

4) Reading comprehension : "Regression vs. Classification in Machine Learning"

Reading: "Regression vs. Classification in Machine Learning"

Comparing regression vs classification in machine learning can sometimes confuse even the most seasoned data scientists. This can eventually make it difficult for them to implement the right methodologies for solving prediction problems. Both regression and classification are types of supervised machine learning algorithms, where a model is trained according to the existing model along with correctly labeled data. But there are also many differences between regression and classification algorithms that you should know in order to implement them correctly and sharpen your machine learning skills. In this blog, we will understand the difference between regression and classification algorithms.

Regression vs Classification in Machine Learning: How they Differ

Some algorithms may need both classification and regression approaches, which is why an in-depth knowledge of both is crucial in the fields of AI and data science. Before we deep dive into understanding the differences between regression and classification algorithms. Let's first understand each algorithm.

What is Regression Machine Learning?

Regression algorithms predict a continuous value based on the input variables. The main goal of regression problems is to estimate a mapping function based on the input and output variables. If your target variable is a quantity like income, scores, height or weight, or the probability of a binary category (like the probability of rain in particular regions), then you should use the regression model. However, there are various types of regressions used by data scientists and ML engineers based on different scenarios. The different types of regression algorithms include:

1. Simple linear regression

With simple linear regression, you can estimate the relationship between one independent variable and another dependent variable using a straight line, given both variables are quantitative.

2. Multiple linear regression

An extension of simple linear regression, multiple regression can predict the values of a dependent variable based on the values of two or more independent variables.

3. Polynomial regression

The main aim of polynomial regression is to model or find a nonlinear relationship between dependent and independent variables.

What is Classification Machine Learning?

Classification is a predictive model that approximates a mapping function from input variables to identify discrete output variables, which can be labels or categories. The mapping function of classification algorithms is responsible for predicting the label or category of the given input variables. A classification algorithm can have both discrete and real-valued variables, but it requires that the examples be classified into one of two or more classes.

The different types of classification algorithms include:

1. Decision tree classification

In this algorithm, a classification model is created by building a decision tree where every node of the tree is a test case for an attribute and each branch coming from the node is a possible value for that attribute.

2. Random forest classification

This tree-based algorithm includes a set of decision trees which are randomly selected from a subset of the main training set. The random forest classification algorithm aggregates outputs from all the different decision trees to decide on the final output prediction, which is more accurate than any of the individual trees.

3. K-nearest neighbor

The K-nearest neighbor algorithm assumes that similar things exist in close proximity to each other. It uses feature similarity for predicting values of new data points. The algorithm helps grouping similar data points together according to their proximity. The main goal of the algorithm is to determine how likely it is for a data point to be a part of the specific group.

Regression vs Classification in Machine Learning: Understanding the Difference

The most significant difference between regression vs classification is that while regression helps predict a continuous quantity, classification predicts discrete class labels. There are also some overlaps between the two types of machine learning algorithms.

A regression algorithm can predict a discrete value which is in the form of an integer quantity

A classification algorithm can predict a continuous value if it is in the form of a class label probability

Let's consider a dataset that contains student information of a particular university. A regression algorithm can be used in this case to predict the height of any student based on their weight, gender, diet, or subject major. We use regression in this case because height is a continuous quantity. There is an infinite number of possible values for a person's height.

On the contrary, classification can be used to analyze whether an email is spam or not spam. The algorithm checks the keywords in an email and the sender's address is to find out the probability of the email being spam. Similarly, while a regression model can be used to predict temperature for the next day, we can use a classification algorithm to determine whether it will be cold or hot according to the given temperature values.



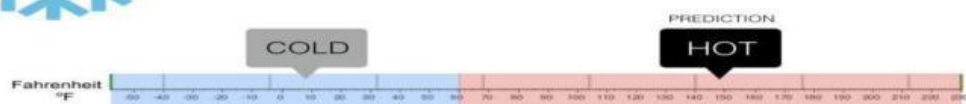
Regression

What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?



Taken from: <https://www.springboard.com/blog/data-science/regression-vs-classification/>

5) True/ False activity

1. **True or False:** Both regression and classification are types of unsupervised machine learning algorithms.
2. **True or False:** Regression algorithms are used when the target variable is a continuous value, while classification algorithms predict discrete class labels.
3. **True or False:** A regression model can only predict continuous values and cannot handle discrete quantities.
4. **True or False:** Decision tree classification is an example of a regression algorithm.
5. **True or False:** The key difference between regression and classification lies in their goals—regression predicts continuous values, while classification predicts discrete class labels.