

What is machine learning regression?

Regression is a method for understanding the relationship between independent variables or features and a dependent variable or outcome. Outcomes can then be predicted once the relationship between independent and dependent variables has been estimated. Regression is a field of study in statistics which forms a key part of forecast models in machine learning. It's used as an approach to predict continuous outcomes in predictive modelling, so has utility in forecasting and predicting outcomes from data. Machine learning regression generally involves plotting a line of best fit through the data points. The distance between each point and the line is minimised to achieve the best fit line.

Alongside classification, regression is one of the main applications of the supervised [type of machine learning](#). Classification is the categorisation of objects based on learned features, whereas regression is the forecasting of continuous outcomes. Both are predictive modelling problems. Supervised machine learning is integral as an approach in both cases, because classification and regression models rely on labelled input and output training data. The features and output of the training data must be labelled so the model can understand the relationship.

Regression analysis is used to understand the relationship between different independent variables and a dependent variable or outcome. Models that are trained to forecast or predict trends and outcomes will be trained using regression techniques. These models will learn the relationship between input and output data from labelled training data. It can then forecast future trends or predict outcomes from unseen input data, or be used to understand gaps in historic data.

As with all supervised machine learning, special care should be taken to ensure the labelled training data is representative of the overall population. If the training data is not representative, the predictive model will be overfit to data that doesn't represent new and unseen data. This will result in inaccurate predictions once the model is deployed. Because regression analysis involves the

relationships of features and outcomes, care should be taken to include the right selection of features too.

What are regression models used for?

Machine learning regression models are mainly used in predictive analytics to forecast trends and predict outcomes. Regression models will be trained to understand the relationship between different independent variables and an outcome. The model can therefore understand the many different factors which may lead to a desired outcome. The resulting models can be used in a range of ways and in a variety of settings. Outcomes can be predicted from new and unseen data, market fluctuations can be predicted and accounted for, and campaigns can be tested by tweaking different independent variables.

In practice, a model will be trained on labelled data to understand the relationship between data features and the dependent variable. By estimating this relationship, the model can predict the outcome of new and unseen data. This could be used to predict missing historic data, and estimate future outcomes too. In a sales environment, an organisation could use regression machine learning to predict the next month's sales from a number of factors. In a medical environment, an organisation could forecast health trends in the general population over a period of time.

Supervised machine learning models are generally used for either classification or regression problems. Classification is when a model is trained to categorise an object based on its features. This could include facial recognition software, or to identify a spam email in a firewall. A model will be trained on labelled input and output data to understand the specific features which classify a labelled object. On the other hand, a regression problem is when a model is used to predict continuous outcomes or values. This could be a model that forecasts salary changes, house prices, or retail sales. The model is trained on labelled input and output data to understand the strength of relationships between data features and output.

Regression is used to identify patterns and relationships within a dataset, which can then be applied to new and unseen data. This makes regression a key

element of [machine learning in finance](#), and is often leveraged to help forecast portfolio performance or stock costs and trends. Models can be trained to understand the relationship between a variety of diverse features and a desired outcome. In most cases, machine learning regression provides organisations with insight into particular outcomes. But because this approach can influence an organisation's decision-making process, the [explainability of machine learning](#) is an important consideration.

Common use for machine learning regression models include:

- Forecasting continuous outcomes like house prices, stock prices, or sales.
- Predicting the success of future retail sales or marketing campaigns to ensure resources are used effectively.
- Predicting customer or user trends, such as on streaming services or ecommerce websites.
- Analysing datasets to establish the relationships between variables and an output.
- Predicting interest rates or stock prices from a variety of factors.
- Creating time series visualisations.

What are the types of regression?

There are a range of different approaches used in machine learning to perform regression. Different popular algorithms are used to achieve machine learning regression. The different techniques may include different numbers of independent variables or process different types of data. Distinct types of machine learning regression models may also assume a different relationship between the independent and dependent variables. For example, linear regression techniques assume that the relationship is linear, so wouldn't be effective with datasets with nonlinear relationships.

Some of the most common regression techniques in machine learning can be grouped into the following types of regression analysis:

- Simple Linear Regression
- Multiple linear regression

- Logistic regression

What is simple linear regression?

Simple Linear regression is a linear regression technique which plots a straight line within data points to minimise error between the line and the data points. It is one of the most simple and basic types of machine learning regression. The relationship between the independent and dependent variables is assumed to be linear in this case. This approach is simple because it is used to explore the relationship between the dependent variable and one independent variable. [Outliers](#) may be a common occurrence in simple linear regression because of the straight line of best fit.

What is multiple linear regression?

Multiple linear regression is a technique used when more than one independent variable is used. Polynomial regression is an example of a multiple linear regression technique. It is a type of multiple linear regression, used when there is more than one independent variable. It achieves a better fit in the comparison to simple linear regression when multiple independent variables are involved. The result when plotted on two dimensions would be a curved line fitted to the data points.

What is logistic regression?

Logistic regression is used when the dependent variable can have one of two values, such as true or false, or success or failure. Logistic regression models can be used to predict the probability of a dependent variable occurring. Generally, the output values must be binary. A sigmoid curve can be used to map the relationship between the dependent variable and independent variables.