SMAA8411 POE Part 1 Submission

Question 1 – Theory Submission (Q2-Q4 in Excel W/Book)

Done by: Kiashen Maharajh (ST10055763)

# Question 1

Q.1.1)

Before delving into lasso and ridge penalties, it is important to understand the broader category they fall under: shrinkage models. Shrinkage models are a form of linear regression used in supervised learning, offering significant advantages over subset selection methods (Medium, 2023).

Subset selection involves choosing a subset of predictors to build a more interpretable model with the goal of reducing prediction error (Hastie, Tibshirani and Friedman, 2017). This approach however often leads to high variance without necessarily improving prediction accuracy. In contrast, shrinkage models utilize all predictors but apply a regularization technique to reduce variability, making them a more robust alternative (Hastie, Tibshirani and Friedman, 2017).

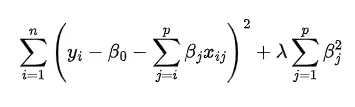
Ridge vs. Lasso Penalty in Shrinkage Models

Shrinkage models, such as Ridge Regression and Lasso, apply penalties to the regression coefficients to improve generalization. While both rely on the ordinary least squares (OLS) framework, their key difference lies in the type of penalty applied (Medium, 2023):

**Ridge Penalty (L2 Regularization)**

Ridge regression utilises L2 penalty by adding a penalty proportional to the squared magnitude of the coefficients (Medium, 2023):

Equation : Ridge Regression (Medium, 2023)

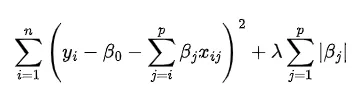


This penalty forces coefficients to shrink toward zero but never sets them exactly to zero. If the penalty term λ is set to zero, ridge regression reduces its effectiveness to OLS. Ridge is particularly useful when all predictors contribute to the response, as it improves model stability through more accurate predictions over OLS’s high variance outputs (Medium, 2023).

**Lasso Penalty (L1 Regularization)**

Lasso on the other hand applies a L1 penalty to the entire dataset proportional to the absolute values of the coefficients (Medium, 2023):

Equation : Lasso Regression (Medium, 2023)



Unlike ridge, this penalty can force some coefficients to be exactly zero, effectively performing automatic feature selection (Medium, 2023). The L1 norm promotes sparsity, meaning that only the most relevant predictors are retained in the model, however all predictors are still considered. This makes lasso particularly useful in high-dimensional settings, where selecting the most important features is crucial for interpretation (Medium, 2023).

**Effect on Estimated Coefficients**

In terms of these regression models effect on estimated coefficients:

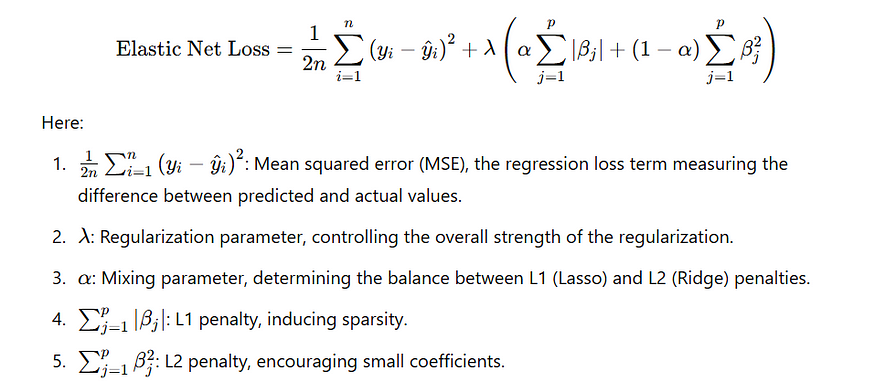
* Ridge regression shrinks all coefficients but retains all predictors, making it ideal when all variables contribute some predictive power (Medium, 2023).
* Whereas lasso performs variable selection by setting some coefficients to exactly zero, making it useful for models where only a subset of predictors is truly relevant (Medium, 2023).

Elastic Net Penalty: A Hybrid Approach

Now that we understand ridge and lasso penalties, we can discuss the elastic net penalty, which a useful linear regression method that combines the advantages of Lasso (L1) and Ridge (L2) regression to create a more versatile solution (Medium, 2024).

The elastic net penalty function is defined as:

Equation : Elastic Net Equation (Medium, 2024)



**Advantages of Elastic Net**

* Balances the strengths of Ridge and Lasso: Elastic Net selects variables like Lasso but also shrinks coefficients together for correlated predictors, like Ridge providing flexibility to adapt to differing situations (Medium, 2024).
* Handles high-dimensional datasets effectively: Traditional regression models struggle with datasets containing a large number of predictors, but Elastic Net is designed to work well in such cases (Medium, 2024).
* Prevents overfitting: Its regularization approach helps reduce error and improve generalization (WallStreetMojo, 2023).

Thus, Elastic Net balances sparsity (Lasso) and stability (Ridge), making it more robust in various real-world applications by ensuring proper fitting and reducing overfitting (WallStreetMojo, 2023).

Q.1.2)

To understand the relationship between regression models and least squares, it is essential to define each concept and explore how they interconnect before identifying their relation.

What are Regression Models?

Regression models can be understood of as statistical methods and techniques used in varying disciplines used to understand the relationship between a dependent variable (response) and one or more independent variables (predictors) (Investopedia, 2024). These models aim to identify patterns and make predictions. One of the most common estimation methods for regression models is linear methods such as ordinary least squares (OLS) estimation which establishes the linear relationship between two specific variables (Investopedia, 2024).

Least Squares as a Method in Regression

Now that we have an understanding of what regression is and how least squares better relate to it, we can get a better understanding of this method of regression. The least squares method is a popular linear regression method where the coefficients are chosen in order to minimize the sum of squared differences between observed values and the predicted values from a regression model (Hastie, Tibshirani and Friedman, 2017). In other words, it is a mathematical, statistical analysis technique where the best fitting curve or line (f(x)) is identified from a set of scatter values (x and y) (GeeksforGeeks, 2024).

Reiterating one of the previously claimed statements, that Ordinary Least Squares (OLS) is one of the most popular methods of linear regression to estimate unknown parameters, in OLS, the objective of the model itself is to find the “best fit” line and minimise the residual sum of squares (RSS), being to minimize the difference between the actual and predicted values to create the most accurate and best fit model (Analytics Yogi, 2023).

The RSS formula is as follows:

RSS = ∑(yi – ŷi)^2 (Analytics Yogi, 2023)

where:

* yi ​is the actual observed value,
* ŷi is the predicted value from the model.

The least squares method is widely used in linear regression but is not directly applicable to logistic regression (Analytics Yogi, 2023).

Linear Regression vs. Logistic Regression

To add onto what was claimed earlier about regression models, they can furthermore be broken down based on the type of dependent variable they predict (ResearchMethod.net, 2024). In the case of this assignment, we will be delving into linear and logistic regression.

1. Linear Regression

* Utilised when the output variable is continuous (e.g., predicting house prices) (GeeksforGeeks, 2025).
* The created model assumes a linear relationship between the independent variables (X) and the dependent variable (Y) (Statology, 2021):

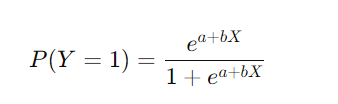
Y = β0 + β1X1 + β2X2 + … + βpXp

* Ordinary Least Squares (OLS) is used to estimate the best fitting coefficients βj to minimise RSS (Statology, 2021).
* Assumptions:
  + Linearity: The relationship between independent and dependent variables is linear (GeeksforGeeks, 2025).
  + Homoscedasticity: The variance of residuals is constant (GeeksforGeeks, 2025).
  + Normality: Residuals (errors) are normally distributed (GeeksforGeeks, 2025).

