Types of Regression Techniques in ML

Veena Ghorakavi

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used, the simplest is linear regression. It tries to fit data with the best hyperplane which goes through the points.

What is Regression Analysis?

Regression Analysis is a statistical process for estimating the relationships between the dependent variables or criterion variables and one or more independent variables or predictors. Regression analysis is generally used when we deal with a dataset that has the target variable in the form of continuous data. Regression analysis explains the changes in criteria in relation to changes in select predictors. The conditional expectation of the criteria is based on predictors where the average value of the dependent variables is given when the independent variables are changed. Three major uses for regression analysis are determining the strength of predictors, forecasting an effect, and trend forecasting.

What is the purpose to use Regression Analysis?

There are times when we would like to analyze the effect of different independent features on the target or what we say dependent features. This helps us make decisions that can affect the target variable in the desired direction. Regression analysis is heavily based on statistics and hence gives quite reliable results due to this reason only regression models are used to find the linear as well as non-linear relation between the independent and the dependent or target variables.

Types of Regression Techniques

Along with the development of the machine learning domain regression analysis techniques have gained popularity as well as developed manifold from just y = mx + c. There are several types of regression techniques, each suited for different types of data and different types of relationships. The main types of regression techniques are:

1. Linear Regression
2. Polynomial Regression
3. Stepwise Regression
4. Decision Tree Regression
5. Random Forest Regression
6. Support Vector Regression
7. Ridge Regression
8. Lasso Regression
9. ElasticNet Regression
10. Bayesian Linear Regression

Linear Regression

Linear regression is used for predictive analysis. Linear regression is a linear approach for modeling the relationship between the criterion or the scalar response and the multiple predictors or explanatory variables. Linear regression focuses on the conditional probability distribution of the response given the values of the predictors. For linear regression, there is a danger of overfitting. The formula for linear regression is:

Syntax:

y = θx + b

where,

- ***θ*** *– It is the model weights or parameters*
- ***b*** *– It is known as the bias.*

This is the most basic form of regression analysis and is used to model a linear relationship between a single dependent variable and one or more independent variables.

Polynomial Regression

This is an extension of linear regression and is used to model a non-linear relationship between the dependent variable and independent variables. Here as well syntax remains the same but now in the input variables we include some polynomial or higher degree terms of some already existing features as well. Linear regression was only able to fit a linear model to the data at hand but with polynomial features, we can easily fit some non-linear relationship between the target as well as input features.

Stepwise Regression

Stepwise regression is used for fitting regression models with predictive models. It is carried out automatically. With each step, the variable is added or subtracted from the set of explanatory variables. The approaches for stepwise regression are forward selection, backward elimination, and bidirectional elimination. The formula for stepwise regression is $b_{j.std} = b_j(s_x * s_y^{-1})$.

Decision Tree Regression

A Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. There is a non-parametric method used to model a decision tree to predict a continuous outcome.

Random Forest Regression

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

Support Vector Regression (SVR)

Support vector regression (SVR) is a type of support vector machine (SVM) that is used for regression tasks. It tries to find a function that best predicts the continuous output value for a given input value.

SVR can use both linear and non-linear kernels. A linear kernel is a simple dot product between two input vectors, while a non-linear kernel is a more complex function that can capture more intricate patterns in the data. The choice of kernel depends on the data's characteristics and the task's complexity.

Ridge Regression

Ridge regression is a technique for analyzing multiple regression data. When multicollinearity occurs, least squares estimates are unbiased. This is a regularized linear regression model, it tries to reduce the model complexity by adding a penalty term to the cost function. A degree of bias is added to the regression estimates, and as a result, ridge regression reduces the standard errors.

$$\text{Cost} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \|i - X\beta\|^2 + \lambda \|\beta\|^2$$

Lasso Regression

Lasso regression is a regression analysis method that performs both variable selection and regularization. Lasso regression uses soft thresholding. Lasso regression selects only a subset of the provided covariates for use in the final model.

This is another regularized linear regression model, it works by adding a penalty term to the cost function, but it tends to zero out some features' coefficients, which makes it useful for feature selection.

ElasticNet Regression

Linear Regression suffers from overfitting and can't deal with collinear data. When there are many features in the dataset and even some of them are not relevant to the predictive model. This makes the model more complex with a too-inaccurate prediction on the test set (or overfitting). Such a model with high variance does not generalize on the new data. So, to deal with these issues, we include both L-2 and L-1 norm regularization to get the benefits of both Ridge and Lasso at the same time. The resultant model has better predictive power than Lasso. It performs feature selection and also makes the hypothesis simpler. The modified cost function for Elastic-Net Regression is given below:

$$\frac{1}{m}\left[\sum_{l=1}^{m}\left(y^{(i)} - h\left(x^{(i)}\right)\right)^2 + \lambda_1 \sum_{j=1}^{n} w_j + \lambda_2 \sum_{j=1}^{n} w_j^2\right]$$

where,

- **w(j)** represents the weight for the j^{th} feature.
- **n** is the number of features in the dataset.
- l**ambda1** is the regularization strength for the L1 norm.
- **lambda2** is the regularization strength for the L2 norm.

Bayesian Linear Regression

As the name suggests this algorithm is purely based on Bayes Theorem. Because of this reason only we do not use the Least Square method to determine the coefficients of the regression model. So, the technique which is used here to find the model weights and parameters relies on features posterior distribution and this provides an extra stability factor to the regression model which is based on this technique.

Last Updated : 06 Apr, 2023

35

Similar Reads

1. ML | Linear Regression vs Logistic Regression
2. ML - Different Regression types
3. Short Circuiting Techniques in Python
4. Variations in different Sorting techniques in Python
5. Python | Thresholding techniques using OpenCV | Set-2 (Adaptive Thresholding)
6. Python | Thresholding techniques using OpenCV | Set-3 (Otsu Thresholding)
7. Impact of AI and ML On Warfare Techniques
8. Feature Encoding Techniques - Machine Learning
9. Different Input and Output Techniques in Python3
10. AI Conversational System - Attack Surface Areas and Effective Defense Techniques

Related Tutorials

1. Data Analysis Tutorial
2. Flask Tutorial
3. Deep Learning Tutorial
4. Top 101 Machine Learning Projects with Source Code
5. Machine Learning Mathematics

Previous

Basic Concept of Classification (Data Mining)

Next

Gradient Descent algorithm and its variants

Article Contributed By :

Veena Ghorakavi
Veena Ghorakavi

Vote for difficulty

Current difficulty : Medium

Easy | Normal | Medium | Hard | Expert

Improved By : ShubhamMaurya3, shruti20, abhishekm482g, rathoadavinash

Article Tags : Picked, Technical Scripter 2018, Machine Learning, Python, Technical Scripter

Practice Tags : Machine Learning, python

Improve Article Report Issue