FLIP ROBO TECHNOLOGIES

ASSIGNMENT – 39 INTERN-26(ANSWERS) USHARANI J

MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is ______?
A) Positive

4. Which of the following will have symmetric relation between dependent variable and independent
variable?

B) Correlation

5. Which of the following is the reason for over fitting condition?

C) Low bias and high variance

6. If output involves label then that model is called as:

B) Predictive modal

7. Lasso and Ridge regression techniques belong to _________?

D) Regularization

8. To overcome with imbalance dataset which technique can be used?

D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary
classification problems. It uses _____ to make graph?

A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the
curve should be less.

A) True

11. Pick the feature extraction from below:

A) Construction bag of words from an email
C) Removing stop words

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear
Regression?

A) We don’t have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.
.

FLIP ROBO TECHNOLOGIES

ASSIGNMENT – 39

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

- It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it.
- It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.
- For regression problems, the increase in flexibility of a model is represented by an increase in its coefficients, which are calculated from the regression line.
- In simple words, "In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables". It maintains accuracy as well as a generalization of the model.

- Why We need Regularization:

- Sometimes what happens is that our Machine learning model performs well on the training data but does not perform well on the unseen or test data. It means the model is not able to predict the output or target column for the unseen data by introducing noise in the output, and hence the model is called an over fitted model.
- Let's understand the meaning of "Noise" in a brief manner:

- By noise we mean those data points in the dataset which don't really represent the true properties of your data, but only due to a random chance.

- So, to deal with the problem of overfitting we take the help of regularization techniques.

Regularization Work in this following ways:

Regularization works by adding a penalty or complexity term or shrinkage term with Residual Sum of Squares (RSS) to the complex model.

Let's consider the Simple linear regression equation:

Here Y represents the dependent feature or response which is the learned relation. Then, Y is approximated to $\beta0 + \beta1X1 + \beta2X2 + \ldots + \beta pXp$

Here, X1, X2, …Xp are the independent features or predictors for Y, and

β0, β1,…..βn represents the coefficients estimates for different variables or predictors(X), which describes the weights or magnitude attached to the features, respectively.

In simple linear regression, our optimization function or loss function is known as the residual sum of squares (RSS).

We choose those set of coefficients, such that the following loss function is minimized:

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 .$$

Fig. Cost Function For Simple Linear Regression.

Now, this will adjust the coefficient estimates based on the training data. If there is noise present in the training data, then the estimated coefficients won't generalize well and are not able to predict the future data.

This is where regularization comes into the picture, which shrinks or regularizes these learned estimates towards zero, by adding a loss function with optimizing parameters to make a model that can predict the accurate value of Y.

14. Which particular algorithms are used for regularization?

Techniques of Regularization

Mainly, there are two types of regularization techniques, which are given below:

- Ridge Regression
- Lasso Regression

Ridge Regression

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

In Statistics, it is known as the L-2 norm.

In this technique, the cost function is altered by adding the penalty term (shrinkage term), which multiplies the lambda with the squared weight of each individual feature. Therefore, the optimization function (cost function) becomes:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = \mathrm{RSS} + \lambda\sum_{j=1}^{p}\beta_j^2$$

In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the magnitudes of the coefficients that help to decrease the complexity of the model.

Usage of Ridge Regression:

- When we have the independent variables which are having high collinearity (problem of multicollinearity) between them, at that time general linear or polynomial regression will fail so to solve such problems, Ridge regression can be used.

- If we have more parameters than the samples, then Ridge regression helps to solve the problems.

Limitation of Ridge Regression:

- Not helps in Feature Selection: It decreases the complexity of a model but does not reduce the number of independent variables since it never leads to a coefficient being zero rather only minimizes it. Hence, this technique is not good for feature selection.

- Model Interpretability: Its disadvantage is model interpretability since it will shrink the coefficients for least important predictors, very close to zero but it will never make them exactly zero. In other words, the final model will include all the independent variables, also known as predictor

Lasso Regression

Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

It is similar to the Ridge Regression except that the penalty term includes the absolute weights instead of a square of weights. Therefore, the optimization function becomes:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|.$$

Fig. Cost Function for Lasso Regression

In statistics, it is known as the L-1 norm.

