**Logistic Regression**

# ****What category of algorithms does logistic regression belong to?****

Looking at the below chart's supervised learning branch, we can see that we have two main categories of problems: regression and classification.

* **Regression**: we use regression algorithms when we have a continuous (numerical) target variable. For example, predicting the price of a house based on its proximity to major amenities.
* **Classification**: used when the target variable is categorical. For example, predicting a win/loss of a game or customer defaulting/not-defaulting on a loan payment. Note, it does not necessarily have to be a binary outcome.

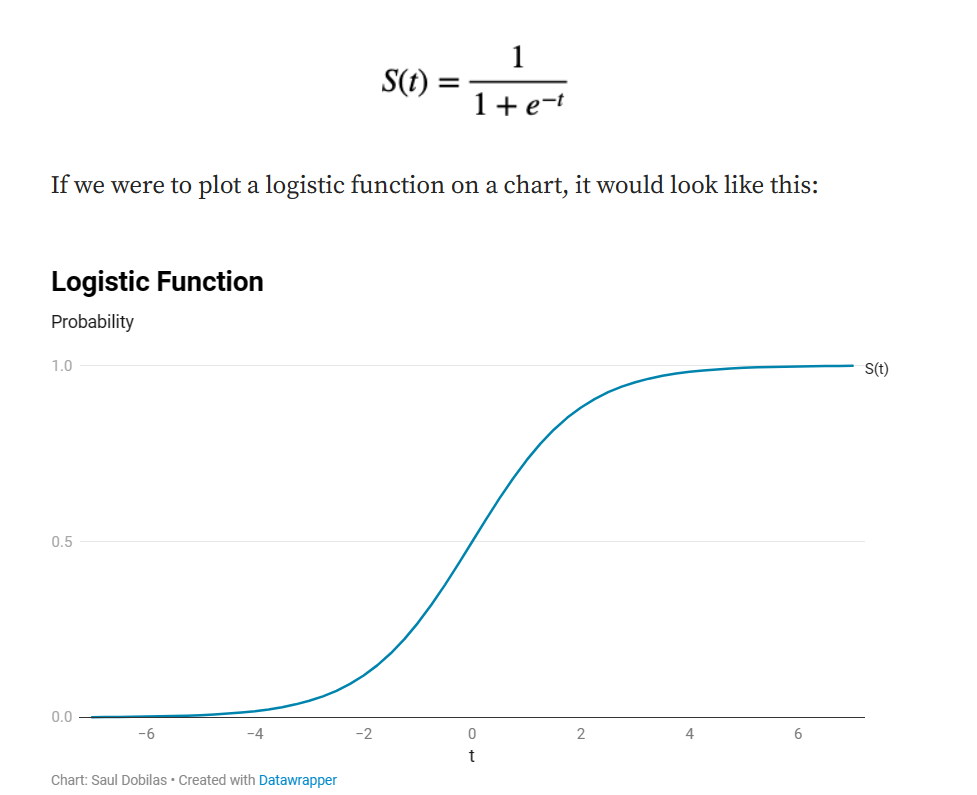
While logistic regression has a “regression” in its name, it actually belongs to the classification algorithms. However, there are some similarities between linear regression and logistic regression, which we will touch upon in the next section.

## **Logistic function**

Let’s now assume that we do not have a ‘final score.’ All we have is an outcome( pass/fail flag). We want to build a logistic regression model where we use ‘hours of study’ to predict a student's likelihood of passing the exam.

As you can see from the table above, there is a strong correlation between ‘hours of study’ and ‘exam outcome,’ although we cannot perfectly separate the two classes. Hence, we want to have a model that gives us a probability of passing the exam given the study hours.

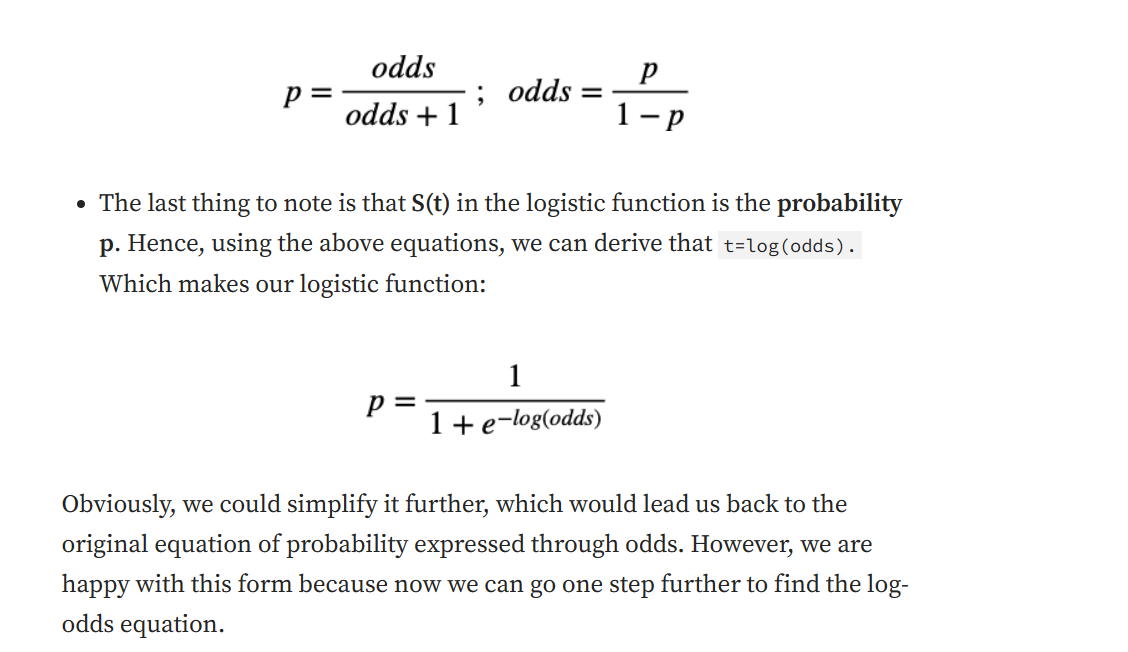
This is done by using a logistic function, also known as a sigmoid function:

