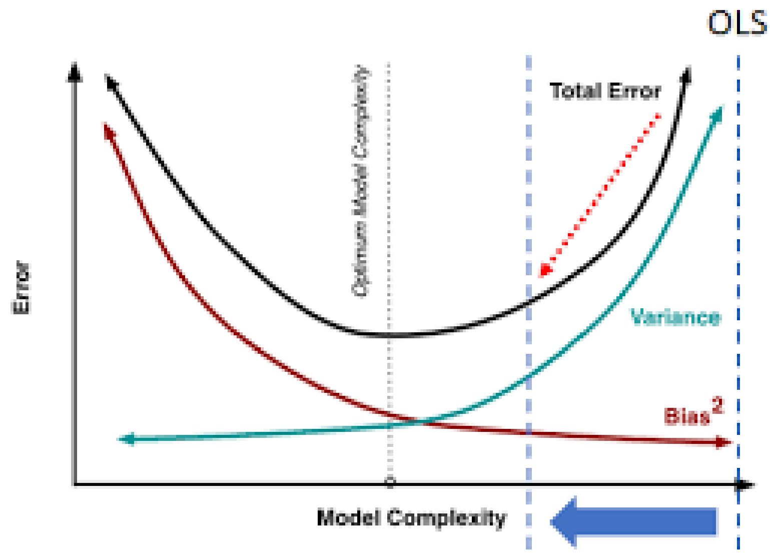
Ridge regression:

Ridge regression is a regularization technique used in data analytics and statistical modeling, particularly in the context of linear regression. It's designed to mitigate the problem of overfitting and improve the stability and generalization performance of linear regression models. Here's an explanation of ridge regression in data analytics:



Objective: The primary objective of ridge regression is to find a linear regression model that fits the data well while also minimizing the sum of the squared values of the model's coefficients.

Key Characteristics and Components:

1. **Regularization Term (L2 Regularization):** Ridge regression introduces a penalty term to the linear regression objective function. This penalty term is proportional to the sum of the squares of the model's coefficients. It is also known as L2 regularization because it involves the Euclidean (L2) norm of the coefficient vector.
    - The regularization term is controlled by a hyperparameter, often denoted as λ (lambda), which determines the strength of regularization. A larger λ results in stronger regularization.
2. **Objective Function:** The ridge regression objective function is defined as follows:

makefileCopy code

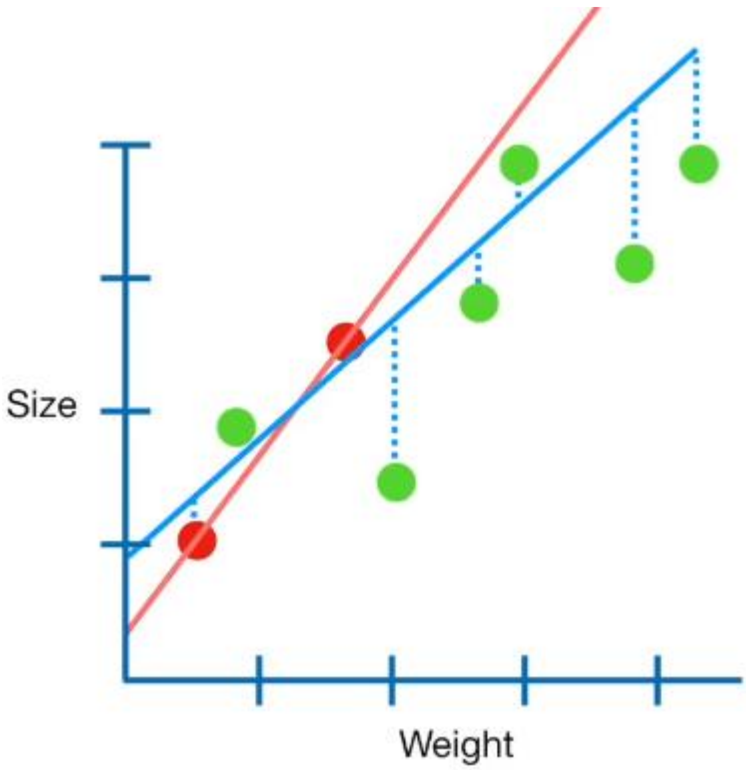
Minimize: Σ(y - ŷ)^2 + λ * Σ(β^2)

- $\Sigma(y - \hat{y})^2$ represents the sum of squared errors between the actual target values (y) and the predicted values ($\hat{y}$) from the linear regression model.
- $\lambda * \Sigma(\beta^2)$ represents the sum of squared coefficients (β) with the regularization term.

Effect and Benefits of Ridge Regression:

1. **Preventing Overfitting:** Ridge regression adds a penalty term that discourages large coefficient values. This constraint on the coefficients helps prevent the model from fitting the noise in the training data, reducing the risk of overfitting.
2. **Shrinking Coefficients:** The penalty term encourages the coefficients to be small but doesn't force them to be exactly zero. As a result, all predictors remain in the model, and their impacts are reduced.
3. **Multicollinearity Handling:** Ridge regression is particularly useful when multicollinearity (high correlation between predictors) exists in the data. It effectively reduces the impact of correlated predictors by distributing the coefficients more evenly.
4. **Improved Stability:** Ridge regression often leads to more stable and robust models because it reduces the sensitivity of the model to variations in the training data.
5. **Hyperparameter Tuning:** The choice of the λ hyperparameter is essential. It can be determined through techniques like cross-validation, where different values of λ are tested, and the one that results in the best model performance on a validation set is selected.

Applications: Ridge regression is widely applied in data analytics, statistics, and machine learning. It is used in various fields, including economics, finance, healthcare, and social sciences, to build predictive models that generalize well and to handle issues like multicollinearity.


Size
Weight