Logistic Regression

Introduction

Logistic regression is a classification algorithm used to assign observations to a discrete set of classes.

Unlike linear regression which outputs continuous number values, logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes.

Comparison to linear regression

LINEAR REGRESSION COULD HELP US PREDICT THE STUDENT'S TEST SCORE ON A SCALE OF 0 - 100. LINEAR REGRESSION PREDICTIONS ARE CONTINUOUS (NUMBERS IN A RANGE).

LOGISTIC REGRESSION COULD HELP USE PREDICT WHETHER THE STUDENT PASSED OR FAILED. LOGISTIC REGRESSION PREDICTIONS ARE DISCRETE (ONLY SPECIFIC VALUES OR CATEGORIES ARE ALLOWED). WE CAN ALSO VIEW PROBABILITY SCORES UNDERLYING THE MODEL'S CLASSIFICATIONS.

Types of logistic regression

- Binary (Pass/Fail)
- Multi (Cats, Dogs, Sheep)

Binary logistic regression

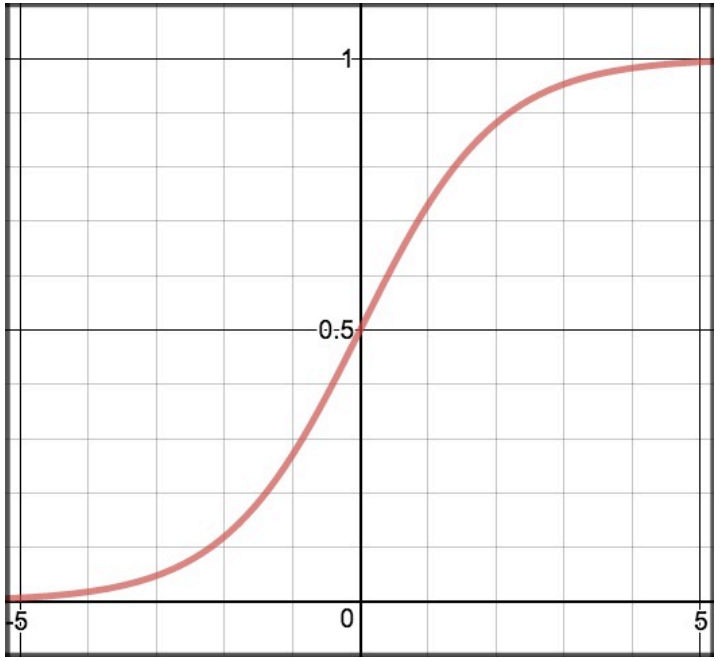
- Say we're given data on student exam results and our goal is to predict whether a student will pass or fail based on number of hours slept and hours spent studying. We have two features (hours slept, hours studied) and two classes: passed (1) and failed (0).

Studied	Slept	Passed
4.85	9.63	1
8.62	3.23	0
5.43	8.23	1
9.21	6.34	0

Sigmoid activation

$$S(z) = \frac{1}{1 + e^{-z}}$$

- In order to map predicted values to probabilities, we use the sigmoid function. The function maps any real value into another value between 0 and 1. In machine learning, we use sigmoid to map predictions to probabilities.



Decision boundary

- Our current prediction function returns a probability score between 0 and 1. In order to map this to a discrete class (true/false, cat/dog), we select a threshold value or tipping point above which we will classify values into class 1 and below which we classify values into class 2.

$$p \geq 0.5, class=1$$

$$p < 0.5, class=0$$

For example, if our threshold was .5 and our prediction function returned .7, we would classify this observation as positive. If our prediction was .2 we would classify the observation as negative. For logistic regression with multiple classes we could select the class with the highest predicted probability.