UNIVERSITY OF VICTORIA

Department of Electrical and Computer Engineering

ECE 503 Optimization for Machine Learning

LABORATORY REPORT

Experiment No: *Lab 1*

Title: *Predicting Fuel Consumption of Automobiles*

Date of Experiment: *Sep.24th, 2025*

Report Submitted on: *Oct.1st, 2025*

To: *Ruilin Wang*

Names: *Da Zhang (V01062902)*

1. Objective

The goal of this lab experiment is to build and employ a linear regression model predicting fuel city-cycle fuel consumption of automobiles in miles per gallon (MPG) by using 6 automobile features, namely, number of cylinders, displacement, horsepower, weight, acceleration, and model year.

2. Introduction

The method we employed in this experiment is a least-squared linear regression model, where the data is split into training set with 314 samples and testing sets with 78 samples, the optimal weight parameter is then computed based on the training data, $\widehat{w^*} = \mathrm{pinv}(X_{\mathrm{tr}}^T)\, y_{\mathrm{tr}}$ *, where pinv is the pseudo-inverse of* $\widehat{X_{\mathrm{tr}}^T}$*, the prediction model is then run separately on the test and training sets, which provides the results* $y_{\mathrm{tr}}^{\mathrm{pred}} = \widehat{X_{\mathrm{tr}}^T}\, \widehat{w^*}$ *and* $y_{\mathrm{te}}^{\mathrm{pred}} = \widehat{X_{\mathrm{te}}^T}\, \widehat{w^*}$*, and the model performance is therefore evaluated based on root-mean-squared error (RMSE) for training and testing sets respectively,* $\mathrm{RMSE}_{\mathrm{tr}} = \sqrt{\frac{1}{314}\sum_{i=1}^{314}\left(y_{\mathrm{tr},i}^{\mathrm{pred}} - y_{\mathrm{tr},i}\right)^2}$ *and* $\mathrm{RMSE}_{\mathrm{te}} = \sqrt{\frac{1}{78}\sum_{i=1}^{78}\left(y_{\mathrm{te},i}^{\mathrm{pred}} - y_{\mathrm{te},i}\right)^2}$*.Lastly, the comparison between the "ground truth"* y_{te} *and the "prediction results of the optimized for testing data"* $y_{\mathrm{te}}^{\mathrm{pred}}$ *is drawn in figure.*

3. Implementation and Results

3.1. Implementation

1. Load the provided MPG dataset so all vehicle features and fuel-efficiency labels are in memory.
2. Exact the MPG targets from row 7 into column vector y, count total samples M, and set the training size P = 314, leaving T = M - P test points.
3. Gather the six predictor features (row 1-6) into X, append a row of ones to form the bias-augmented design matrix Xh, and split it into training set Xh_tr (1:P) and target vector y_tr.
4. Slice the remaining columns of Xh and y into Xh_te and y_te to act as the test set.
5. Compute the optimal augmented weight vectors wh_star by taking the pseudoinverse of Xh_tr and multiplying by y_tr, i.e., solving the least-squared fit.
6. Generate training predictions y_tr_pred and test predictions y_te_pred.
7. Quantify fit quality with root-mean-sqaured errors, RMSE_tr and RMSE_te, using the respective sample counts (314 train, 78 test).
8. Plot the ground-truth test MPG values against the predicted ones.

3.2. MATLAB code

```
% 3.1 From the course website download D_mpg.mat
    load /Users/adamcheung/Desktop/D_mpg.mat


    % 3.2 Prepare matrix Xh_tr and y_tr
    y = D_mpg(7,:)';
    M = length(y);
    P = 314;
    T = M - P;
    X = D_mpg(1:6,:);
    Xh = [X;ones(1,M)];
    Xh_tr = Xh(:,1:P);
    y_tr = y(1:P);


    % 3.3 Prepare test data X
    Xh_te = Xh(:,P+1:M);
    y_te = y(P+1:M);


    % 3.4 Compute optimal parameter wh_star
    wh_star = pinv((Xh_tr)') * y_tr;


    % 3.5 Apply optimized model to the training and test data respectively
    y_tr_pred = Xh_tr' * wh_star; % Optimized model on training data
    y_te_pred = Xh_te' * wh_star; % Optimized model on testing data


    RMSE_tr = sqrt(1/314 * (sum((y_tr_pred - y_tr).^2))); % Root-mean-sqaured
error for training data
```

```
    RMSE_te = sqrt(1/78 * (sum((y_te_pred - y_te).^2))); % Root-mean_sqaured
error for testing data


    % 3.6 plot y_te in blue and y_te_pred in red

    figure;

    plot(1:length(y_te), y_te, 'b-o', 'LineWidth', 1.5); hold on;

    plot(1:length(y_te_pred), y_te_pred, 'r-*', 'LineWidth', 1.5);

    legend('Ground Truth (y\_te)', 'Prediction (y\_te\_pred)');

    xlabel('Test Samples {1,2,...78}');

    ylabel('Miles Per Gallon (MPG)');

    title('Ground Truth vs Prediction on Test Samples');

    grid on;
```

3.3 Results

Results for 3.5. Report numerical results in terms of $RMSE_{tr}$ and $RMSE_{te}$

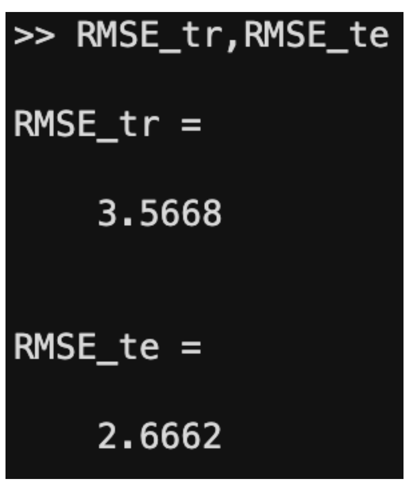

Fig 1. Numerical results of $RMSE_{tr}$ and $RMSE_{te}$

Results for 3.6.1. Plot the "ground truth" y_{te} as a blue curve and its prediction $y_{\text{te}}^{\text{pred}}$ as a red curve

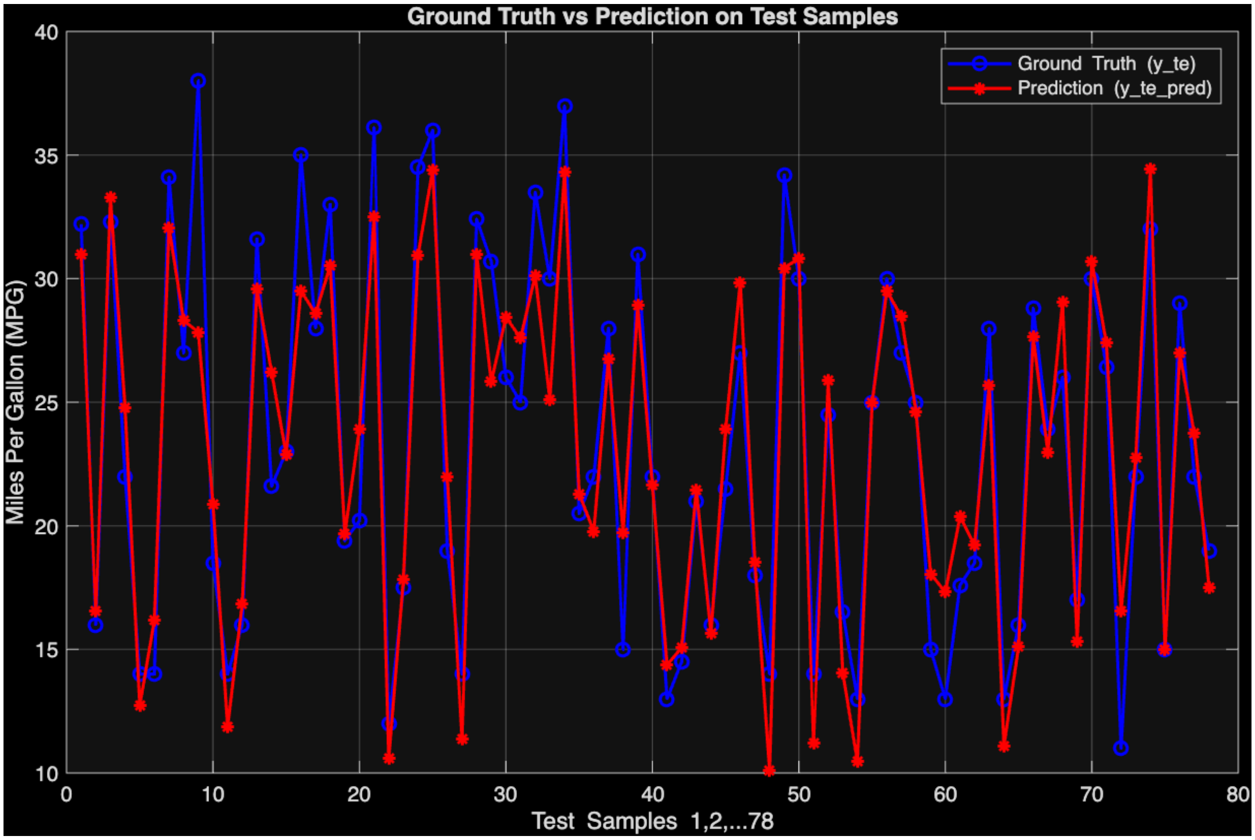

Fig 2. Ground Truth vs. Prediction on Test Samples

Results for 3.6.2. Comment on the visual inspection of the figure.

The model's prediction performance is good, with no large bias and a similar variance to the true outputs – the red curve tracks the blue one quite closely, following the same general trend, peaks, and valleys. However, there are some local mismatches, for example, the prediction sometimes slightly overshoots or undershoots the actual values, especially at the extreme highs and lows of MPG. Overall, the prediction line overlaps substantially with the true line, it suggests low generalization error, which is consistent with RMSE values for training and testing, indicating no overfitting.

4. Discussion

Since MPG ranges from 9 to 46.6, which is about 37.6 units wide. Translating RMSE for training, 3.5668 and RMSE for testing, 2.662, into a fraction of the range, which are 9.8% and 7.1%. Errors are less than 10% of the total variation, which are relatively small. Looking closer to the RMSE *results, the training RMSE provides the information that, on average, the model's prediction deviate from the actual training MPG values by about 3.57 units, which shows that the model achieves a decent fit on the training set but does not perfectly capture all variations. The testing RMSE is lower than the training RMSE, which is a bit surprising, since the expected behaviour should be training RMSE > testing RMSE. This could suggest that the model generalizes well and is not overfitting the training data, and the model has already "seen" the training data patterns during its learning phase or might also be due to the specific distribution of the 78 test samples, for instance, they contain fewer extreme or noisy points.*

From the plotted curves, the red prediction line generally follow the trend of the blue "ground truth" curve. Deviations occur in some regions, but overall alignment is quite strong, which confirms that linear regression is capturing the primary relationships between features and MPG.

However, since fuel consumption may depend on nonlinear interactions, for example, weight and horsepower combined, a more complex model such as polynomial regression or regularized regression could further be employed and evaluated to improve prediction accuracy.

5. Conclusion

In this experiment, we successfully implemented a least-squared linear regression model for predicting automobile fuel consumption (MPG). The model achieves a RMSE of 3.5668 on training data and 2.6662 on test data. Visual inspection of the comparison between the "ground truth" and the prediction on test samples confirms that predictions follow the ground truth reasonably well, though limitation may exist due to the linear nature of the model.

6. References

None.