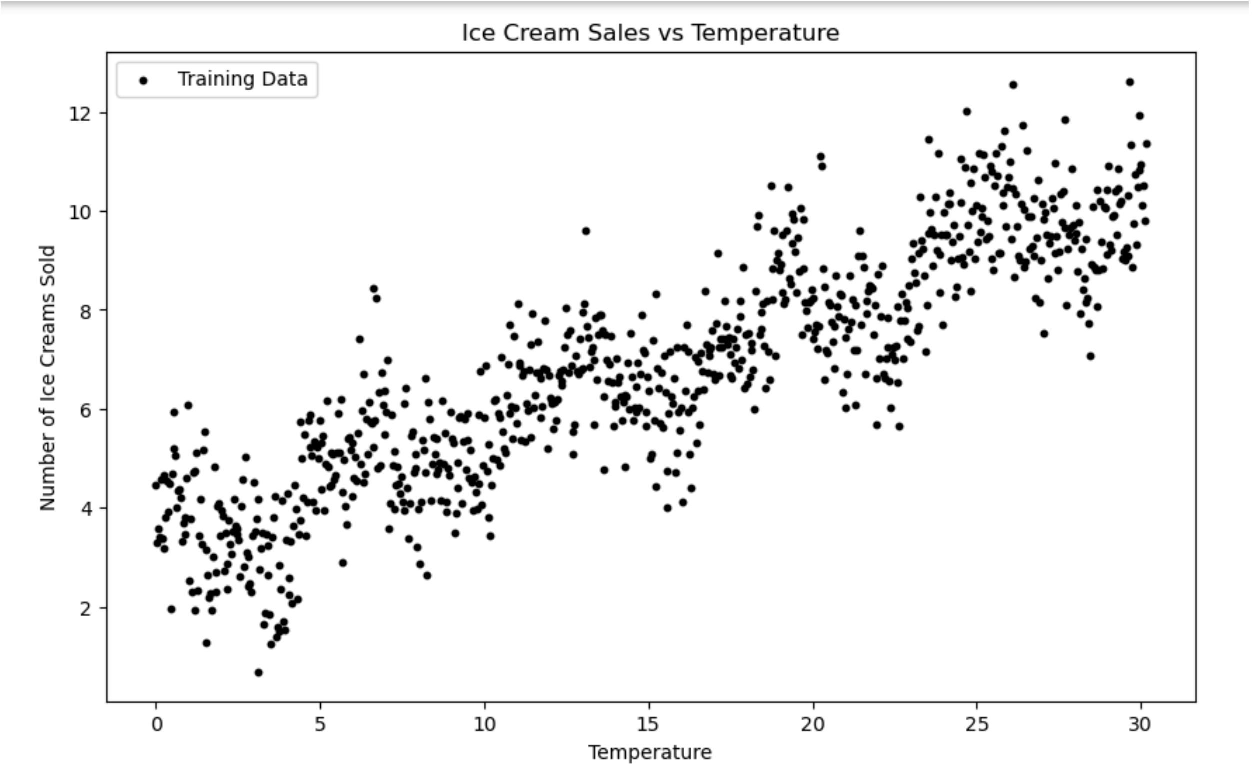
1. Introduction

- This report analyzes the relationship between temperature and ice cream sales based on data provided in `train.txt` and `test.txt`. Using various statistical and machine learning techniques, we build and evaluate models to predict ice cream sales based on temperature. The objective is to identify the best predictive model using both linear and non-linear approaches, including feature engineering and regularization techniques (Lasso and Ridge).

Data Description:

- The data is sourced from `train.txt`, and `test.txt` containing two columns: temperature (independent variable) and ice cream sales (dependent variable). We aim to find the best linear and non-linear models, including Lasso and Ridge regression, to maximize prediction accuracy on sales.