In this technique, the L1 penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero which means there is a complete removal of some of the features for model evaluation when the tuning parameter λ is sufficiently large. Therefore, the lasso method also performs Feature selection and is said to yield sparse models.

Limitation of Lasso Regression:

- Problems with some types of Dataset: If the number of predictors is greater than the number of data points, Lasso will pick at most n predictors as non-zero, even if all predictors are relevant

- Multicollinearity Problem: If there are two or more highly collinear variables then LASSO regression selects one of them randomly which is not good for the interpretation of our model.

Key Differences between Ridge and Lasso Regression

Ridge regression helps us to reduce only the overfitting in the model while keeping all the features present in the model. It reduces the complexity of the model by shrinking the coefficients whereas Lasso regression helps in reducing the problem of overfitting in the model as well as automatic feature selection.

Lasso Regression tends to make coefficients to absolute zero whereas Ridge regression never sets the value of coefficient to absolute zero.

Mathematical Formulation of Regularization Techniques

Now, we are trying to formulate these techniques in mathematical terms. So, these techniques can be understood as solving an equation,

For ridge regression, the total sum of squares of coefficients is less than or equal to s and for Lasso regression, the total sum of modulus of coefficients is less than or equal to s.

Here, s is a constant which exists for each value of the shrinkage factor λ.

These equations are also known as constraint functions.

Let's take an example to understand the mathematical formulation clearly,

For Example, Consider there are 2 parameters for a given problem

Ridge regression:

According to the above mathematical formulation, the ridge regression is described by $\beta 1^2 + \beta 2^2 \leq s$.

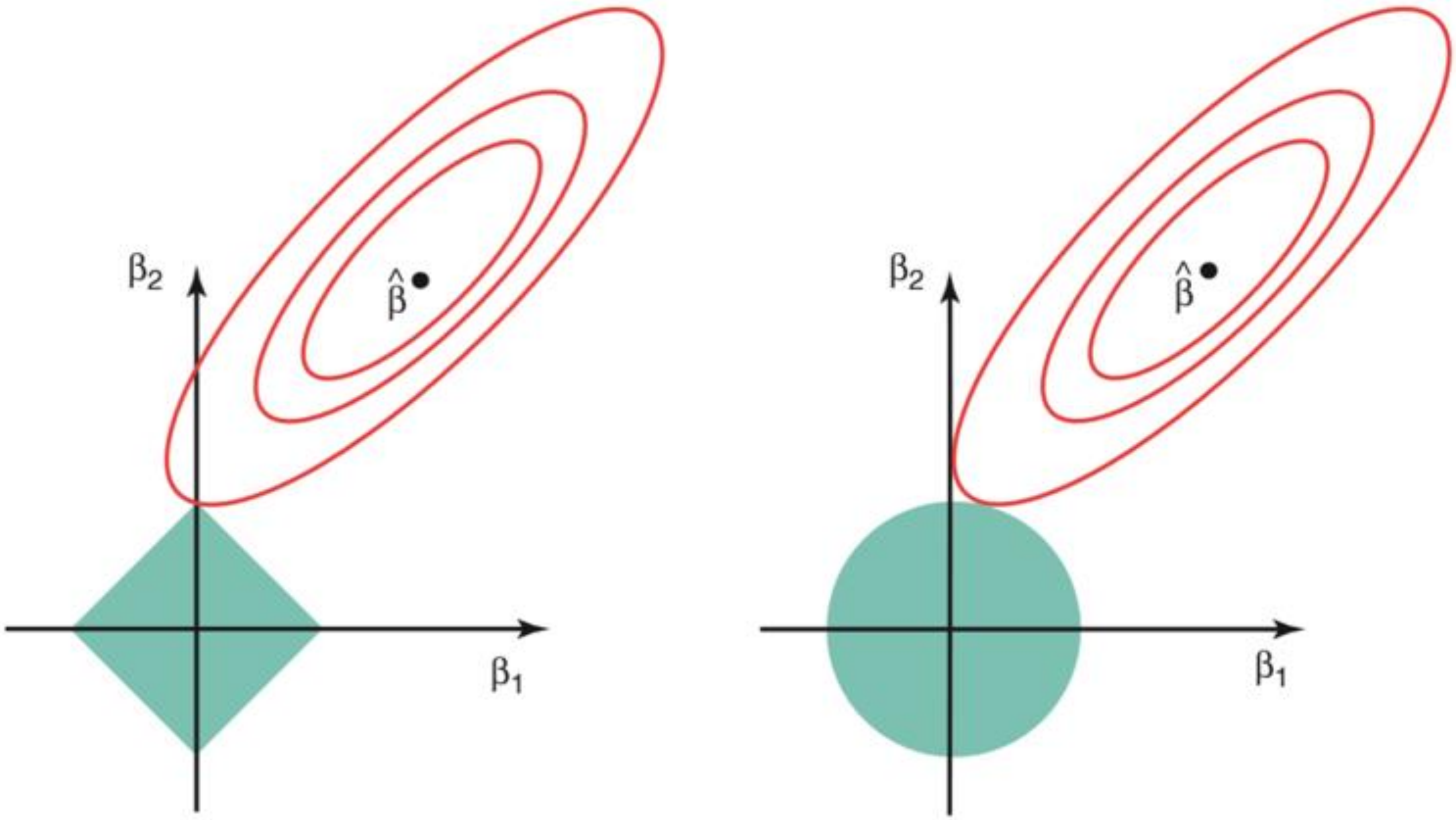
This implies that ridge regression coefficients have the smallest RSS (loss function) for all points that lie within the circle given by $\beta 1^2 + \beta 2^2 \leq s$.