2. Logistic Regression

* Used when the response variable is categorical (e.g., predicting whether an email is spam or not) (Statology, 2021).
* Instead of modelling a direct relationship between X and Y, logistic regression estimates the probability that an outcome belongs to a certain category (Statology, 2021).
* The model uses the logistic (sigmoid) function to ensure that the output is between 0 and 1 (GeeksforGeeks, 2025):

Equation : Logistic Regression Equation (ResearchMethod.net, 2024)



* Maximum Likelihood Estimation (MLE) is used instead of least squares because OLS does not work well when predicting probabilities (Statology, 2021).

Furthermore, please find a table breaking down their core components below:

Table : Key Differences: Linear vs Logistic Regression Table (Statology, 2021)

|  |  |  |
| --- | --- | --- |
| Feature | Linear Regression | Logistic Regression |
| Response Variable | Continuous (e.g. price, height) | Categorical (e.g., Yes/No, 0/1) |
| Mathematical Model | Predicts a numeric outcome directly | Predicts probability and converts it to a category |
| Estimated Method | Ordinary Least Squares (OLS) | Maximum Likelihood Estimation (MLE) |
| Output Values | Can be any real number | Restricted to range [0,1] -> 0 least likely, 1 most likely |
| Common Use Cases | Predicting numeric values | Binary classification problems |

Conclusion

* Least squares is a fundamental technique for estimating regression coefficients, mainly used in linear regression (Analytics Yogi, 2023).
* Logistic regression does not use least squares; instead, it relies on maximum likelihood estimation (MLE) to predict probabilities rather than continuous values (Statology, 2021).
* The choice between linear vs. logistic regression depends on the nature of the dependent variable and the type of relationship between inputs and outputs (Statology, 2021).

Q.1.3)

In machine learning, selecting the appropriate level of model complexity is essential for building good quality models that generalize well to unseen data (Hastie, Tibshirani, & Friedman, 2017). This process is closely tied to managing bias, variance, and model complexity, as a balance between these three factors determines whether a model performs accurately or not. If this balance is poorly managed, the model may either underfit (failing to capture key patterns) or overfit (performing well on training data but poorly on new data) (Medium, 2024).

To determine the right level of model complexity for a given problem, it is critical to understand the bias-variance trade-off. Bias-variance trade-off is a core concept that describes the relationship between the 3 aforementioned concepts (bias, variance and model complexity) and governs model performance and influences prediction error, making it imperative to understand how different models perform with unseen data in a given scenario (Medium, 2024). We will now look at these key concepts further.

Understanding Bias and Variance

* Bias refers to the error introduced by approximating a complex real-world problem with a simplified model. High-bias models make strong assumptions and often underfit the data, missing important patterns (Medium, 2024).
* Variance refers to how much a model's predictions change with small variations in the training data. High-variance models are often overly complex, capturing noise in the data and resulting in overfitting (Medium, 2024).

Using these key terms, the following scenarios can be understood:

Table : Table of Scenarios (Bias and Variance) (Medium, 2024)

|  |  |  |  |
| --- | --- | --- | --- |
| Scenario | Bias | Variance | Outcome |
| High Bias | High | Low | Underfitting |
| High Variance | Low | High | Overfitting |
| Optimal Trade-off | Balanced | Balanced | Best Generalisation |

The Role of Model Complexity

Model complexity refers to how flexible or powerful a model is in fitting the data. This is done through assessing how well a model fits a use case scenario through capturing underlying patterns in the data (Medium, 2024):

* Low complexity simple models (e.g., linear regression with nonlinear data) may lack the capacity to capture underlying relationships. This results in high bias and making strong assumptions about data through underfitting, resulting in inaccurate and error prone data prediction (Medium, 2024).
* High complexity complex models (e.g., deep neural networks or models with many predictors on a small dataset) can capture more patterns and fit training data well, but may fit the sample data too well, capture noise as if it were signal. This results in high variance and fails to generalise and predict new data leading to poor performance and inaccurate predictions through overfitting (Medium, 2024).

The goal of model complexity is to find the sweet spot: a model complex enough to capture meaningful patterns but simple enough to generalize well. Identifying the right level of complexity and finding a balance between these concepts is crucial for building a thorough and well-defined model which will generalise accurate predictions on new, unseen data and finding the best fit and most accurate model for the scenario at hand (Medium, 2024).

How to Determine Optimal Model Complexity

Now that we understand that finding a balance of model complexity to avoid undefining and overfitting and creating a model that is most accurate at predictions is crucial, there are a few key strategies that can be employed to ensure that this balance can be achieved. We will go into four of these key concepts below:

1. Cross-Validation

* Split the data into training and validation sets (e.g., using k-fold cross-validation) to assess how well a model performs on unseen data (Medium, 2024).
* Evaluate model performance at different complexity levels (Medium, 2024).
* Select the model that minimizes validation error, not just training error — this helps prevent overfitting (Medium, 2024).

2. Learning Curves

* Plot model performance (training vs. validation error) as a function of the training set size or model complexity. The model complexity-error curve can show and track these changes (LinkedIn, 2023).
* This visualization helps detect:
  + Underfitting: Both errors are high and close (LinkedIn, 2023).
  + Overfitting: Training error is low, but validation error is high (LinkedIn, 2023).
* Through visualisation of these key concepts, the point at which test error reaches its minimum with optimal complexity can be observed and the best fitting and most balanced model can be found (LinkedIn, 2023).

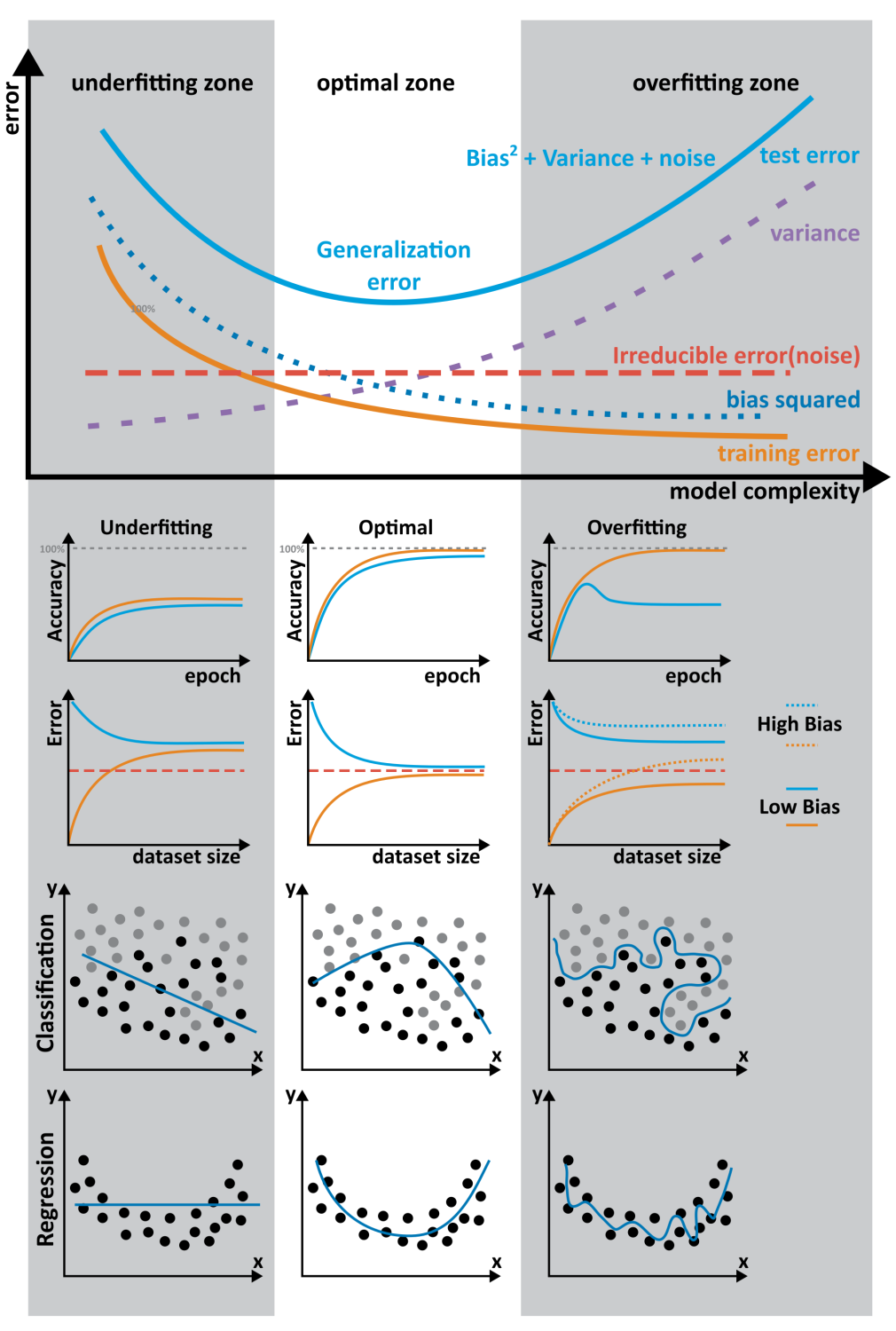


Figure : Visual Representation Showing the Model-Complexity Learning Curve (LinkedIn, 2023)

3. Regularization Techniques

* In order to control complexity, we can add penalty terms to the loss function to reduce overfitting without entirely removing features (LinkedIn, 2023):
  + Ridge Regression (L2): Shrinks all coefficients though parameter regularity (LinkedIn, 2023).
  + Lasso Regression (L1): Shrinks some coefficients to zero, performing feature selection and parameter nulling (LinkedIn, 2023).
  + Elastic Net: Combines both.
* Regularization helps control complexity and improves model generalization.

4. Model Selection Criteria

* To obtain the best fitting model and finding optimal model complexity, we can use performance metrics such as:
  + AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion): Lower values indicate better balance between fit and complexity (LinkedIn, 2023).
* This will ensure unnecessary complexity can be removed the best fitting model can be found and utilised to offer the most accurate predictions.

Q.1.4)

In stats, 2-tailed, left tailed and upper tailed tests can be understood of as hypothesis testing techniques whereby, we can test a population parameter (Statology, 2021).

A graphical representation of these tests can be found below along with their rejection regions:

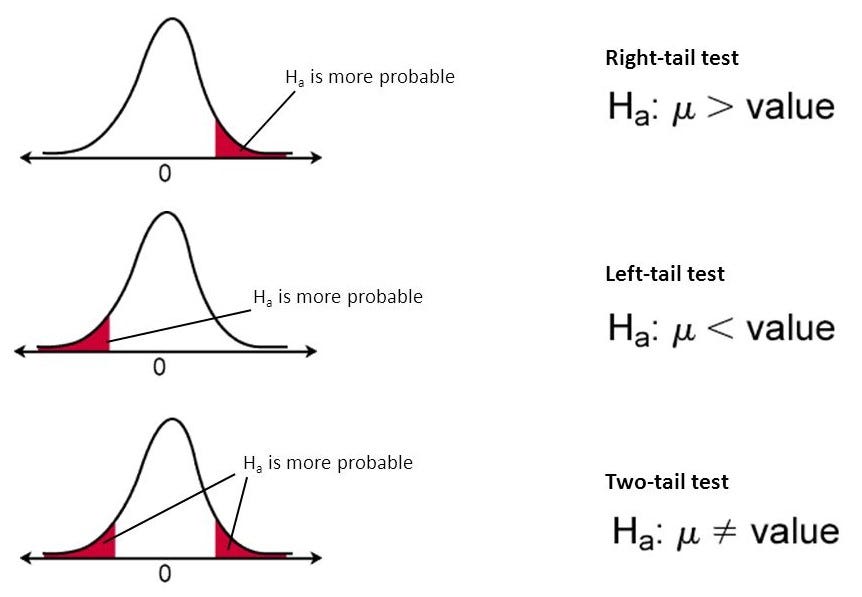


Figure : 2-Tailed Test vs Left Tailed Test vs Upper Tailed Test (Towards Data Science, 2019)

Before delving into the types of tests, we need to understand what the region of rejection is. This can also be referred to as the critical region and is representative of the area where the null hypothesis is rejected (Statistics How To, n/d)

1. Two-Tailed Test

* **Used when**: Testing whether a value is significantly different (either higher or lower) from the hypothesized value. It executes null hypothesis testing so if the sample being tested falls into these outliers, the alternative hypothesis is accepted instead of the null hypothesis (Investopedia, 2024).
* **Rejection region**: On both ends (tails) of the normal distribution (Investopedia, 2024).
* S**ignificance level (α)** is split equally between the left and right tails (e.g., α/2 in each tail if α = 0.05 → 0.025 in each tail) (Investopedia, 2024).
* **Null Hypothesis (H0​)**: The value is equal to the hypothesized value (Investopedia, 2024).
* **Alternative Hypothesis (H1​)**: The value is not equal (≠) to the hypothesized value (Investopedia, 2024).
* Reject H0**​** if the test statistic falls into either tail (Investopedia, 2024).

2. Left-Tailed Test (Lower-Tailed Test)

* **Used when**: One-tailed testing if a value is significantly less than the hypothesized value (Statology, 2021). This null hypothesis test, therefore, if the hypothesized value falls into this left region, the null hypothesis will be rejected and the alternative hypothesis will be accepted (W3Schools, n/d).
* **Rejection region**: Only in the left tail of the distribution (Statology. 2021).
* **Entire significance level (α)** is in the left tail (Statology, 2021).
* **Null Hypothesis (H0)**: The value is greater than or equal to the hypothesized value (Statology, 2021).
* **Alternative Hypothesis (H1)**: The value is less than the hypothesized value (Statology, 2021).
* Reject H0​ if the test statistic falls into the left tail (below critical value) (Statology. 2021).

3. Right-Tailed Test (Upper-Tailed Test)

* **Used when**: One-tailed testing if a value is significantly greater than the hypothesized value (Statology, 2021). This null hypothesis test, therefore, if the hypothesized value falls into this right region, the null hypothesis will be rejected and the alternative hypothesis will be accepted (Statistics How To, n/d)
* **Rejection region**: Only in the right tail of the distribution (Statology. 2021).
* **Entire significance level (α)** is in the right tail (Statology. 2021).
* **Null Hypothesis (H0​)**: The value is less than or equal to the hypothesized value (Statology. 2021).
* **Alternative Hypothesis (H1)**: The value is greater than the hypothesized value (Statology. 2021).
* Reject H0**​** if the test statistic falls into the right tail (above critical value) (Statology. 2021).

As can be seen, these 3 tests each have their own use cases, and all prove useful in the realm of statistics and testing.

# Question 2 – 4

Please Refer to: “SMAA8411 A1 Submission - ST10055763 (K Maharajh)”. The individual sheets hold the answers to the relevant questions and furthermore referencing for Q2-4 can be found here.

# References (Q1 Only)

Analytics Yogi, 2023. *Ordinary Least Squares Method: Concepts & Examples*. [online] Available at: <<https://vitalflux.com/ordinary-least-squares-method-concepts-examples/>> [Accessed 27 March 2025].

GeeksforGeeks, 2024. *Least Square Method | Definition Graph and Formula*. [online] Available at: <<https://www.geeksforgeeks.org/least-square-method/>> [Accessed 26 March 2025].

