DS085- Supervised Regression Models - Linear Regression - Ridge & lasso model

By: eng. Esraa Madhi

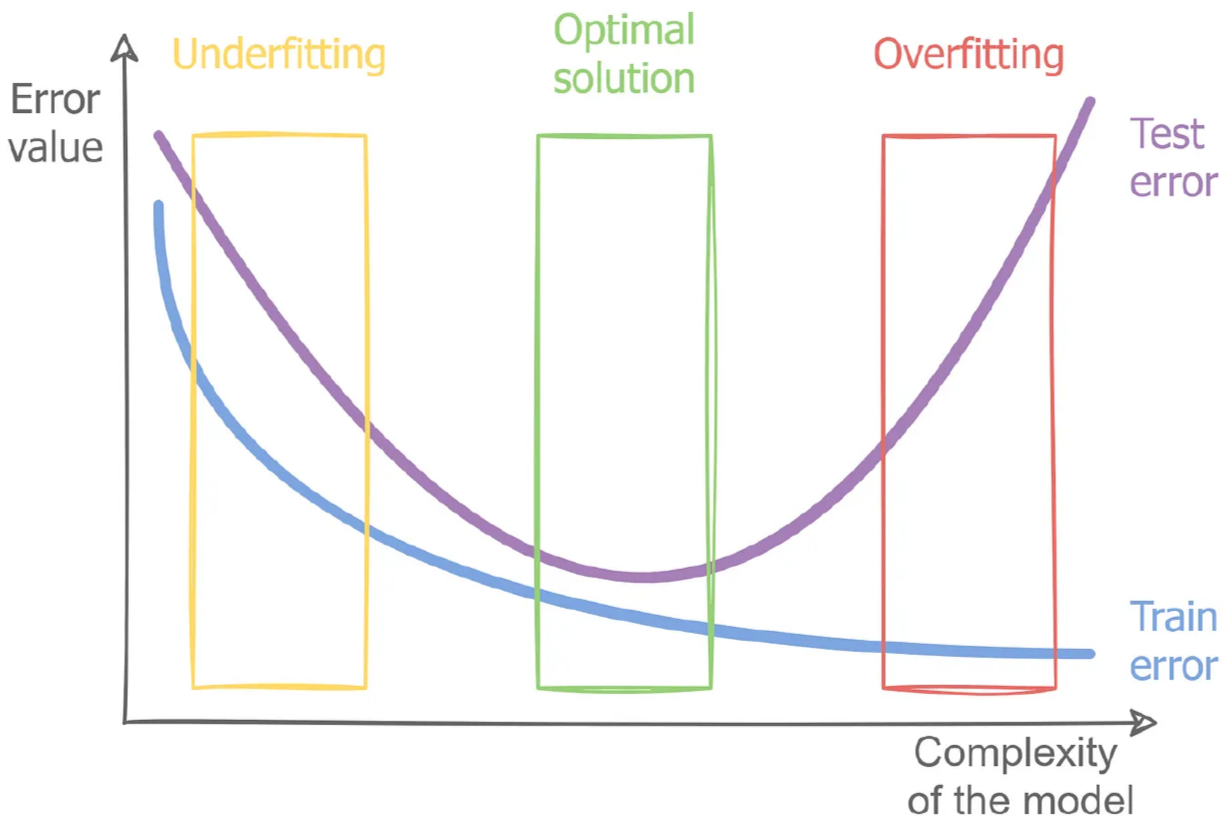
A key concept in Regression is Fitting a line to a set of data points .

This process involves finding parameters such as **intercept** and **slope**, which describe how well the line fits the data.

Model learn and remember all noise in dataset → **give large weight for features (big coefficients)**
Give large weight for features (big coefficients) → model complexity increases
Model complexity increases → overfitting
Overfitting → the model not generalised well (remember not learn)
A model not generalised well → increase error on unseen data

Different Regularization techniques are used to optimize this fit (finding best **intercept** and **slope** to minimize the loss within the specified green region below **).**

What is Regularization? It is a technique used in machine learning to penalize complex models to protect them from overfitting, to enhance the generalization of the model on new data



By doing this, regularization helps to prevent models from over-interpreting the noise and randomness found in data sets.

The two main types of regularization are

1. Lasso Regularization (**L1 Regularization**)
2. Ridge Regularization (**L2 Regularization**)

Both Lasso and Ridge regression are extensions of linear regression that include regularization terms **(use the same line equation for prediction).** The main difference between them lies in the type of regularization they apply, which affects how they handle **redundant or correlated features.**

How Ridge Regression & Lasso Regression works?

