Assignment 2: Build Non-Linear Models Part 1

Anuj Bansal

Data Sciences, National University

DDS-8555 v1: Predictive Analysis

Dr. Mohammad

August 2nd, 2025

This week’s report goes over understanding approaches for feature selection when building regression models (Regression in machine learning, 2025).  We studied stepwise method where we can use a forward method by adding features and backward method removing features. We learned about different metrices used to evaluate models when running these approaches. We learned about ridge regression, Lasso regression and elastic net as different models that can be used to find the optimal feature subset. Lastly, we studied about Principal component regression (PCR) which is used to reduce dimensionality and also handle correlated features.

Using the data from Kaggle, we built 3 different models on the data and compared their performance (Table 1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Regularization** | **Feature Selection** | **Dimensionality Reduction** | **MSE** | **R²** |
| Ridge | L2 | No | No | 4.0953 | 0.6013 |
| Lasso | L1 | Yes | No | 4.4996 | 0.5620 |
| PCR | None (PCA) | Indirectly | Yes | 4.0953 | 0.6013 |

Table : Performance comparison

Ridge and PCR both achieve the lowest MSE, meaning their predictions are closest to the actual values. Lasso’s higher MSE suggests it may have removed useful features, slightly reducing accuracy. Thus, Ridge and PCR are equally effective for this dataset, while Lasso trades some accuracy for feature selection. Similarly, we see from R squared values that Ridge and PCR have higher R values and thus explain more variance in the target than Lasso, so they fit the data better.

Finally, the results were submitted to Kaggle ([Regression with an Abalone Dataset | Kaggle](https://www.kaggle.com/competitions/playground-series-s4e4/submissions))

# References

Gareth James, D. W. (2023). *An Introduction to Statistical Learning: with Applications in Python.* Springer. Retrieved from https://www.statlearning.com/

Geek for Geek. (2025, jul 12). *Random Forest Regression in Python*. Retrieved from geek for geeks: https://www.geeksforgeeks.org/machine-learning/random-forest-regression-in-python/

*Regression in machine learning*. (2025, Jan 13). Retrieved from geek for geeks: https://www.geeksforgeeks.org/machine-learning/regression-in-machine-learning/