications:

Features Used: season, yr, mnth, hr, temp, atemp, hum

Optimal Learning Rate with Initial Learning Rate of 0.001

1000 Max Iterations

90/10 Train/Test split

R Squared Score (Training Set): 0.35

R Squared Score (Test Set): 0.35

Mean Squared Error (Training Set): 0.66

Mean Squared Error (Test Set): 0.63

**Questions**

**Are you satisfied that you have found the best solution? Explain.**

We are not fully satisfied that we have found the best solution overall, as the results from both the training and test sets were underwhelming. However, we believe we have found the best solution within the constraints of a linear regression model, having undertaken various steps in each training run to get better results, such as: feature selection, hyperparameter tuning, data preprocessing, outlier removal, and changes in train/test split percentages to optimize performance. Despite these efforts, the final R-squared score of 0.39 and MSE of 0.61 on the test set fell short of our goal (R-squared ≥ 0.5). This suggests that while we have maximized the linear regression approach, a non-linear model might better capture the complexities of predicting count of bike rentals, as the relationship does not appear to be linear. Additionally, the low correlation values, with a maximum of approximately 0.4 (temp and count), likely contributed to the model’s limited performance.

**Error Plots:**

**A graph with numbers and lines

Description automatically generated**

**A graph with a line graph

Description automatically generated**

**Features vs Output Variable Plots:**

**A graph of blue dots

Description automatically generated**

**A graph with blue dots

Description automatically generated**

**A graph with numbers and a line

Description automatically generatedA graph of a number of blue dots

Description automatically generatedA graph showing the temperature of a temperature

Description automatically generatedA graph of a graph showing a diagram

Description automatically generated with medium confidence**

**A graph of a diagram

Description automatically generated with medium confidence**