Professional Certificate in Machine Learning and Artificial Intelligence

Module 13: Logistic Regression

Learning outcomes:

- Distinguish logistic regression from linear regression in terms of its categorical output and its best use cases for binary classification.
- Describe how adjustments to parameters impact the shape of logistic functions.
- Determine the optimal number of statistically relevant predictors to add to a logistic regression function.
- Apply a maximum likelihood method to fit a logistic regression to a real-life data set.
- Compare the advantages and disadvantages of applying logistic regression relative to other classification methods.

Logistic regression

Logistic regression is a classification method that functions as a generalised linear model, extending linear regression to a categorical output response variable.

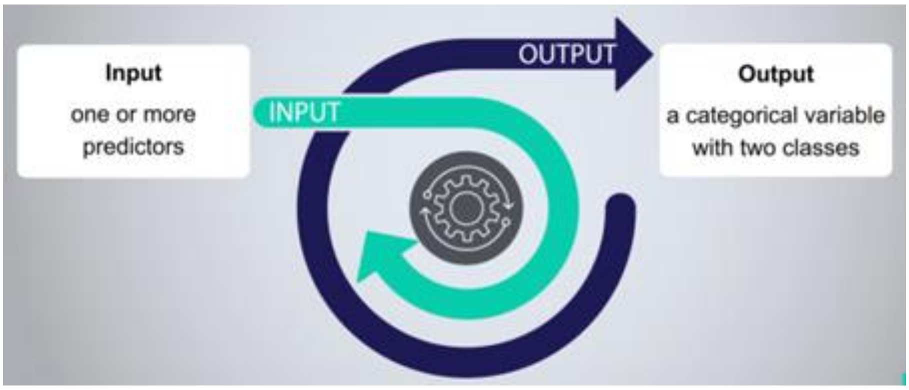
.

Advantages

- It has deep statistical properties
- It models the probability of belonging to a particular category
- It is a parametric method

Inputs and outputs

Logistic regression can have more than one predictor in its input but only two classes in its output, such as, on - off, yes - no, etc.

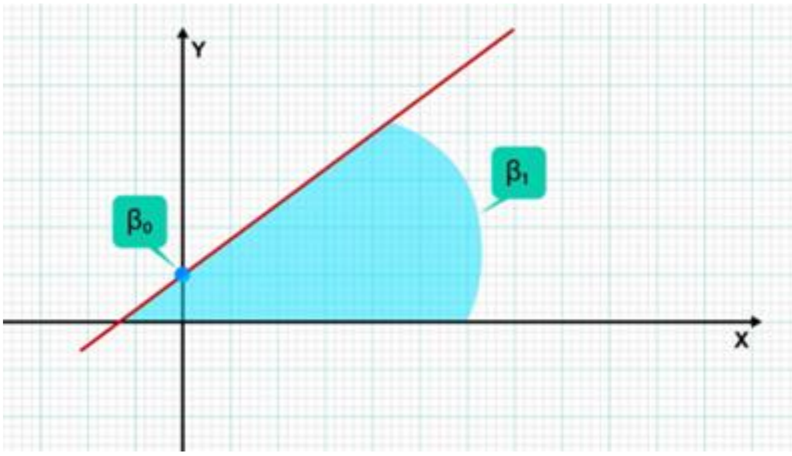


In logistic regression, we cannot have uncertain categories. For more than two categories, we would use linear discriminant analysis

Linear vs logistic regression

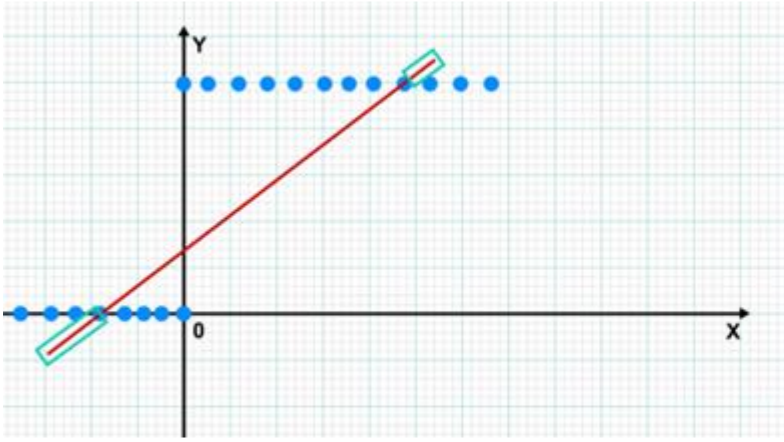
Imagine you have the following set of input, output, and parameters.

- Input: X
- Output: Y
- Parameters: β_0 and β_1

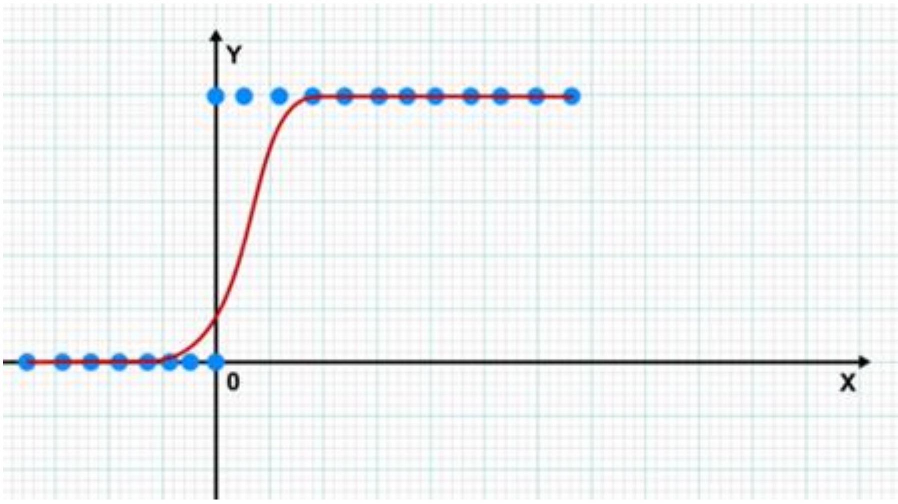


We see that the linear regression we get is a line.

