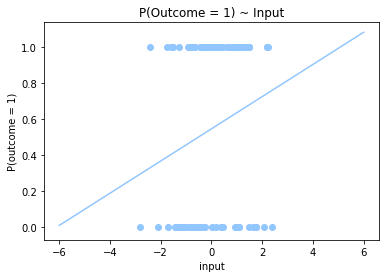
Logistic Regression



Consider the following data above. There is an outcome category 1/0 that you'd like to predict based on some input data. Using some statistical models we can try to predict the probability of outcome = 1, P(outcome = 1).

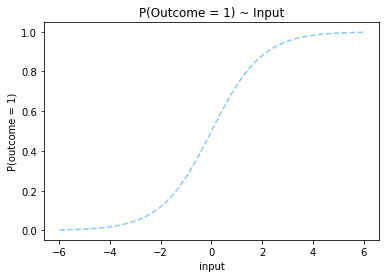
In linear regression you'd try to identify a linear relationship between two continuous variables in order to model the relationship the two variables have with each other. This usually results in a line.

In the situation above, you could try to fit a linear regression model, but you may encounter some issues.



Some of the problems this method has is that the predicted values are not bounded by (0,1). If we're modeling probabilities, predicted probability values less than zero or greater than one do not make sense. Also, truncating these values would not make sense, since a value of 1 or 0 would mean complete certainty (which we do not have).

Therefore, we need to find a statistical model that is naturally bounded between (0,1). Logistic regression provides a solution since it uses the logistic (or sigmoid) distribution to estimate probability values between (0,1)



Fitting this type of statistical model to our data results in a function that looks like

