Module-04, Python for Machine Learning

Regression Algorithms

Dostdar Ali
Instructor

Data science and Artificial Intelligence
3-Months Course
at
Karakaroum international Univrsity

February 18, 2024

Table of Contents

Regression Algorithms

- We learned in the previous section that the task of classification is to assign categorical, unordered labels to instances. A second type of supervised learning is the prediction of continuous outcomes, which is also called regression analysis.
- In regression analysis, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome), and we try to find a relationship between those variables that allows us to predict an outcome.
- The term "regression" was devised by Francis Galton in his article Regression towards Mediocrity in Hereditary Stature in 1886. Galton described the biological phenomenon that the variance of height in a population does not increase over time. He observed that the height of parents is not passed on to their children, but instead, their children's height regresses toward the population mean.

Regression Algorithms

- We learned in the previous section that the task of classification is to assign categorical, unordered labels to instances. A second type of supervised learning is the prediction of continuous outcomes, which is also called regression analysis.
- In regression analysis, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome), and we try to find a relationship between those variables that allows us to predict an outcome.
- The term "regression" was devised by Francis Galton in his article Regression towards Mediocrity in Hereditary Stature in 1886. Galton described the biological phenomenon that the variance of height in a population does not increase over time. He observed that the height of parents is not passed on to their children, but instead, their children's height regresses toward the population mean.

Regression Algorithms

- We learned in the previous section that the task of classification is to assign categorical, unordered labels to instances. A second type of supervised learning is the prediction of continuous outcomes, which is also called regression analysis.
- In regression analysis, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome), and we try to find a relationship between those variables that allows us to predict an outcome.
- The term "regression" was devised by Francis Galton in his article Regression towards Mediocrity in Hereditary Stature in 1886. Galton described the biological phenomenon that the variance of height in a population does not increase over time. He observed that the height of parents is not passed on to their children, but instead, their children's height regresses toward the population mean.

Regression Algorithms

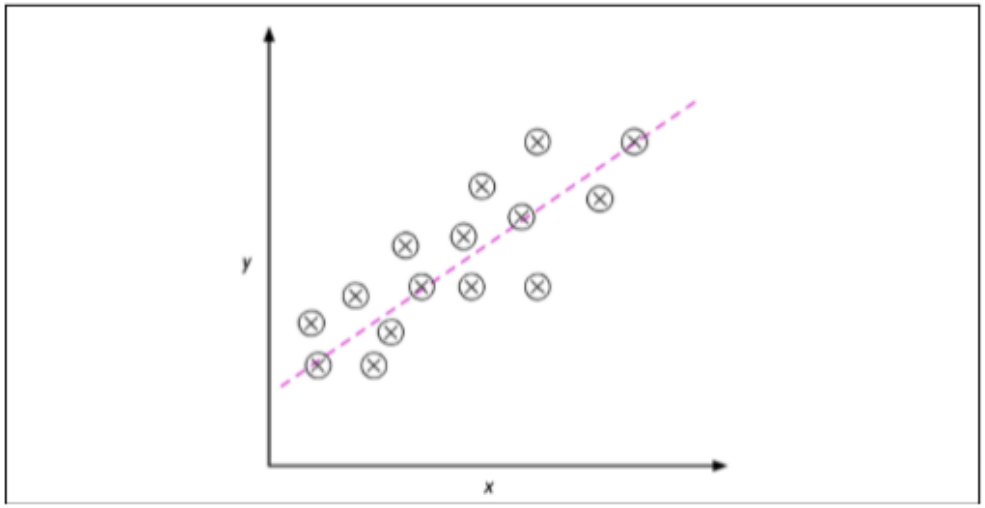
Note

In the field of machine learning, the predictor variables are commonly called "features," and the response variables are usually referred to as "target variables."

There is many types of regression algorithms commonly two is well known Linear and multi Linear regression algorithms.

Simple Linear Regression

- The following figure illustrates the concept of linear regression. Given a feature variable, x, and a target variable, y, we fit a straight line to this data that minimizes the distance—most commonly the average squared distance—between the data points and the fitted line. We can now use the intercept and slope learned from this data to predict the target variable of new data:
- Visual



Simple Linear Regression

The simple linear regression model is represented as:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where:

$$\begin{aligned} Y &: \text{Dependent variable (response)} \\ X &: \text{Independent variable (feature)} \\ \beta_0 &: \text{Intercept} \\ \beta_1 &: \text{Slope coefficient} \\ \varepsilon &: \text{Error term} \end{aligned}$$

The predicted value ($\hat{Y}$) is given by:

$$\hat{Y} = \beta_0 + \beta_1 X$$

Multi Linear Regression

Definition

Multiple linear regression is a statistical method used to model the relationship between two or more independent variables (features) and a dependent variable (response). It is an extension of simple linear regression, where only one independent variable is considered.

- The model can be estimated using various methods, such as the least squares method, which minimizes the sum of squared differences between the observed and predicted values. Multiple linear regression assumes that there is a linear relationship between the independent variables and the dependent variable and that the errors are normally distributed and have constant variance.

Multi Linear Regression

Definition

Multiple linear regression is a statistical method used to model the relationship between two or more independent variables (features) and a dependent variable (response). It is an extension of simple linear regression, where only one independent variable is considered.

- The model can be estimated using various methods, such as the least squares method, which minimizes the sum of squared differences between the observed and predicted values. Multiple linear regression assumes that there is a linear relationship between the independent variables and the dependent variable and that the errors are normally distributed and have constant variance.

Multi Linear Regression

The multiple linear regression model is represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Where:

$$\begin{aligned} Y &: \text{Dependent variable (response)} \\ X_1, X_2, \ldots, X_n &: \text{Independent variables (features)} \\ \beta_0 &: \text{Intercept} \\ \beta_1, \beta_2, \ldots, \beta_n &: \text{Coefficients for the respective independent variables} \\ \varepsilon &: \text{Error term} \end{aligned}$$

The predicted value ($\hat{Y}$) is given by:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Examples of Regressions

- let's assume that we are interested in predicting the math SAT scores of students. If there is a relationship between the time spent studying for the test and the final scores, we could use it as training data to learn a model that uses the study time to predict the test scores of future students who are planning to take this test.

Great Job
Thank yo