RIDGE REGRESSION

Introduction to Ridge Regression

Ridge regression, also known as **Tikhonov regularization**, is a technique used to analyze multiple regression data that suffer from multicollinearity. Multicollinearity occurs when predictor variables are highly correlated, making it difficult to estimate the relationship between each predictor and the response variable accurately. Ridge regression addresses this problem by adding a penalty to the regression model to shrink the regression coefficients, which helps reduce model complexity and prevent overfitting.

Key Concepts of Ridge Regression

1. **Ordinary Least Squares (OLS) Regression Limitations**:
   - o **Multicollinearity**: When two or more predictors in a model are correlated, it becomes difficult to determine their individual effect on the response variable. This leads to large variances in the estimated regression coefficients, making the model unstable.
   - o **Overfitting**: In cases where the model has many predictors, OLS regression can overfit the data, capturing noise rather than the underlying trend.
2. **Ridge Regression Solution**:
   - o **Regularization**: Ridge regression introduces a regularization parameter that adds a penalty for large coefficients in the regression model. This penalty helps to reduce the magnitude of the coefficients, leading to a simpler, more generalizable model.
   - o **L2 Penalty**: Ridge regression uses an L2 penalty, which is the sum of the squares of the coefficients.

Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.

Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called L2 **regularization**.

In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called **Ridge Regression penalty**. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.

The equation for the cost function in ridge regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}\beta_j^2$$

o In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model.

o As we can see from the above equation, if the values of **λ tend to zero, the equation becomes the cost function of the linear regression model.** Hence, for the minimum value of λ, the model will resemble the linear regression model.

o A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used. o It helps to solve the problems if we have more parameters than samples.

Advantages of Ridge Regression:

o **Reduced Variance**: By shrinking the coefficients, ridge regression reduces the model variance, making it less sensitive to small changes in the training data. o **Improved Prediction**: Ridge regression often improves prediction accuracy, especially when dealing with multicollinearity or high-dimensional data.

Conclusion

Ridge regression is a powerful technique for handling multicollinearity and overfitting in regression models. By adding an L2 penalty to the regression, it reduces the magnitude of the coefficients, making the model more robust and generalizable. The choice of the regularization parameter λ\\lambda λ is crucial and is typically selected using cross-validation to balance model complexity and prediction accuracy.

Bayesian Linear Regression

Linear regression is a popular regression approach in machine learning. Linear regression is based on the assumption that the underlying data is normally distributed and that all relevant predictor variables have a linear relationship with the outcome. But In the real world, this is not always possible, it will follows these assumptions, Bayesian regression could be the better choice.
Bayesian regression employs prior belief or knowledge about the data to "learn" more about it and create more accurate predictions. It also takes into account the data's uncertainty and leverages prior knowledge to provide more precise estimates of the data. As a result, it is an ideal choice when the data is complex or ambiguous.
Bayesian regression uses a Bayes algorithm to estimate the parameters of a linear regression model from data, including prior knowledge about the parameters. Because of its probabilistic character, it can produce more accurate estimates for regression parameters than ordinary least squares (OLS) linear regression, provide a measure of uncertainty in the estimation, and make stronger conclusions than OLS. Bayesian regression can also be utilized for related regression analysis tasks like model selection and outlier detection.

Bayesian Regression

Bayesian regression is a type of linear regression that uses Bayesian statistics to estimate the unknown parameters of a model. It uses Bayes' theorem to estimate the likelihood of a set of parameters given observed data. The goal of Bayesian regression is to find the best estimate of the parameters of a linear model that describes the relationship between the independent and the dependent variables.
The main difference between traditional linear regression and Bayesian regression is the underlying assumption regarding the data-generating process. Traditional linear regression assumes that data follows a Gaussian or normal distribution, while Bayesian regression has stronger assumptions about the nature of the data and puts a prior probability distribution on the parameters. Bayesian regression also enables more flexibility as it allows for additional parameters or prior distributions, and can be used to construct an arbitrarily complex model that explicitly expresses prior beliefs about the data. Additionally, Bayesian regression provides more accurate predictive measures from fewer data points and is able to construct estimates for uncertainty around the estimates. On the other hand, traditional linear regressions are easier to implement and generally faster with simpler models and can provide good results when the assumptions about the data are valid.
Bayesian Regression can be very useful when we have insufficient data in the dataset or the data is poorly distributed. The output of a Bayesian Regression model is obtained from a probability distribution, as compared to regular regression techniques where the output is just obtained from a single value of each attribute.

Some Dependent Concepts for Bayesian Regression
The important concepts in Bayesian Regression are as follows:

Bayes Theorem
Bayes Theorem gives the relationship between an event's prior probability and its posterior probability after evidence is taken into account. It states that the conditional probability of an event is equal to the probability of the event given certain conditions multiplied by the prior probability of the event, divided by the probability of the conditions.

i.e .
Where P(A|B) is the probability of event A occurring given that event B has already occurred, P(B|A) is the probability of event B occurring given that event A has already occurred, P(A) is the probability of event A occurring and P(B) is the probability of event B occurring. **Maximum Likelihood Estimation (MLE)**
MLE is a method used to estimate the parameters of a statistical model by maximizing the likelihood function. it seeks to find the parameter values that make the observed data most probable under the assumed model. MLE does not incorporate any prior information or assumptions about the parameters, and it provides point estimates of the parameters

Maximum A Posteriori (MAP) Estimation
MAP estimation is a Bayesian approach that combines prior information with the likelihood function to estimate the parameters. It involves finding the parameter values that maximize the posterior distribution, which is obtained by applying Bayes' theorem. In MAP estimation, a prior distribution is specified for the parameters, representing prior beliefs or knowledge about their values. The likelihood function is then multiplied by the prior distribution to obtain the joint distribution, and the parameter values that maximize this joint distribution are selected as the MAP estimates. MAP estimation provides point estimates of the parameters, similar to MLE, but incorporates prior information.

Need for Bayesian Regression
There are several reasons why Bayesian regression is useful over other regression techniques. Some of them are as follows:

1. Bayesian regression also uses the prior belief about the parameters in the analysis. which makes it useful when there is limited data available and the prior knowledge are relevant. By combining prior knowledge with the observed data, Bayesian regression provides more informed and potentially more accurate estimates of the regression parameters.

2. Bayesian regression provides a natural way to measure the uncertainty in the estimation of regression parameters by generating the posterior distribution, which captures the uncertainty in the parameter values, as opposed to the single point estimate that is produced by standard regression techniques. This

distribution offers a range of acceptable values for the parameters and can be used to compute trustworthy intervals or Bayesian confidence intervals.

3. In order to incorporate complicated correlations and non-linearities, Bayesian regression provides flexibility by offering a framework for integrating various prior distributions, which makes it capable to handle situations where the basic assumptions of standard regression techniques, like linearity or homoscedasticity, may not be true. It enables the modeling of more realistic and nuanced relationships between the predictors and the response variable.

4. Bayesian regression facilitates model selection and comparison by calculating the posterior probabilities of different models.

5. Bayesian regression can handle outliers and influential observations more effectively compared to classical regression methods. It provides a more robust approach to regression analysis, as extreme or influential observations have a lesser impact on the estimation.