2. Initial Data Analysis and Plotting



Data plotting

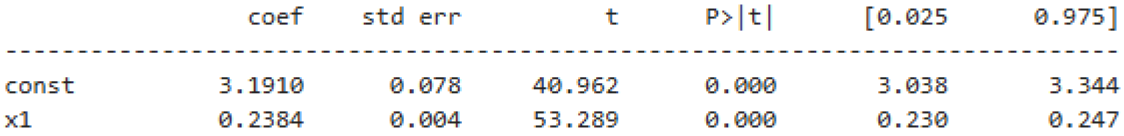
- Figure above: Temperature vs Ice cream sales
- The plot suggests a positive trend between temperature and sales. This relationship will be explored through linear and non linear modeling techniques.

3. Linear Regression Model

- Model description
    - I start by fitting linear regression model y=β0+β1x to understand baseline relationship between temperature and ice cream sales (code for it)

- Results and Interpretation
    - Coefficients:
        - Estimated $\hat{\beta}0$ (Intercept): 3.191037800253055
          Estimated $\hat{\beta}1$ (Slope): 0.23839763045936505
    - Statistical Inference:
        - **Null Hypothesis (H_0)**: The coefficient of the predictor variable is zero, meaning that the predictor does not significantly contribute to the model.
        - **Alternative Hypothesis (H_1)**: The coefficient of the predictor variable is not zero, indicating that the predictor significantly contributes to the model
        - The OLS summary (Table 1) reveals a statistically significant positive relationship between temperature and ice cream sales, as both the intercept ($\hat{\beta}0$) and slope ($\hat{\beta}1$) have p-values less than 0.05, indicating they are significantly different from zero.

          This table below confirms the linear relationship, supporting the hypothesis that temperature positively influences ice cream sales.



```
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.1910      0.078     40.962      0.000       3.038       3.344
x1             0.2384      0.004     53.289      0.000       0.230       0.247
```

4. Feature Engineering and Model Selection with Non-linear Features

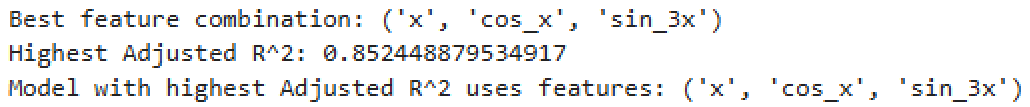
- Feature Selection Approach:
    - To improve the model beyond a simple linear fit, I included non-linear transformations of temperature, such as cos(x), log(x), cos(4x), sin(3x), sin(5x), and sin(2x)×cos(2x). Given that the feature set was relatively small, I evaluated all possible combinations of these features to identify the best subset.

- Metric Used:
    - Adjusted R^2 was employed as the metric for feature selection because of its balance between model complexity and explained variance. Unlike standard R^2 , adjusted R^2 accounts for the number of predictors in the model, ensuring that added features improve the model's fit without simply increasing complexity.

- **Alternative Metrics for Model Evaluation**

While Adjusted R² was chosen for feature selection due to its balance between explained variance and model complexity, other metrics could be used depending on the analysis goals:

    - **Mean Squared Error (MSE)**:
        - MSE measures the average squared difference between actual and predicted values, providing a sense of overall prediction error. Unlike Adjusted R², MSE is more sensitive to large errors, making it useful in assessing models’ predictive accuracy.
    - **Mean Absolute Error (MAE)**:
        - MAE calculates the average of absolute errors, offering an alternative to MSE that is less sensitive to outliers. It is helpful in cases where large errors should not disproportionately affect the evaluation metric.
    - **Cross-Validation Error (e.g., RMSE in K-Fold Cross-Validation)**:
        - Cross-validation provides a robust estimate of model performance on unseen data. Using metrics like RMSE (Root Mean Squared Error) in cross-validation helps assess how well a model generalizes to new data. This approach is particularly valuable for avoiding overfitting.
- **Selection Process and Results:**
    - The exhaustive feature selection process revealed that the optimal feature combination is: x,cos(x),sin(3x) with an adjusted R^2 of 0.852. This combination showed the highest adjusted R^2. (screenshot with results below)



```
Best feature combination: ('x', 'cos_x', 'sin_3x')
Highest Adjusted R^2: 0.852448879534917
Model with highest Adjusted R^2 uses features: ('x', 'cos_x', 'sin_3x')
```

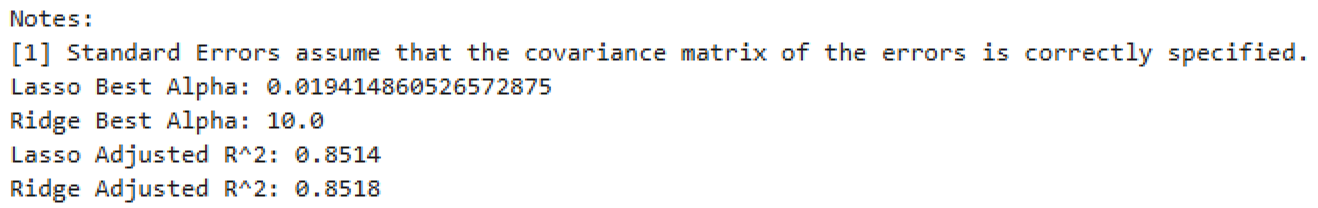
5. Lasso and Ridge Regression with All Features

- **Model Setup and Selection**
    - Lasso and Ridge regression models are trained with all features, using cross-validation to find the optimal alpha values for each model.
    - Code for Lasso and Ridge Regression
    - `lasso_model, ridge_model = train_lasso_ridge(x_train, y_train)`

- **Results and Interpretation**
    - **Lasso Best Alpha**: 0.019
    - **Ridge Best Alpha**: 10.0
    - **Lasso Adjusted R2R^2R2**: 0.8514
    - **Ridge Adjusted R2R^2R2**: 0.8518

Both regularized models perform similarly, with Ridge having a slightly higher adjusted R^2.

Screenshot below: Best Alpha Values and Adjusted R^2 for Lasso and Ridge



```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Lasso Best Alpha: 0.019414860526572875
Ridge Best Alpha: 10.0
Lasso Adjusted R^2: 0.8514
Ridge Adjusted R^2: 0.8518
```

6. Model Evaluation on Test Data

- **Test Data Prediction**
  - We evaluate all models using the test dataset `test.txt`.
- **Code for Model Evaluation**
  - evaluate_on_test_data(x_test, y_test, best_model, all_features[list(best_features)], model_name="Linear Regression with Best Features")
  - evaluate_on_test_data(x_test, y_test, lasso_model, all_features, model_name="Lasso Regression with All Features")
  - evaluate_on_test_data(x_test, y_test, ridge_model, all_features, model_name="Ridge Regression with All Features")

