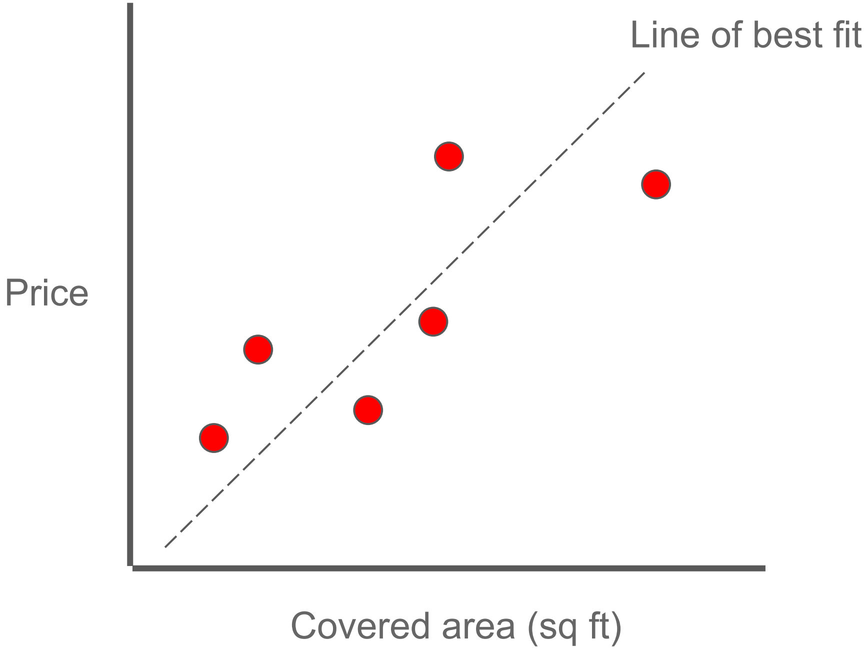
Logistic Regression

Linear Regression

Predict square_foot vs home prices



1. Is there relationship between covered area and home price?

2. How strong is the relationship (R^2 value)

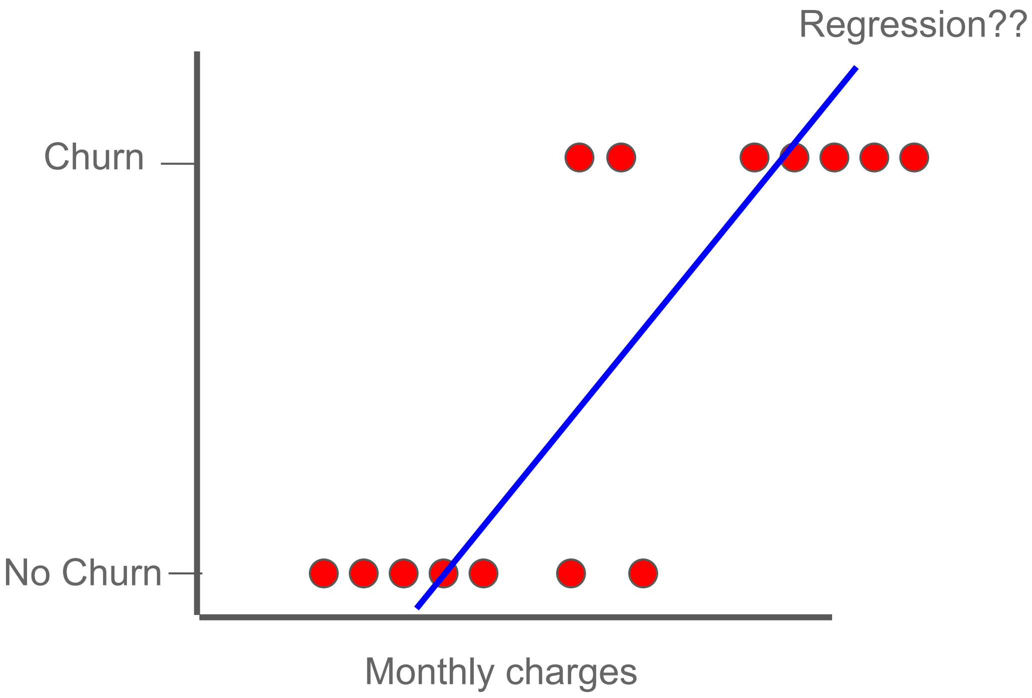
3. Is the relationship significant (p-value)

4. Predict home price for a given covered area

Regression equation: Price = $\beta_0 + \beta_1$ * area

Logistic Regression

Outcome is binary

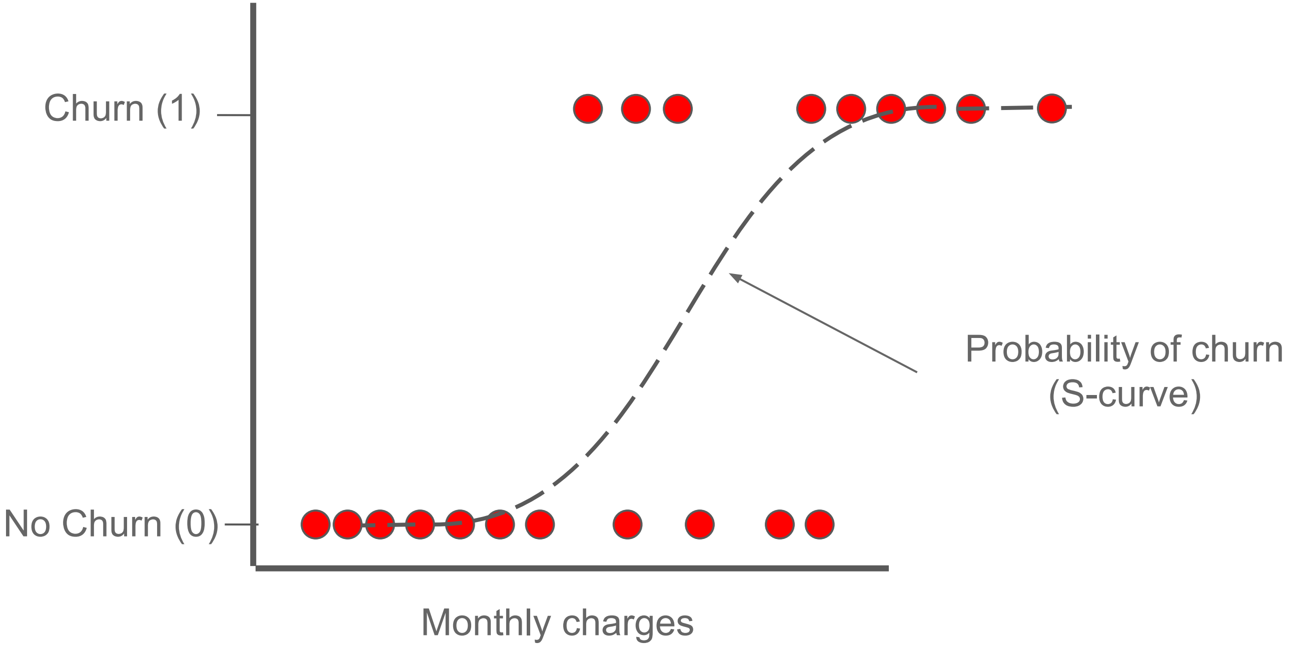


Why not use linear regression?.

1. Binary data does not have normal distribution which is an condition for most types of linear regression
2. Predicted target values can be beyond 0 and 1 in linear regression.
   a. For probability, values outside of 0 and 1 does not make sense.
3. Probabilities are often not linear
   a. Can be ‘U’ shaped i.e. extreme values at the ends.

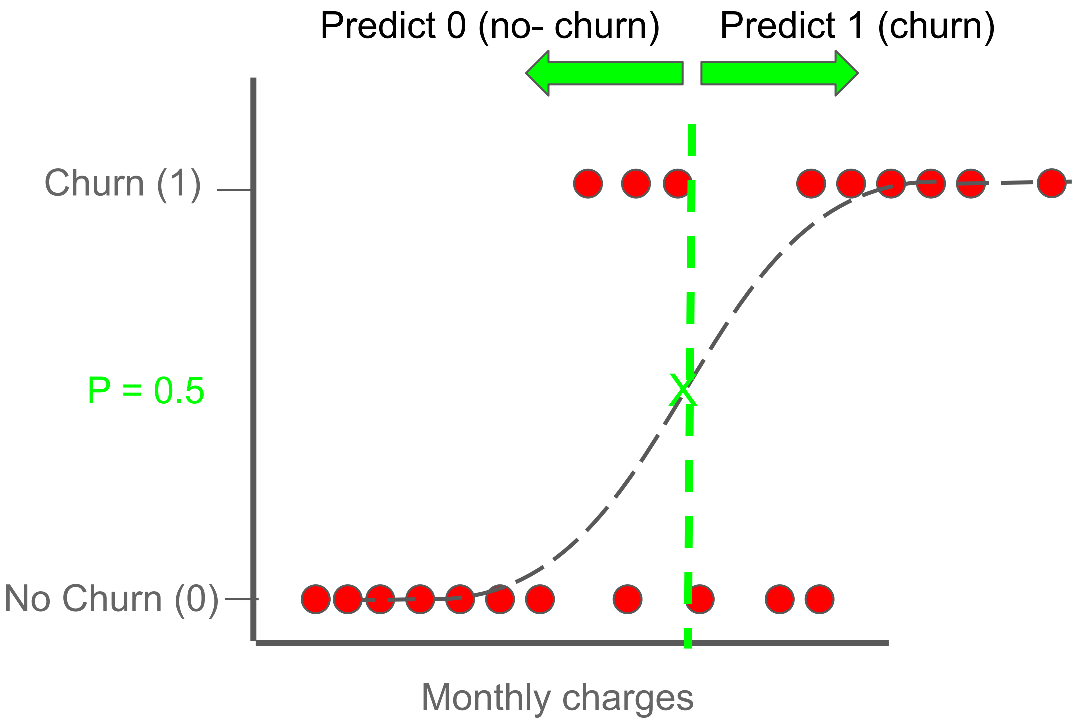
Logistic Regression

Outcome is binary



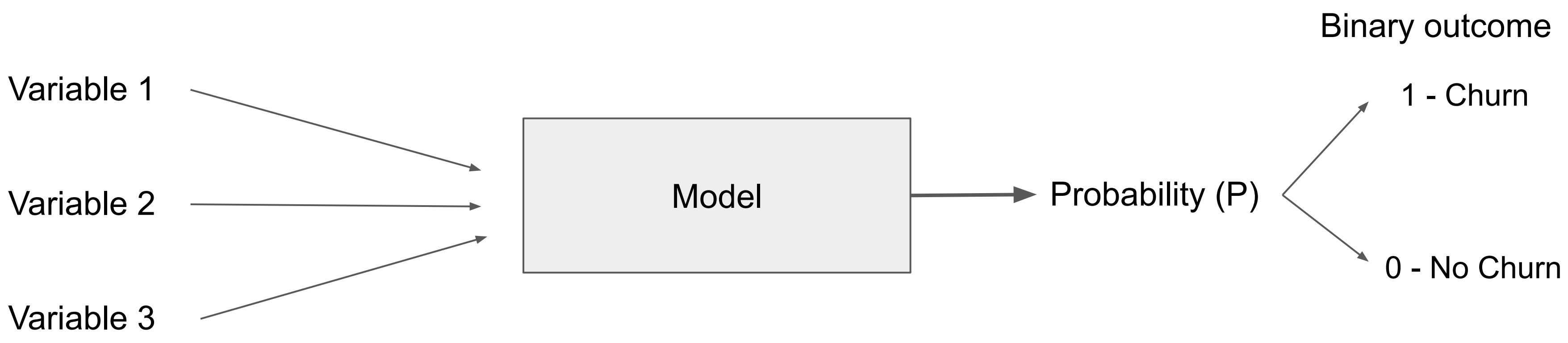
$$p = \frac{1}{1 + e^{-(b_0 + b_1 x)}}$$

Logistic Regression

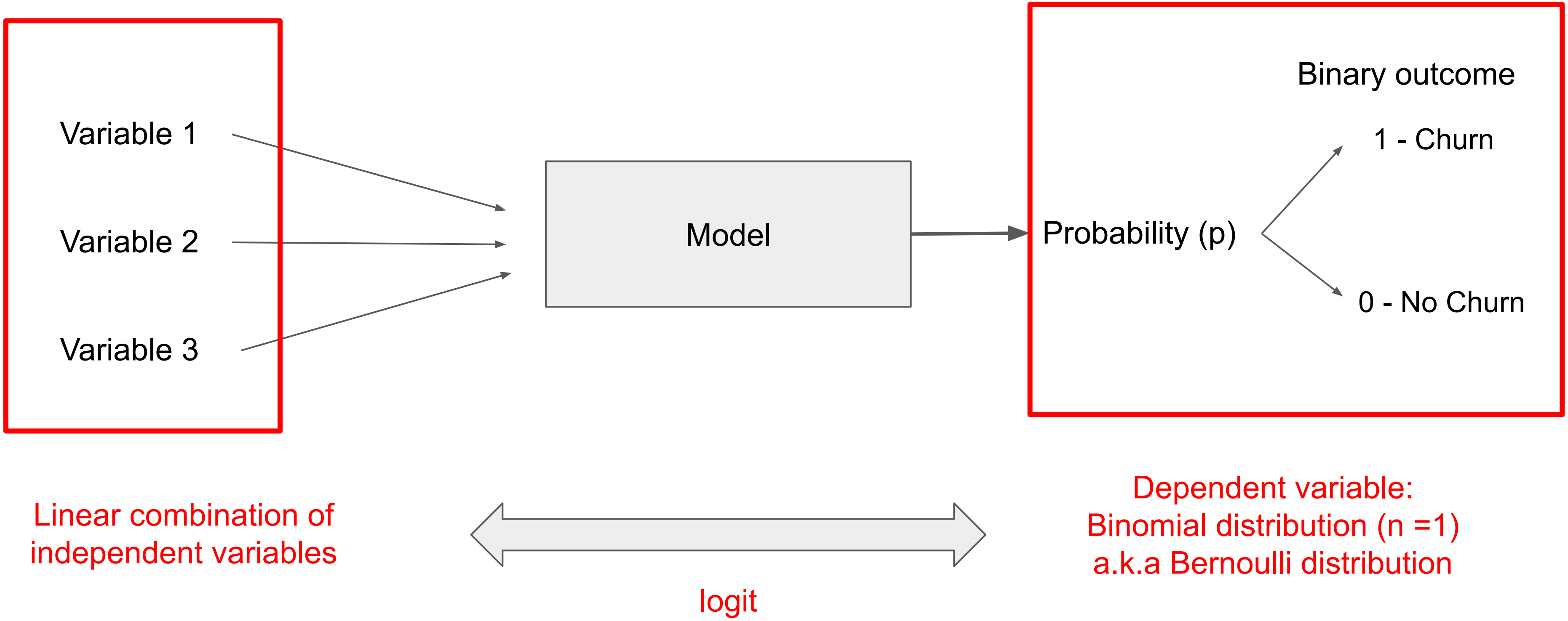


Logistic Regression

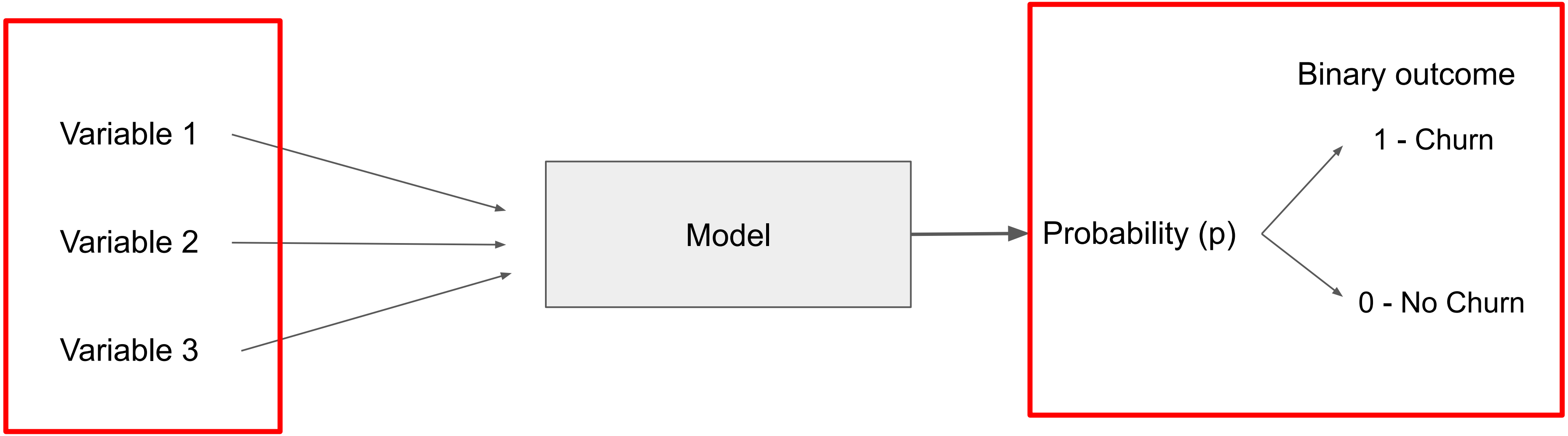
- Model a probability of certain event happening based on linear combination of independent variables



Logistic Regression



Logistic Regression



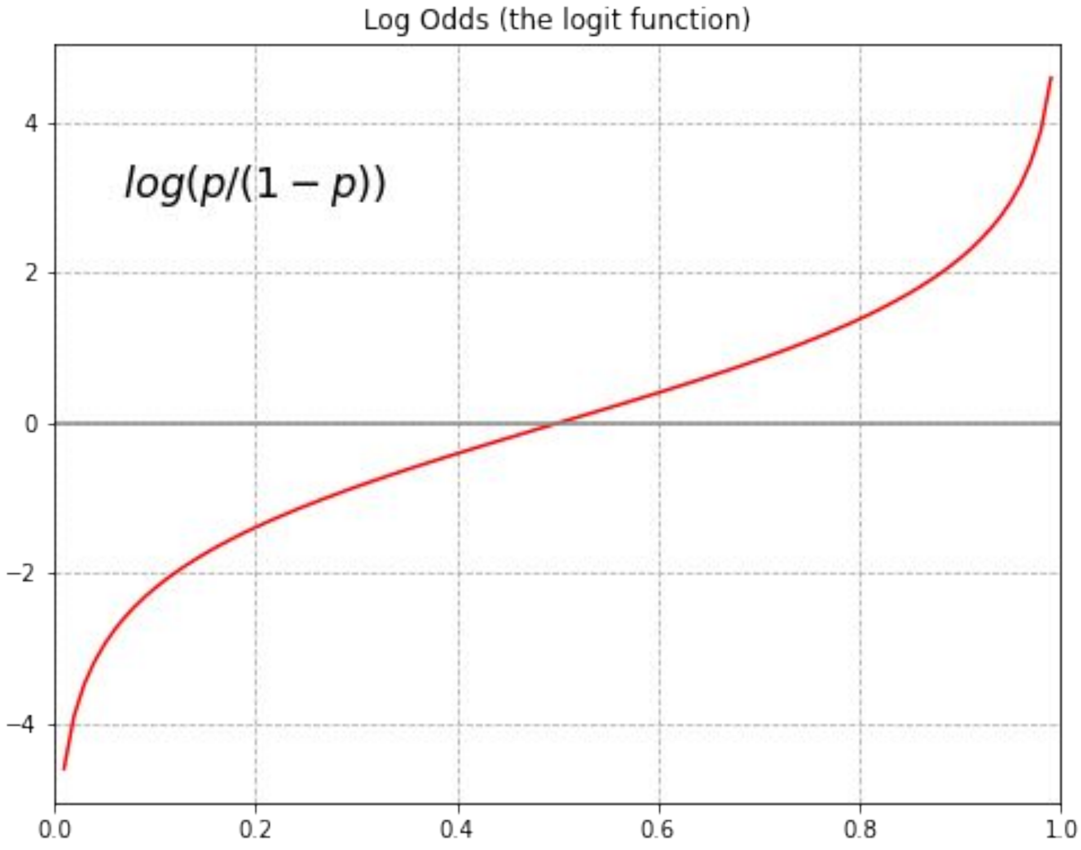
Linear combination of independent variables

logit

$$\text{logit}\,(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1$$

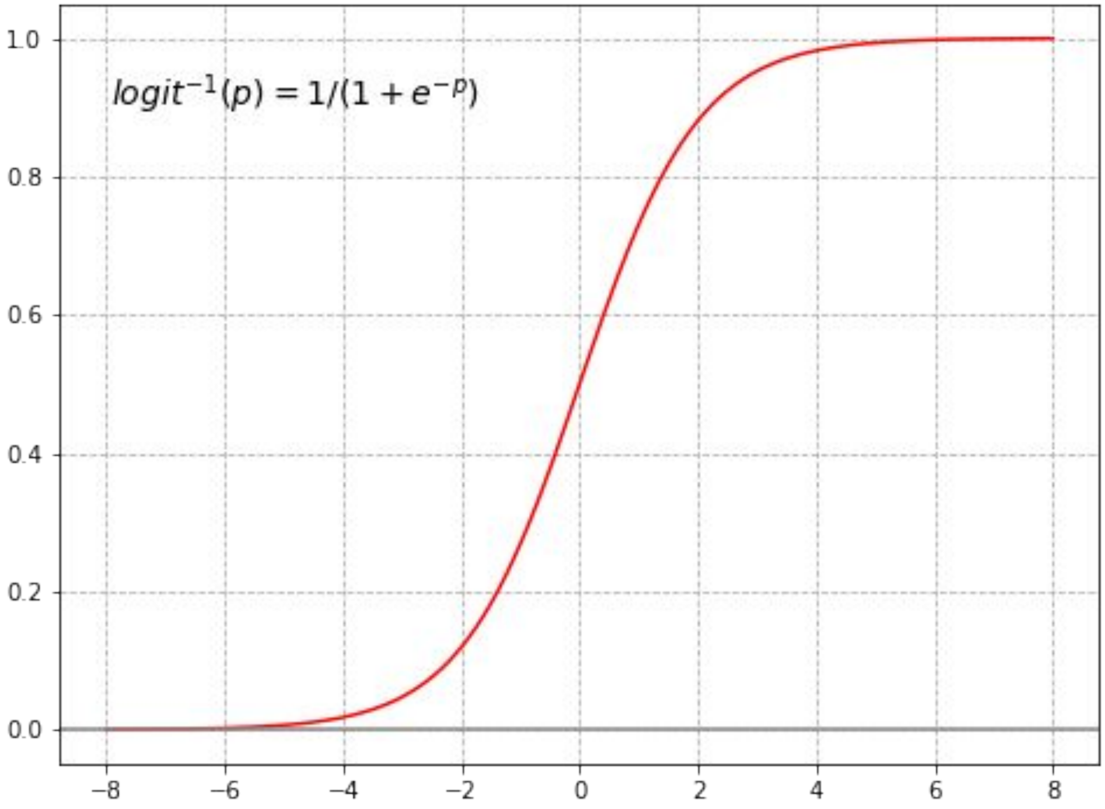
Dependent variable:
Binomial distribution (n =1)
aka. Bernoulli distribution

Logit



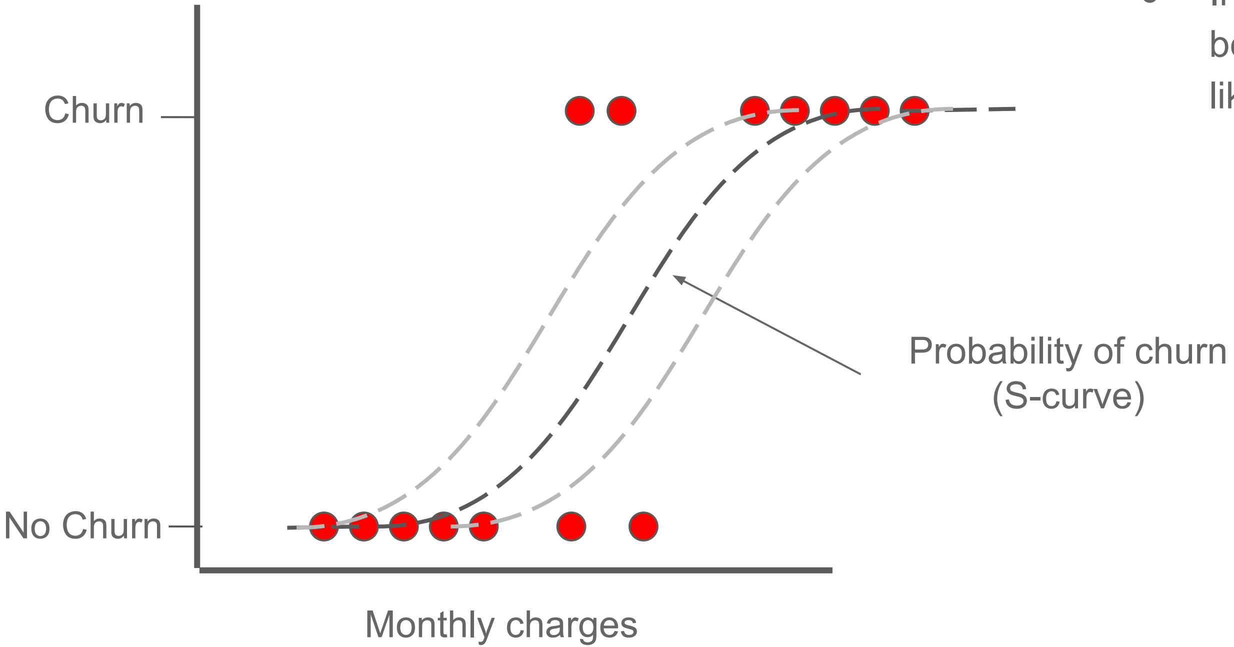
Inverse logit function (Sigmoid function)

Inverse logit gives the probability (on y-axis)



Logistic Regression

Outcome is binary



- In regression we find the line of best fit using ‘least squares’ method
- In logistic regression we find the best curve using MLE (maximum likelihood estimation)

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x)}}$$

Extra: Logit function

Logit = maps the linear combination of independent variables to the bernoulli's probability distribution in domain 0 to 1.

Logit = $\log_e(\text{odds}) = \log_e(p/(1-p))$

odds = P(occurring) / P(not occurring) = p / (1-p)

Toss a fair coin

odds = 0.5 / (1-0.5) = 1 i.e. Odd of landing tails vs heads is 1:1 for fair coin