Regularization Methods

By

Maysam Asser 202200276

Riwan Ashraf 202201726

Pousy Ayman 202202076

Prepared for Dr. Ahmed Abdelsamea

December 2024

Contents:

1. **Theoretical definition on each regularization method.**
2. **Implementation of the designated method using its original definition.**
3. **Finding the optimal regularization parameter (λ) in each case using cross validation, providing the appropriate diagrams.**
4. **Compare the results between the 3 different methods.**
5. **Compare your results with the Python built-in library in each case.**
6. **Draw conclusions with the pros and cons of each method.**

1. Theoretical Definitions:

Ridge Regression: used when data has many predictors, sometimes more than the number of observations. It is similar to linear regression but adds a penalty to the model for having large coefficients. This helps prevent overfitting and makes the model better at making predictions on new data.

$$E = \sum_{i=0}^{n}(y - \hat{y})^2 + \lambda \sum m^2$$

Lasso Regression: used to simplify models by minimizing some coefficients to be zero. It is used for selecting the most important features in the data.

$$E = \sum_{i=0}^{n}(y - \hat{y})^2 + \lambda \sum |m|$$

ElasticNet Regression:this is a mix of Ridge and LASSO. It combines their strengths by minimizing the coefficients (like Ridge) and selecting important features (like LASSO). It works well when the data has many predictors that are related to each other.

$$E = \sum_{i=0}^{n}(y - \hat{y})^2 + \lambda_1 \sum m^2 + \lambda_2 \sum |m|$$

2. Implementation of the designated method using its original definition.

Implementation is in a colab notebook provided.

https://colab.research.google.com/drive/1bpm1nkFPyFgmpdZQnX_FzPOL7yup8jzN#scrollTo=WZIcoV_9cLpv

3. Finding the optimal regularization parameter (λ) in each case using cross validation, providing the appropriate diagrams.

Ridge Regression:

1st dataset:

```
Optimal lambda (from-scratch): 1.7782794100389228
Optimal lambda (sklearn): 1.7782794100389228
```

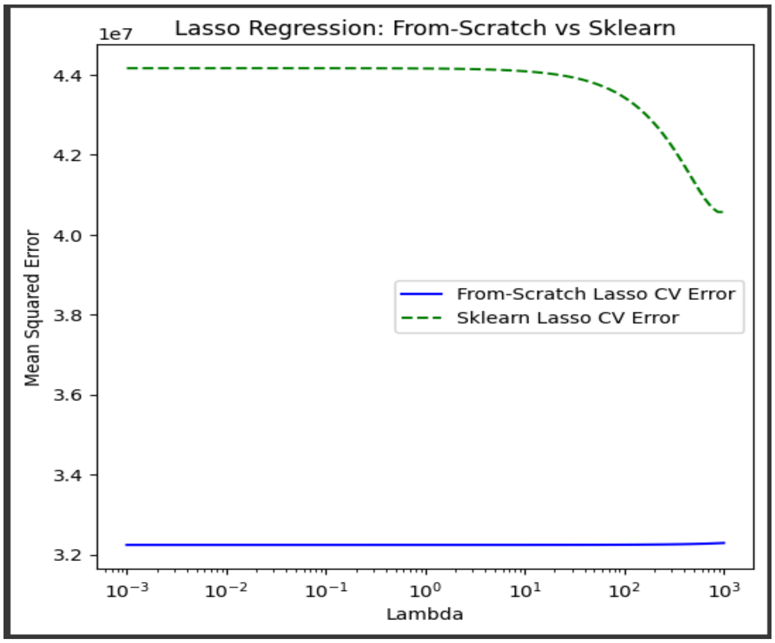
2nd dataset:

```
Optimal lambda (from-scratch): 10.0
Optimal lambda (sklearn): 10.0
```

Lasso Regression:

