LASSO REGRESSION

Lasso (Least Absolute Shrinkage and Selection Operator) regression is a linear regression technique that adds an L1 penalty to the model's cost function, shrinking some coefficients to zero. This allows it to perform automatic feature selection and reduce model complexity, making it effective for preventing overfitting, especially with high-dimensional datasets. It works by penalizing the sum of the absolute values of the regression coefficients, effectively setting the coefficients of less important variables to zero.

Lasso was originally formulated for linear regression models. This simple case reveals a substantial amount about the estimator. These include its relationship to ridge regression and best subset selection and the connections between lasso coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates do not need to be unique if covariates are collinear.

How Lasso regression works:

Adds a penalty term: It modifies the standard linear regression by adding a penalty based on the absolute value of the coefficients (|wi|) to the sum of squared residuals (RSS).

Minimizes a new cost function: The goal is to minimize the sum of the squared errors plus a penalty term. The penalty term is $\lambda \sum_{i=1}^{n} |w_i|$, where λ is a tuning parameter and wi are the coefficients.

- **Controls complexity**: The λ (lambda) parameter controls the strength of the penalty. A higher λ increases the penalty, leading to more coefficients being shrunk towards zero.

- **Performs feature selection**: When the penalty term is strong enough, it forces the coefficients of irrelevant or less important features to become exactly zero. This process automatically removes these features from the model, creating a simpler and more interpretable model.

Key benefits:

- **Prevents overfitting**: By simplifying the model, Lasso regression can improve generalization and reduce the risk of the model performing poorly on new data.
- **Automatic feature selection**: It identifies and removes less important features, which is particularly useful in high-dimensional datasets where many variables are not relevant to the outcome.
- **Improves interpretability**: By eliminating unnecessary variables, the resulting model is simpler and easier to understand and interpret

Ridge Regression

Ridge regression is a regularization technique that modifies linear regression to reduce the impact of multicollinearity and prevent overfitting. It achieves this by adding a penalty term (L2 regularization) to the cost function that shrinks the regression coefficients toward zero, making the model more stable and better at generalizing to new data.

How it works:

- **Standard vs. Ridge:** Standard linear regression can become unstable when independent variables are highly correlated (multicollinearity), leading to unreliable coefficient estimates. Ridge regression addresses this by shrinking these coefficients, making them less sensitive to small changes in the data.
- **L2 regularization:** It adds a penalty term proportional to the square of the magnitude of the coefficients to the standard least squares cost function.
- **Lambda (λ):** A tuning parameter called lambda (λ) controls the strength of the penalty.
  - If $\lambda=0$, ridge regression is the same as standard linear regression.
  - As λ increases, the coefficients are shrunk further towards zero.

Benefits:

- **Reduces multicollinearity:** It handles highly correlated independent variables effectively.
- **Prevents overfitting:** By penalizing large coefficients, it reduces model variance and improves generalization.
- **Keeps all predictors:** Unlike some other methods, it does not set coefficients to exactly zero, so all variables are retained in the model, though their influence is reduced.
- **Improves stability:** It provides more stable coefficient estimates, especially when there are more predictors than observations.

Finding the right λ: The optimal value of λ is typically found using cross-validation, which involves testing different values to find the one that minimizes prediction error.

Imagine your model is *overreacting* to tiny details in the data (like memorizing noise). Ridge regression "calms it down" by shrinking the model's weights (coefficients) toward zero. Think of it like adjusting a volume knob to get the perfect sound level—not too loud (overfitting), not too quiet (underfitting).

This penalty discourages the model from using large values for the coefficients (the numbers multiplying the features). It forces the model to keep these coefficients small. By making the coefficients smaller and closer to zero, ridge regression simplifies the model and reduces its sensitivity to random fluctuations or noise in the data. This makes the model less likely to overfit and helps it perform better on new, unseen data, improving its overall accuracy and reliability.