Results and Comparison

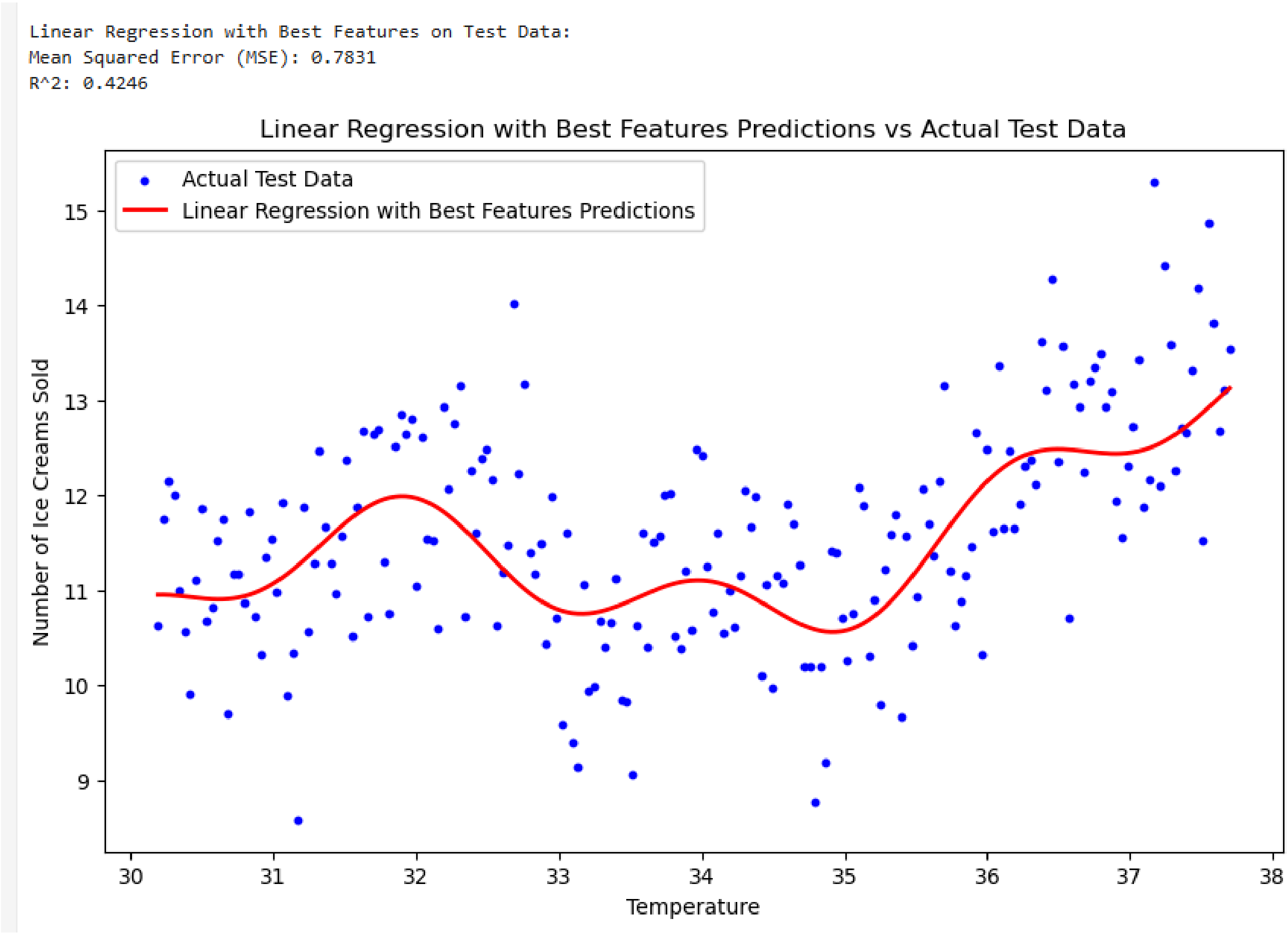
The Mean Squared Error (MSE) and R^2 values for each model on the test data are summarized below.

- Linear Regression with Best Features: MSE = 0.7831, R^2 = 0.4246
- Lasso Regression with All Features: MSE = 0.7890, R^2 = 0.4203
- Ridge Regression with All Features: MSE = 0.7797, R^2 = 0.4271

Visual Comparison of Predictions

Each figure below compares the predicted vs. actual ice cream sales for Linear Regression, Lasso, and Ridge models. Also for metrics we now use R^2 and not adjusted R^2 because we need to understand how well the model performed on new data, not just how it fits the training data Figure. Highest model R^2 was shown by ridge Regression r^2=0.4271 also it showed least mean square error. All screenshots provided.