****

## **Odds**

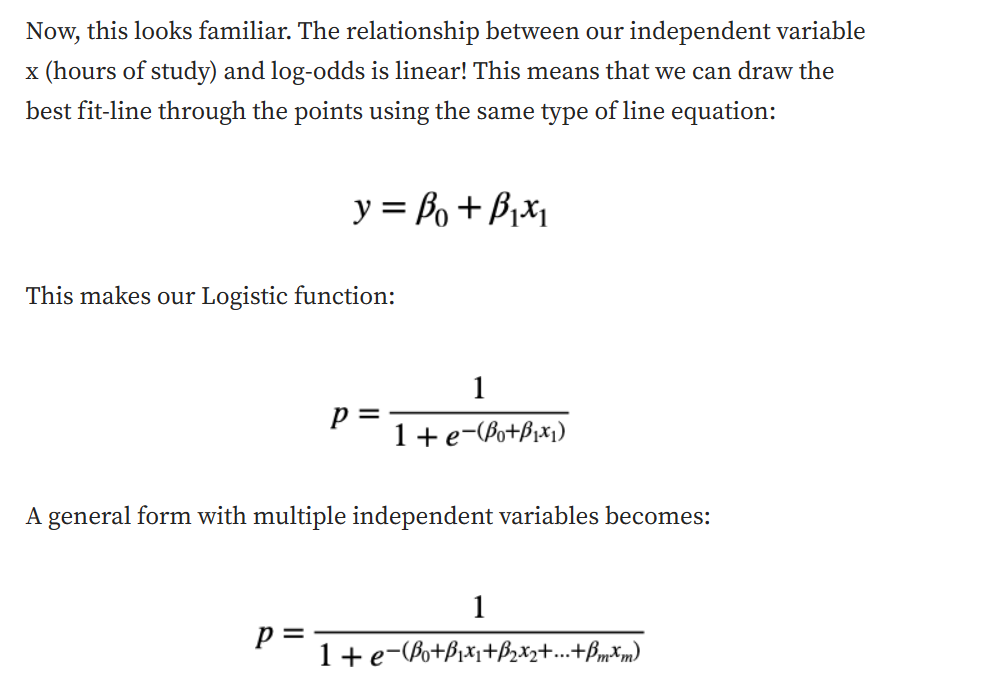
To understand how the data is mapped to the logistic function, we first need to learn about the relationship between probability, odds, and log-odds.

* **Odds** — this is simply a ratio between the number of events (in this case, exam passes) and non-events (exam failures). Say, if you had 5 pupils that spent 7 hours each studying for an exam with 3 pupils passing and 2 failing it, the odds of passing would be 3:2, which is 1.5 in decimal notation.
* **Log-odds**— is just a natural logarithm of odds. So if, the odds are 3:2 = 1.5, then log(odds) = log(1.5) = 0.405...
* **Probability vs. odds** — you can easily convert between probability and odds. So if, the odds are 3:2, then the probability is 3/5=0.6.You can use the following equations to convert between probability and odds:

****

## **Log-odds equation**

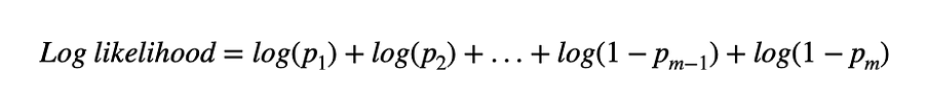
Let’s use another example to plot the data onto a graph to understand how the log-odds equation is created.

****

## **Maximum Likelihood Estimation (MLE)**

When you build logistic regression models, the algorithm's goal is to find the coefficients β(0), β(1), etc. Unlike linear regression, though, it is not done by minimizing squared residuals but finding the maximum likelihood instead.

Maximum likelihood is most often expressed through a log-likelihood formula:

****

where ***p*** is the probability for points with an actual outcome of event ("pass") and ***1-p*** is the probability for points with an actual outcome of non-event ("fail").

There are multiple methods available to maximize the log-likelihood. Some of the most commonly used ones would be **gradient descent** and **Newton–Raphson**.

In general, methods used to find the coefficients for the logistic function go through an iterative process of selecting a candidate line and calculating the log-likelihood. This is continued until the convergence is achieved and the maximum likelihood is found.