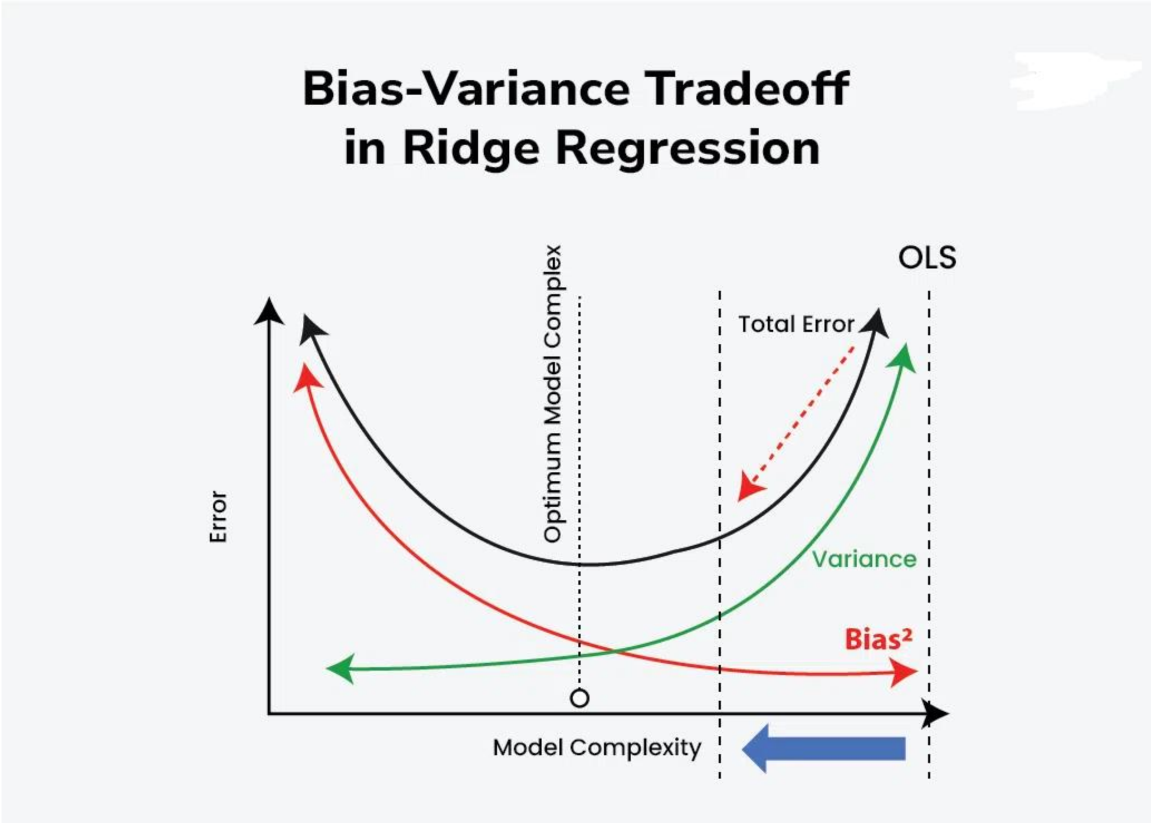
For Example - We are predicting house prices based on multiple features such as square footage, number of bedrooms, and age of the house:
Price=1000 Size−500·Age+Noise
Ridge might adjust it to:
Price=800·Size−300·Age+Less Noise
As lambda increases the model **places more emphasis on shrinking the coefficients of highly correlated features**, making their impact smaller and more stable. This reduces the effect of multicollinearity by preventing large fluctuations in coefficient estimates due to correlated predictors.

Mathematical Formulation of Ridge Regression Estimator
Consider the multiple linear regression model:.
$y=X\beta+\epsilon$

where:
- y is an n×1 vector of observations,
- X is an n×p matrix of predictors,
- β is a p×1 vector of unknown regression coefficients,
- ϵ is an n×1 vector of random errors.

The ordinary least squares (OLS) estimator of *β* is given by:
$\hat{\beta}^{OLS}=(X'X)^{-1}X'y$
In the presence of multicollinearity, X′X is nearly singular, leading to unstable estimates. ridge regression addresses this issue by adding a penalty term kI, where k is the ridge parameter and I is the identity matrix. The ridge regression estimator is:
$\hat{\beta}_k=(X'X+kI)-1X'y$
This modification stabilizes the estimates by shrinking the coefficients, improving generalization and mitigating multicollinearity effects.

Bias-Variance Tradeoff in Ridge Regression
Ridge regression allows control over the bias-variance trade-off. Increasing the value of λ increases the bias but reduces the variance, while decreasing λ does the opposite. The goal is to find an optimal λ that balances bias and variance, leading to a model that generalizes well to new data.
As we increase the penalty level in ridge regression, the estimates of β gradually change. The following simulation illustrates how the variation in β is affected by different penalty values, showing how estimated parameters deviate from the true values.



Applications of Ridge Regression

- **Forecasting Economic Indicators:** Ridge regression helps predict economic factors like GDP, inflation, and unemployment by managing multicollinearity between predictors like interest rates and consumer spending, leading to more accurate forecasts.
- **Medical Diagnosis**: In healthcare, it aids in building diagnostic models by controlling multicollinearity among biomarkers, improving disease diagnosis and prognosis.
- **Sales Prediction**: In marketing, ridge regression forecasts sales based on factors like advertisement costs and promotions, handling correlations between these variables for better sales planning.
- **Climate Modeling:** Ridge regression improves climate models by eliminating interference between variables like temperature and precipitation, ensuring more accurate predictions.
- **Risk Management**: In credit scoring and financial risk analysis, ridge regression evaluates creditworthiness by addressing multicollinearity among financial ratios, enhancing accuracy in risk management.