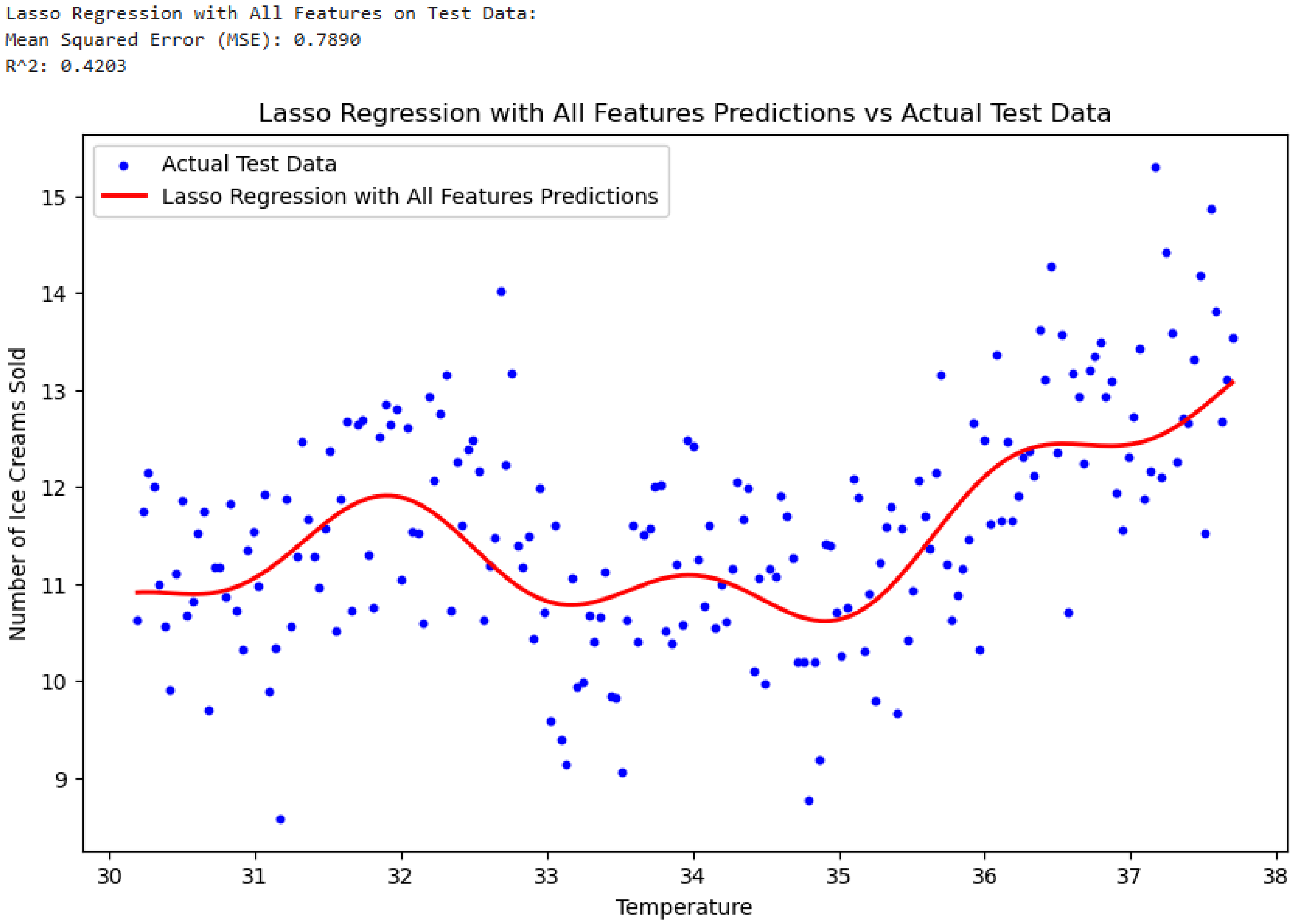
Linear Regression with Best Features on Test Data (Screenshot below)



```
Linear Regression with Best Features on Test Data:
Mean Squared Error (MSE): 0.7831
R^2: 0.4246
```