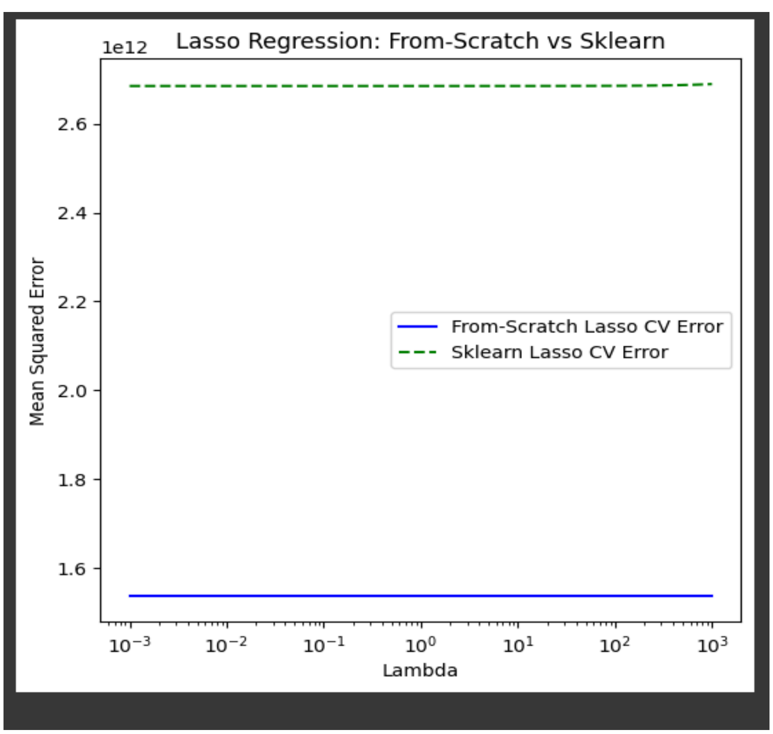
1st dataset:

```
Optimal lambda (from-scratch): 0.001
Optimal lambda (sklearn): 1000.0
```



2nd dataset:

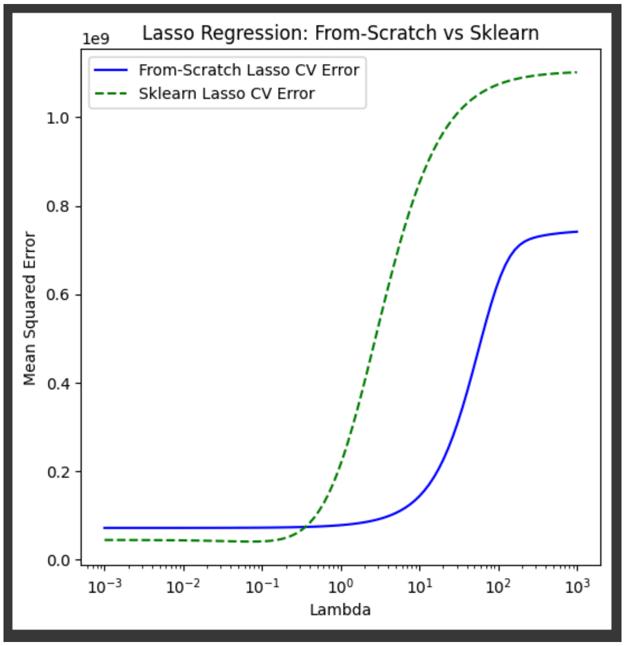
```
Optimal lambda (from-scratch): 1000.0
Optimal lambda (sklearn): 0.003511191734215131
```



ElasticNet Regression:

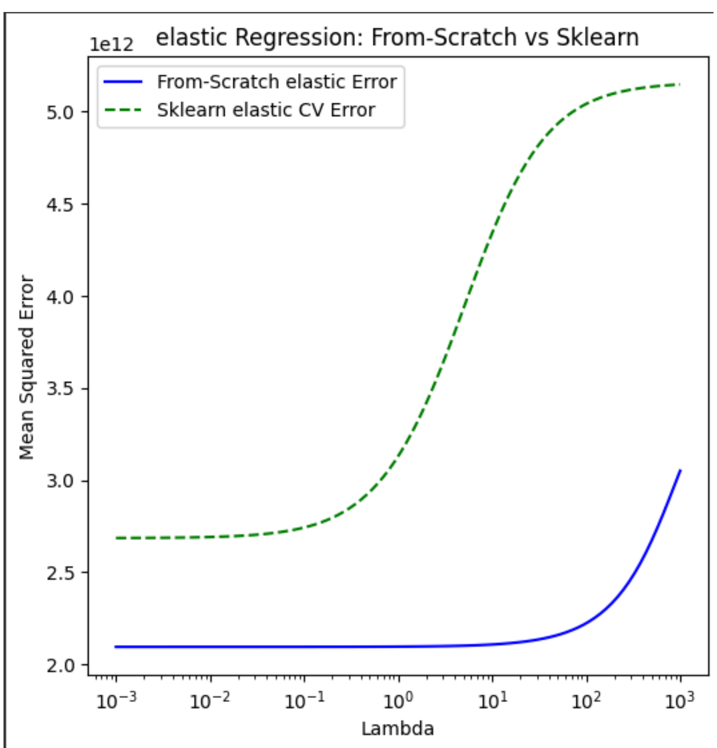
1st dataset:

```
Optimal lambda (from-scratch): 0.001
Optimal lambda (sklearn): 0.07564633275546291
```



2nd dataset:

```
Optimal lambda (from-scratch): 0.001
Optimal lambda (sklearn): 0.001
```



4. Compare the results between the 3 different methods.

Ridge Regression:

1st dataset:

```
MSE (from-scratch): 153838229.7706959
```

2nd dataset:

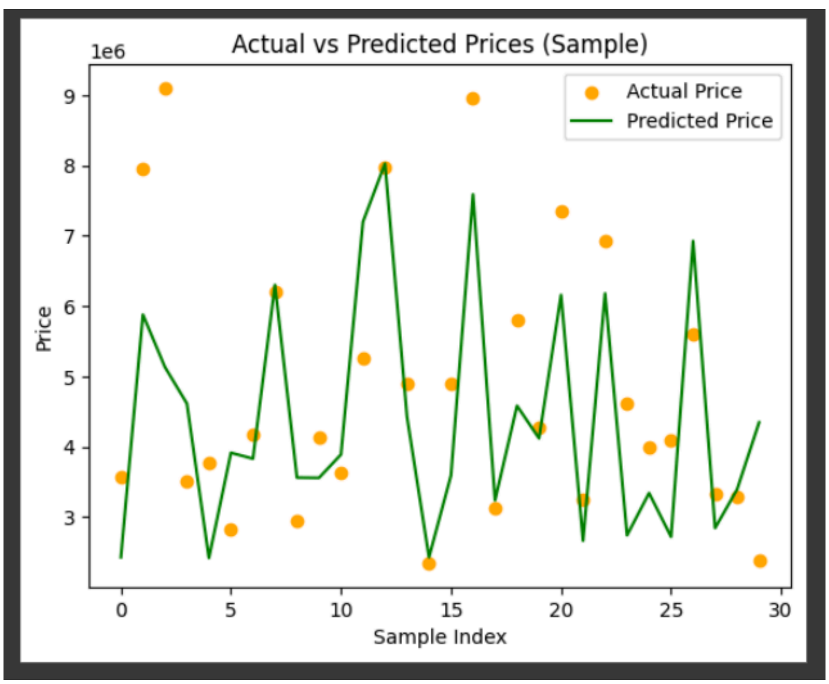
```
MSE (from-scratch): 1340472302491.9648
```

Lasso Regression:

1st dataset:


```
Mean Squared Error from scratch: 40634782.39975862
```

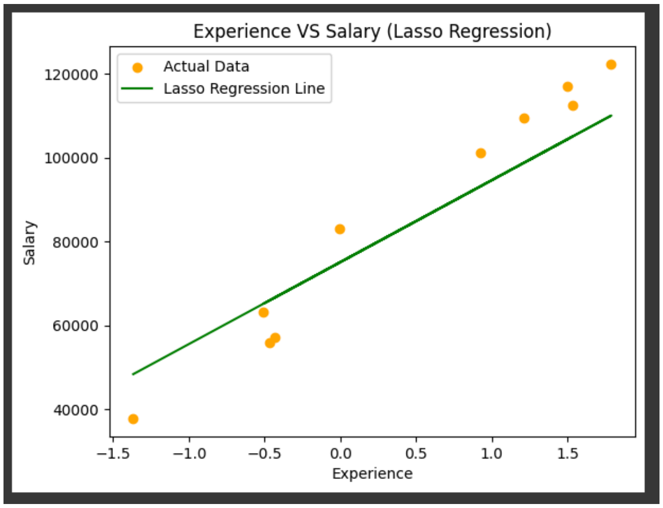
2nd dataset:




```
Mean Squared Error from scratch: 1456612374243.3464
```

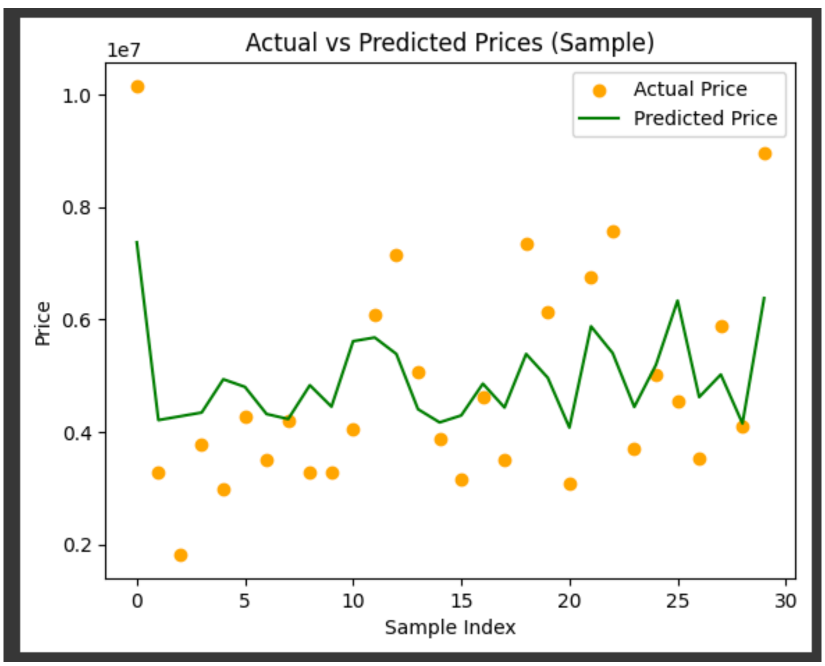
ElasticNet Regression:

1st dataset:




```
Mean Squared Error from scratch: 91453913.37635908
```

2nd dataset:




```
Mean Squared Error from scratch: 1812655763246.8105
```

5. Compare your results with the Python built-in library in each case.

Ridge Regression:

1st dataset:

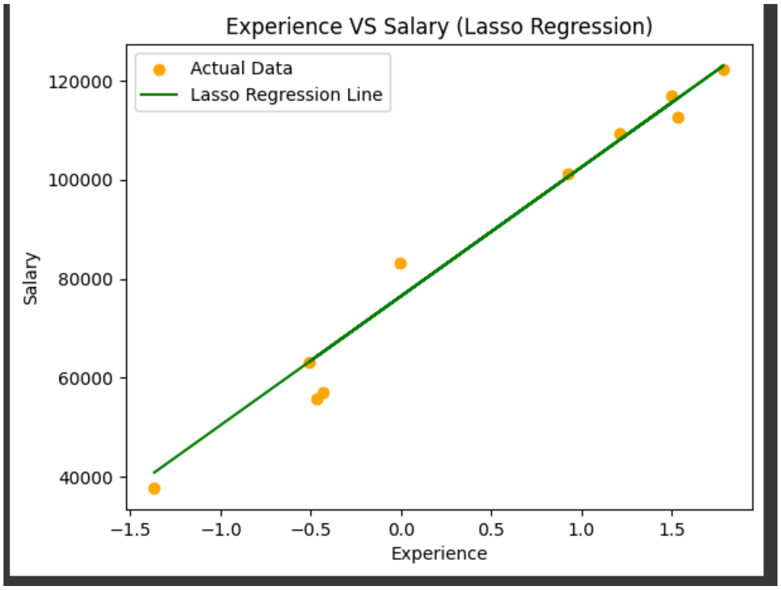
```
MSE (sklearn): 33447993.1324752
```

2nd dataset:

```
MSE (sklearn): 1117458845860.744
```

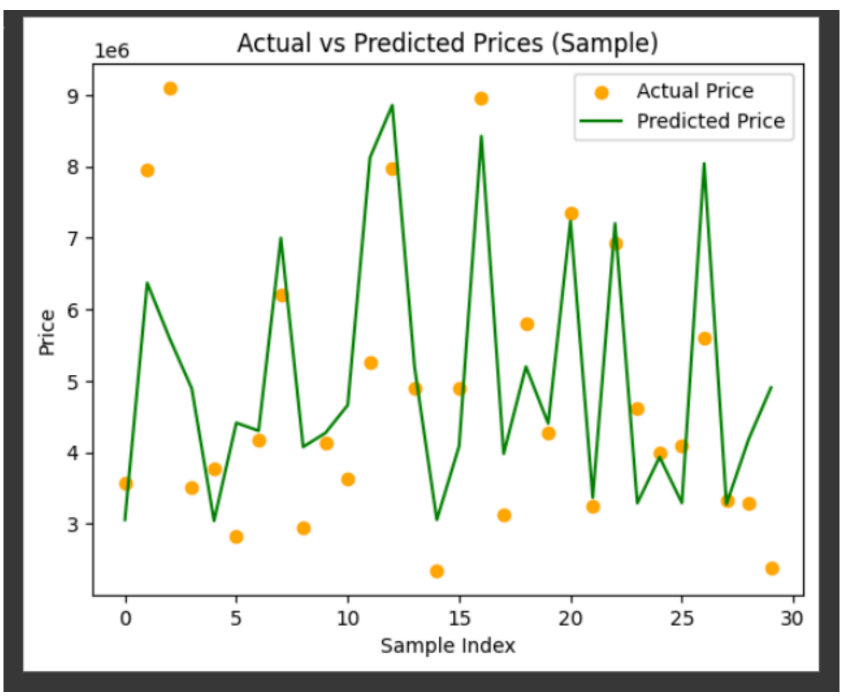
Lasso Regression:

1st dataset:




```
Mean Squared Error built in: 21026177.177518718
```

2nd dataset:



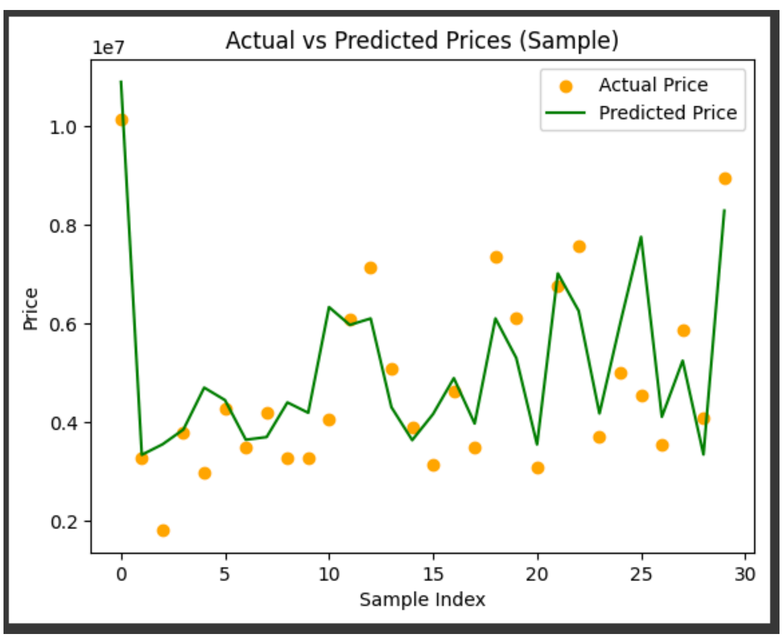

```
Mean Squared Error built in: 1239484544545.601
```

ElasticNet Regression:

1st dataset:


```
Mean Squared Error built in: 26062534.06867128
```

2nd dataset:




```
Mean Squared Error built in: 1226726669918.2717
```

6. Draw conclusions with the pros and cons of each method.

Ridge Regression:

pros:

1. **Good for Related Features:** Handles features that are similar or related to each other.
2. **Keeps All Features**: Doesn't remove any features, just makes some less important.
3. **Prevents Overfitting**: Stops the model from memorizing because of training data, so it works better on new data.
4. **Fast to Use**: Quicker to run than methods that remove features.

cons:

1. **Doesn't Remove Features:** Keeps everything, which can make the model harder to understand.
2. **Adds Bias**: May hurt accuracy if the dataset is small or the penalty is too strong.

Lasso Regression

pros:

1. **Feature Selection:** Automatically removes unimportant features by setting their values to zero.
2. **Simplifies the Model**: Makes the model easier to understand by keeping only the important features.
3. **Good for Many Features**: Works well when there are lots of features, but only a few are important.

cons:

1. **May Remove Useful Features**: If the penalty is too strong, it might remove features that still have some value.
2. **Takes Longer to Compute**: Slower than Ridge because it has to decide which features to remove.
3. **Adds Bias**: Like Ridge, it introduces some bias, and removing features completely is a disadvantage.

Elastic Net Regression

Pros:

1. **Handles Correlated Features**: Works well with predictors that are related, by balancing between Lasso and Ridge .
2. **Feature Selection**: Removes unimportant features like Lasso, keeping the model simpler and easier to understand.
3. **Prevents Overfitting**: Combines lasso and ridge penalties to make the model work well with new data.
4. **Flexible**: need to be modified to act more like Lasso or Ridge, depending on the data.

Cons:

1. **Needs Tuning**:by adjusting two settings (alpha and lambda), which consumes time.
2. **Only used for Linear Data**: Works only when the data has a straight-line relationship, not with complex patterns.

References

Tavishi. (2023, June). Regularization techniques: Ridge, Lasso and elastic net. Medium. https://medium.com/@tavishi.1402/regularization-techniques-lasso-and-ridge-90d3cc73ca4c

van Wieringen, W. N. (2015). Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.

GeeksforGeeks. (2024, September 18). *Ridge Regression vs Lasso Regression*. GeeksforGeeks. https://www.geeksforgeeks.org/ridge-regression-vs-lasso-regression/#:~:text=Lasso%20regression%20is%20better%20than,all%20features%20in%20the%20model.

Datasets:

1-Salary

https://www.kaggle.com/datasets/harsh45/random-salary-data-of-employes-age-wise

2-Housing

https://www.kaggle.com/datasets/yasserh/housing-prices-dataset/data