CPSC 483 Assignment-1 Report - Tommy Vu
Assignment #1

This assignment can be completed individually or by a team.
Total score: 100
Due date: 9/29

Objective: Linear Regression Modeling

(1) (a) Write a program to learn the best polynomial model using the least square approach by only increasing or decreasing the polynomial order (without using any method to improve the accuracy of a model). (b) What's the size of training and testing data in terms of the percentage of the data used? (c) What are the training and testing errors in terms of RMSE and R2 and the training time of the model? (d) What's the total number of terms and the polynomial order? (e) Show the training results and justify why you think the model is the best.

1.
(b) The size of training and testing data is .5 and .2.
(c) The R^2 value of training was 0.75229 and the RMSE was 0.12531, whereas the testing values were 0.75333 and 0.12514. Finally, the training time took 0.74 seconds.
(d) The model was of order 10, so there are 50 parameters.
(e)

```
Order: 1, Training Time: 0.03 secs, Testing RMSE: 0.13619, Training RMSE: 0.13635, Testing R^2: 0.70784, Training R^2: 0.70674
Order: 2, Training Time: 0.07 secs, Testing RMSE: 0.12272, Training RMSE: 0.12276, Testing R^2: 0.76276, Training R^2: 0.76229
Order: 3, Training Time: 0.15 secs, Testing RMSE: 0.12161, Training RMSE: 0.12167, Testing R^2: 0.76704, Training R^2: 0.76648
Order: 4, Training Time: 0.22 secs, Testing RMSE: 0.12081, Training RMSE: 0.12088, Testing R^2: 0.77012, Training R^2: 0.76953
Order: 5, Training Time: 0.27 secs, Testing RMSE: 0.12212, Training RMSE: 0.12219, Testing R^2: 0.76508, Training R^2: 0.76448
Order: 6, Training Time: 0.39 secs, Testing RMSE: 0.12231, Training RMSE: 0.12238, Testing R^2: 0.76438, Training R^2: 0.76376
Order: 7, Training Time: 0.46 secs, Testing RMSE: 0.12253, Training RMSE: 0.12261, Testing R^2: 0.76351, Training R^2: 0.76286
Order: 8, Training Time: 0.54 secs, Testing RMSE: 0.12279, Training RMSE: 0.12288, Testing R^2: 0.76250, Training R^2: 0.76181
Order: 9, Training Time: 0.64 secs, Testing RMSE: 0.12479, Training RMSE: 0.12496, Testing R^2: 0.75471, Training R^2: 0.75369
Order: 10, Training Time: 0.74 secs, Testing RMSE: 0.12514, Training RMSE: 0.12531, Testing R^2: 0.75333, Training R^2: 0.75229
```

Results listed are the training time, testing RMSE, training RMSE, testing R^2, and training R^2. This model is the best one because the R^2 value is a prime indicator of whether or not the model is a good fit, and the results showcase that this model is better fitted for the data as compared to the other models.

(2) (a) Write a program without using any library or package to learn the best polynomial model
using the gradient descent method discussed in class by only increasing or decreasing the polynomial order. (b) What's the size of training and testing data in terms of the percentage of the data used? (c) What are the initial weights and the learning rate? (d) What are the training and testing errors in terms of RMSE and R2 and the training time of the model? (e) What are the total number of terms and the polynomial order? (f) Show the training results and justify why you think the model was the best.

(b) Size of training and testing data are 5 and 0.5.
(c) The initial weights and learning rates are
(d) The training and testing errors in terms of RMSE and R2 are RMSE Train : 0.51, RMSE Test :0.71, R^2 Train: -3.12, R^2 Test: -7.03. Training time of the model was 141.80 seconds
(e) Total number of terms was 5.
The data above indicates that gradient descent is used for finding optimal values for linear regression.

(3) (a) Write a program for a feature scaling and run one of the programs written in either (1) or
(2) to learn the best polynomial model with the scaled data set. (b) What is the feature scaling method used? Show an example row of before and after scaling. (c) What are the training and testing errors in terms of RMSE and R2 and the training time? (d) Show the training results with the scaled data and compare the accuracies to determine whether or not scaling impacts the model's accuracy.

3.
(b) The feature scaling method used was the fit transform method.

```
Row 10,000 before scaling:
             T     P        TC          SV
10000  273.15  17.7  0.029331  402.411212

Row 10,000 after scaling:
         T    P        TC        SV
10000  0.2  0.0  0.186876  0.376271
```

(c) The training and testing errors in terms of RMSE and R2 and training time are similar to the previous.
(d) Here are the training results.

```
Training Results:
Order: 4, Training Time: 0.31 secs, Testing RMSE: 0.12079, Training RMSE: 0.12086, Testing R^2: 0.77018, Training R^2: 0.76959
```

According to the results, scaling makes it so that the results of the model are more accurate and precise than the initial result.

(4) (a) Write programs to learn the best polynomial model using LASSO, Ridge, and Elastic net
discussed in class. (b) What are the training and testing errors in terms of RMSE and R2 for the model learned and the training time by each method. (c) What is λ selected for each method with a brief justification of λ? (d) What are the total number of terms and the polynomial order for each model, and what features can be removed and why? (e) Select the best model from the three models with a brief justification of the selection. (f) Does the regularization help prevent overfitting and ultimately improve accuracy? Justify your answer.

(b) For Elastic Net: Training Time: 3.76 secs, Testing RMSE: 0.25197, Training RMSE: 0.25178, Testing R^2: -0.00002, Training R^2: 0.00000

For Lasso: Training Time: 3.30 secs, Testing RMSE: 0.25197, Training RMSE: 0.25178, Testing R^2: -0.00002, Training R^2: 0.00000

For Ridge: Training Time: 0.77 secs, Testing RMSE: 0.12139, Training RMSE: 0.12145, Testing R^2: 0.76789, Training R^2: 0.76733

(c) For λ, it is important to use the right balance between simplicity and training-data fit. If the value is too high, then the data will be underfitted and the model won't learn enough to make useful predictions. If the value is too low, the data will be overfitted and will be unable to generalize new data.

(d) The total number of terms are 9 and 45 terms.

(e) The best model is ridge because it executes in the fastest time. It also appears to be more accurate in terms of R^2 as well.

(f) Yes, regularization helps prevent overfitting because it improves the model accuracy. It also improves accuracy because it prevents the loss of important data as a result of underfitting.

(5) (a) Write a program to perform K-fold cross-validation for the three models in (4). (b) What value do you choose for K and why? What's the best model selected based on cross-validation? (c) Does the cross-validation help select the best model? Justify your answer.

(b) K values used were 5 and 10. This is because these values are not too affected by high bias or high variance. The best model based on cross-validation is the elastic net model.

(c) Cross-validation does aid in selecting the best model because it allows us to compare the different models and identify which is the better algorithm based on the accuracy.