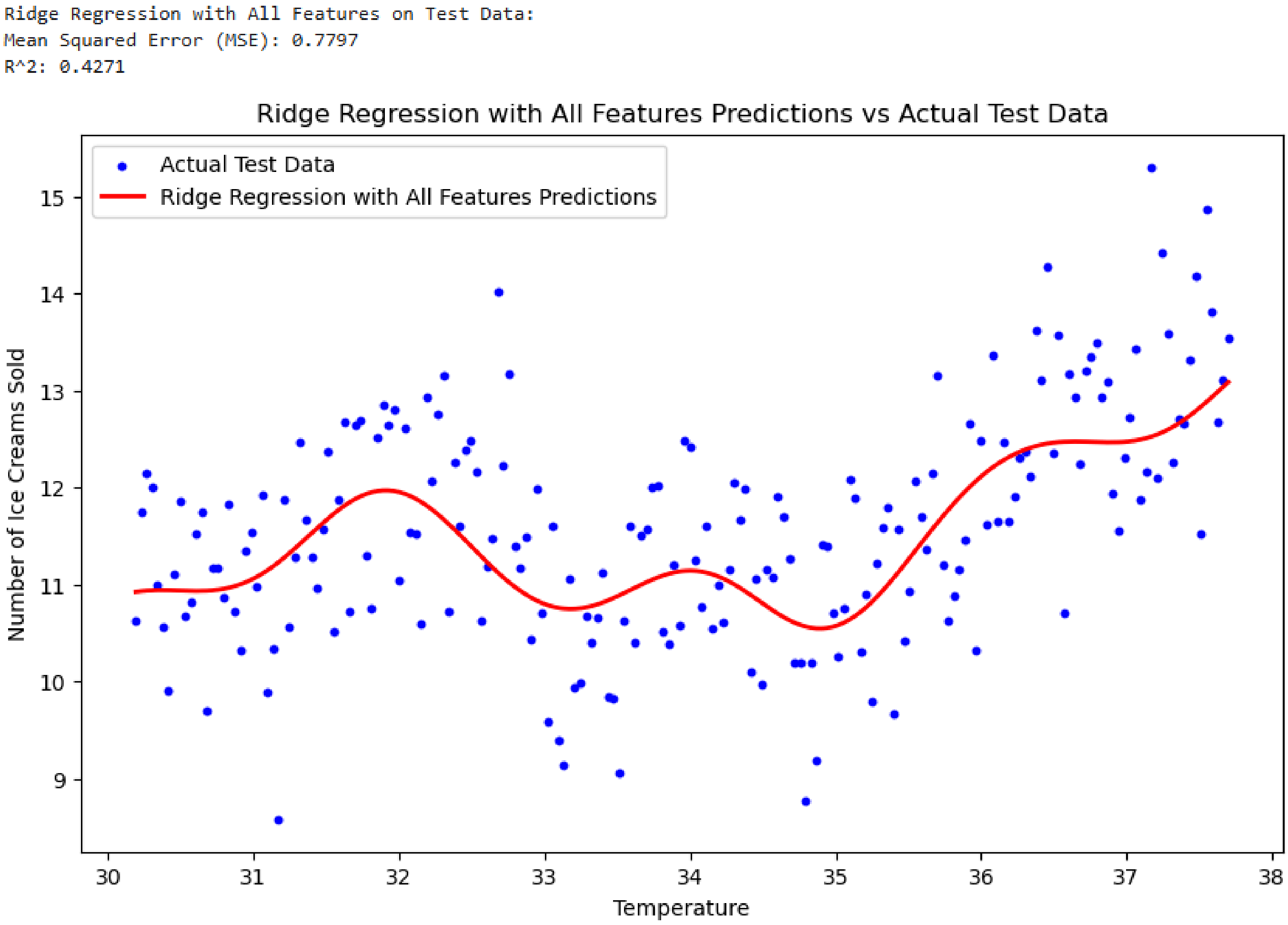
Linear Regression with Best Features Predictions vs Actual Test Data
● Actual Test Data
Linear Regression with Best Features Predictions
Number of Ice Creams Sold
15
14
13
12
11
10
9
30
31
32
33
34
35
36
37
38
Temperature

Figure 3: Lasso Regression with All Features on Test Data



```
Lasso Regression with All Features on Test Data:
Mean Squared Error (MSE): 0.7890
R^2: 0.4203
```

Figure 4: Ridge Regression with All Features on Test Data



```
Ridge Regression with All Features on Test Data:
Mean Squared Error (MSE): 0.7797
R^2: 0.4271
```

7. Conclusion

Summary of Findings

- The linear regression model confirmed a positive relationship between temperature and ice cream sales.
- Adding non-linear features improved the model, with the best feature combination achieving an adjusted R^2 of 0.852.
- Lasso and Ridge models achieved comparable adjusted R^2 values with all features not selected, but Ridge provided a slightly better generalization on test data.