However, if we regress 'a set' of points (more than one) where each of them is taking either output zero or output one, we get information as to what is happening in only two points, and the regression line moves past both zero and one.



Logistic regression helps to show a smoother switch between outputs by modelling more of the data points in the new function.



Adjusting parameters of logistic functions

We can measure the probability that the output of a logistic function is going to be 'one' and not 'zero' given certain inputs, with the following functions.

For functions with one predictor

$$P_r(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

Parameters

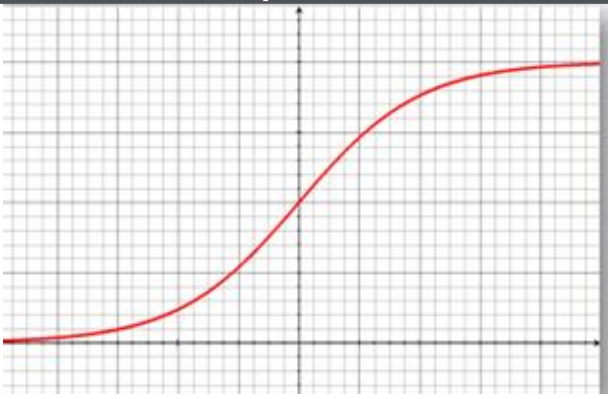
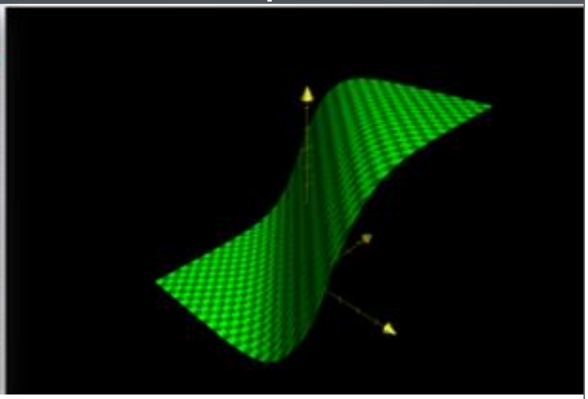
β_0 - In a linear regression, β_0 moves the linear function up and down, while in logistic regression, β_0 moves the function left or right.

β_1 - If you increase the value of β_1, it steepens the curve of the function. A steep curve denotes that you are surer of where the output switches from 0 to 1, or from on to off, etc.

For functions with two predictors

$$P_r(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X_1+\beta_1 X_2}}{1 + e^{\beta_0+\beta_1 X_1+\beta X_2}}$$

When you move from 1 predictor to two predictors, that is, from 1 to 2 dimensions, the logistic function does not change much, however, the switch looks different.

Logistic regression	
One predictor	**Two predictors**
	

For functions with three predictors

$$P_r(Y = 1|X) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3}}{1 + e^{\beta_0+\beta_1X_1+\beta_1X_2+\beta_3X_3}}$$

For functions with predictor interactions

$$P_r(Y = 1|X) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\beta_{12}X_1X_2}}{1 + e^{\beta_0+\beta_1X_1+\beta_2X_2+\beta_{12}X_1X_2}}$$

Optimising logistic functions

Statistical properties of logistic regression

Inputs

- We can have more than one predictor.
- To avoid complications, however, it is better not to use predictors or interacting terms, unless necessary.

Outputs

- We usually have a binary output, such as ‘yes/no’.
- There are extensions to multiple class classification, but they’re not commonly used.
- For multiple-class classification, discriminate analysis or another tool is recommended.

Parameters

- For a single input, we had β_0 and β_1
- For multiple inputs, we would have multiple parameters
- For good fit and statistical properties, use maximum likelihood to select the parameters

Linear function inside the logistic model

$$\Pr(Y=1|X)=\frac{\boxed{e^{\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3}}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3}}$$

$$\log\left(\frac{\boxed{\Pr(Y=1|X)}}{1-\Pr(Y=1|X)}\right)=\beta_0+\beta_1X$$

The advantages and trade-offs of logistic regression

Advantages

- It's relatively simple.
- According to Occam’s razor, we must use simpler things, if possible.
- It’s a generalised linear model.
- Linear models are easier to use.
- We can test input predictors and check whether they're statistically relevant.

- We can measure the accuracy of the coefficient estimates by computing standard errors.

Nearest neighbour vs logistic regression

Nearest neighbour	Logistic regression
It is non-parametric.	It requires us to do the maximum likelihood step.
It only outputs labels.	It provides greater data accuracy.

Naive Bayes vs logistic regression

Naive Bayes	Logistic regression
It is generative.	It is discriminative.
It assumes features are conditionally independent.	It is better at collinearity because of predictor interactions.