Logistic Regression

Supervised- categorical- classification

What is a logistic regression?
A: A generalization of the linear regression model to classification problems.

In linear regression, we used a set of input variables to predict the value of a continuous response variable.

In logistic regression, we use a set of input variables to predict probabilities of class membership.

These probabilities can then mapped to class labels, thus predicting the class for each observation.

When performing linear regression, we use the following function:
$Y = b_0 + b_1 x$

When performing logistic regression, we use the following form:

$Pie = Pr(y=1|z) = e^{b_0+b_1 x} / 1 + e^{b_0+b_1 x}$

In order to interpret the outputs of a logistic function we must understand the difference between probability and odds.

The odds of an event are given by the ratio of the probability of the event by its complement:

Odds = pie/ 1-pie

You're trying to determine whether a customer will convert or not. The customer conversion rat is 33.33% what are the odds that a customer will convert

Odds = .333 / 1-.3333 = .5

Notice if we take the logarithm of the odds, we return a linear equation.

$B_0 + b_1 x$

The logistic function takes on an "S" shape where y is bounded by [0,1]

Changing the b0 value it shifts the function horizontally
Changing the b1 value changes the slope of the curve