Lasso Regression:

According to the above mathematical formulation, the equation becomes, $|\beta 1|+|\beta 2| \leq s$.

This implies that the coefficients for lasso regression have the smallest RSS (loss function) for all points that lie within the diamond given by $|\beta 1|+|\beta 2| \leq s$.

The image below describes these equations:



Description About Image: The given image shows the constraint functions (in green areas), for lasso (in left) and ridge regression (in right), along with contours for RSS (red ellipse).

Points on the ellipse describe the value of Residual Sum of Squares (RSS) which is calculated for simple linear regression.

For a very large value of s, the green regions will include the center of the ellipse with itself, which makes the coefficient estimates of both regression techniques equal to the least-squares estimates of simple linear regression. But, the given image shown does not describe this case. In that case, coefficient estimates of lasso and ridge regression are given by the first point at which an ellipse interacts with the constraint region.

Ridge Regression: Since ridge regression has a circular type constraint region, having no sharp points, so the intersection with the ellipse will not generally occur on the axes, therefore, the ridge regression coefficient estimates will be exclusively non-zero.

Lasso Regression: Lasso regression has a diamond type constraint region that has corners at each of the axes, so the ellipse will often intersect the constraint region at axes. When this happens, one of the coefficients (from collinear variables) will be zero and for higher dimensions having parameters greater than 2, many of the coefficient estimates may equal zero simultaneously.

15. Explain the term error present in linear regression equation?

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.

KEY TAKEAWAYS

- An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.
- The error term is a residual variable that accounts for a lack of perfect goodness of fit.
- Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.

Linear Regression, Error Term

Linear regression is a form of analysis that relates to current trends experienced by a particular security or index by providing a relationship between a dependent and independent variables, such as the price of a security and the passage of time, resulting in a trend line that can be used as a predictive model.

A linear regression exhibits less delay than that experienced with a moving average, as the line is fit to the data points instead of based on the averages within the data. This allows the line to change more quickly and dramatically than a line based on numerical averaging of the available data points.

The error calculated in a linear regression model is as follows:

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. Calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.

FLIP ROBO TECHNOLOGIES