GeeksforGeeks, 2025. *Linear Regression in Machine learning.* [online] Available at: <<https://www.geeksforgeeks.org/ml-linear-regression/>> [Accessed 30 March 2025].

Hastie, T., Tibshirani, R. and Friedman, J., 2017. *Springer Series in Statistics: The Elements of Statistical Learning - Data Mining, Inference, and Prediction.* 2ed. New York: Springer Science+Business Media.

Investopedia, 2024. *Regression: Definition, Analysis, Calculation, and Example*. [online] Available at: <<https://www.investopedia.com/terms/r/regression.asp>> [Accessed 26 March 2025].

Investopedia. 2024. *What Is a Two-Tailed Test? Definition and Example.* [online] Available at: <<https://www.investopedia.com/terms/t/two-tailed-test.asp>> [Accessed 03 April 2025].

Medium, 2023. *Supervised Learning in R: Shrinkage Methods*. [online] Available at: <<https://medium.com/@ozturkfemre/supervised-learning-in-r-shrinkage-methods-abe727eb9599>> [Accessed 22 March 2025].

Medium, 2024. *Elastic Net Regression —Combined Features of L1 and L2 regularization.* [online] Available at: <<https://medium.com/@abhishekjainindore24/elastic-net-regression-combined-features-of-l1-and-l2-regularization-6181a660c3a5>> [Accessed 22 March 2025].

Medium. 2024. *Achieving Optimal Fit: A Guide to Model Complexity and Common Pitfalls in Machine Learning.* [online] Available at: <<https://medium.com/@vijaysada29/achieving-optimal-fit-a-guide-to-model-complexity-and-common-pitfalls-in-machine-learning-72a00f43cd87>> [Accessed 28 March 2025].

ResearchMethod.net, 2024. *Regression Analysis – Methods, Types and Examples.* [online] Available at: <<https://researchmethod.net/regression-analysis/>> [Accessed 30 March 2025].

Statology. 2020. *Introduction to Linear Discriminant Analysis.* [online] Available at: <<https://www.statology.org/linear-discriminant-analysis/>> [Accessed 28 March 2025].

Statology, 2021. *Logistic Regression vs. Linear Regression: The Key Differences.* [online] Available at: <<https://www.statology.org/logistic-regression-vs-linear-regression/>> [Accessed 30 March 2025]

Statology. 2021. *How to Identify a Left Tailed Test vs. a Right Tailed Test.* [online] Available at: <<https://www.statology.org/left-tailed-test-vs-right-tailed-test/>> [Accessed 01 April 2025].

Statology. 2021. *How to Identify a Left Tailed Test vs. a Right Tailed Test.* [online] Available at: <<https://www.statology.org/left-tailed-test-vs-right-tailed-test/>> [Accessed 03 April 2025].

Statistics How To. n/d. *Rejection Region (Critical Region) for Statistical Tests.* [online] Available at: <<https://www.statisticshowto.com/rejection-region/>> [Accessed 03 April 2025].

Statistics How To. n/d. *Left Tailed Test or Right Tailed Test? How to Decide.* [online] Available at: <<https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/how-to-decide-if-a-hypothesis-test-is-a-left-tailed-test-or-a-right-tailed-test/>> [Accessed 03 April 2025].

Towards Data Science. 2019. *Everything You Need To Know about Hypothesis Testing – Part I.* [online] Available at: <<https://towardsdatascience.com/everything-you-need-to-know-about-hypothesis-testing-part-i-4de9abebbc8a/>> [Accessed 01 April 2025]

WallStreetMojo, 2023. *Elastic Net.* [online]. Available at: <https://www.wallstreetmojo.com/elastic-net/> [Accessed 23 March 2025].

W3Schools. n/d. *Statistics - Hypothesis Testing a Mean (Left Tailed).* [online] Available at: <<https://www.w3schools.com/statistics/statistics_hypothesis_testing_mean_left.php>> [Accessed 03 April 2025].