- Ridge Regression & Lasso Regression adds a regularization term to the linear Regression objective function (loss function).
    - Lasso Regression for Regularization, or L1 regularization, **adds a penalty equal to the absolute value** of the weights associated with each feature variable.

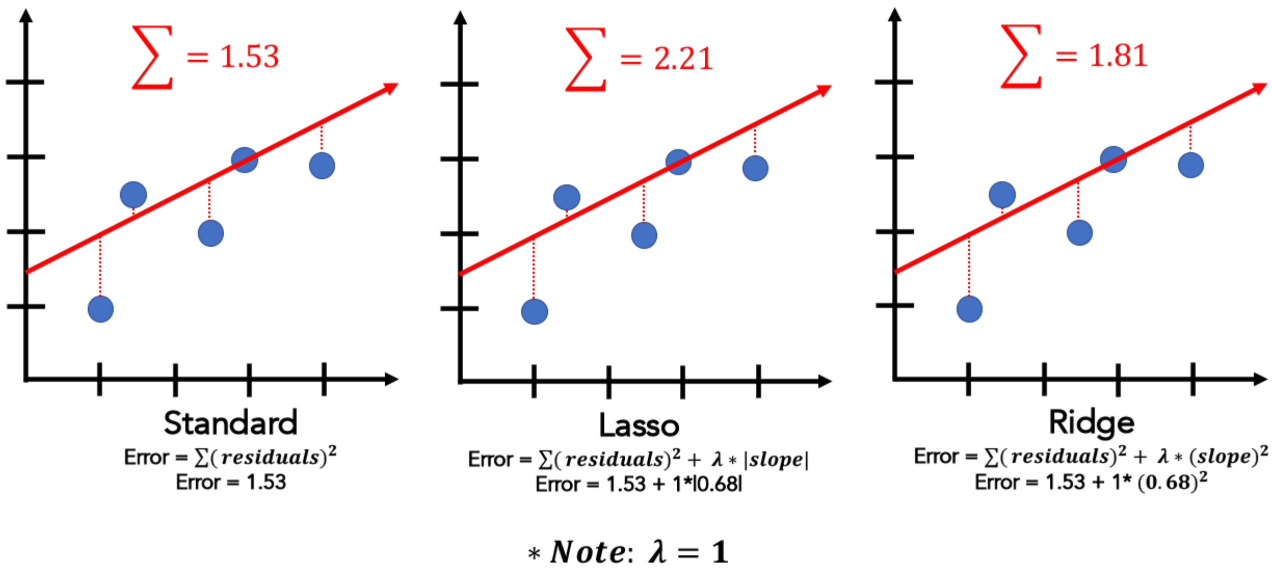
      $$E = \frac{1}{n}\sum_{i=0}^{n}(y_i - \overline{y}_i)^2 + \alpha\sum|b|$$

      (b) are the coefficients.
      (α) is the regularization parameter, a hyperparameter that controls the strength of the penalty.

    - Ridge Regularization, also known as L2 regularization, **adds a penalty equal to the square** of the weights associated with each feature variable.

      $$E = \frac{1}{n}\sum_{i=0}^{n}(y_i - \overline{y}_i)^2 + \alpha\sum b^2$$

      (b) are the coefficients.
      (α) is the regularization parameter, a hyperparameter that controls the strength of the penalty.



Note: Both methods should be tuned using **cross-validation for optimal results.**

What are the main differences between Ridge Regression & Lasso Regression?

While both methods aim to reduce coefficients' magnitudes, they differ in terms of how they do so – Lasso uses L1 regularization while Ridge uses L2 regularization.

1. **Penalty Type**:
    - Ridge: L2 penalty (squared magnitude of coefficients).
    - Lasso: L1 penalty (absolute magnitude of coefficients).
2. **Effect type:**
    - Ridge regression is particularly effective **in cases where there is multicollinearity in the data,** i.e., when features are highly correlated. It

reduces model complexity by penalizing large coefficients, thus mitigating the risk of overfitting.
    - Lasso can **force certain features' coefficients to be zero,** thus performing **feature selection** alongside regularization, while Ridge does not.

Lastly, it is important to consider which technique is more suitable for a given problem since some scenarios require one approach over the other.

https://youtu.be/Xm2C_gTAl8c

https://youtu.be/Xm2C_gTAl8c

Resources:

- https://thaddeus-segura.com/lasso-ridge/
- https://medium.com/@devsachin0879/ridge-regression-and-lasso-regression-a-beginners-guide-b3e33c77678
- https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/
- https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression