Logistic Regression

Logistic Regression

- We've explored how to use Linear Regression and its many variations to predict a continuous label.
- But how can we predict a categorical label?

Logistic Regression

- We've explored how to use Linear Regression and its many variations to predict a continuous label.
- But how can we predict a categorical label?
    - Logistic Regression

Logistic Regression

- Logistic Regression
  - Don't be confused by the use of the term "regression" in its name!
  - Logistic Regression is a **classification** algorithm designed to predict **categorical** target labels.

Logistic Regression

- Logistic Regression Section Overview
  - Transforming Linear Regression to Logistic Regression
  - Mathematical Theory behind Logistic Regression
  - Simple Implementation of Logistic Regression for Classification Problem

Logistic Regression

- Logistic Regression Section Overview
  - Interpreting Results
    - Odds Ratio and Coefficients
    - Classification Metrics
      - Accuracy
      - Precision
      - Recall
    - ROC Curves

Logistic Regression

- Logistic Regression Section Overview
  - Multiclass Classification with Logistic Regression
  - Logistic Regression Project
  - Logistic Regression Project Solutions

Logistic Regression

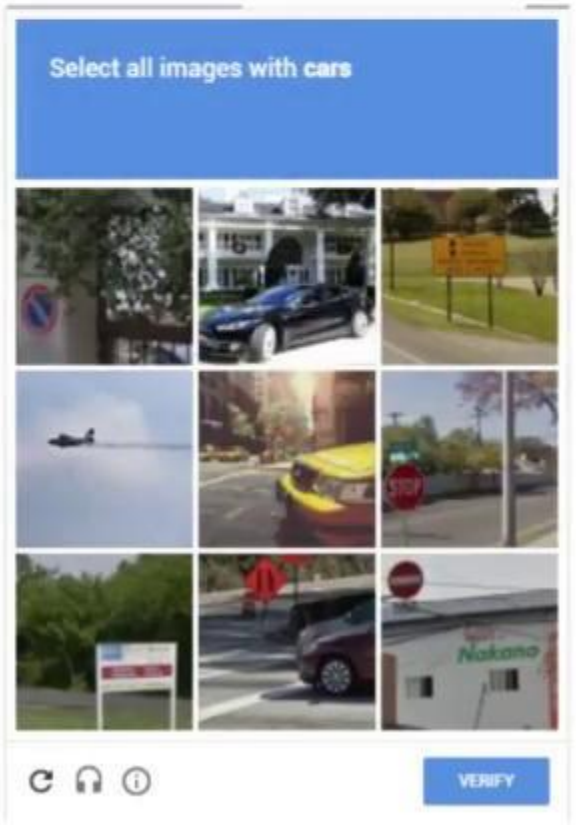
- Logistic Regression will allow us to predict a categorical label based on historical feature data.
- The categorical target column is two or more discrete class labels.

Logistic Regression

- Classification algorithms predict a class or category label:
  - Class 0: Car Image
  - Class 1: Street Image
  - Class 2: Bridge Image

Logistic Regression

- You may not have realized you are helping Google label class data!



Logistic Regression

- Keep in mind, any continuous target can be converted into categories through discretization.
  - Class 0: House Price $0-100k
  - Class 1: House Price $100k-200k
  - Class 2: House Price <$200k

Logistic Regression

- Classification algorithms also often produce a **probability** prediction of belonging to a class:
  - Class 0: 10% Probability
  - Class 1: 85% Probability
  - Class 2: 5% Probability

Logistic Regression

- Classification algorithms also often produce a **probability** prediction of belonging to a class:
  - Class 0: 10% Probability - Car Image
  - Class 1: 85% Probability - Street Image
  - Class 2: 5% Probability - Bridge Image
    - Model reports back prediction of Class 1, image is a street.

Logistic Regression

- Also note our prediction **ŷ** will be a category, meaning we won't be able to calculate a difference based on **y-ŷ.**
  - **Car Image - Street Image** does not make sense.
- We will need to discover a completely different set of error metrics and performance evaluation!

Logistic Regression Theory and Intuition

Part One: The Logistic Function

Logistic Regression

- Logistic Regression works by transforming a Linear Regression into a classification model through the use of the logistic function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- Let's begin by understanding the history and motivation behind the logistic function (a.k.a the sigmoid function):

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- Note:
  - For now, we're only referring to the logistic function itself, not the logistic regression model!

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- 1830-1850: Under guidance of Adolphe Quetelet, Pierre François Verhulst developed the logistic function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

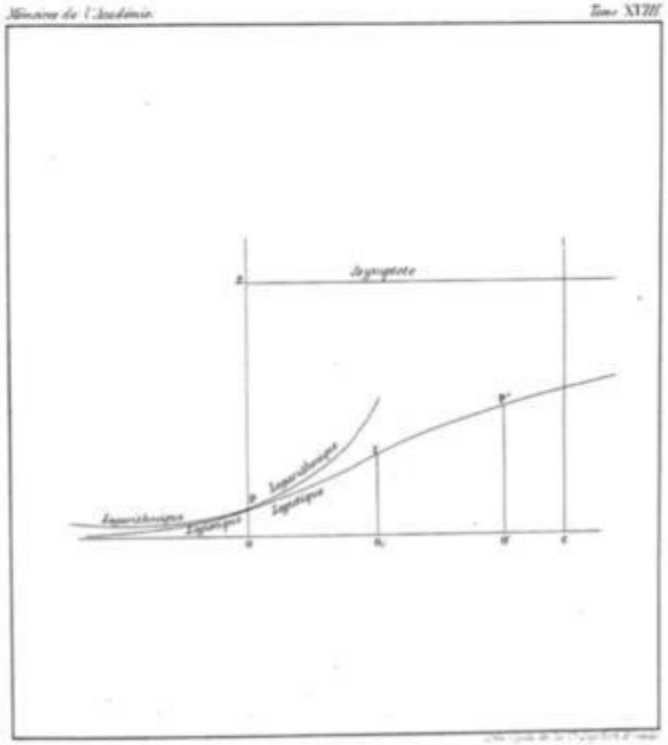
Logistic Regression

- 1883: Logistic function was independently developed in chemistry as a model of autocatalysis by Wilhelm Ostwald.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

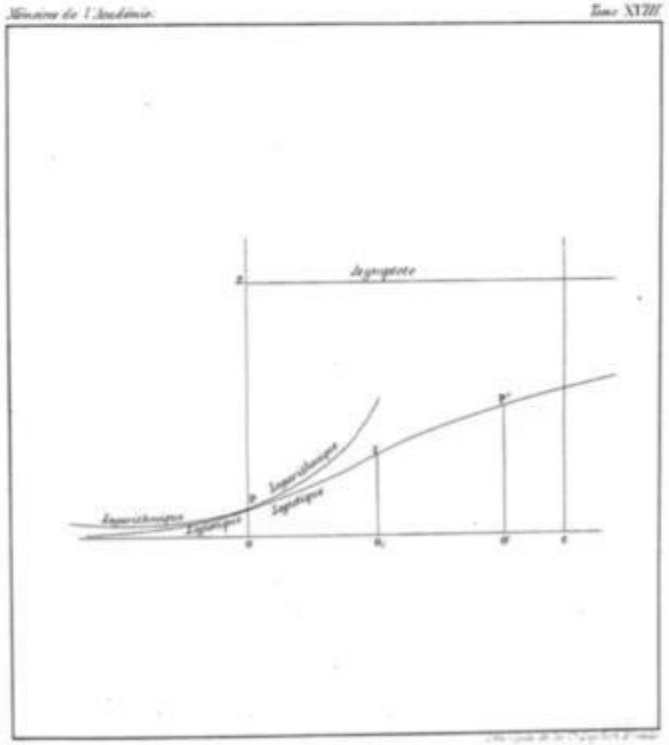
Logistic Regression

- 1830-1850: Developed for the purposes of modeling population growth.



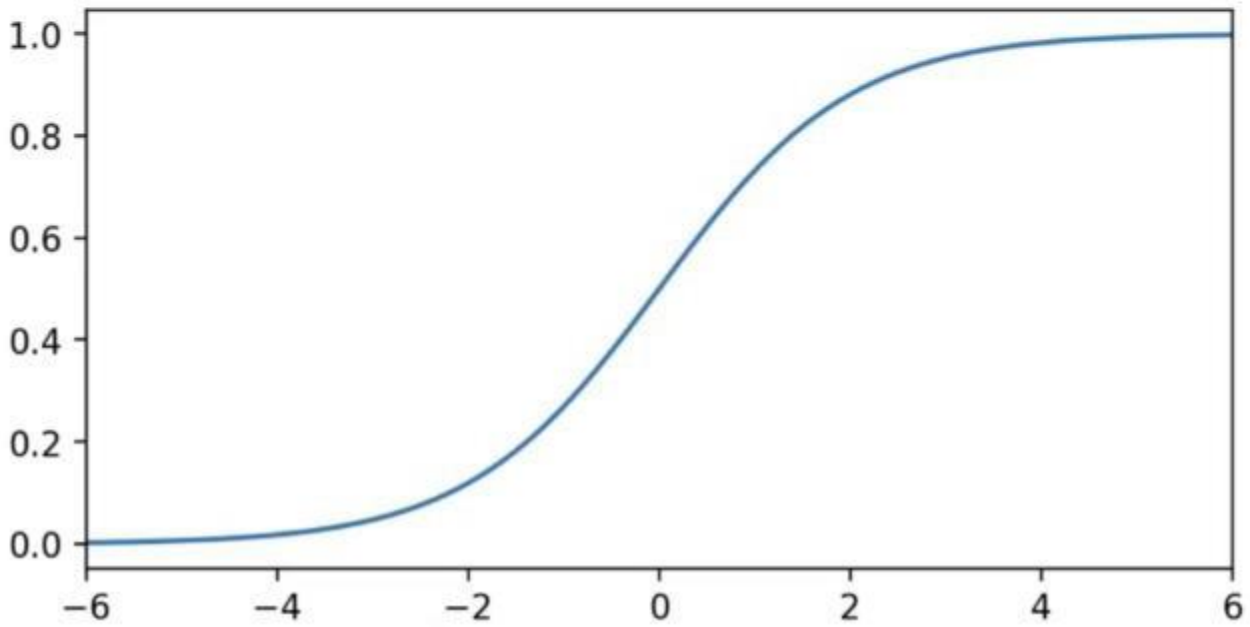
Logistic Regression

- Why the need for a logistic function versus a logarithmic function?



Logistic Regression

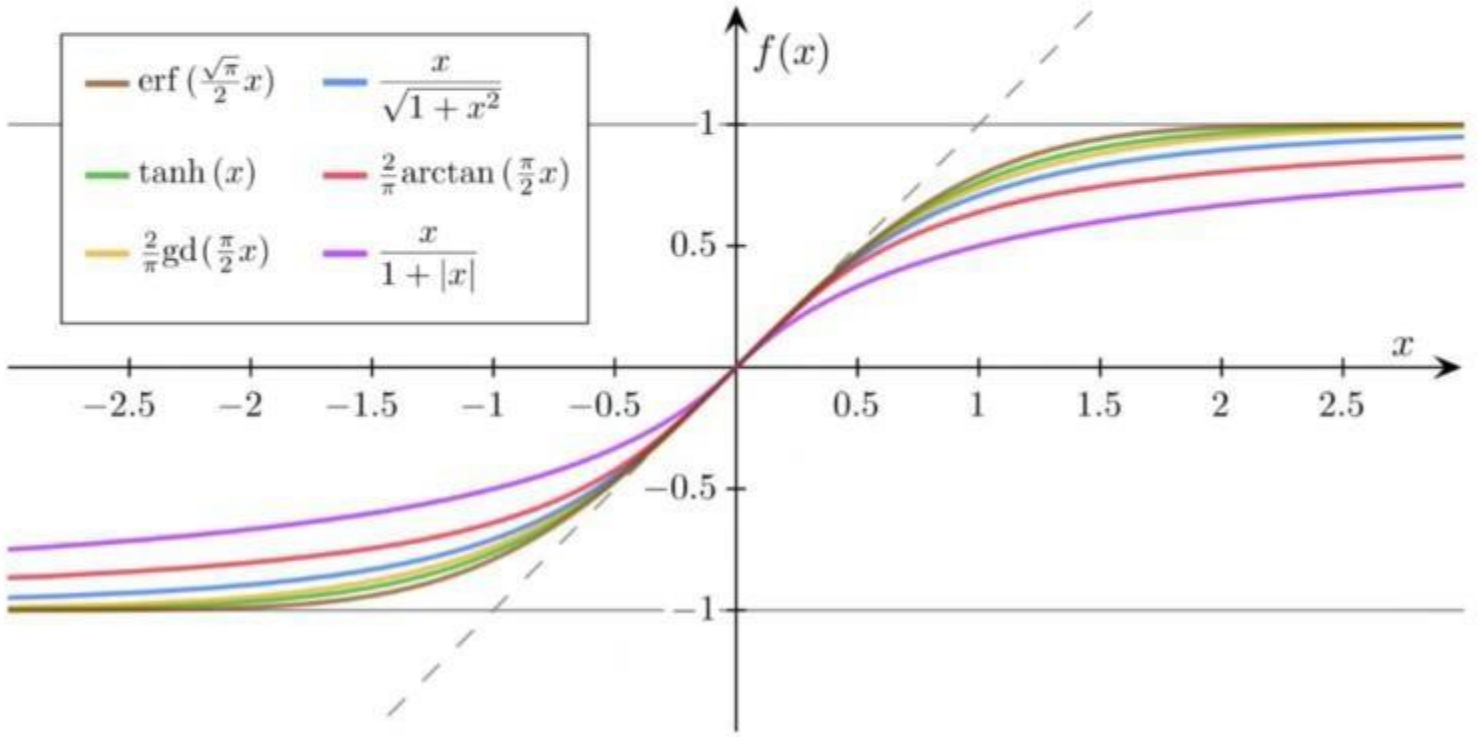
- Why the need for a logistic function versus a logarithmic function?



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

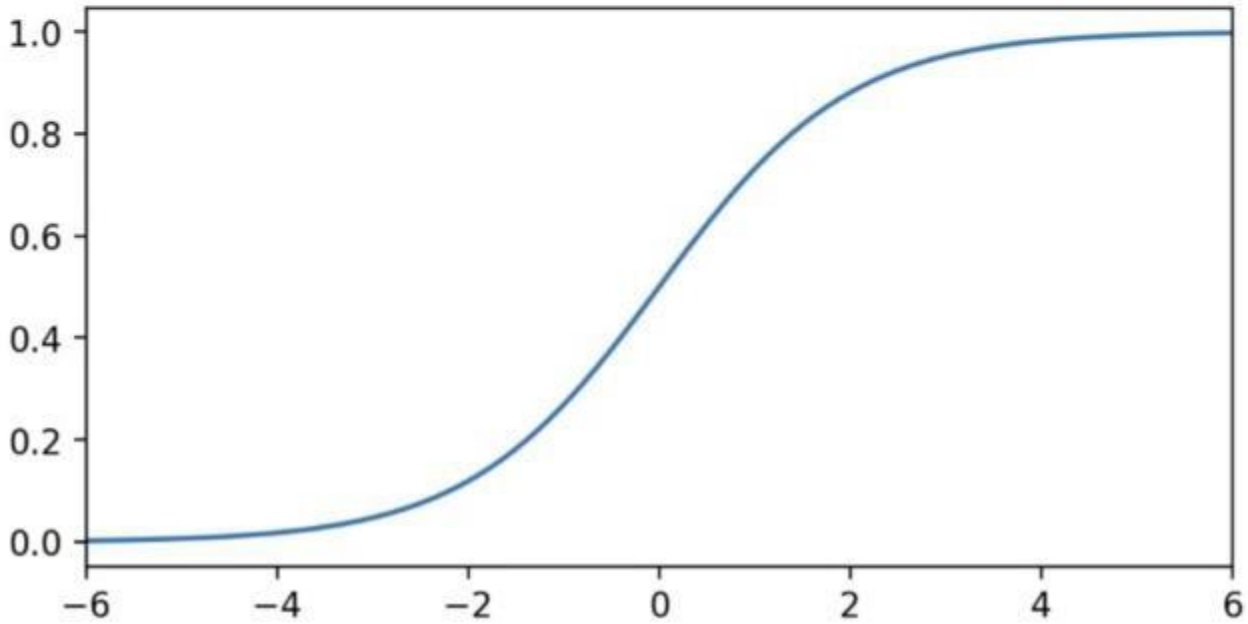
Logistic Regression

- Note: There is a “family” of logistic functions.



Logistic Regression

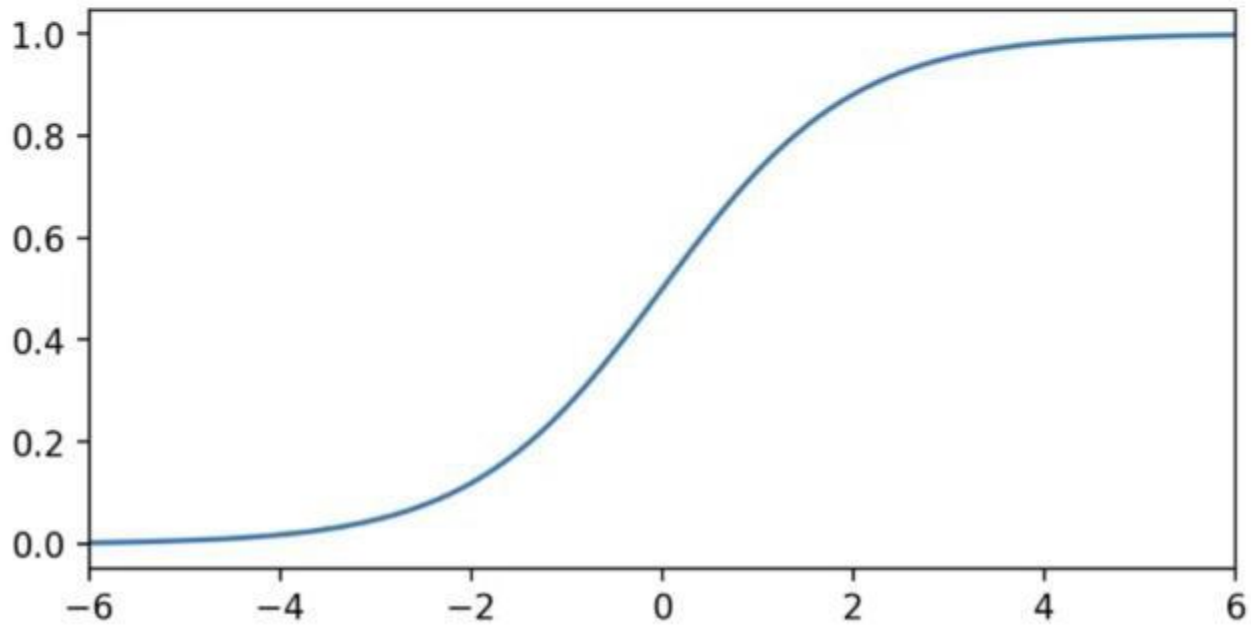
- Notice the “leveling off” behavior of the curve.



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

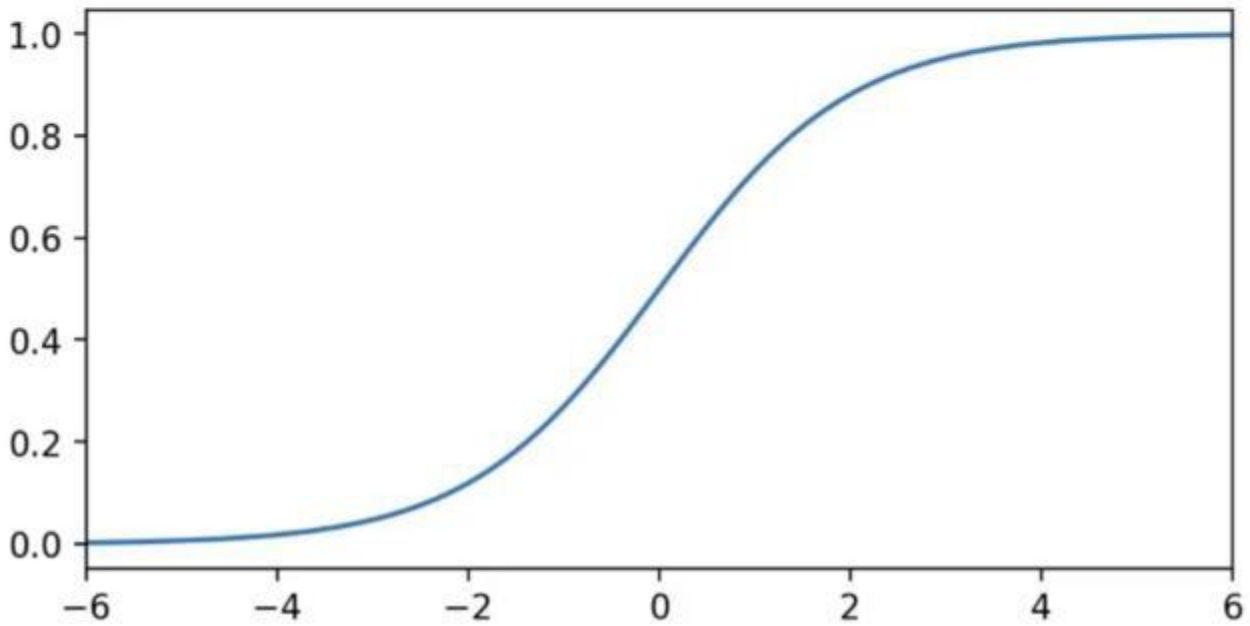
- Also notice **any** value of **x** will have an output range between 0 and 1.



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

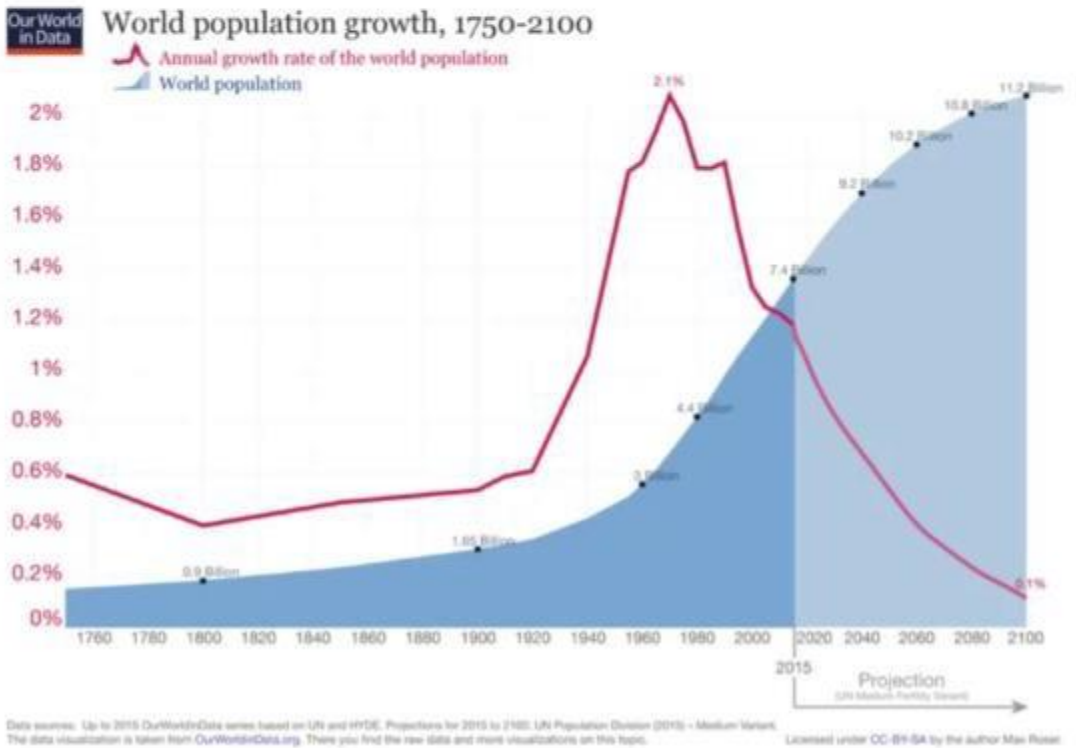
- Many natural real world systems have a “carrying capacity” or a natural limiting factor.



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

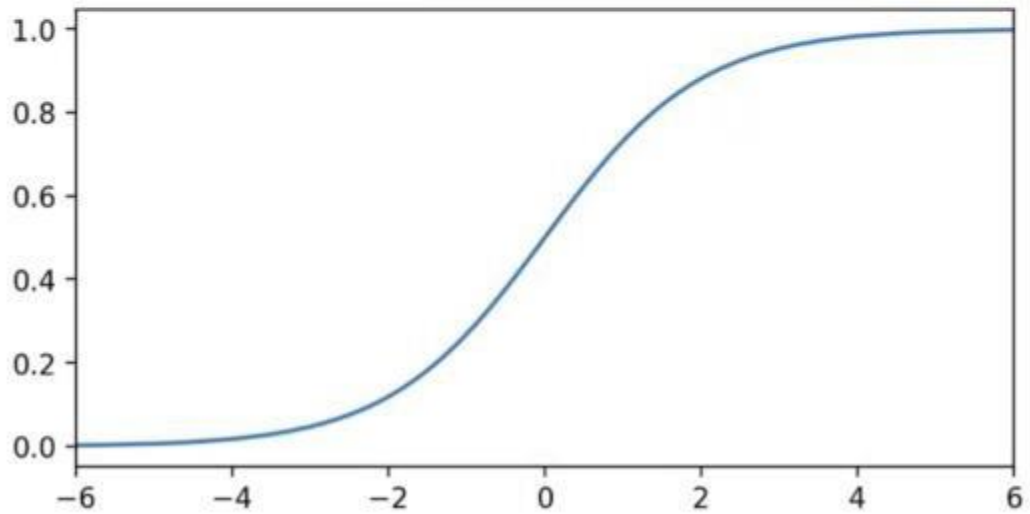
Logistic Regression

- Many natural real world systems have a “carrying capacity” or a natural limiting factor.



Logistic Regression

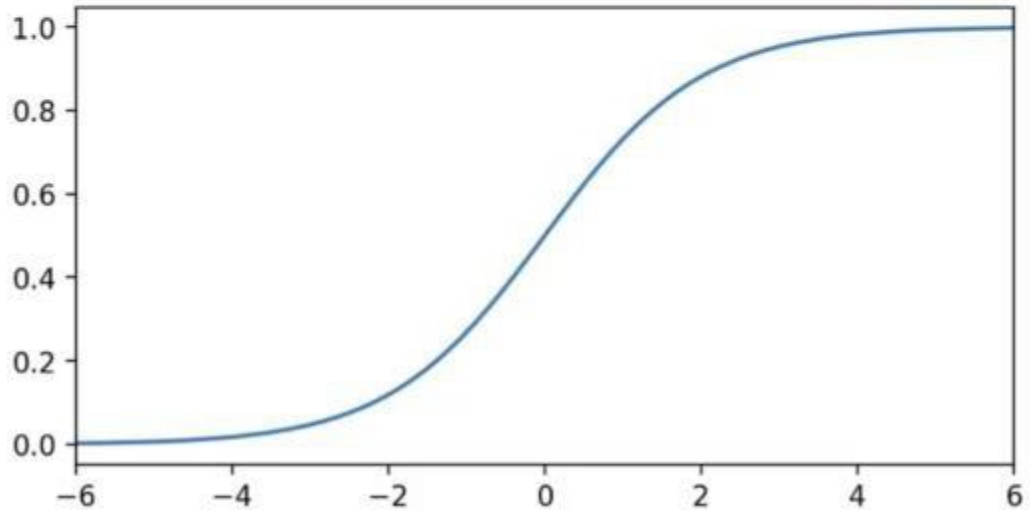
- 1940s: Using the logistic function for statistical modeling was developed by Joseph Berkson.



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- **1944:** *"Application of the logistic function to bio-assay"* **in the Journal of the American Statistical Association**



$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression Theory and Intuition

Part Two:
Linear to Logistic Intuition

Logistic Regression

- Let's explore how to convert a Linear Regression model used for a **regression task** into a Logistic Regression model used for a **classification task.**
- Imagine a dataset with a single feature (previous year's income) and a single target label (loan default)

Logistic Regression

- Our data set:

Income	Loan Paid
-5	0
-4	0
-2	0
-1	0
0	0
2	1
3	1
4	1
5	1

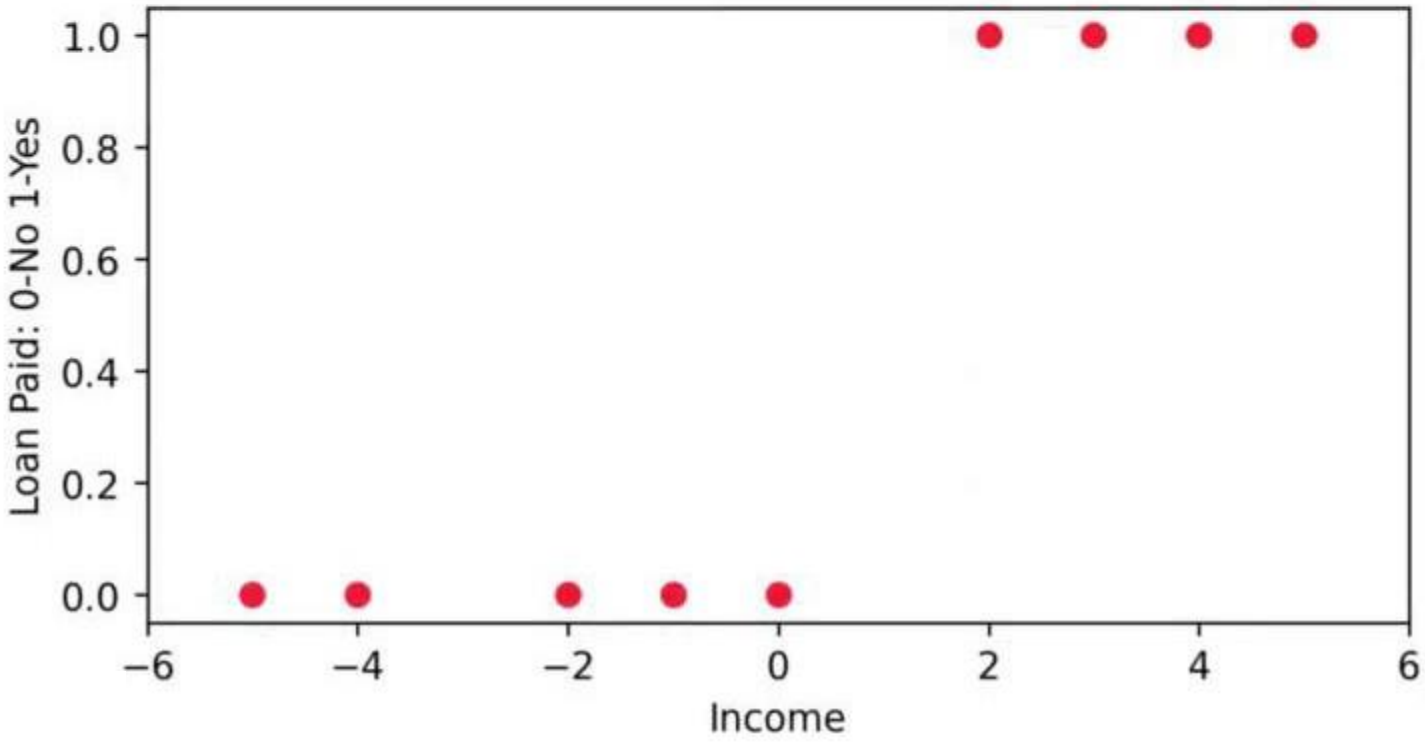
Logistic Regression

- Our data set:

Income	Loan Paid
-5	0
-4	0
-2	0
-1	0
0	0
2	1
3	1
4	1
5	1

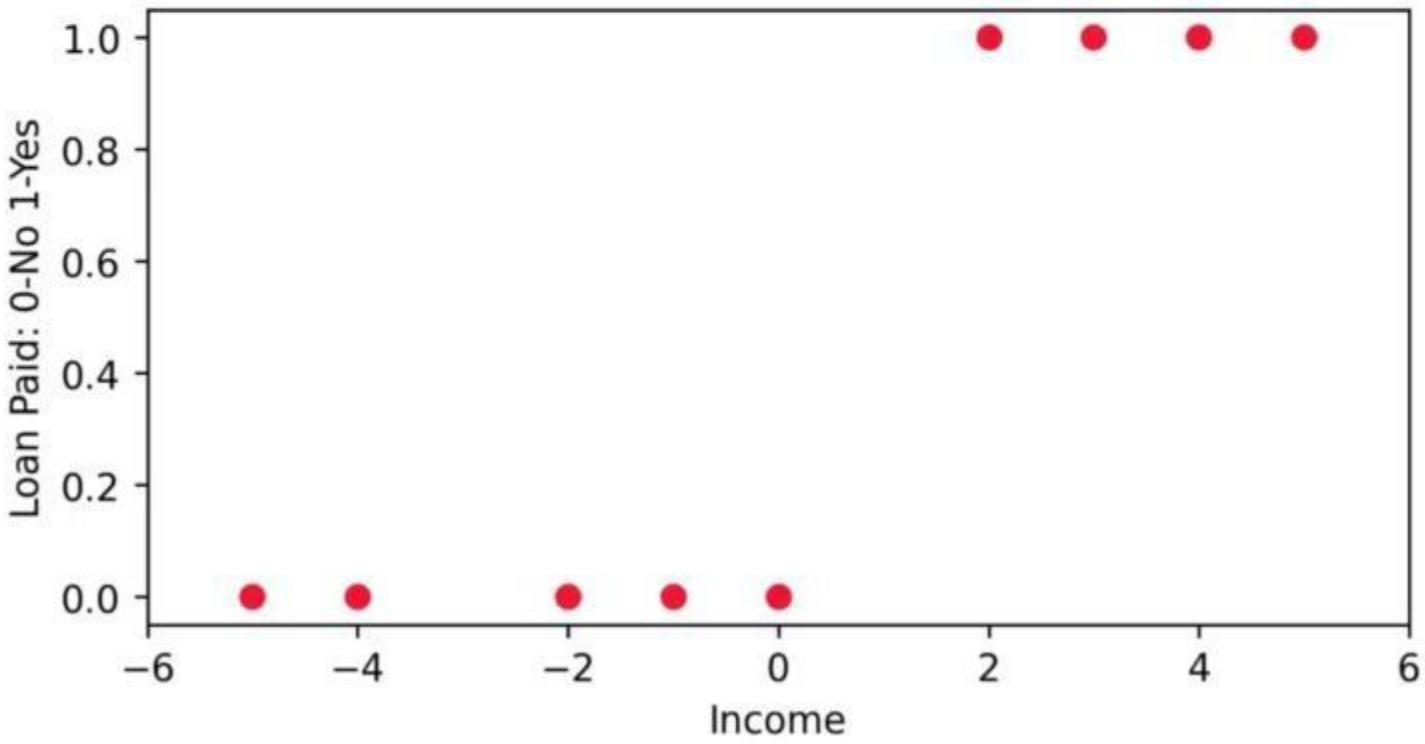
Logistic Regression

- Let's begin by plotting income versus default:



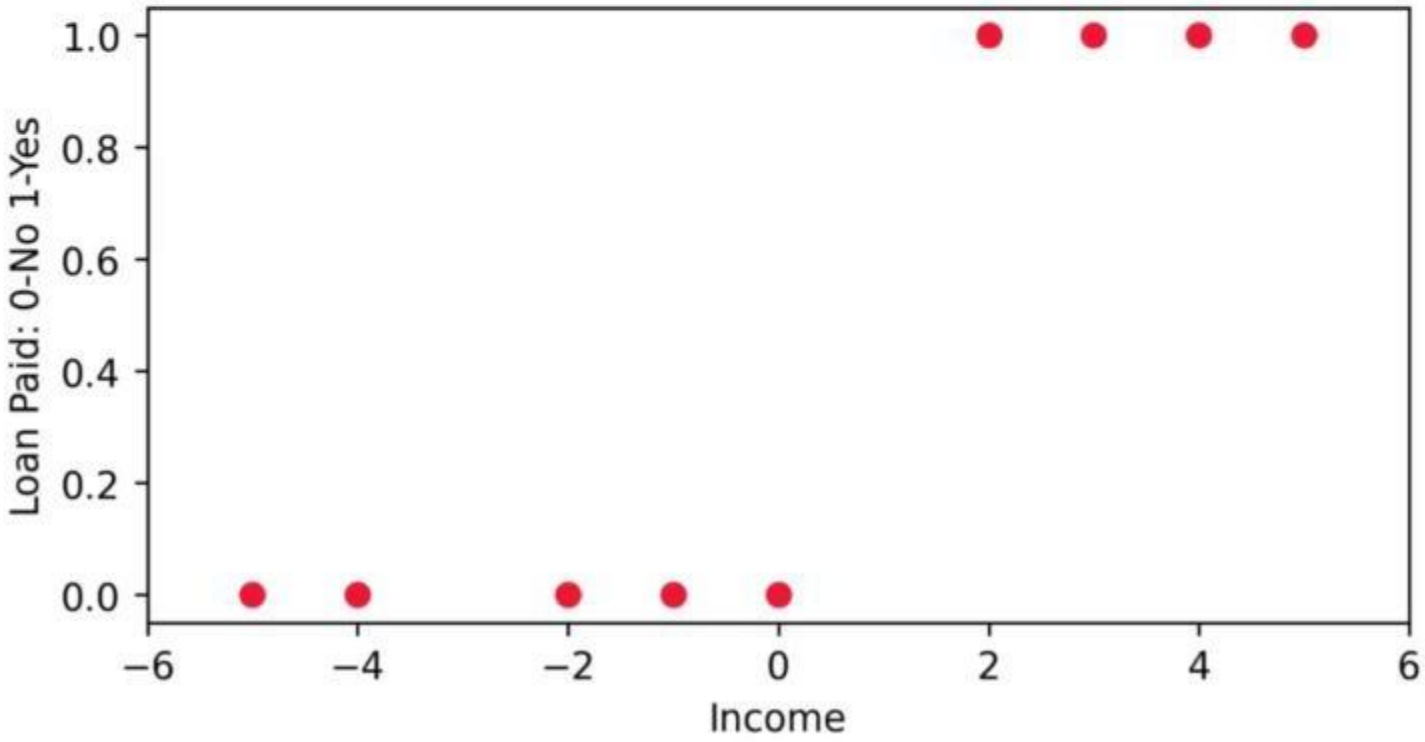
Logistic Regression

- Notice that people with negative income tend to default on their loans.



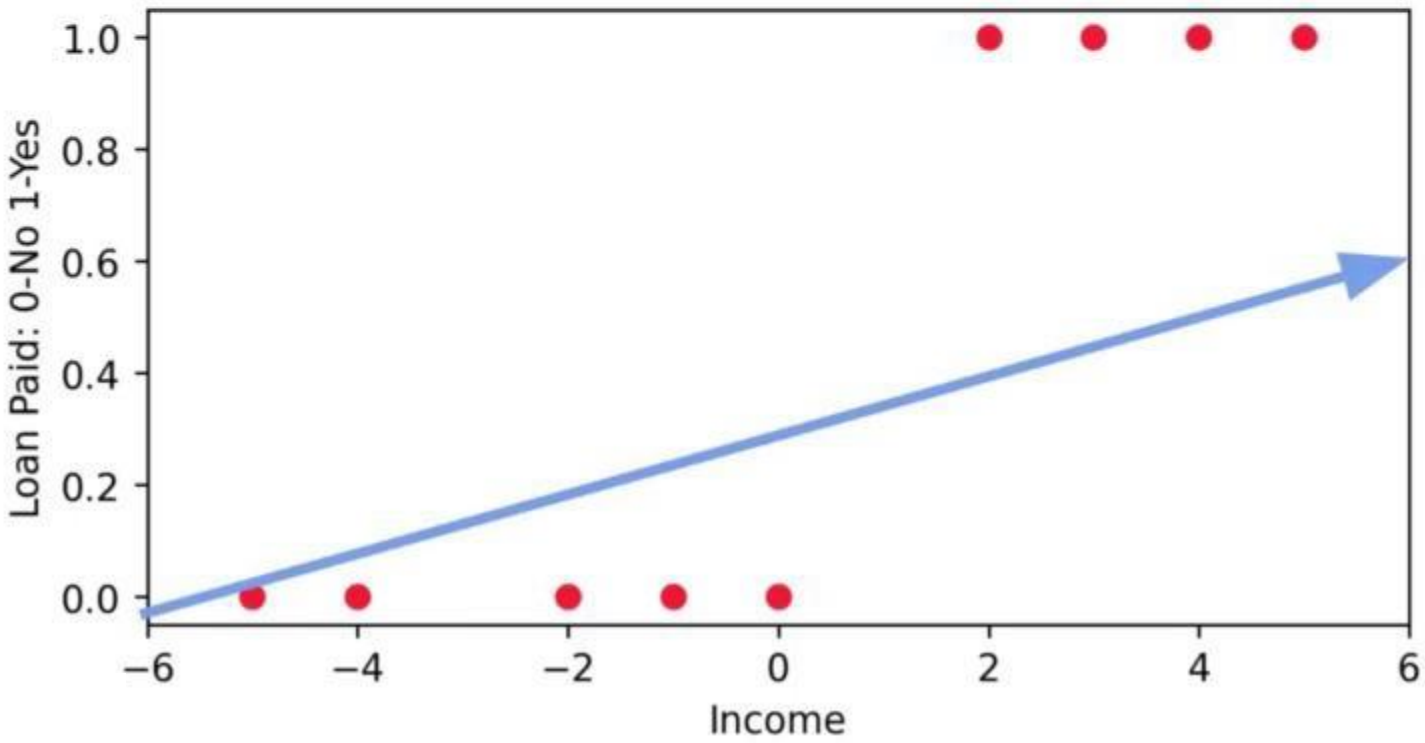
Logistic Regression

- What if we had to predict default status given someone's income?



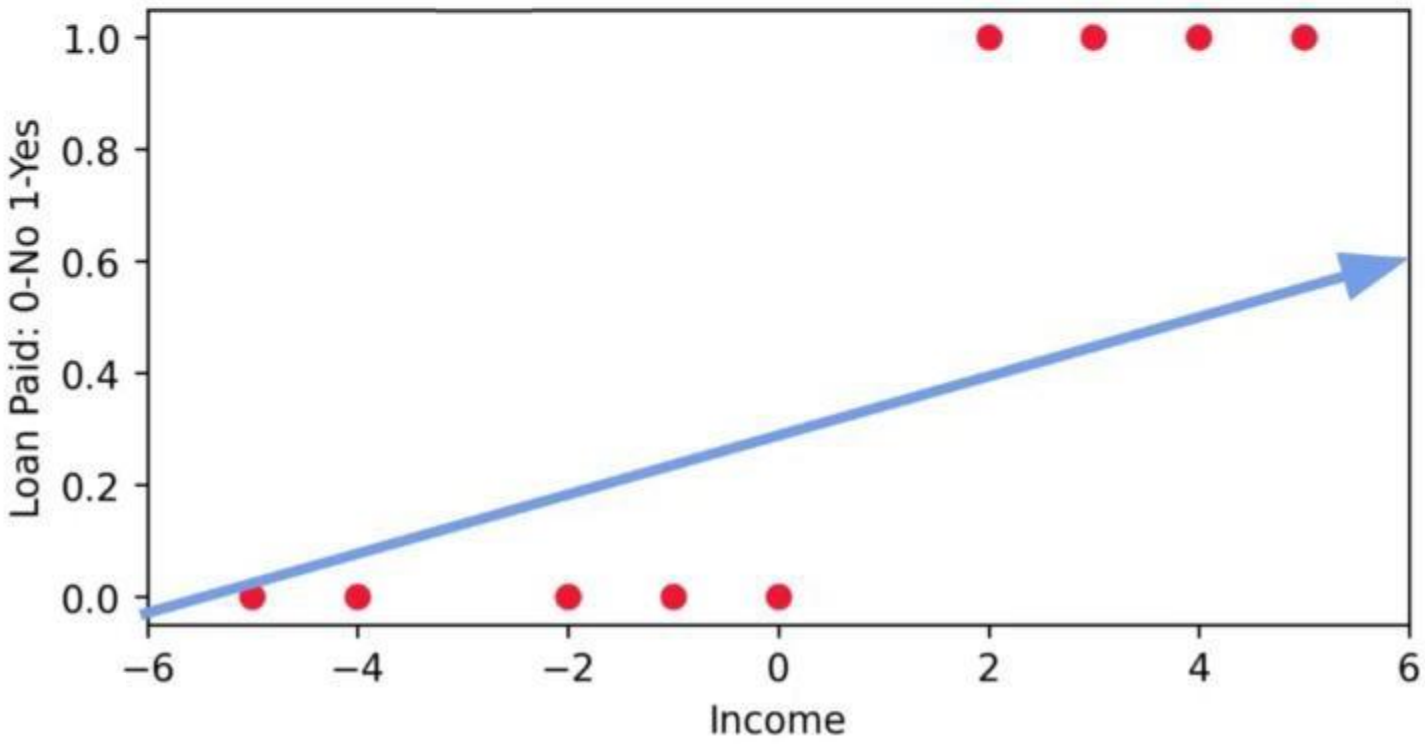
Logistic Regression

- Fitting a Linear Regression would not work (recall Anscombe's quartet):



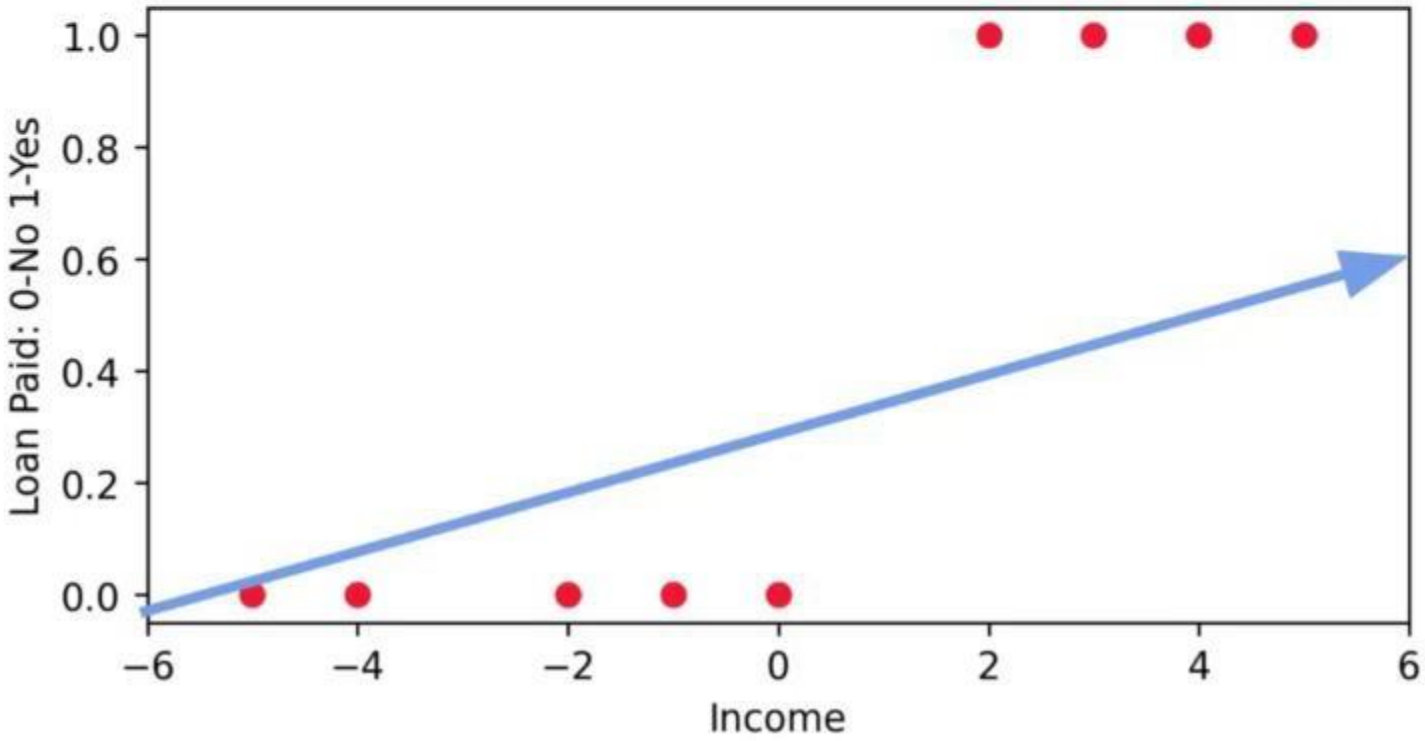
Logistic Regression

- Linear Regression easily distorted by only having 0 and 1 as possible y training values.



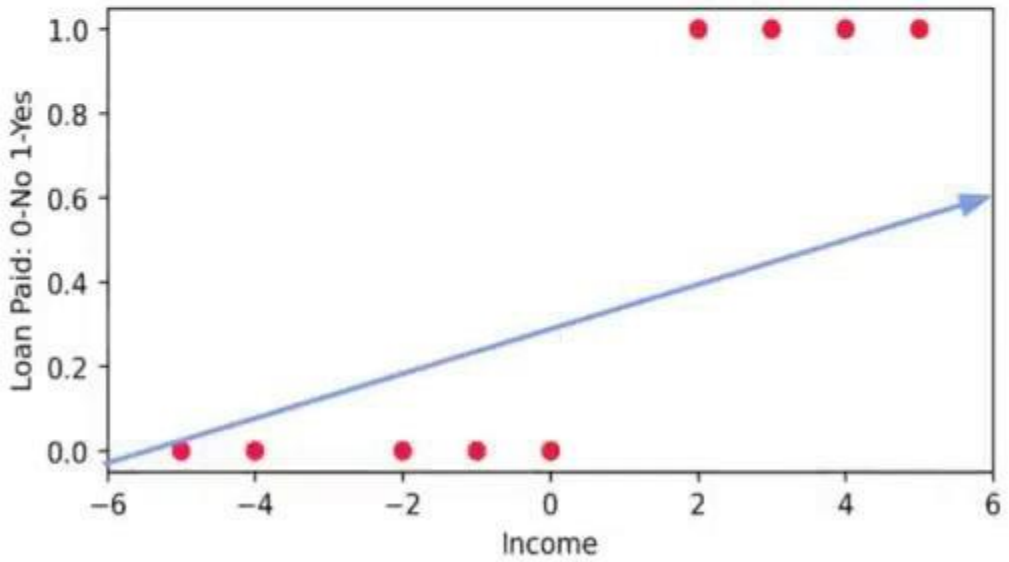
Logistic Regression

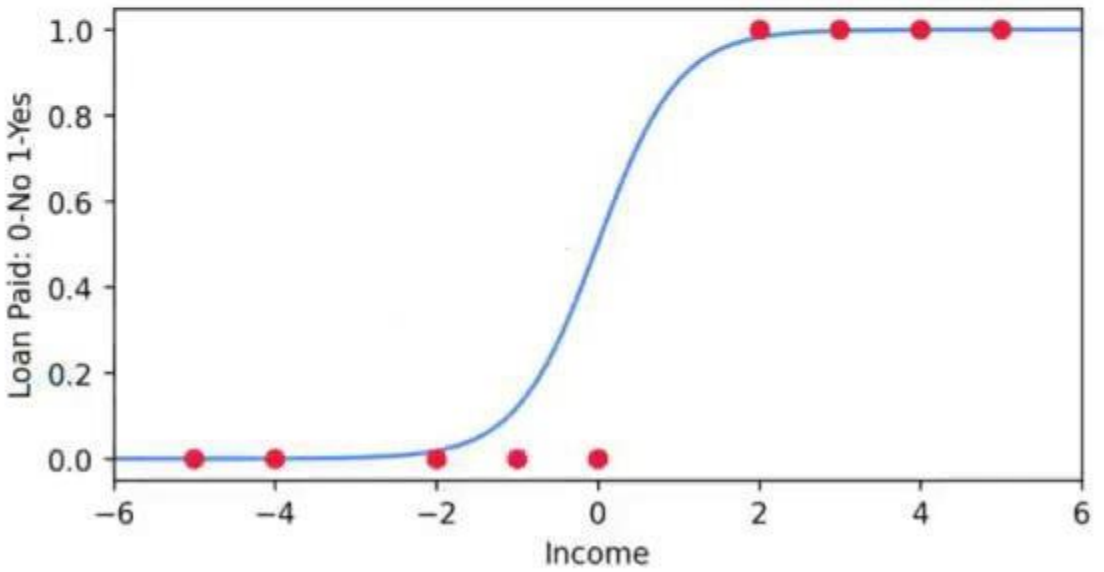
- Also would be unclear how to interpret predicted y values between 0 and 1.



Logistic Regression

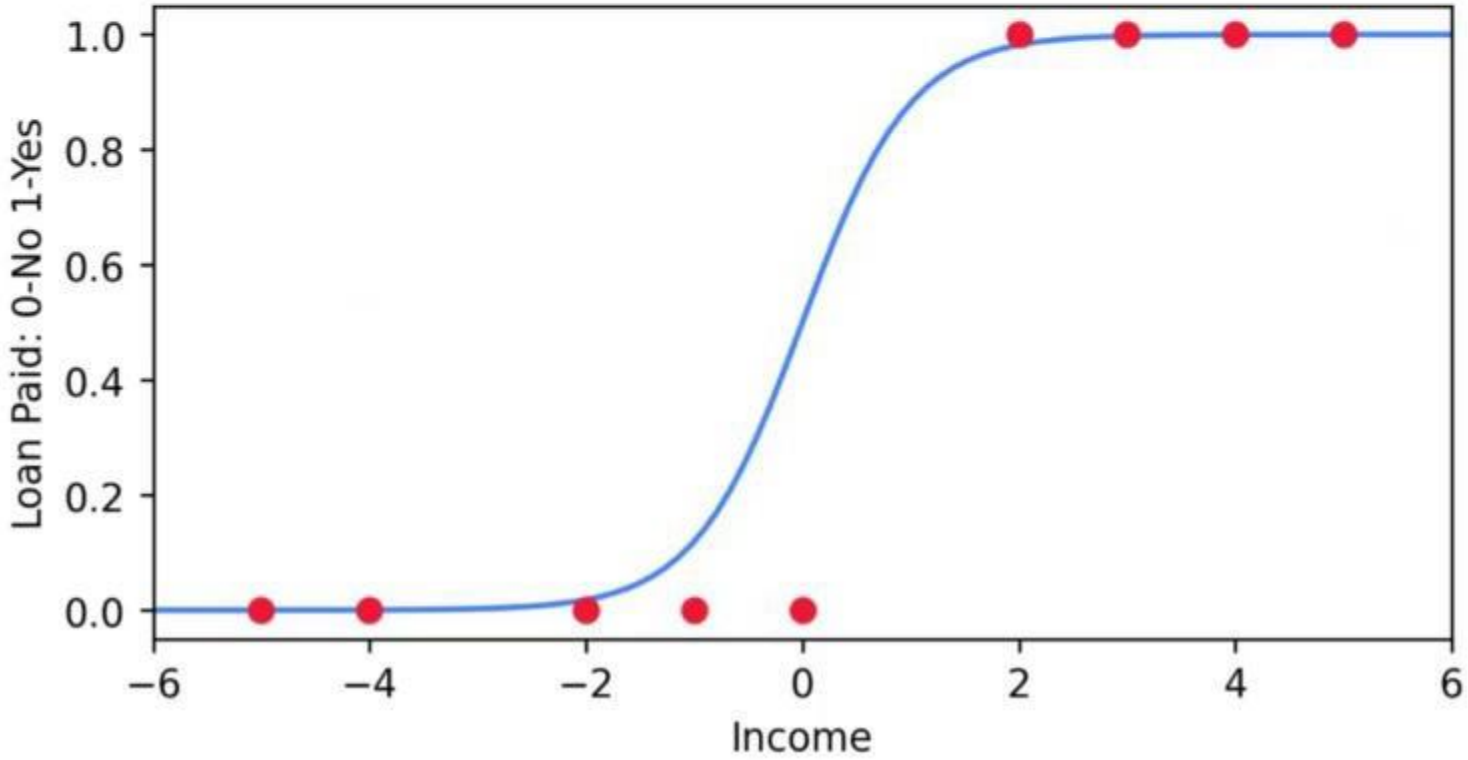
- We could make use of the Logistic Function for a conversion!





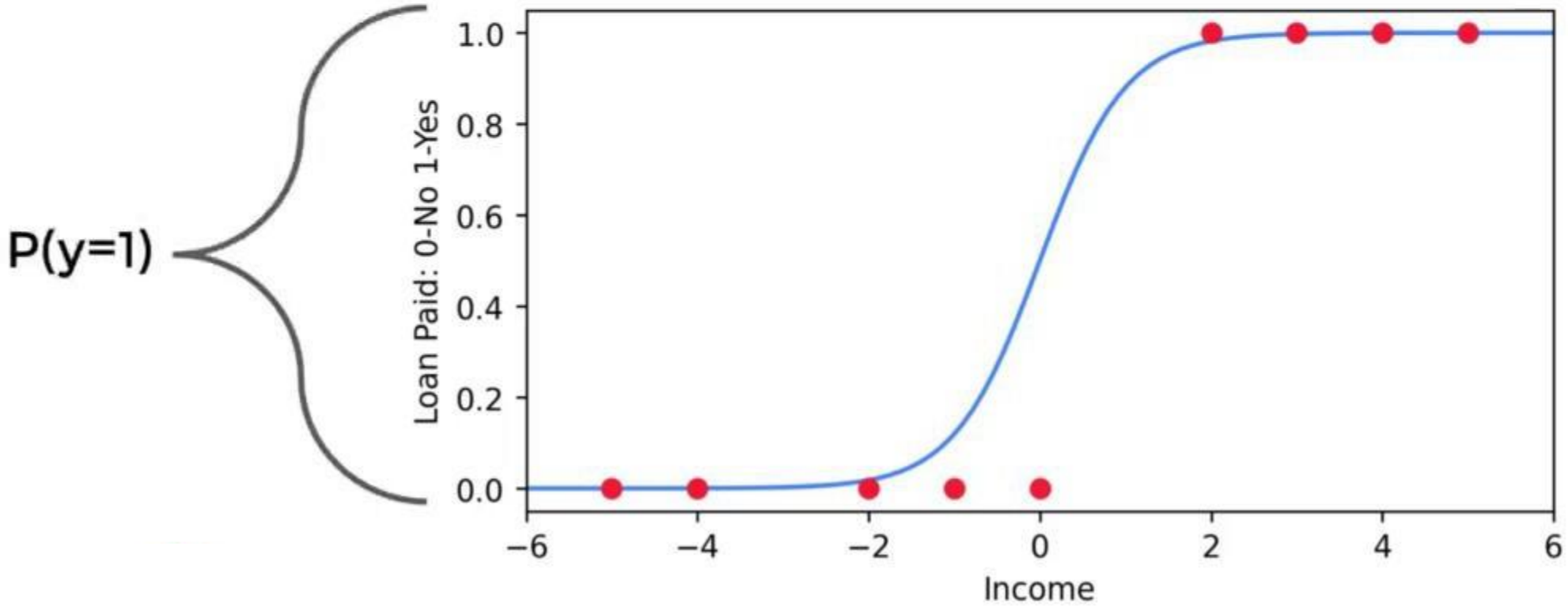
Logistic Regression

- Let's first focus on what this Logistic Regression would look like.



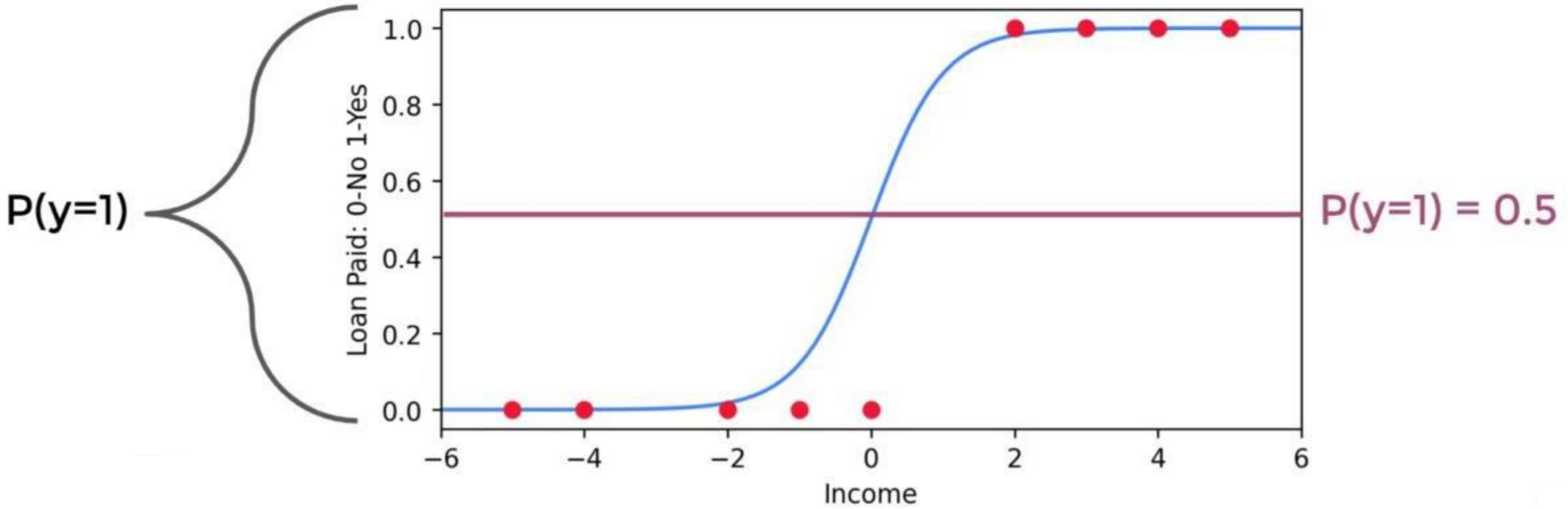
Logistic Regression

- Treat the y-axis as a probability of belonging to a class:



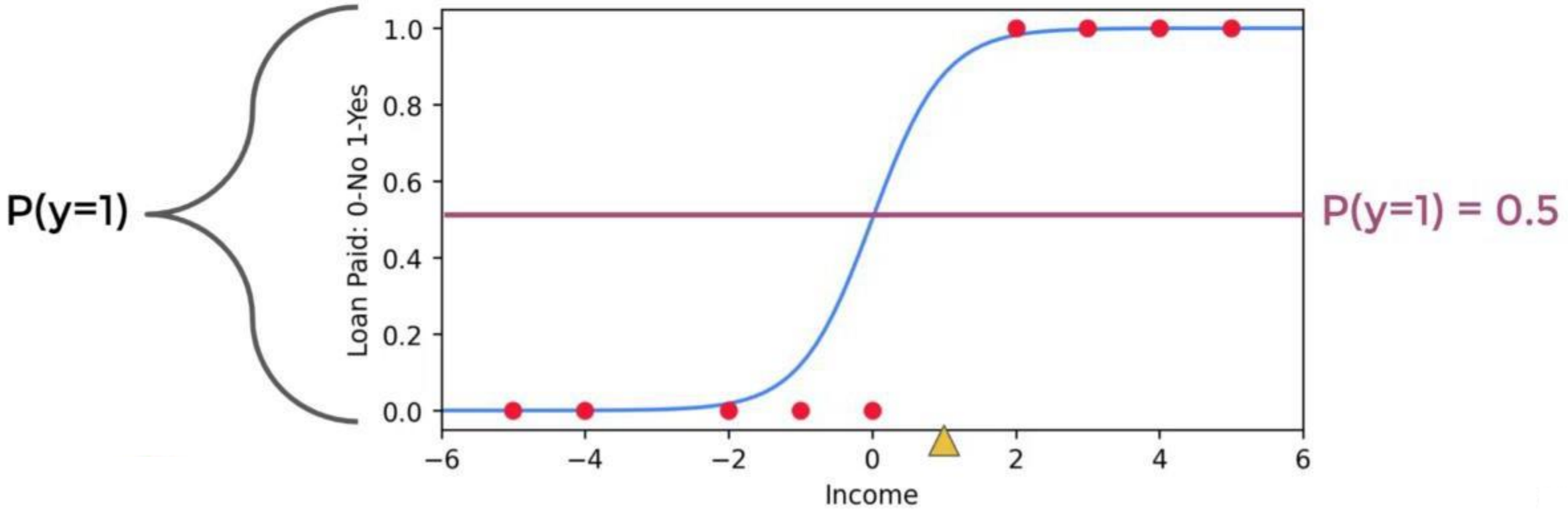
Logistic Regression

- Treating P(y=1) >= 0.5 as a cut-off for classification:



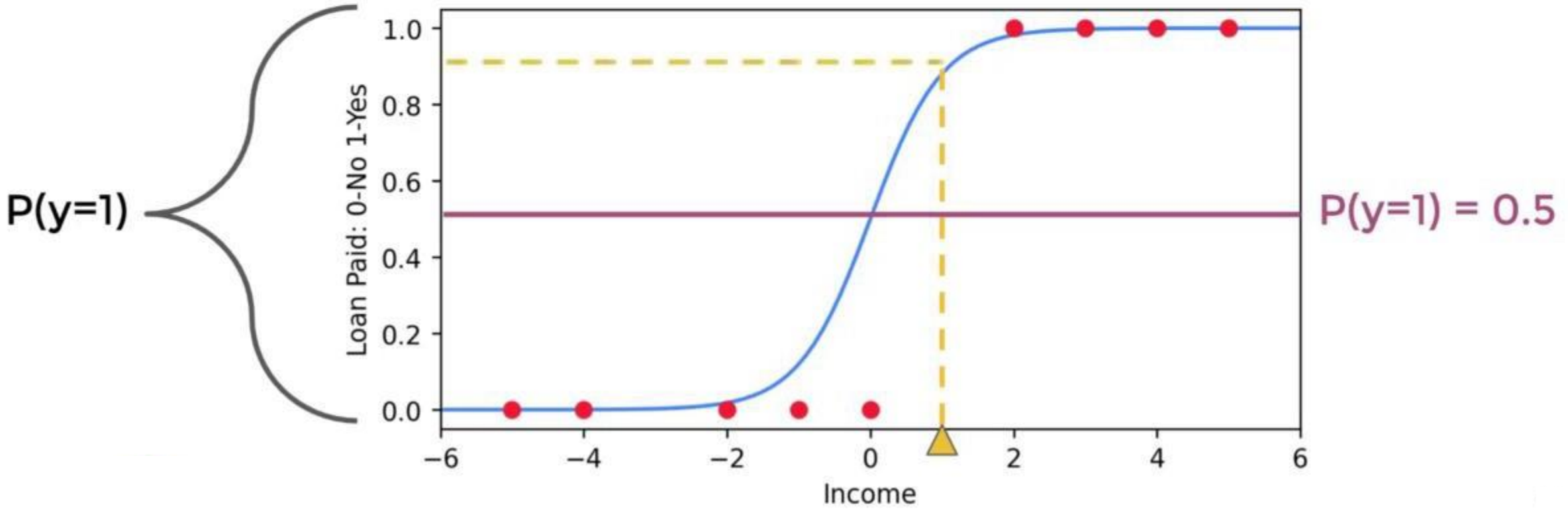
Logistic Regression

- For example, a new person with an income of 1:



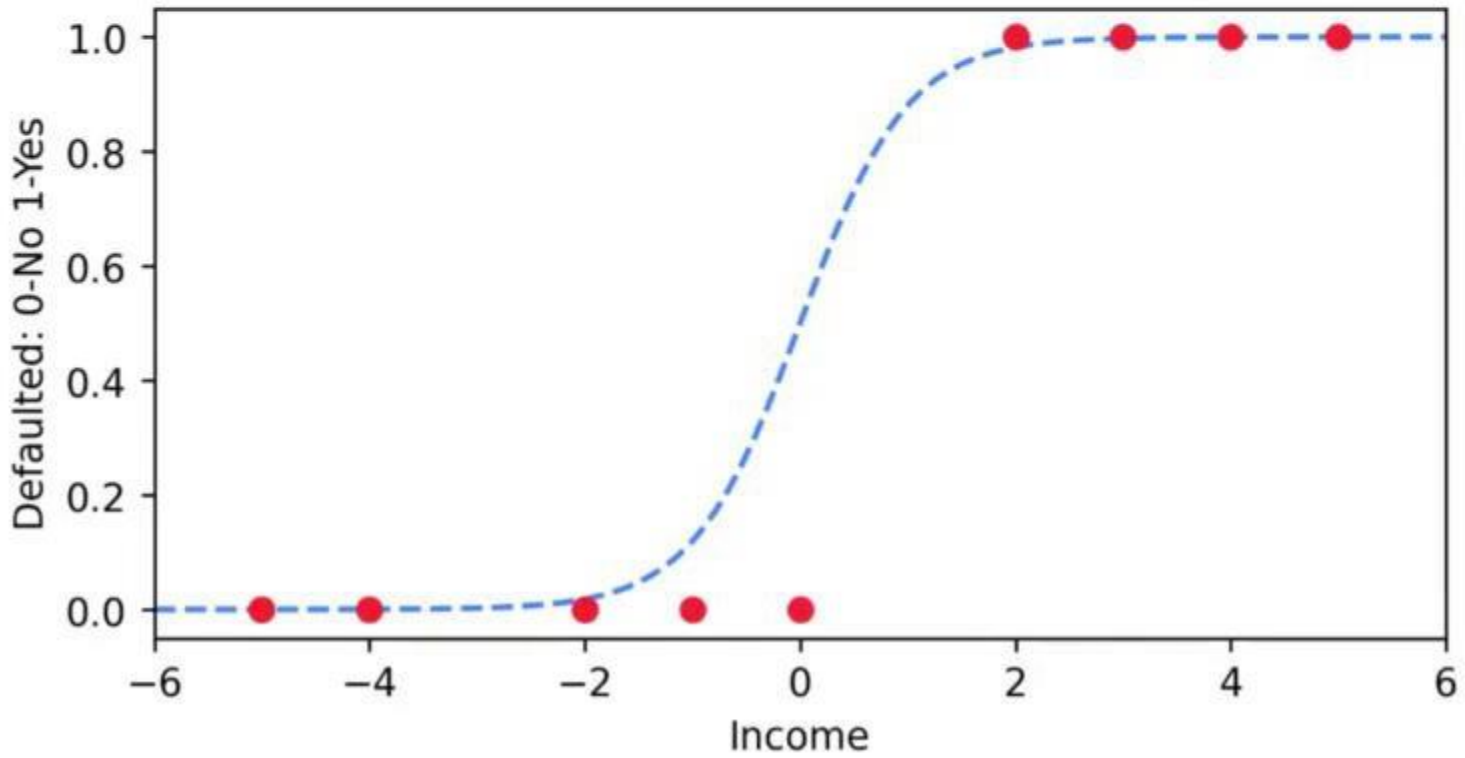
Logistic Regression

- Predict a 90% probability of paying off loan, return prediction of Loan Paid = 1.



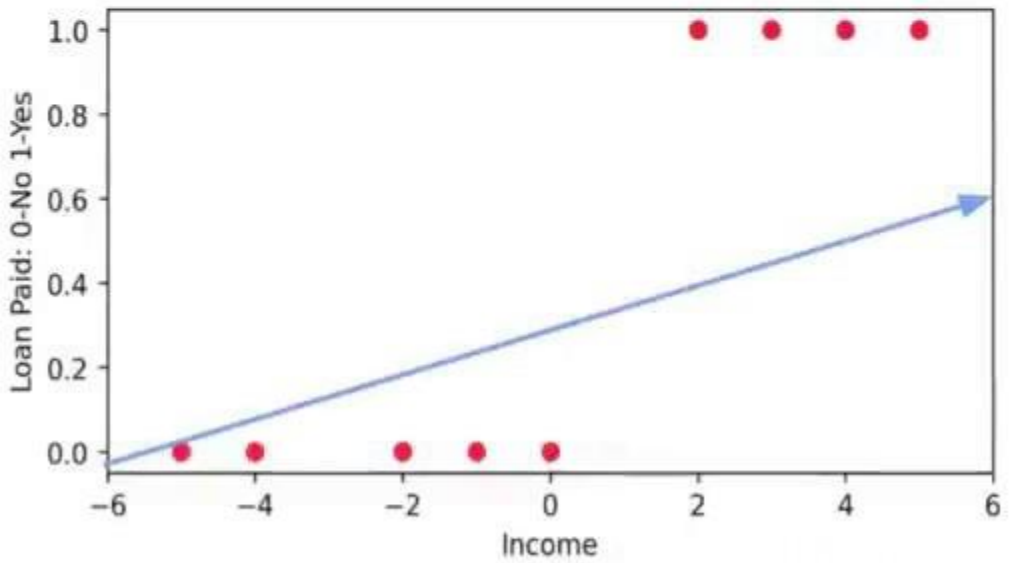
Logistic Regression

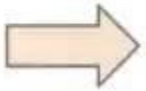
- But how do we actually create this line?

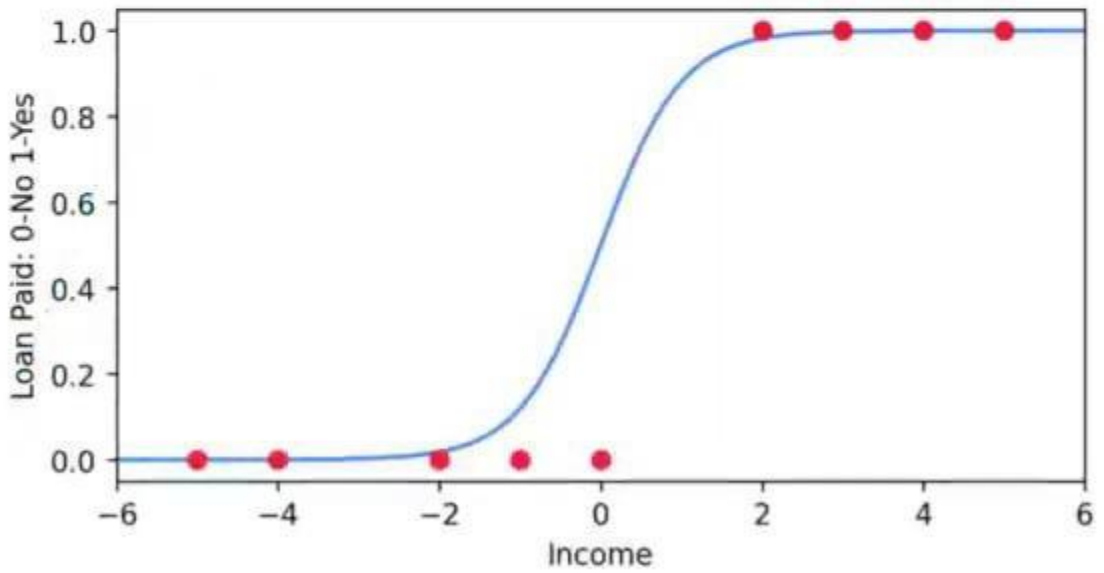


Logistic Regression

- Fortunately, the mathematics of the conversion are quite simple!





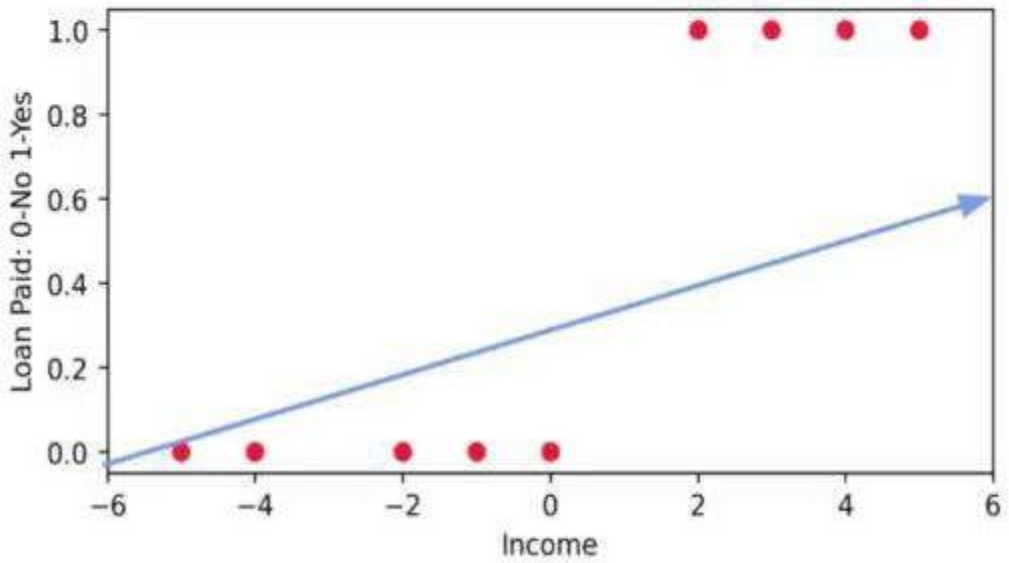


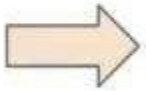
Logistic Regression Theory and Intuition

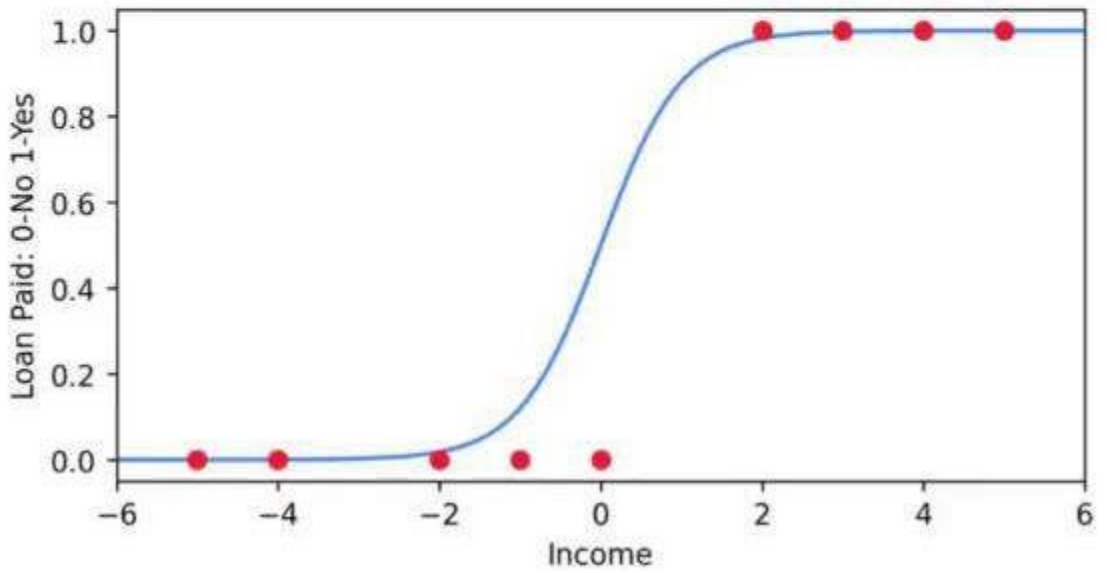
Part Two: Linear to Logistic Math

Logistic Regression

- Let's go through the math of converting Linear Regression to Logistic Regression.

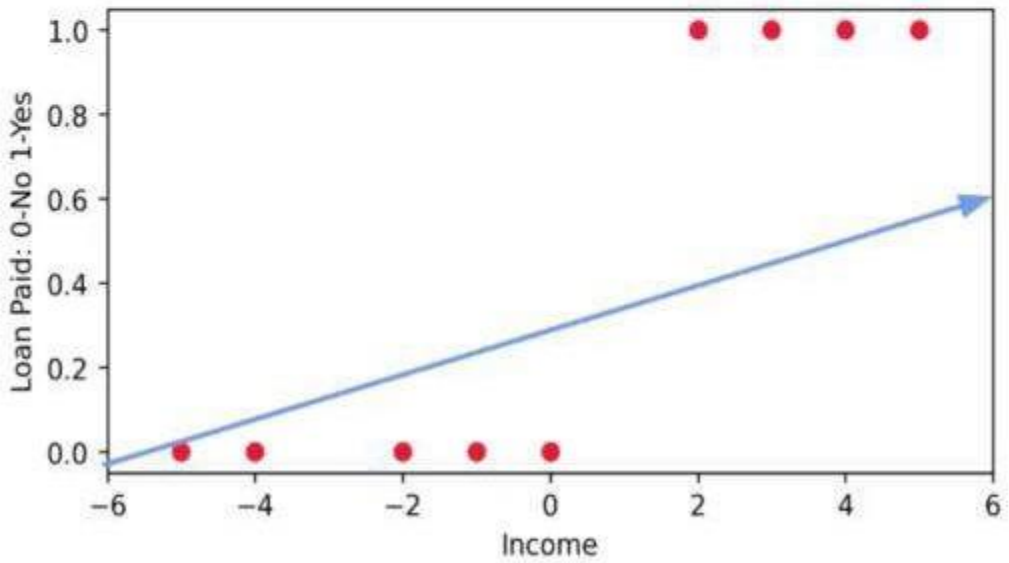


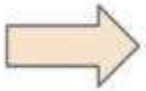


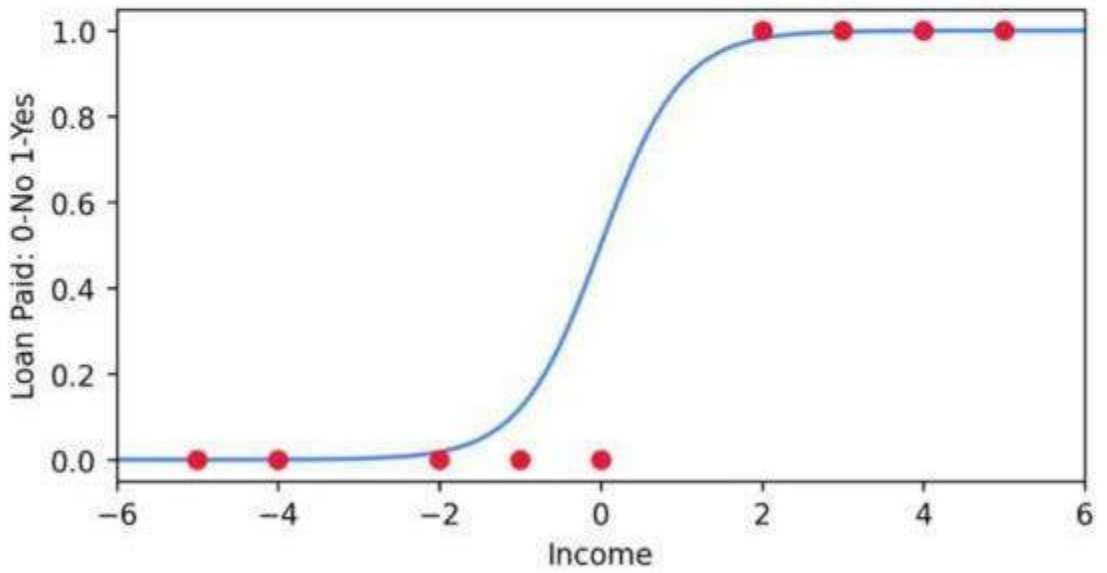


Logistic Regression

- Relevant ISLR Reading:
  - Section 4.3 Logistic Regression







Logistic Regression

- We already know the Linear Regression equation:

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

Logistic Regression

- We also know the Logistic function transforms any input to be between 0 and 1

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- All we need to do is plug the Linear Regression equation into the Logistic function to create a Logistic Regression!

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Logistic Regression

- Simply put in terms of the logistic function:

$$\hat{y} = \sigma(\beta_0 x_0 + \cdots + \beta_n x_n)$$

$$\hat{y} = \sigma\left(\sum_{i=0}^{n} \beta_i x_i\right)$$

Logistic Regression

- Writing it out fully:

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic Regression

- How do we interpret the coefficients and their relation to $\hat{y}$?

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic Regression

- First we need to understand the term **odds.**
- A term you may be familiar with from gambling **odds**.

Logistic Regression

- In gambling odds are often referred to in the sense of N to 1.
- But where does this actually come from?

Logistic Regression

- The odds of an event with probability **p** is defined as the chance of the event happening divided by the chance of the event not happening:

$$\frac{p}{1-p}$$

Logistic Regression

- The odds of an event with probability **p** is defined as the chance of the event happening divided by the chance of the event not happening:

$$\frac{p}{1-p}$$

Logistic Regression

- Imagine an event with **50%** probability of occuring. This is **0.5/1-0.5** which is **0.5/0.5** , the same as **1/1** or **1 to 1 odds of occurring.**

$$\frac{p}{1-p}$$

Logistic Regression

- Taking the formula below, we can rearrange it to show that it is equivalent to modelling the log of the odds as a linear combination of the features.

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic Regression

- This will allow us to solve for the coefficients and feature x in terms of **log odds.**

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic Regression

- Solving for **log odds:**

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic Regression

- Solving for **log odds:**

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

$$\hat{y} + \hat{y} e^{-\sum_{i=0}^{n} \beta_i x_i} = 1$$

Logistic Regression

- Solving for **log odds:**

$$\hat{y} + \hat{y}e^{-\sum_{i=0}^{n} \beta_i x_i} = 1$$

Logistic Regression

- Solving for **log odds:**

$$\hat{y} + \hat{y}e^{-\sum_{i=0}^{n} \beta_i x_i} = 1$$

$$\hat{y}e^{-\sum_{i=0}^{n} \beta_i x_i} = 1 - \hat{y}$$

Logistic Regression

- Solving for **log odds:**

$$\hat{y} + \hat{y}e^{-\sum_{i=0}^{n} \beta_i x_i} = 1$$

$$\hat{y}e^{-\sum_{i=0}^{n} \beta_i x_i} = 1 - \hat{y}$$

$$\frac{\hat{y}}{1 - \hat{y}} = e^{\sum_{i=0}^{n} \beta_i x_i}$$

Logistic Regression

- Solving for **log odds:**

$$\frac{\hat{y}}{1-\hat{y}} = e^{\sum_{i=0}^{n} \beta_i x_i}$$

Logistic Regression

- Solving for **log odds:**

$$\frac{\hat{y}}{1-\hat{y}} = e^{\sum_{i=0}^{n} \beta_i x_i}$$

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

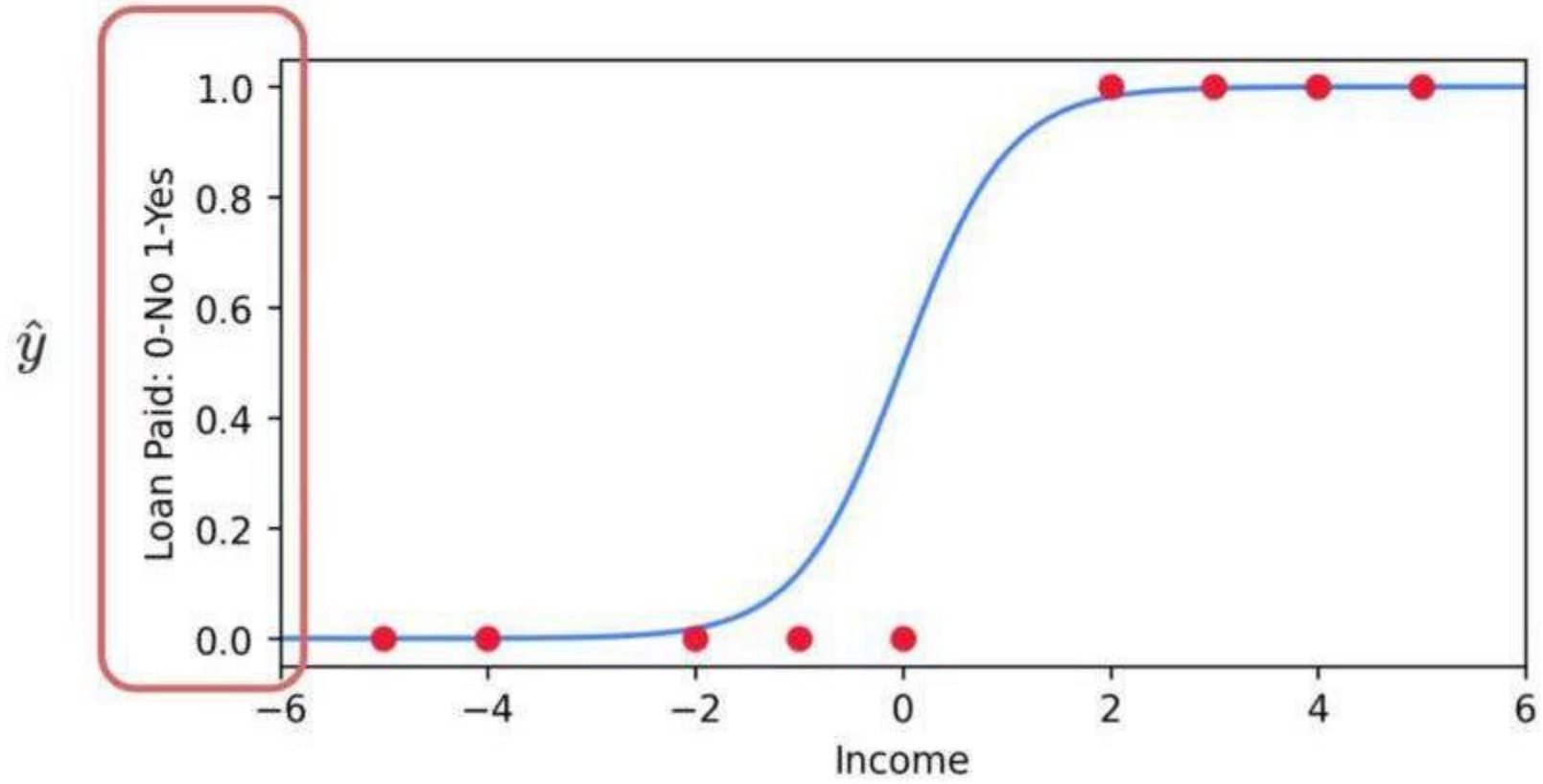
Logistic Regression

- What would the function curve look like in terms of log odds?

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

Logistic Regression

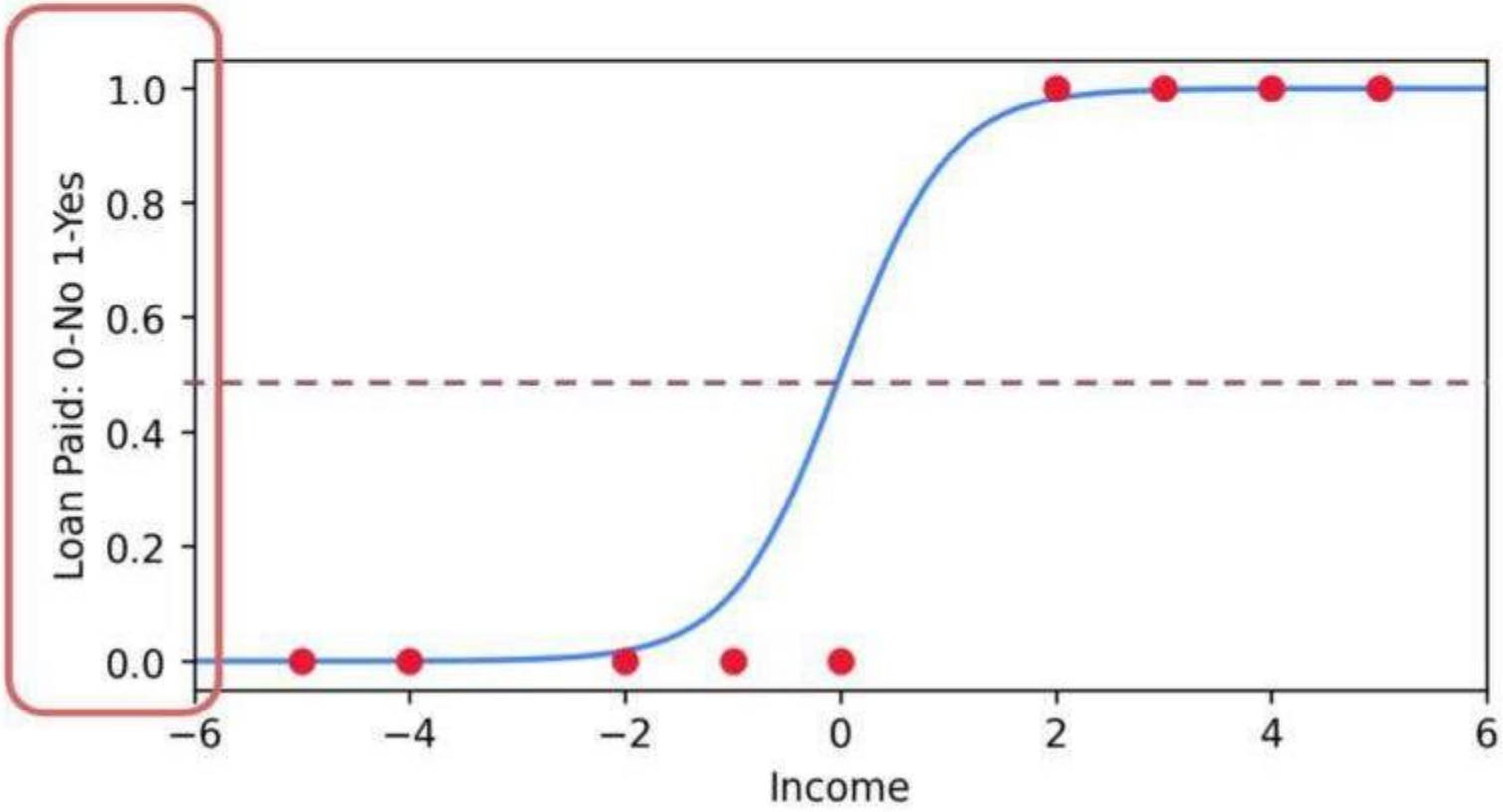
- What would the function curve look like in terms of log odds?



Logistic Regression

- Consider p=0.5

$$ln(\frac{0.5}{1-0.5}) = 0$$



Logistic Regression

- Consider p=0.5, halfway point now at 0.

$$ln(\frac{0.5}{1-0.5}) = 0$$

Logistic Regression

- As p goes to 1 then log odds becomes ∞

$$\lim_{p \to 1} ln(\frac{p}{1-p}) = \infty$$

$$ln(\frac{0.5}{1-0.5}) = 0$$

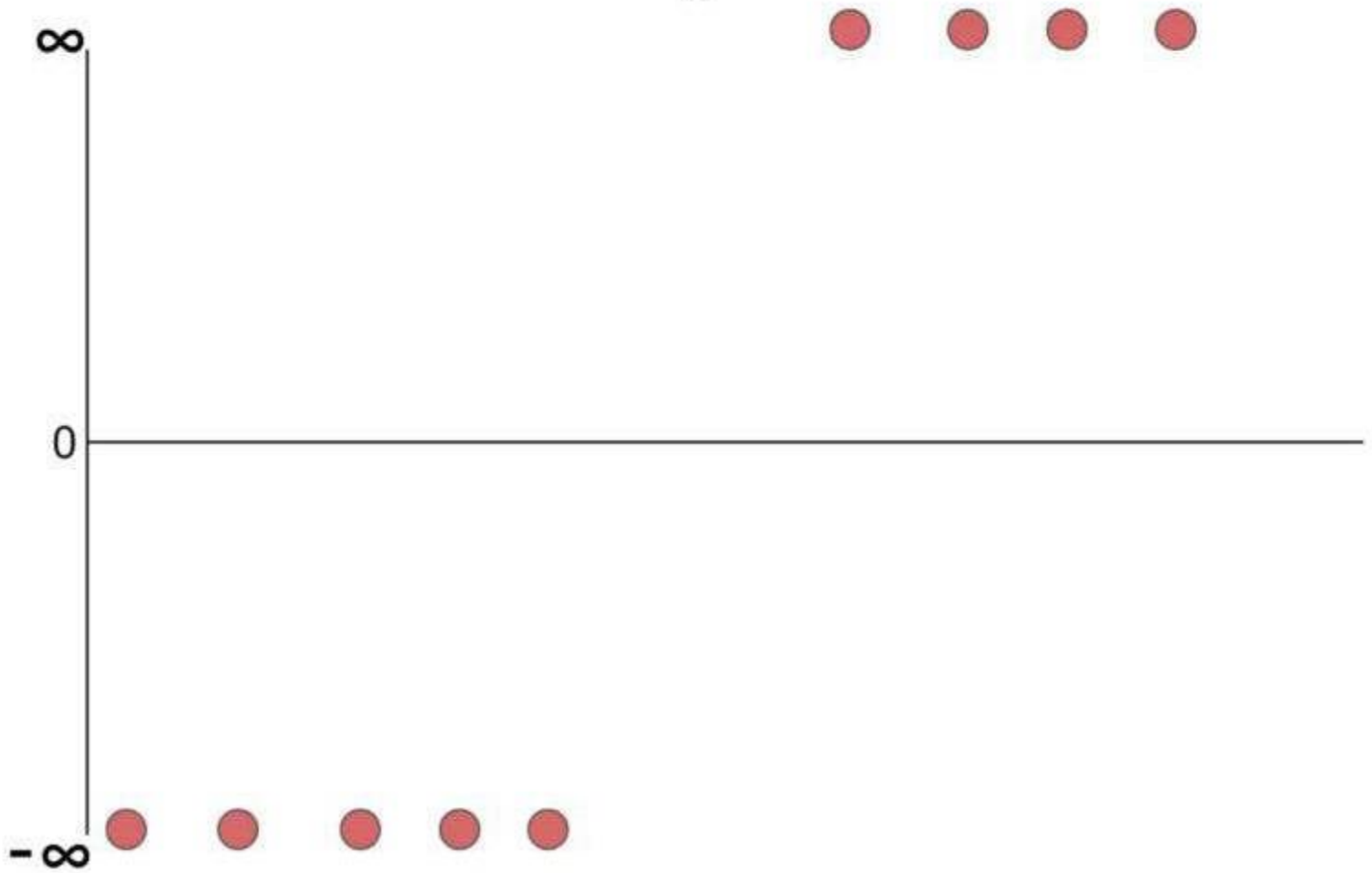
Logistic Regression

- Class points now at infinity

$$\lim_{p \to 1} ln(\frac{p}{1-p}) = \infty$$

$$ln(\frac{0.5}{1-0.5}) = 0$$

$$\lim_{p \to 0} ln(\frac{p}{1-p}) = -\infty$$



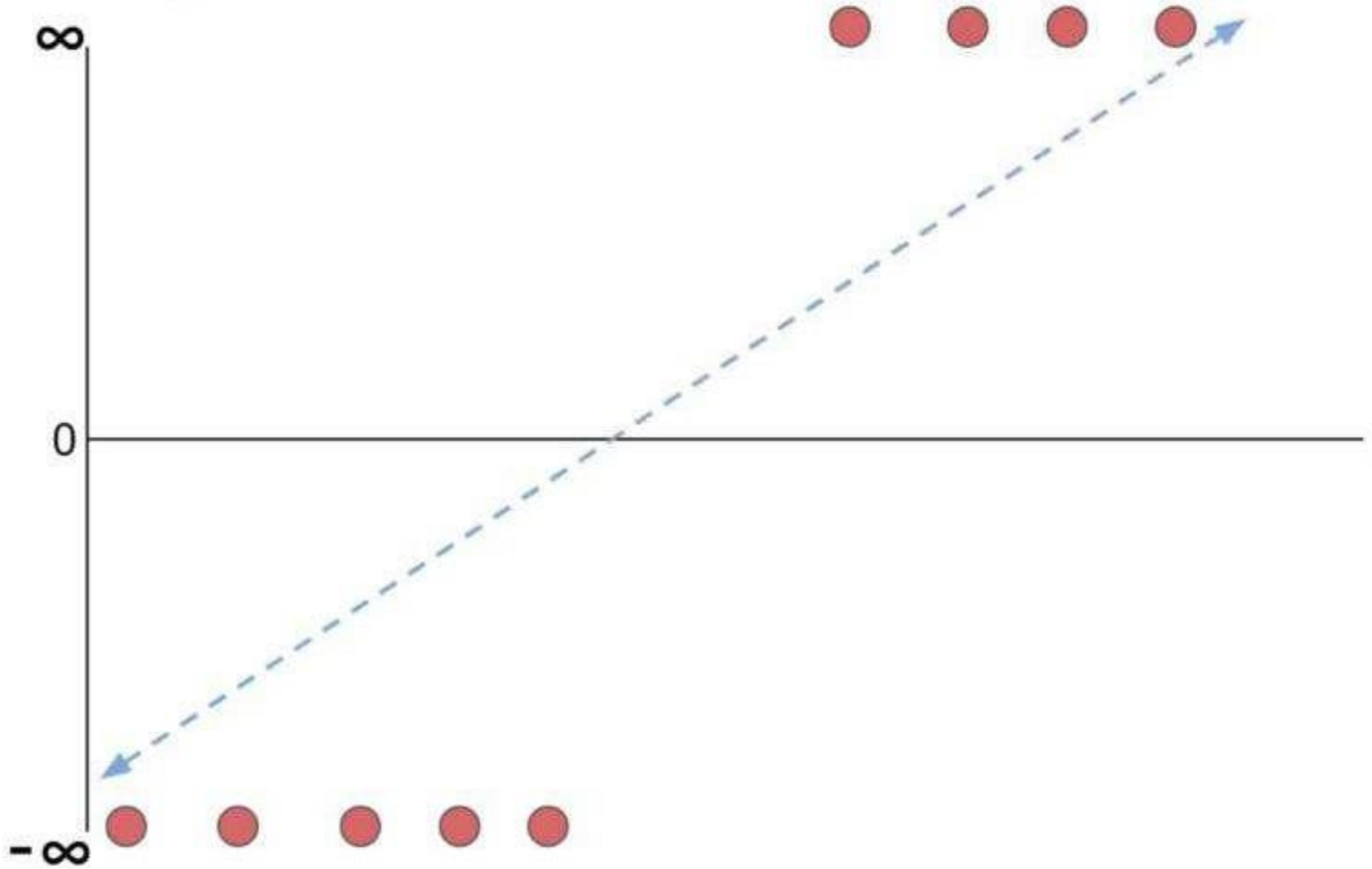
Logistic Regression

- On log scale logistic function is straight line

$$\lim_{p \to 1} ln(\frac{p}{1-p}) = \infty$$

$$ln(\frac{0.5}{1-0.5}) = 0$$

$$\lim_{p \to 0} ln(\frac{p}{1-p}) = -\infty$$



Logistic Regression

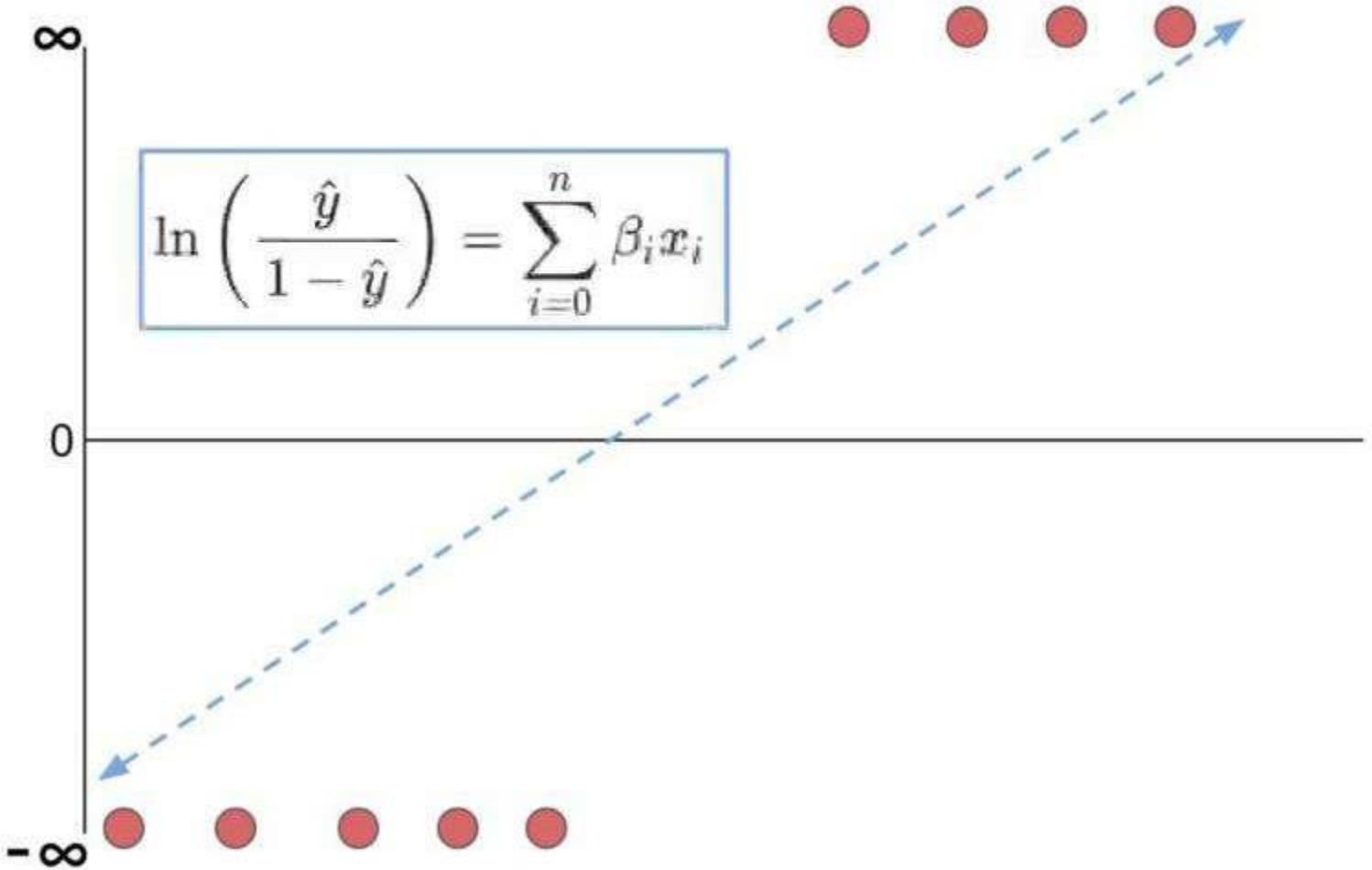
- Coefficients in terms of change in log odds.

$$\lim_{p \to 1} ln(\frac{p}{1-p}) = \infty$$

$$ln(\frac{0.5}{1-0.5}) = 0$$

$$\lim_{p \to 0} ln(\frac{p}{1-p}) = -\infty$$

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$



Logistic Regression

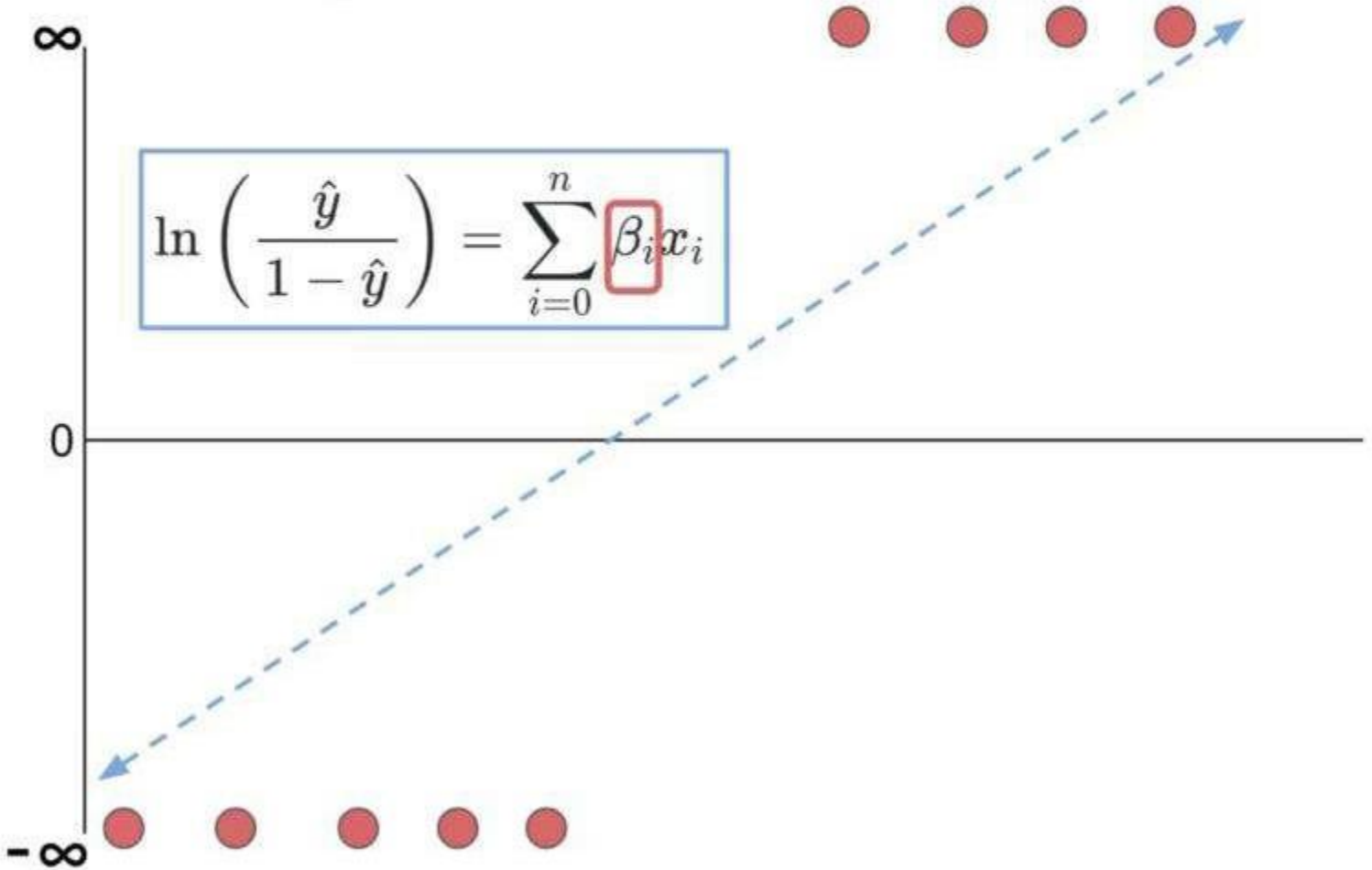
- Is β simple to interpret? Not really...

$$\lim_{p \to 1} ln(\frac{p}{1-p}) = \infty$$

$$ln(\frac{0.5}{1-0.5}) = 0$$

$$\lim_{p \to 0} ln(\frac{p}{1-p}) = -\infty$$

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$



Logistic Regression

- Since the log odds scale is nonlinear, a β value can not be directly linked to "one unit increase" as it could in Linear Regression.

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

Logistic Regression

- There are some straightforward insights we can gain however...

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

Logistic Regression

- Sign of Coefficient
  - Positive β indicates an increase in likelihood of belonging to 1 class with increase in associated **x** feature.
  - Negative β indicates an decrease in likelihood of belonging to 1 class with increase in associated **x** feature.

Logistic Regression

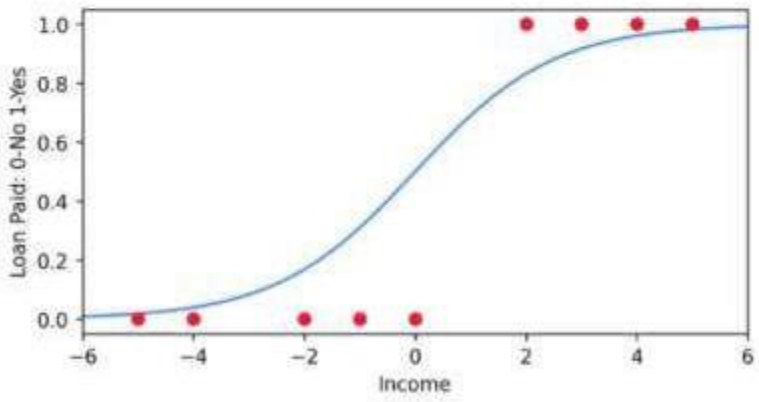
- Magnitude of Coefficient
  - Harder to directly interpret magnitude of β directly, especially when we could have discrete and continuous x feature values.
  - We can however begin to use **odds ratio**, essentially comparing magnitudes against each other.

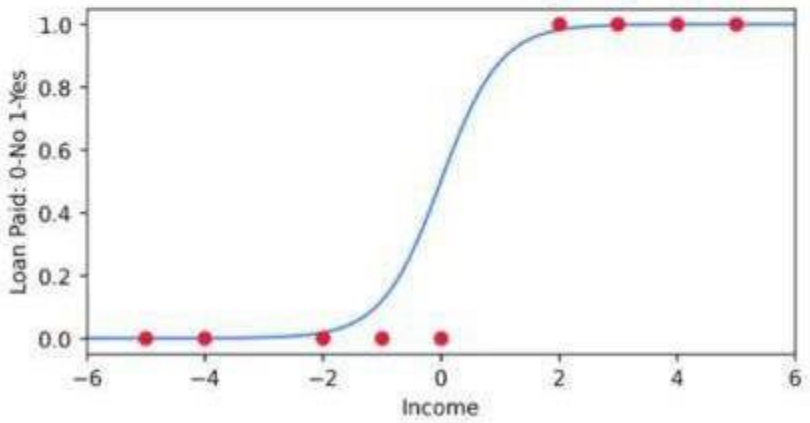
Logistic Regression

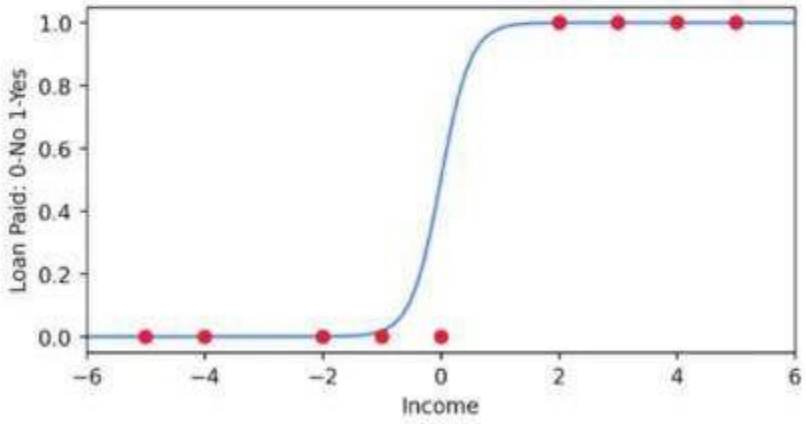
- Magnitude of Coefficient
  - Comparing magnitudes of coefficients against each other can lead to insight over which features have the strongest effect on prediction output.

Logistic Regression

- The last mathematical topic we need to discuss concerning Logistic Regression is how we actually fit this curve!







Logistic Regression Theory and Intuition

Part Three: Finding the Best Fit

Logistic Regression

- Logistic Regression uses Maximum Likelihood to find the best fitting model.
- This lecture will give you an intuition of how this method works.
- We'll also then display the cost function and gradient descent that is solved for by the computer.

Logistic Regression

- Quick Note: ISLR Section 4.3.2

default status. In other words, we try to find $\hat{\beta}_0$ and $\hat{\beta}_1$ such that plugging these estimates into the model for $p(X)$, given in (4.2), yields a number close to one for all individuals who defaulted, and a number close to zero for all individuals who did not. This intuition can be formalized using a mathematical equation called a *likelihood function*:

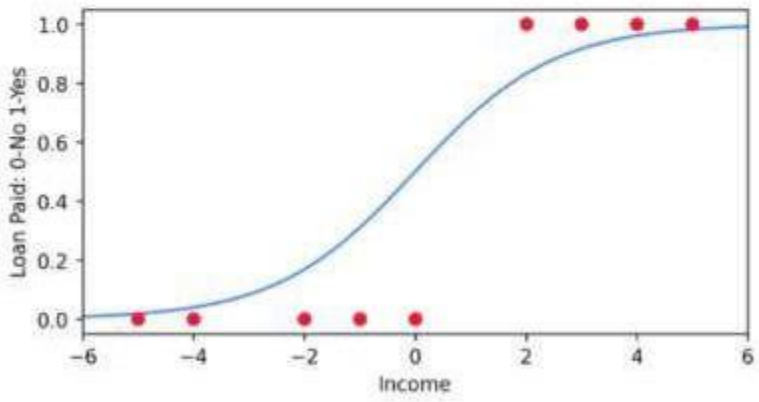
$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} (1 - p(x_{i'})). \tag{4.5}$$

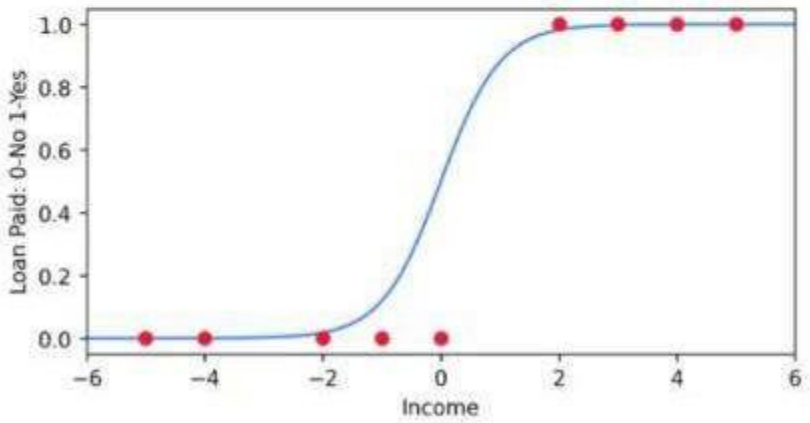
The estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are chosen to *maximize* this likelihood function.

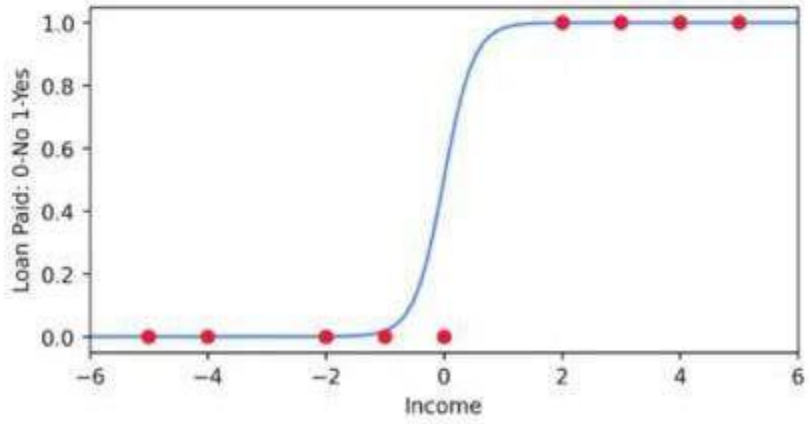
Maximum likelihood is a very general approach that is used to fit many of the non-linear models that we examine throughout this book. In the linear regression setting, the least squares approach is in fact a special case of maximum likelihood. The mathematical details of maximum likelihood are beyond the scope of this book. However, in general, logistic regression

Logistic Regression

- Here we see three different Logistic Regression curves with different β values.
- How do we measure which is the best fit?

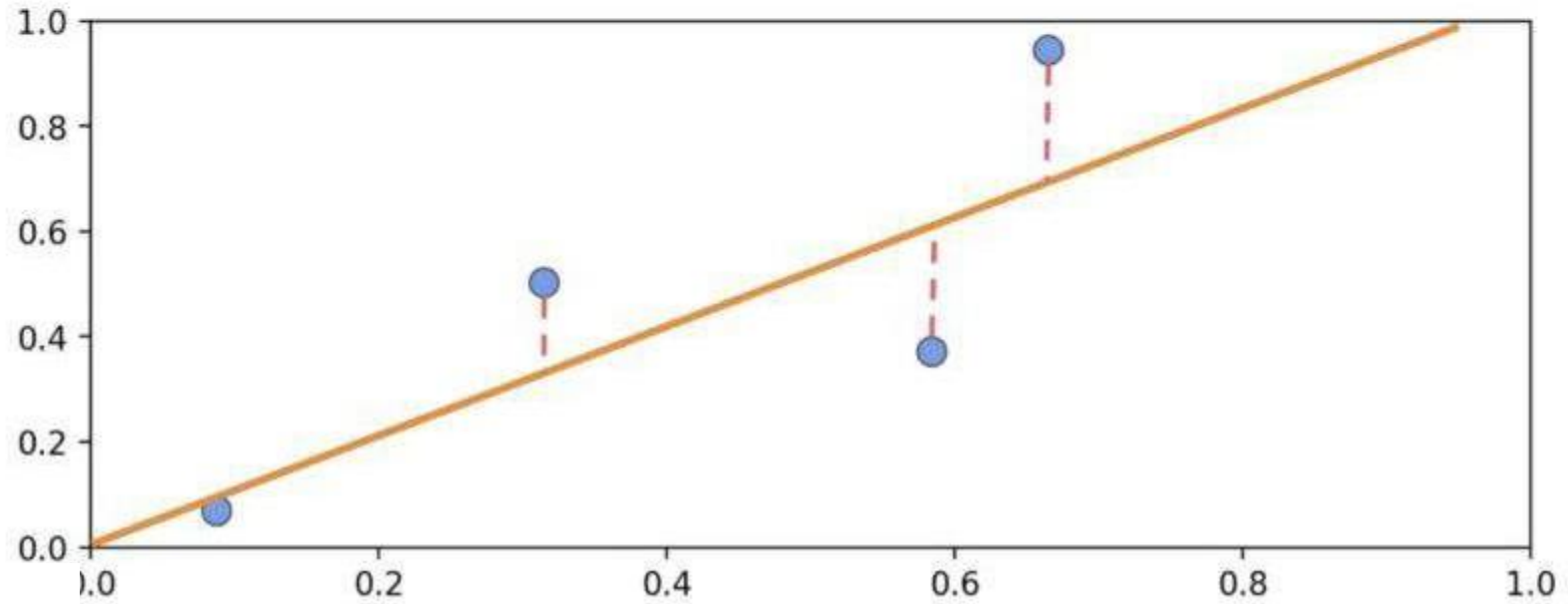






Logistic Regression

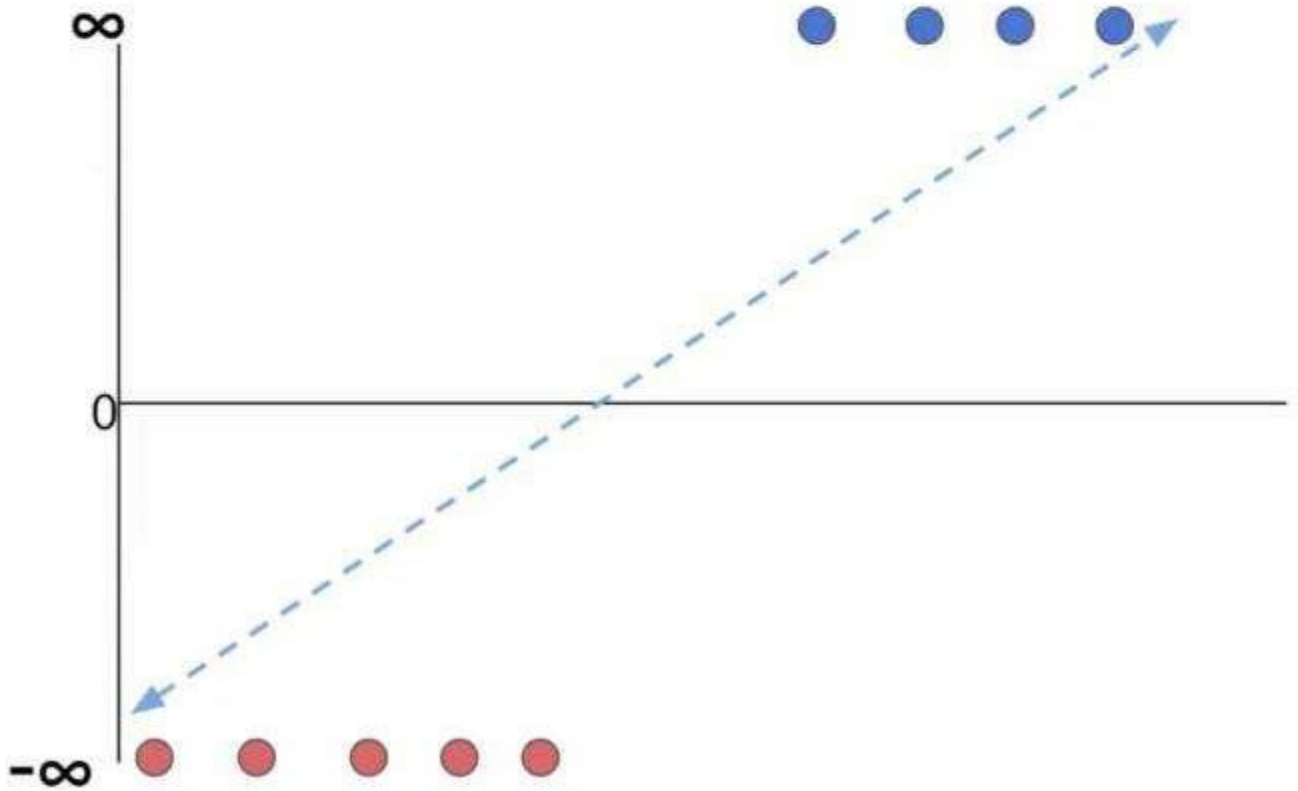
- Recall in Linear Regression we seek to minimize the Residual Sum of Squares (RSS).



Logistic Regression

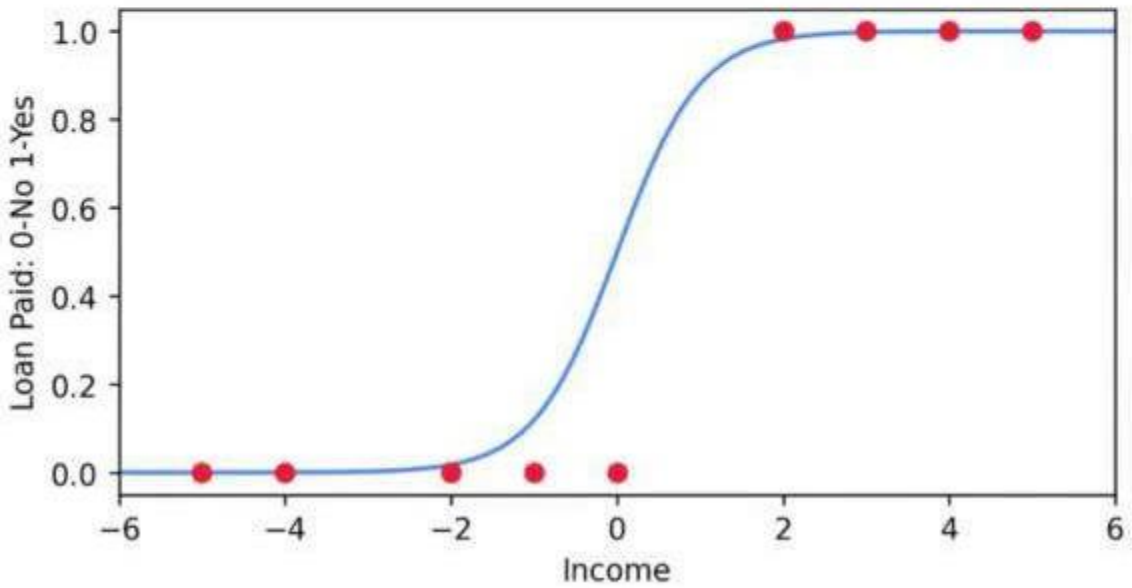
- Unfortunately, even in log odds targets are at infinity, making RSS unfeasible.

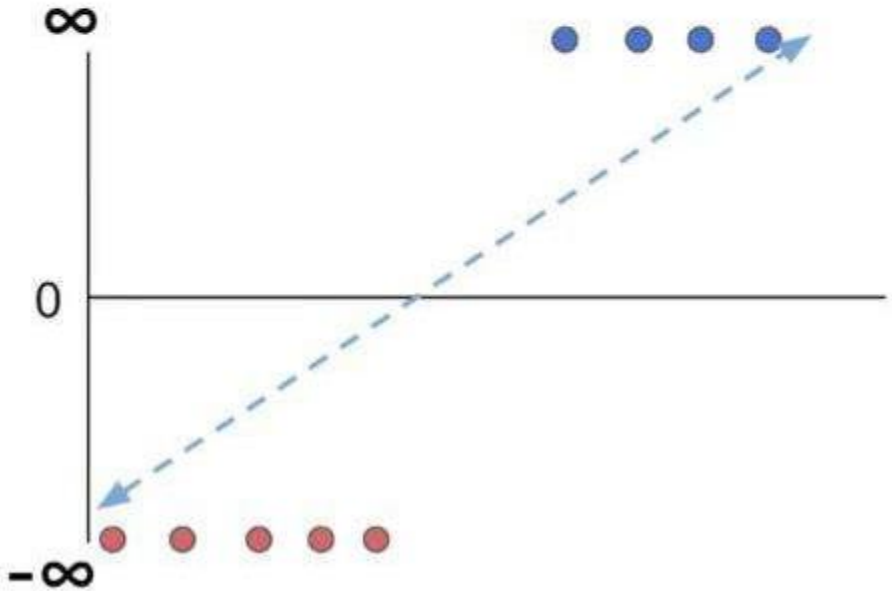
$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$



Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.





Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

$$\ln\left(\frac{p}{1-p}\right) = \ln(odds)$$

Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

$$\ln(\frac{p}{1-p}) = \ln(odds)$$

$$\frac{p}{1-p} = e^{\ln(odds)}$$

$$p = (1-p)e^{\ln(odds)}$$

Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

$$p = (1 - p)e^{\ln(odds)}$$

$$p = e^{\ln(odds)} - pe^{\ln(odds)}$$

Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

$$p = (1 - p)e^{\ln(odds)}$$

$$p = e^{\ln(odds)} - pe^{\ln(odds)}$$

$$p + pe^{\ln(odds)} = e^{\ln(odds)}$$

Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

$$p = e^{\ln(odds)} - pe^{\ln(odds)}$$

$$p + pe^{\ln(odds)} = e^{\ln(odds)}$$

$$p(1 + e^{\ln(odds)}) = e^{\ln(odds)}$$

$$p = \frac{e^{\ln(odds)}}{1 + e^{\ln(odds)}}$$

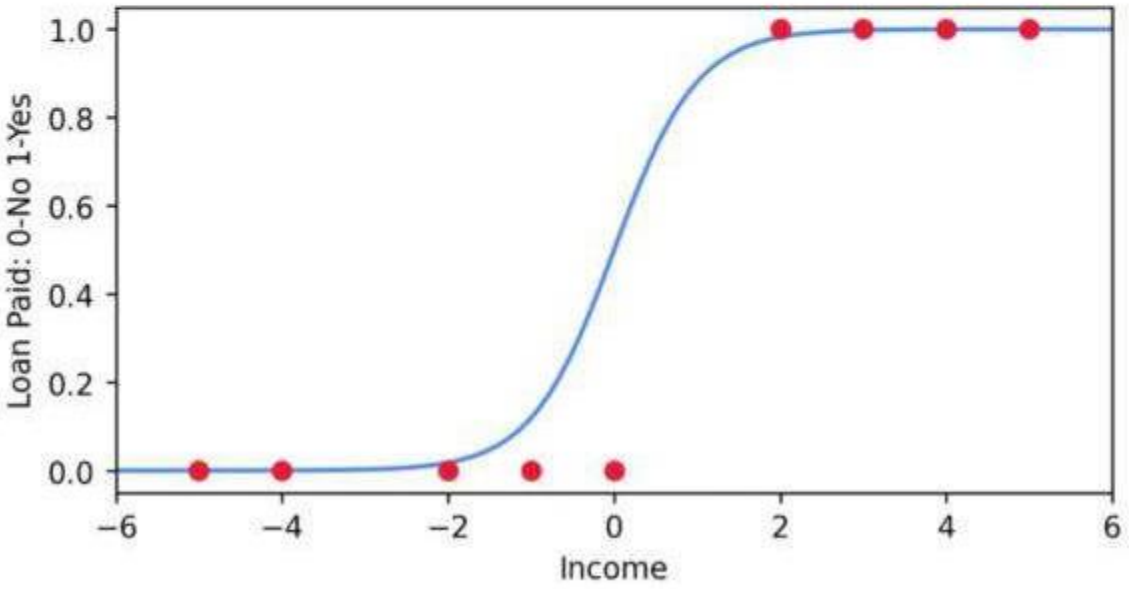
Logistic Regression

- The first step for maximum likelihood is to go from log odds back to probability.

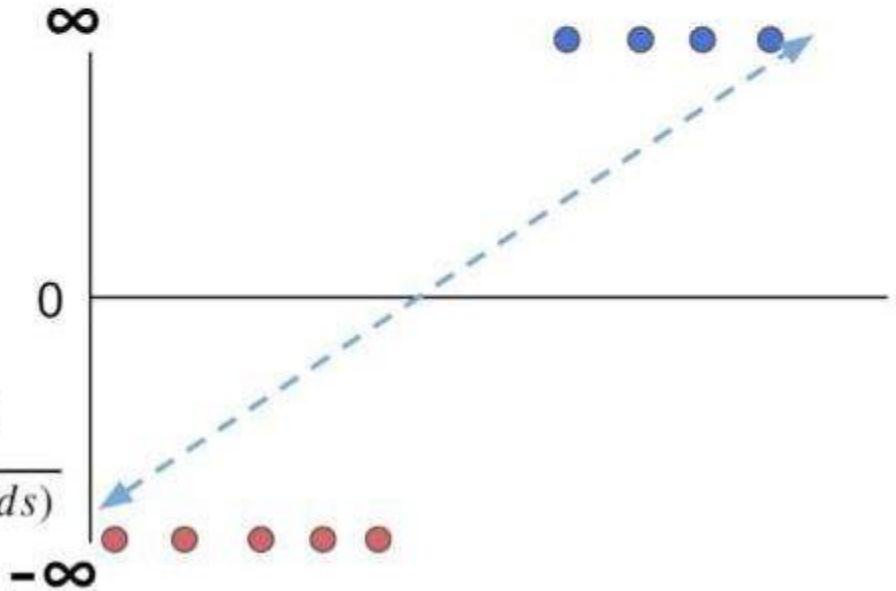
$$p = \frac{e^{\ln(odds)}}{1 + e^{\ln(odds)}}$$

Logistic Regression

- We are now able to convert ln(odds) into a probability.

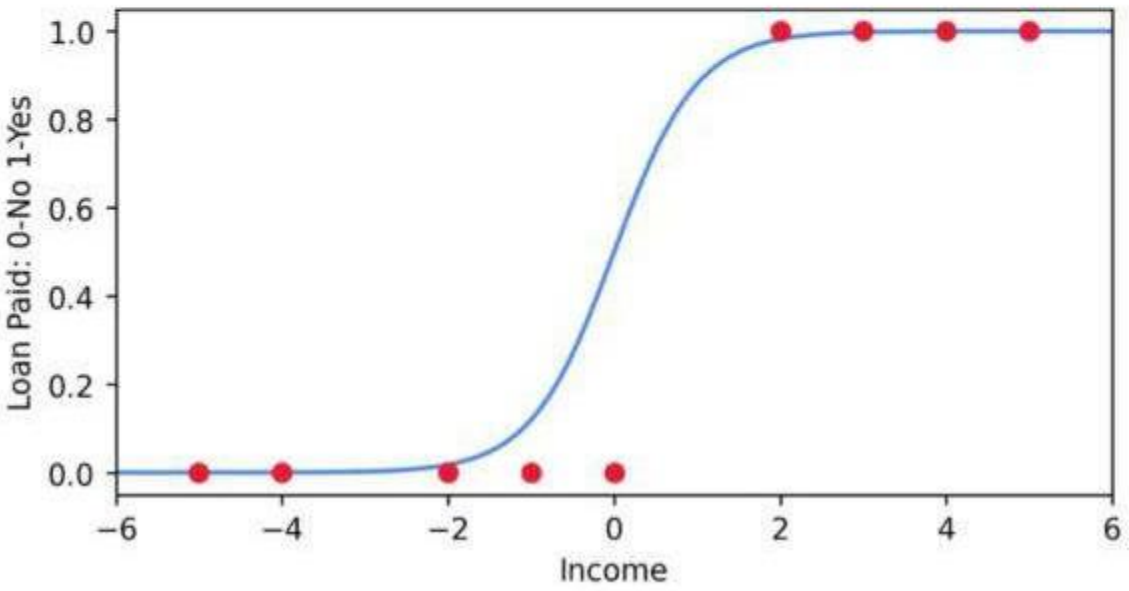


$$p = \frac{e^{\ln(odds)}}{1 + e^{\ln(odds)}}$$

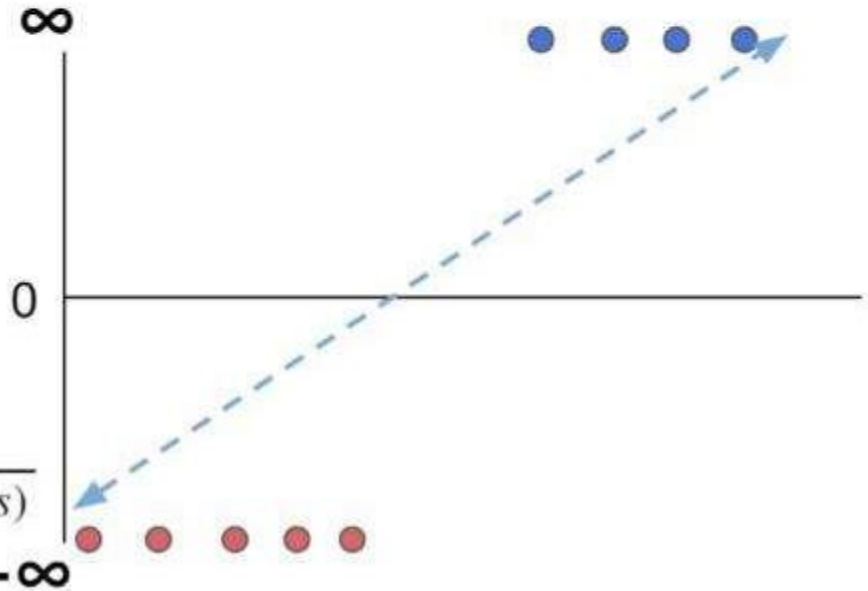


Logistic Regression

- Let's now explore the intuition behind **maximum likelihood.**

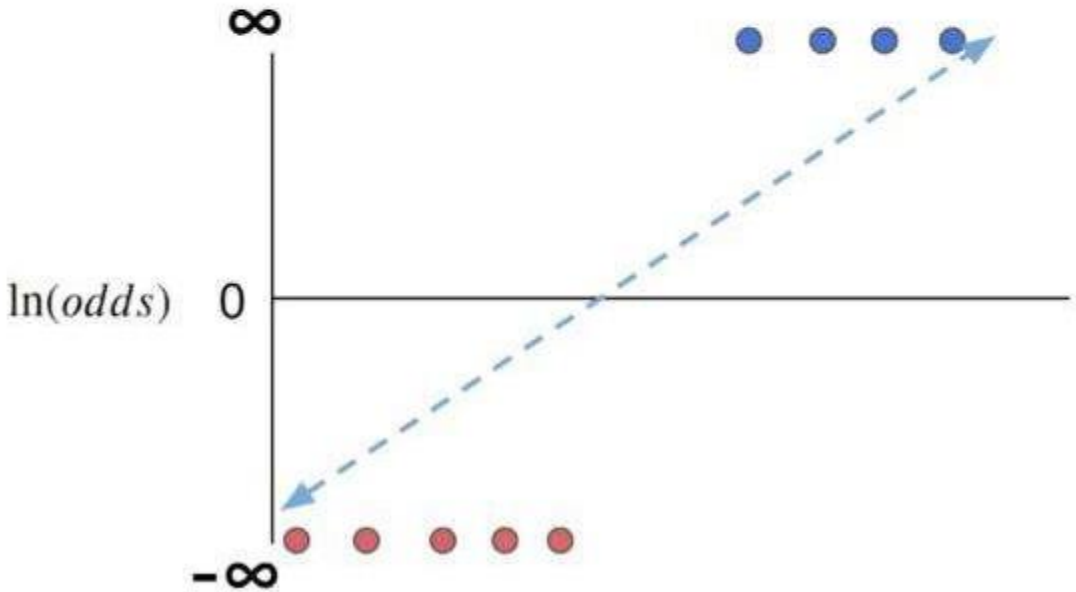


$$p = \frac{e^{\ln(odds)}}{1 + e^{\ln(odds)}}$$



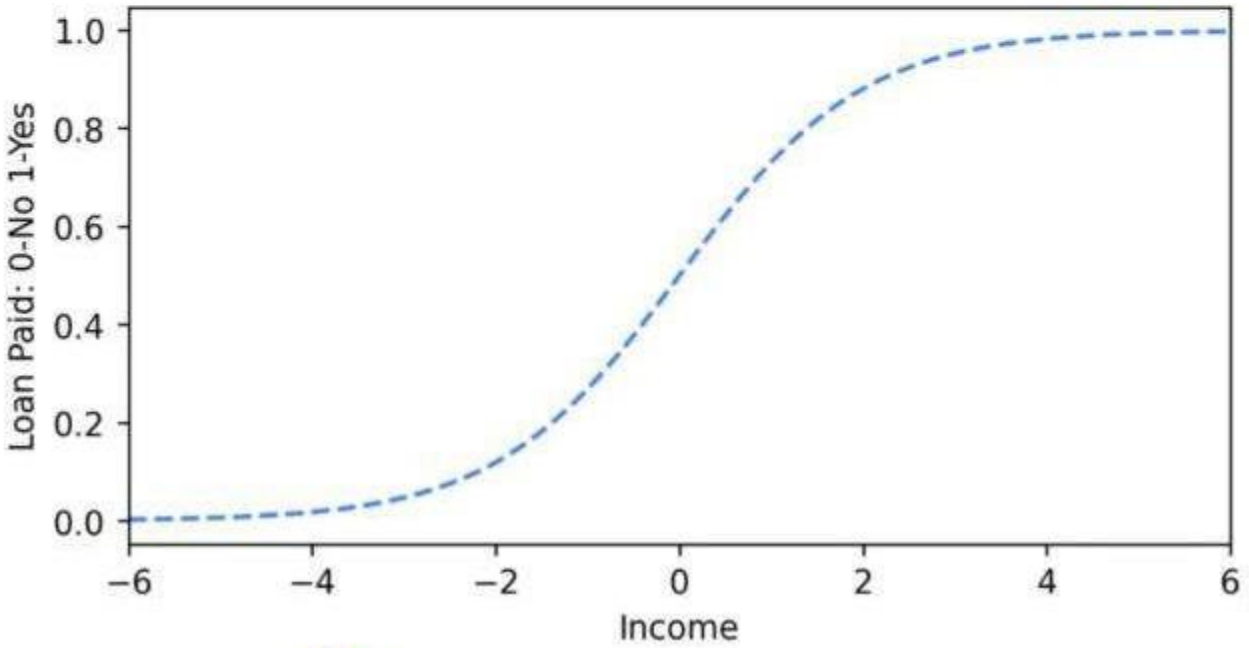
Logistic Regression

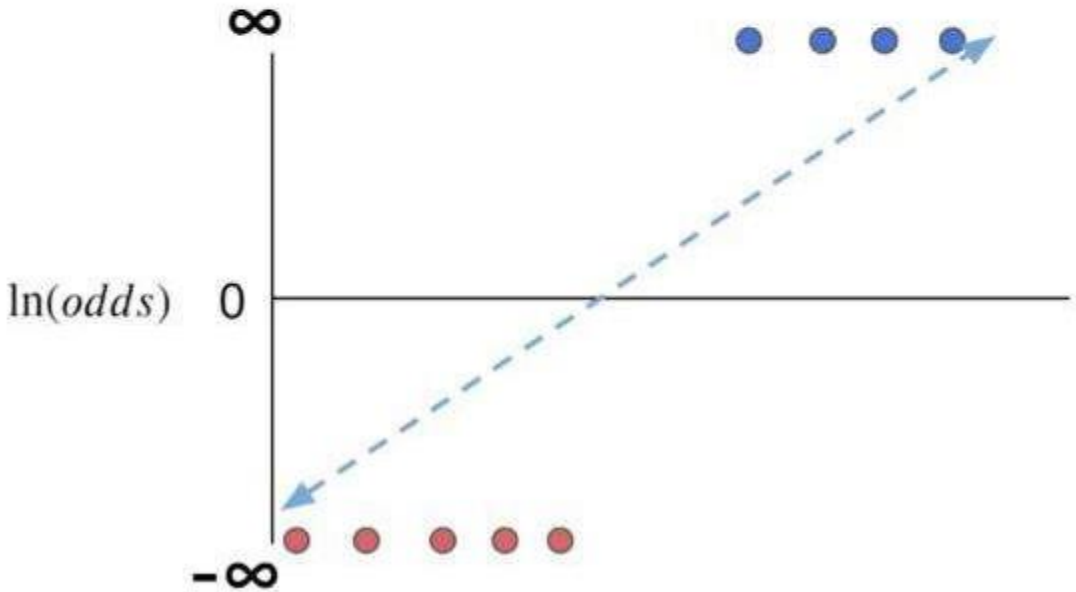
- We choose a line in the log(odds) axis and project the points on to the line:



Logistic Regression

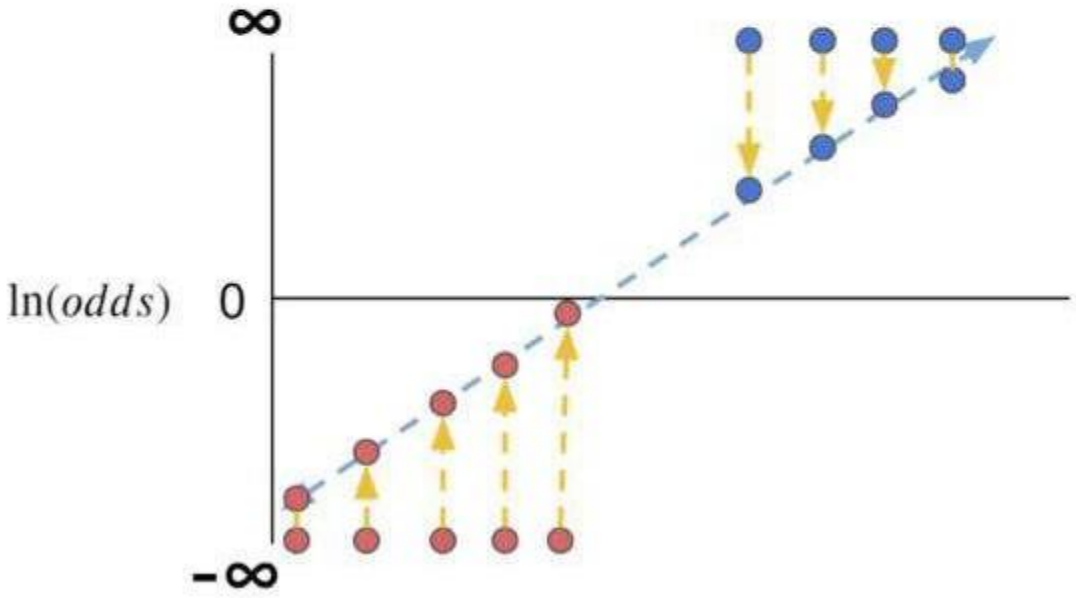
- We also know this line has a form on the probability y-axis.





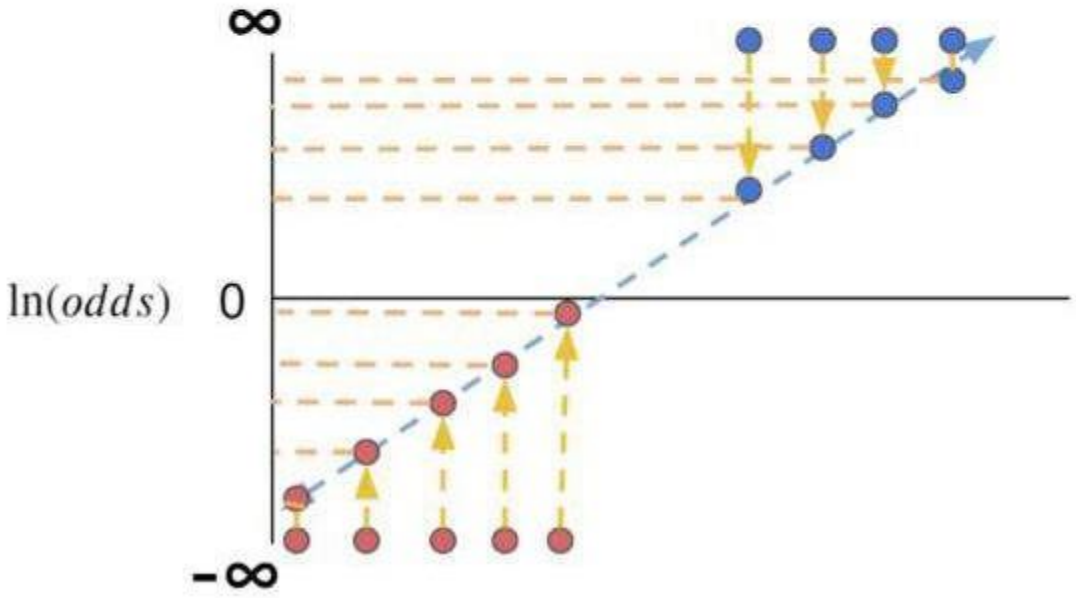
Logistic Regression

- We choose a line in the log(odds) axis and project the points on to the line:



Logistic Regression

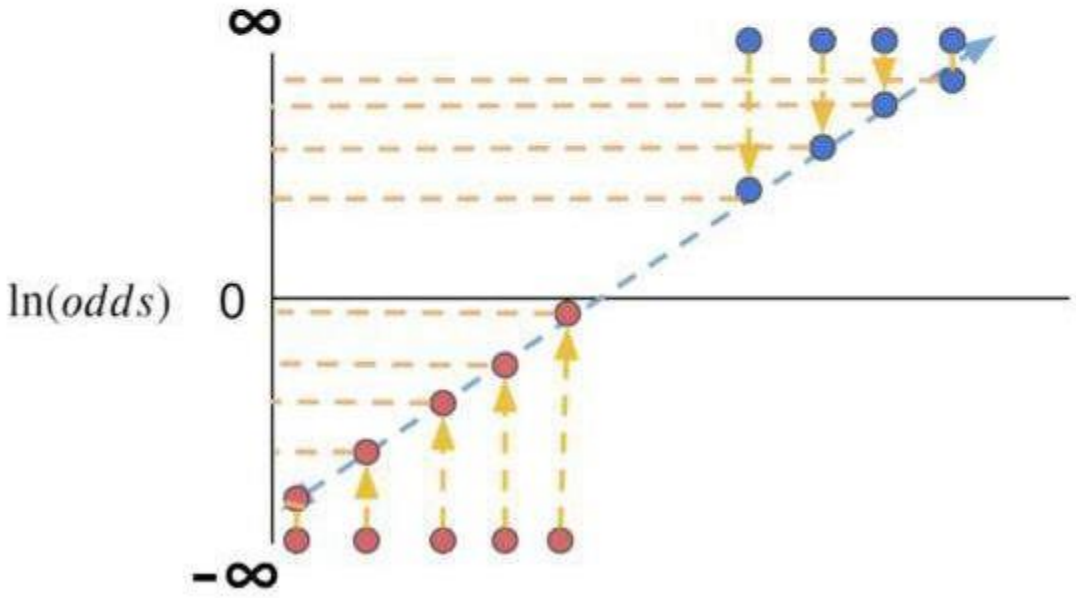
- Calculate the log odds for the projected points on this line.



Logistic Regression

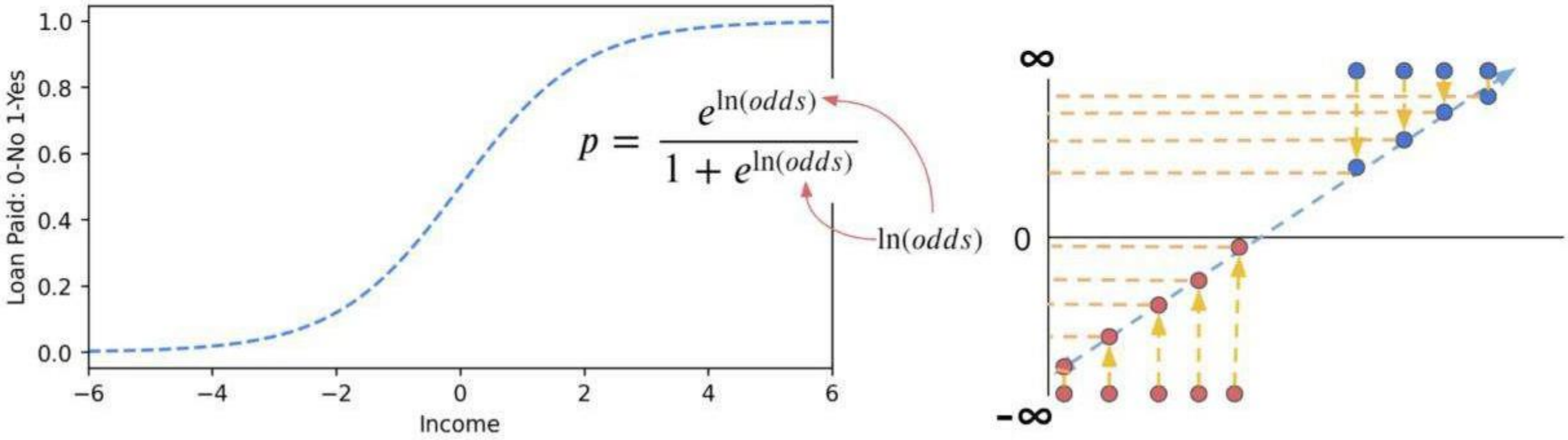
- Plot these values as probabilities on the logistic regression model.

$$p = \frac{e^{\ln(odds)}}{1 + e^{\ln(odds)}}$$



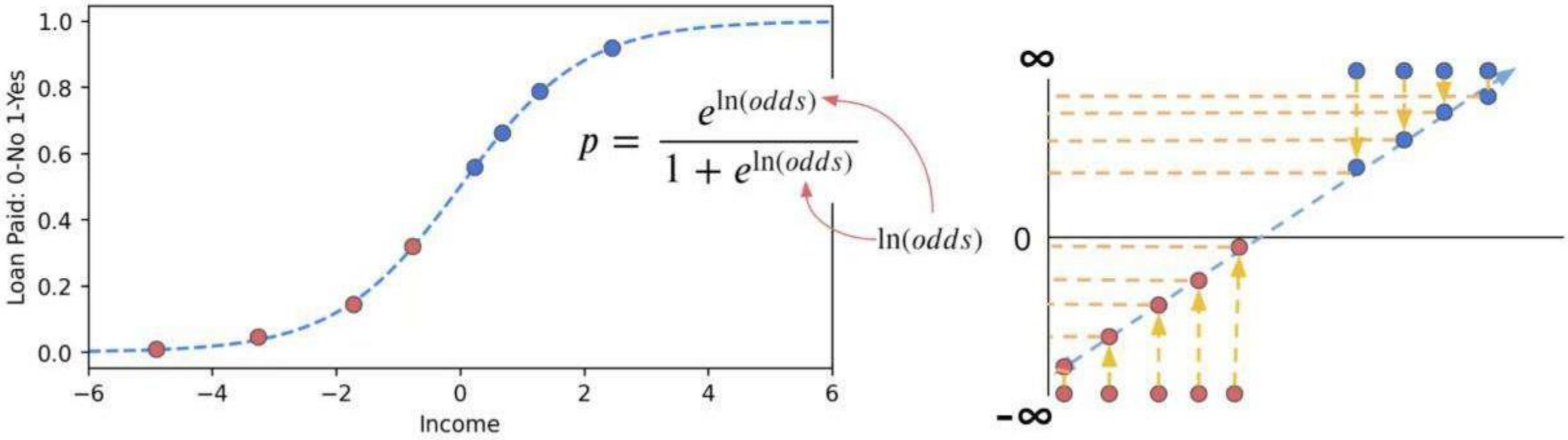
Logistic Regression

- Plot these values as probabilities on the logistic regression model.



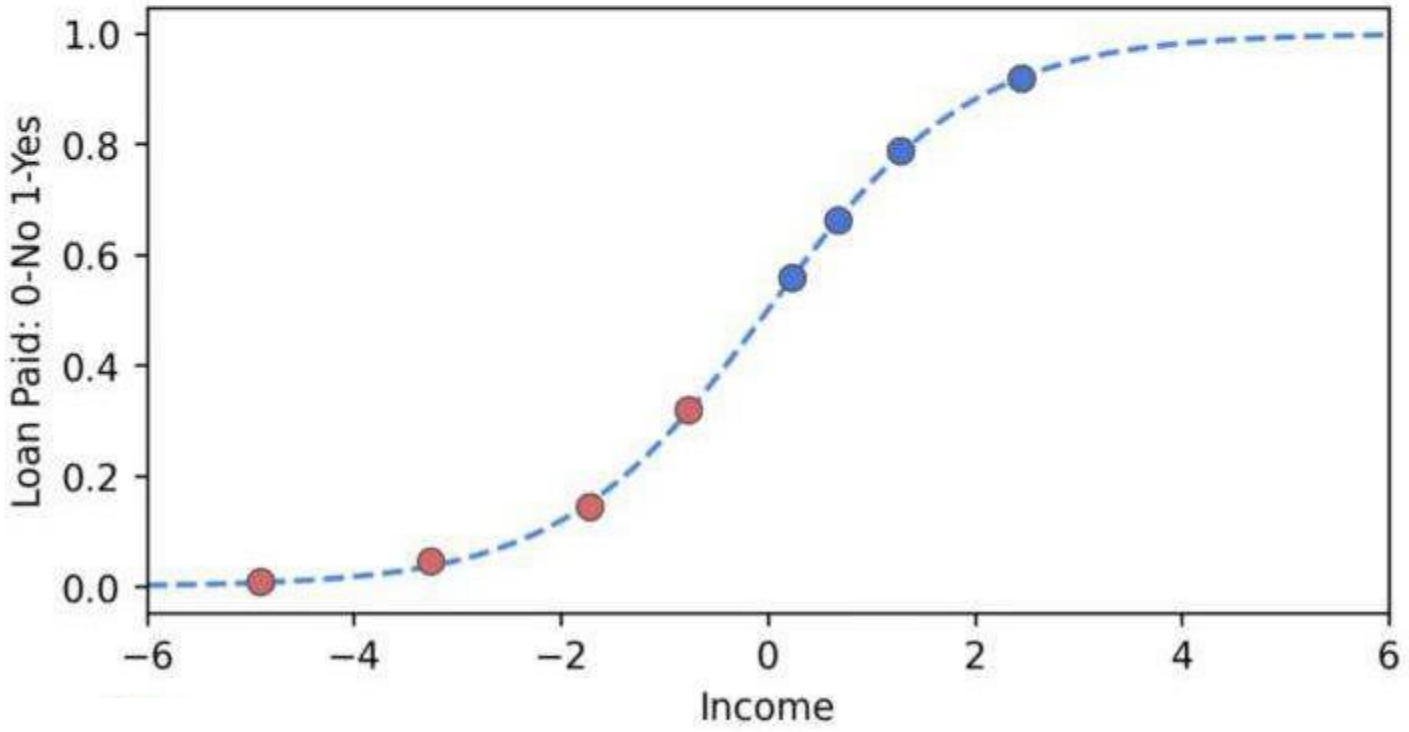
Logistic Regression

- We now measure the likelihood of these probabilities.



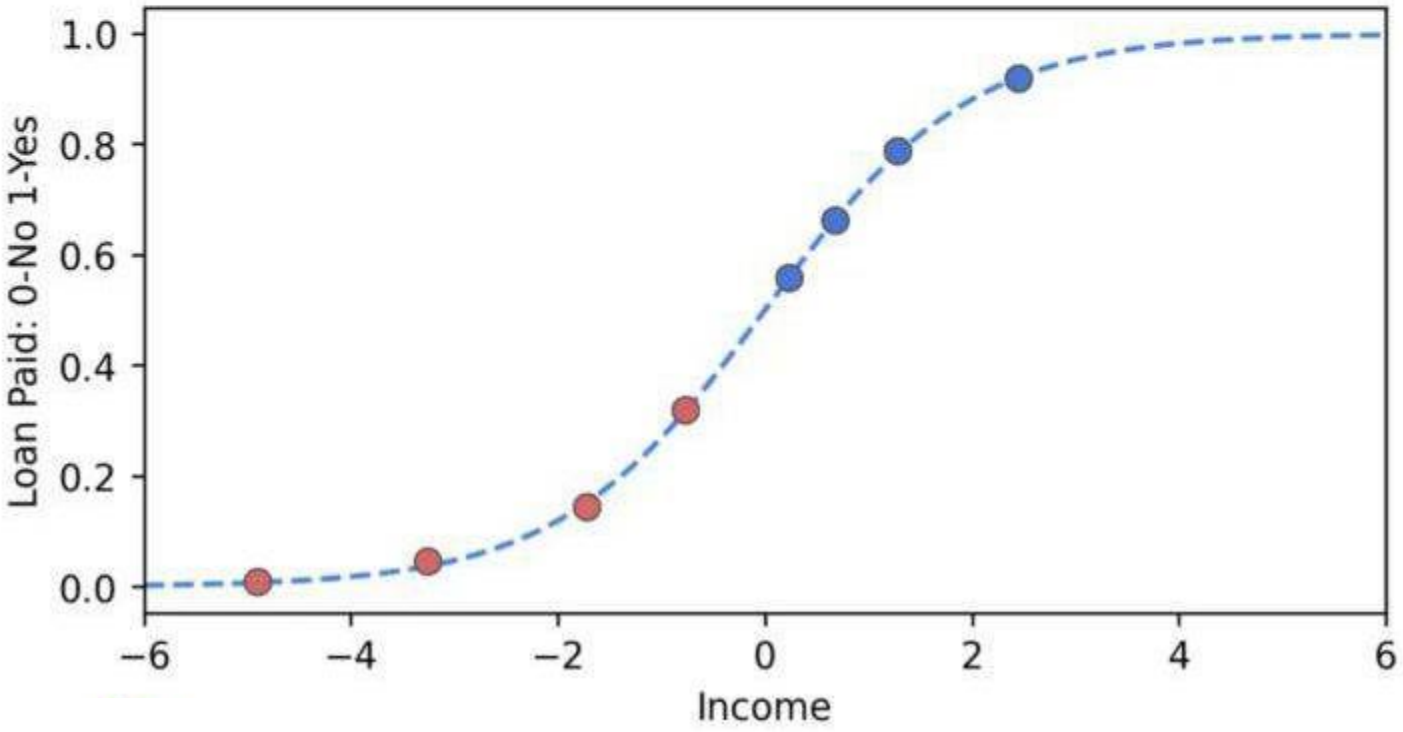
Logistic Regression

- We now measure the likelihood of these probabilities.



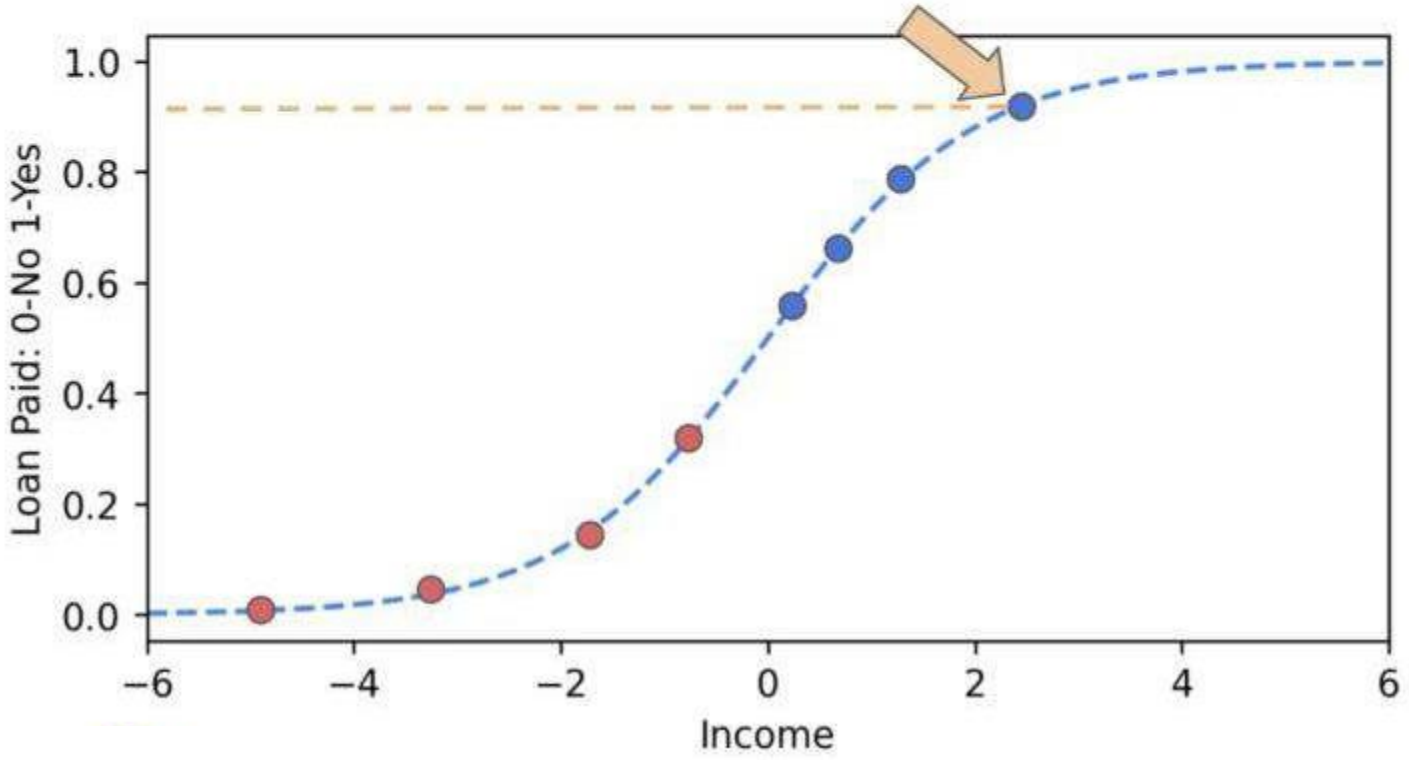
Logistic Regression

- Likelihood = Product of probabilities of belonging to class 1.



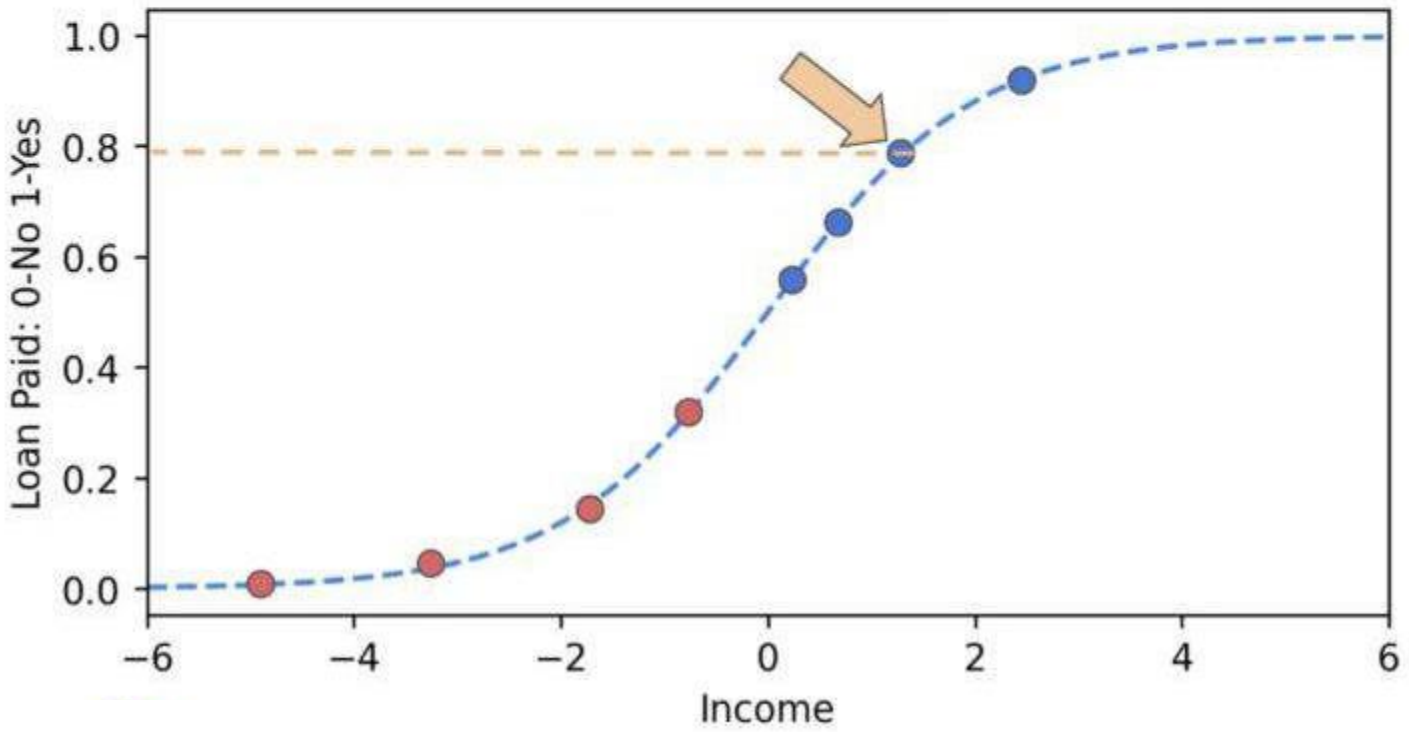
Logistic Regression

- Likelihood = 0.9 ...



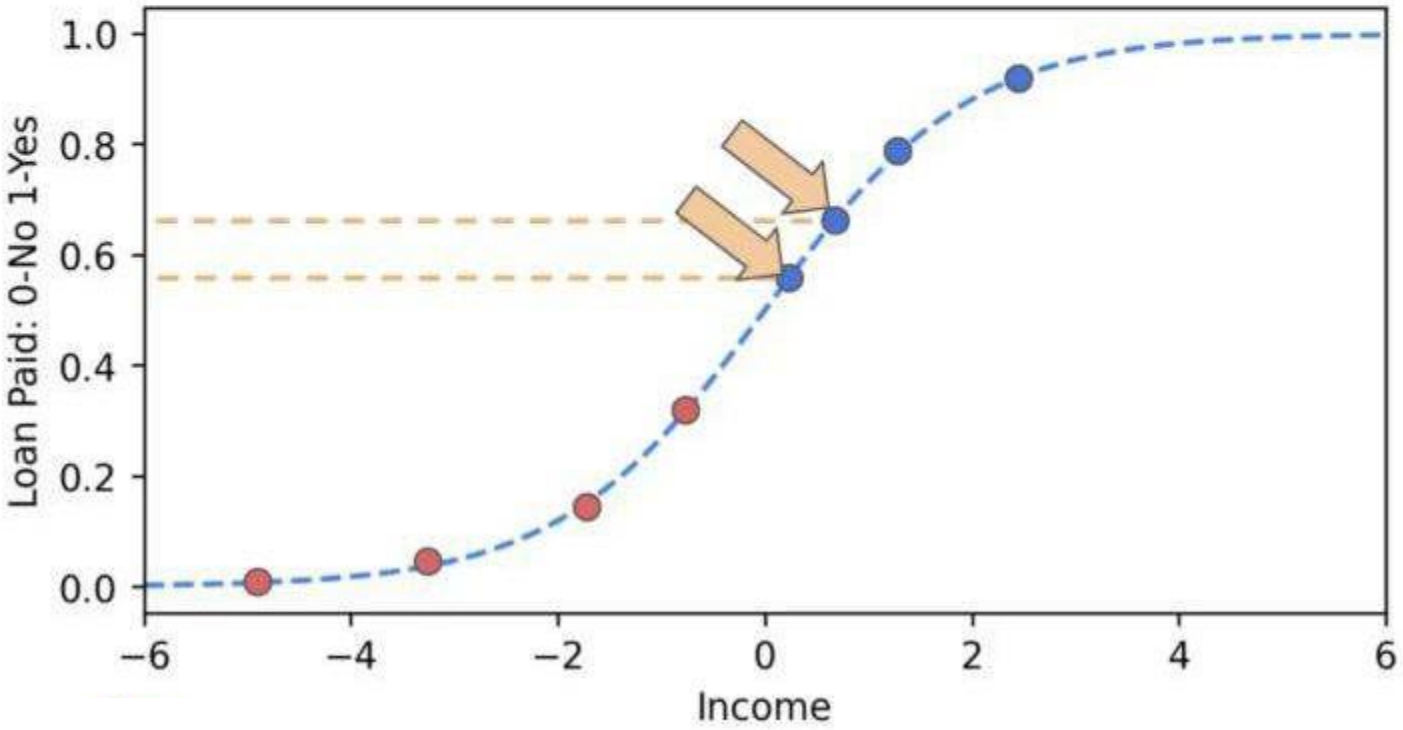
Logistic Regression

- Likelihood = 0.9 × 0.8 × ...



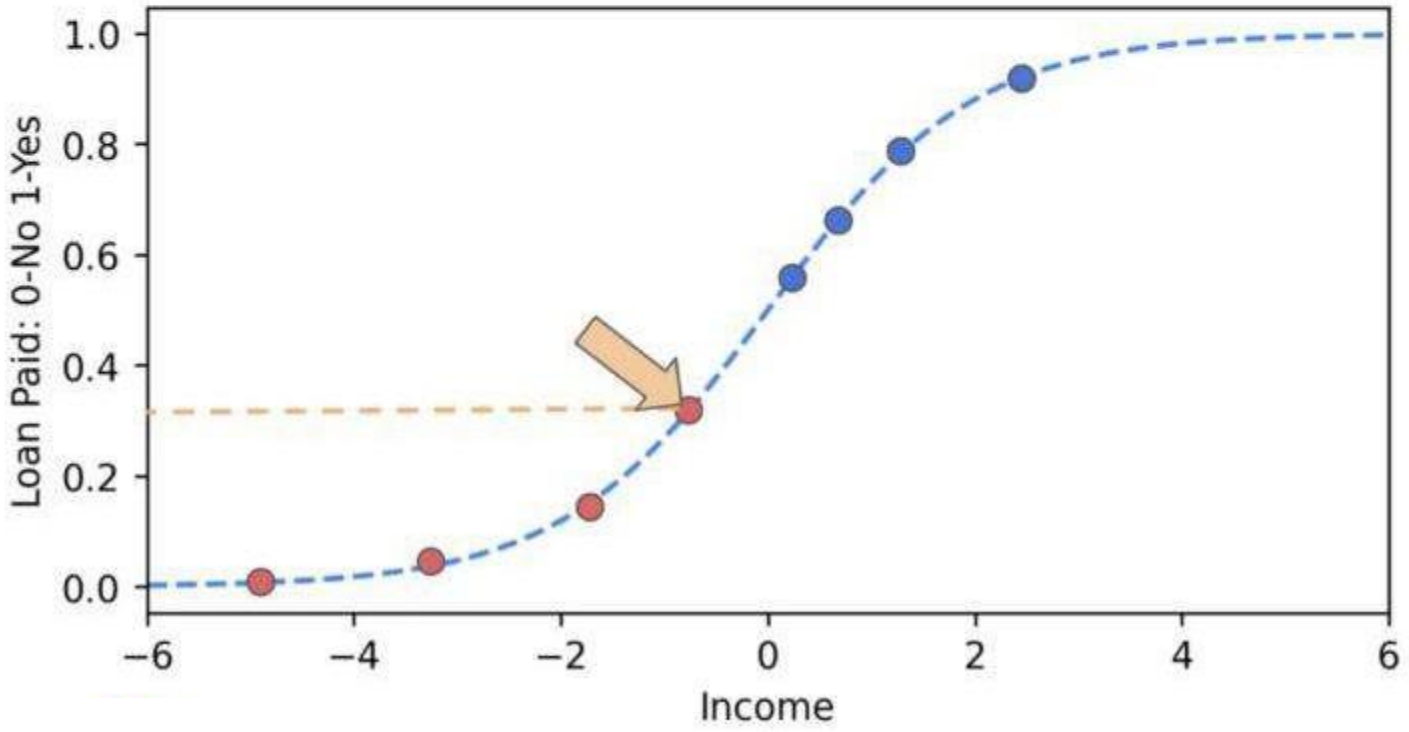
Logistic Regression

- Likelihood = 0.9 × 0.8 × 0.65 × 0.55 × ...



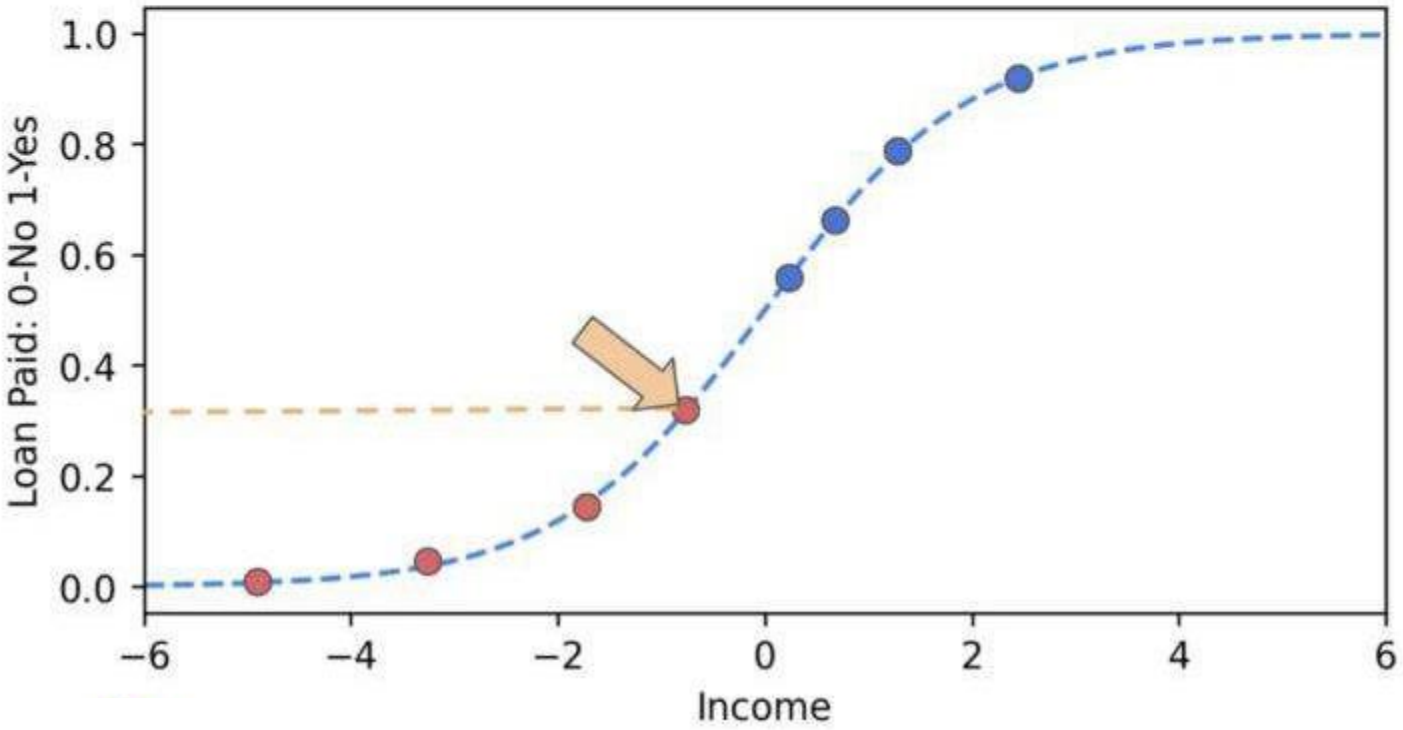
Logistic Regression

- Likelihood = 0.9 × 0.8 × 0.65 × 0.55 × (1-p) ×...



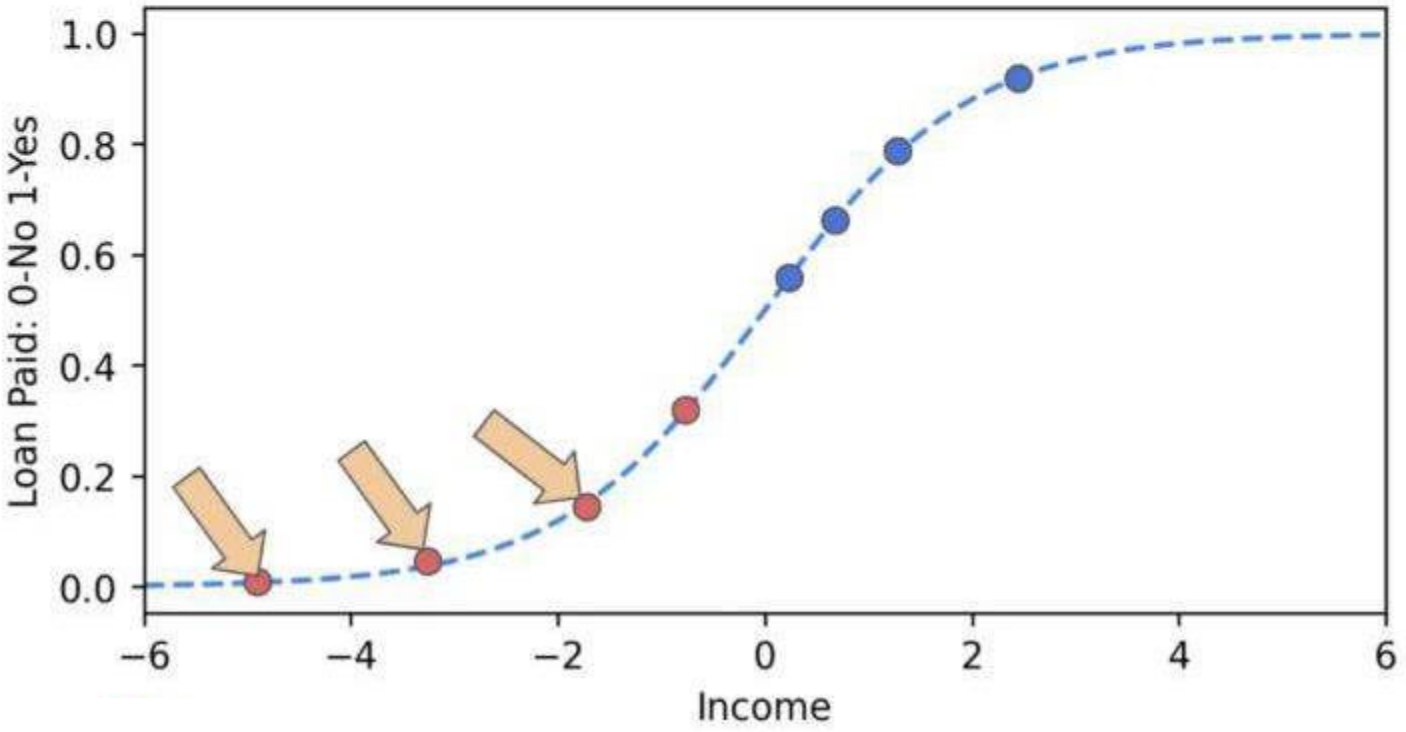
Logistic Regression

- Likelihood = 0.9 × 0.8 × 0.65 × 0.55 × (1-0.3) ×...



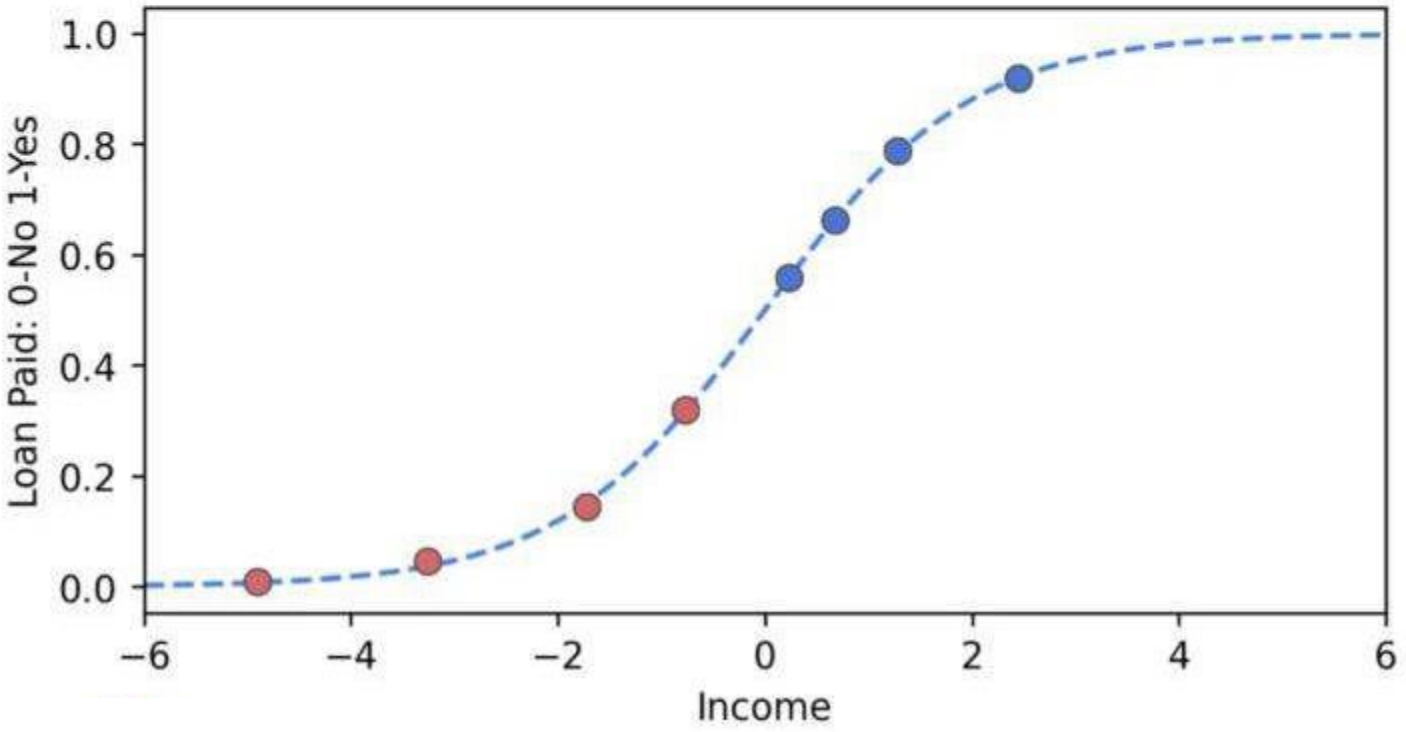
Logistic Regression

- Likelihood = 0.9 × 0.8 × 0.65 × 0.55 × (1-0.3) × (1-0.2) × (1-0.08) × (1-0.02)



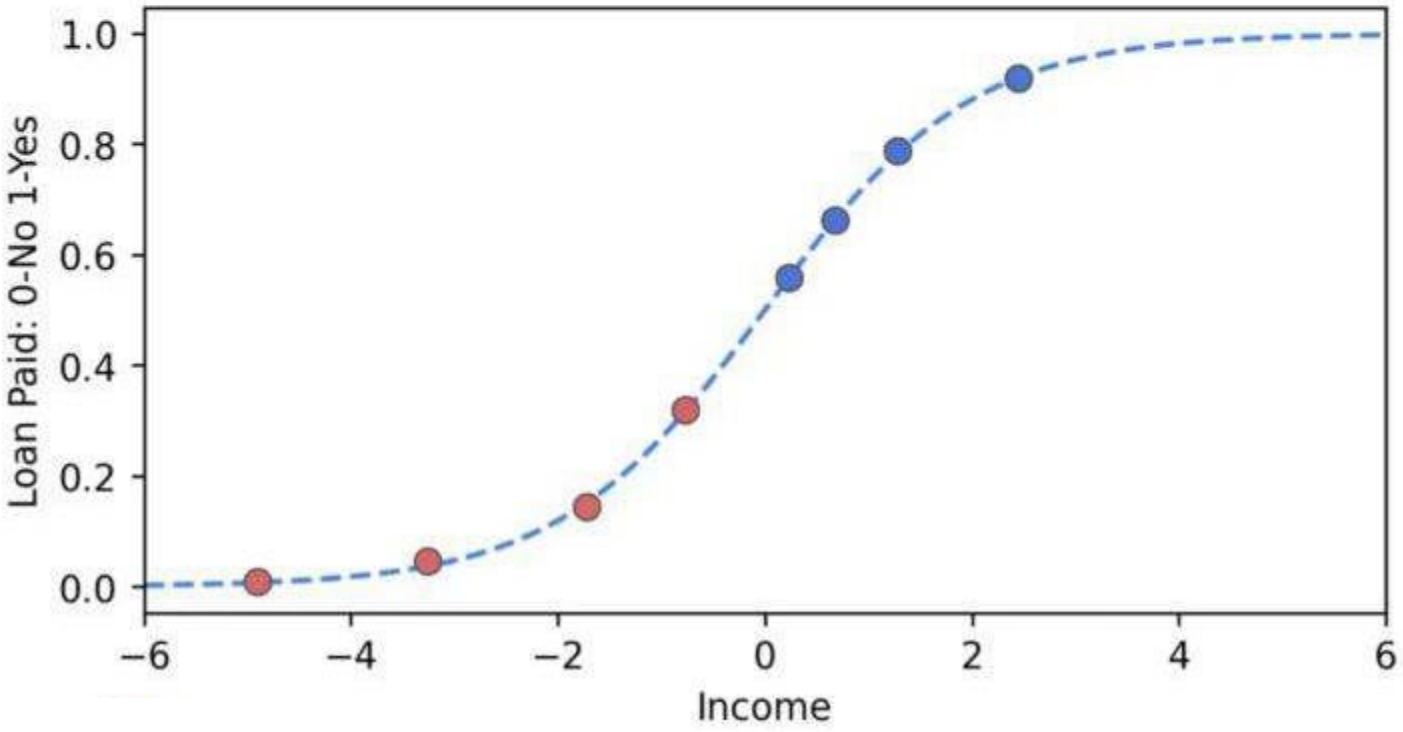
Logistic Regression

- Likelihood = 0.129



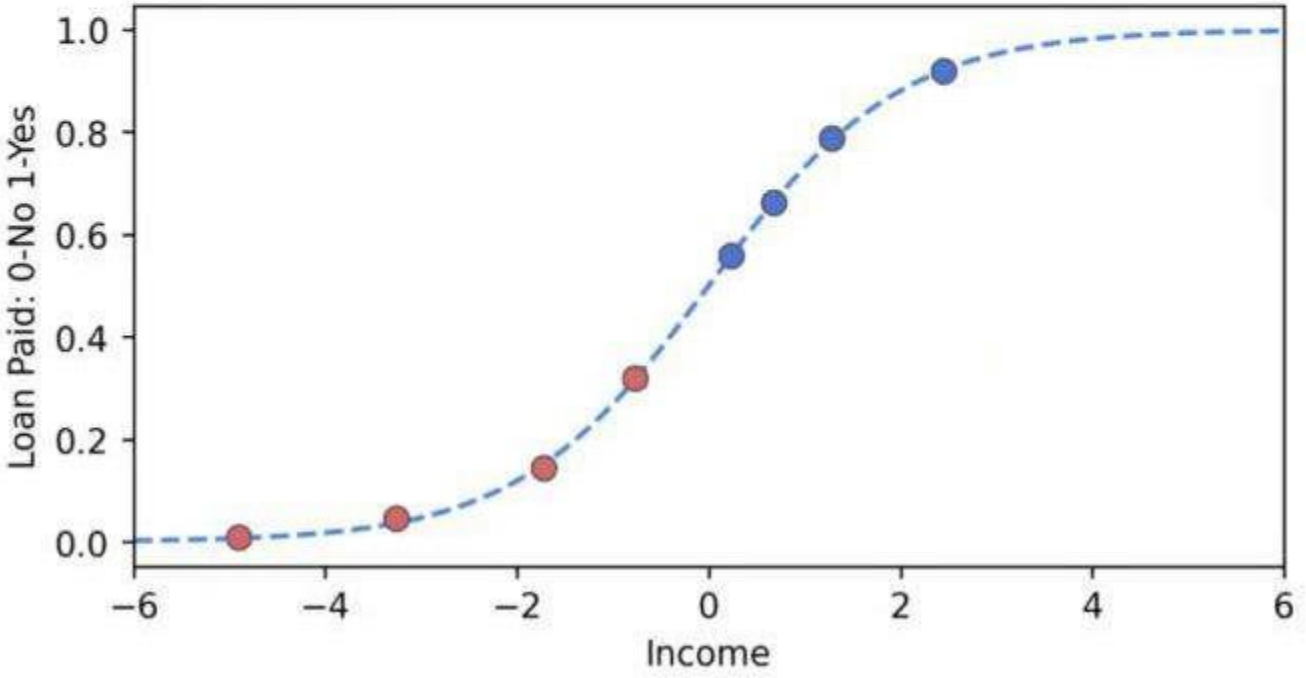
Logistic Regression

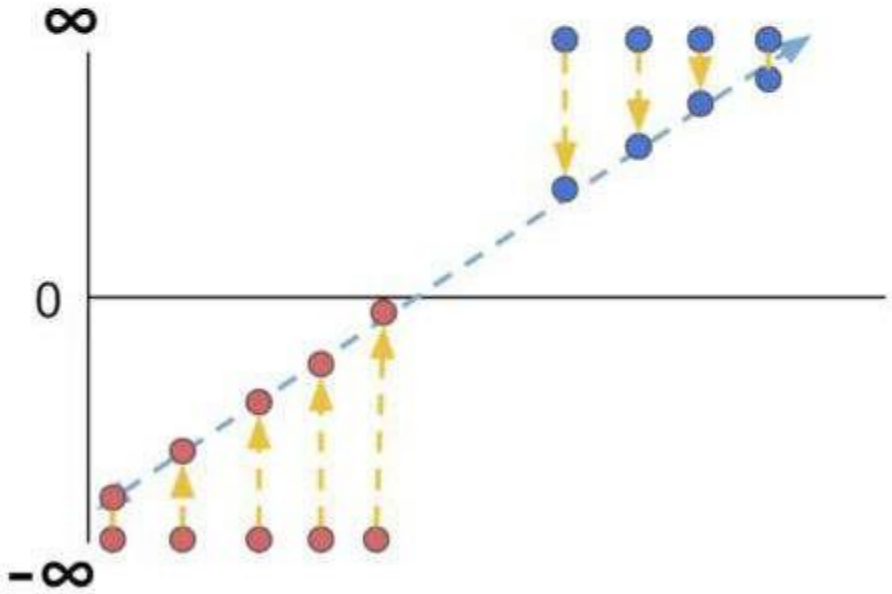
- Note in practice we actually maximize the **log** of the likelihoods. (e.g. ln(0.9)×ln(0.8)×…)



Logistic Regression

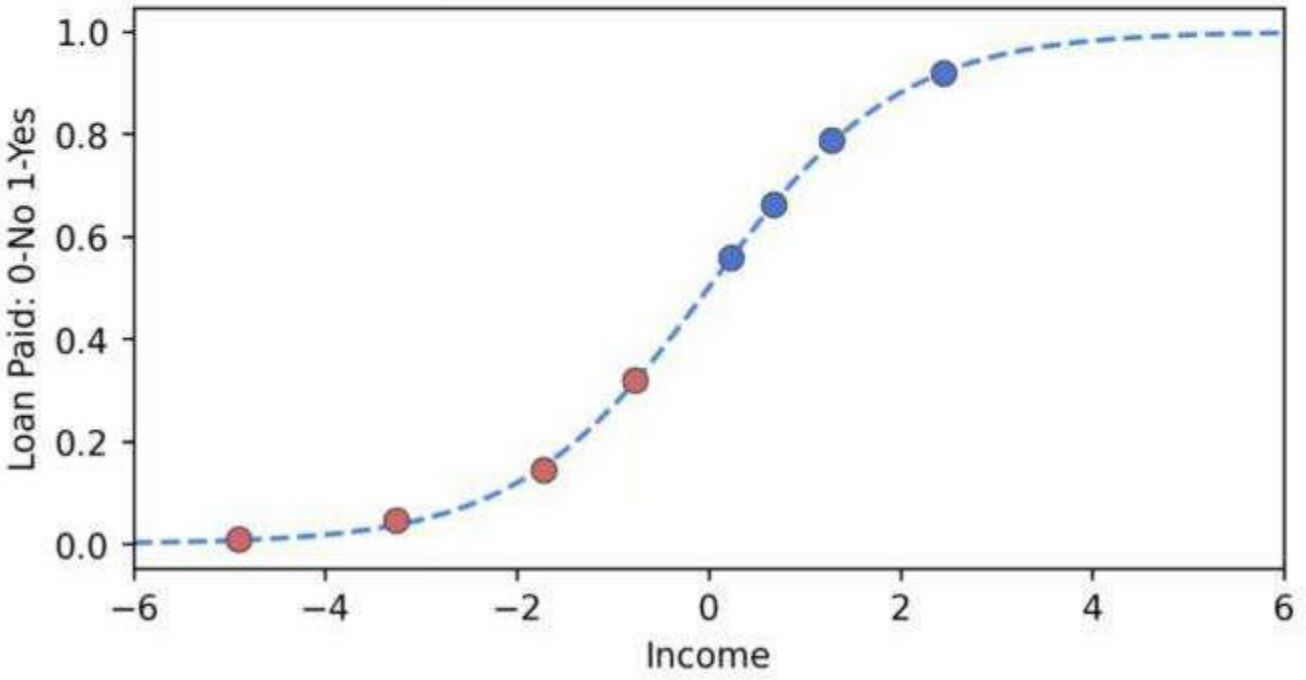
- There is some set of coefficients that will maximize these log likelihoods.

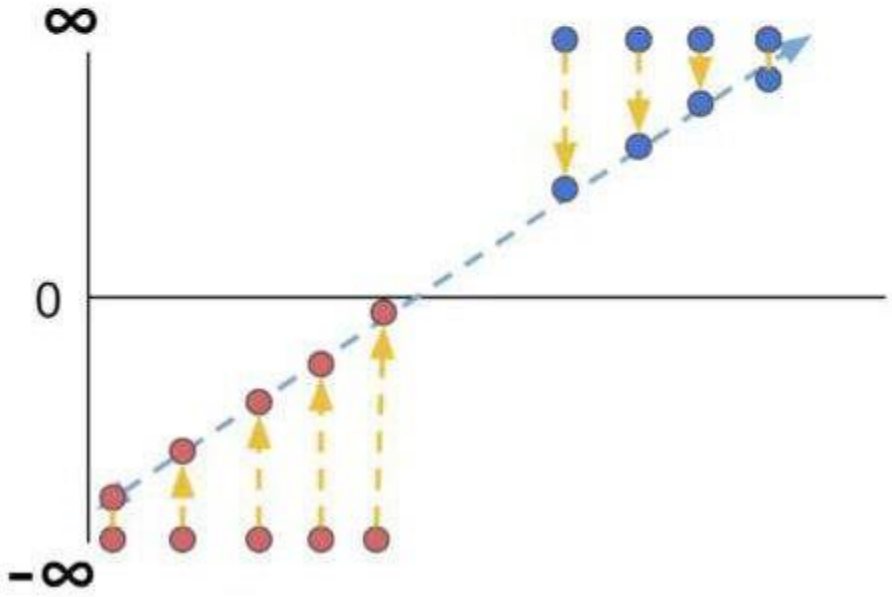




Logistic Regression

- Choose best coefficient values in log odds terms that creates maximum likelihood.





Logistic Regression

- While we are trying to **maximize** the likelihood, we still need something to **minimize**, since the computer's gradient descent methods can only search for minimums.

Logistic Regression

- In terms of a cost function, we seek to minimize the following (log loss):

$$J(\mathbf{x}) = -\frac{1}{m}\sum_{j=1}^{m} y^j \log\left(\hat{y}^j\right) + (1-y^j)\log\left(1-\hat{y}^j\right)$$

$$J(\mathbf{x}) = -\frac{1}{m}\sum_{j=1}^{m}\left(y^j \log\left(\frac{1}{1+e^{-\sum_{i=0}^{n}\beta_i x_i^j}}\right) + (1-y^j)\log\left(1-\frac{1}{1+e^{-\sum_{i=0}^{n}\beta_i x_i^j}}\right)\right)$$

Logistic Regression

- Just as with Linear Regression, gradient descent can solve this for us!

$$J(\mathbf{x}) = -\frac{1}{m}\sum_{j=1}^{m} y^j \log\left(\hat{y}^j\right) + (1-y^j)\log\left(1-\hat{y}^j\right)$$

$$J(\mathbf{x}) = -\frac{1}{m}\sum_{j=1}^{m}\left(y^j \log\left(\frac{1}{1+e^{-\sum_{i=0}^{n}\beta_i x_i^j}}\right) + (1-y^j)\log\left(1-\frac{1}{1+e^{-\sum_{i=0}^{n}\beta_i x_i^j}}\right)\right)$$

Logistic Regression

- Don’t worry about fully understanding this gradient descent.
- In practice we never have to implement it ourselves.
- Main takeaway should be the relationship between log odds and probability.

Logistic Regression

- Now that we have an intuition of what happens “behind the scenes”, let’s explore Logistic Regression with Python!