Datascience fundamentals

Week 5: Logistic Regression

Fernando Lovera
Mario Verstraeten

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? With logistic regression

Logistic Regression is a classification algorithm designed to predict categorical target labels.

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.
- Classification algorithms predict a class or category label:

  Class 0: Car Image

  Class 1: Street Image

  Class 2: Bridge Image

- 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:

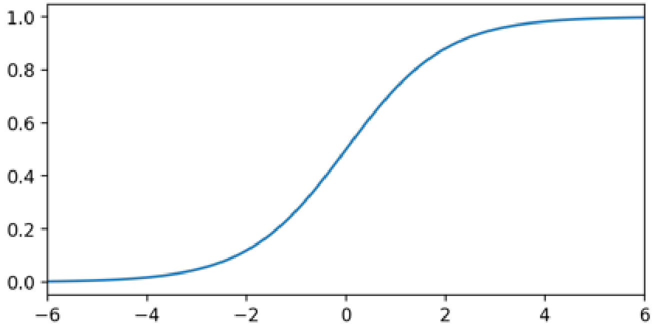
Class 0: 10% Probability - Car Image

Class 1: 85% Probability - Street Image

Class 2: 5% Probability - Bridge Image

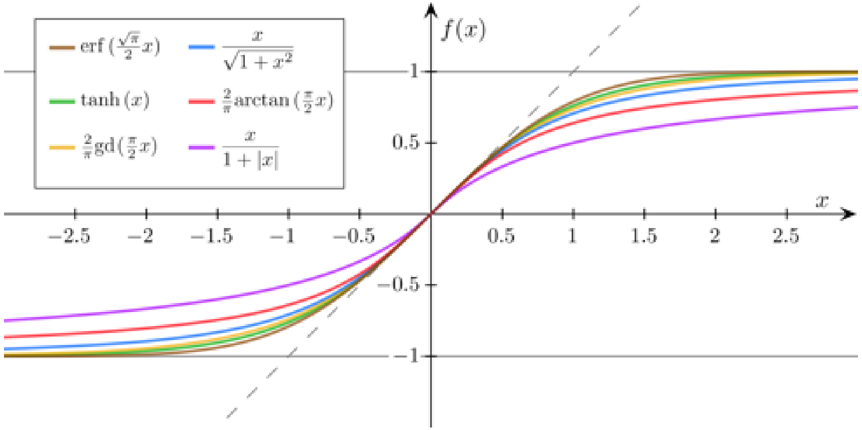
Model reports back prediction of Class 1, image is a street.

Logistic regression



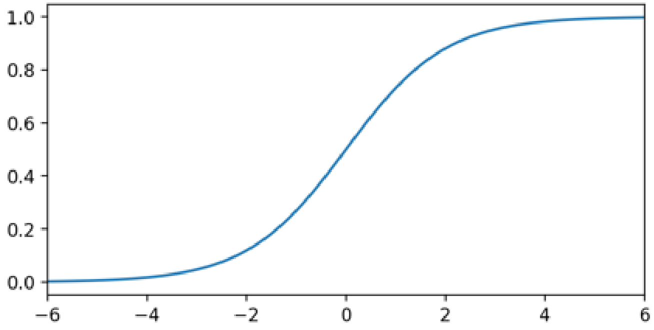
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Notice the “leveling off” behavior of the curve.



There is a “family” of logistic functions.

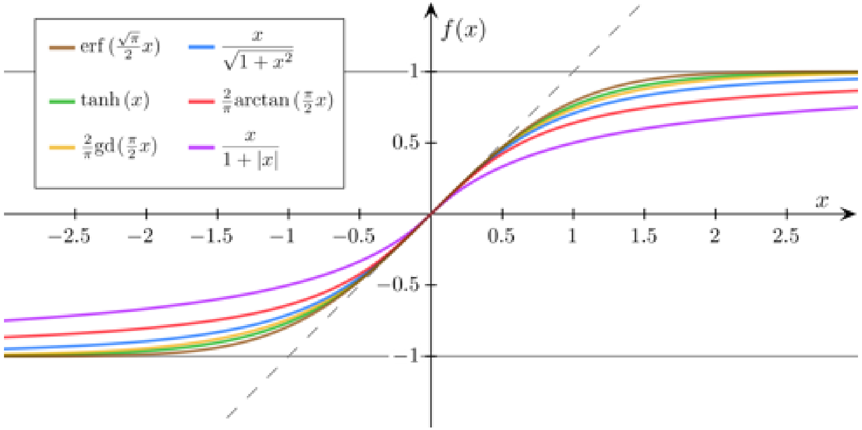
Logistic regression



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

Notice the "leveling off" behavior of the curve.

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



There is a "family" of logistic functions.

Logistic regression

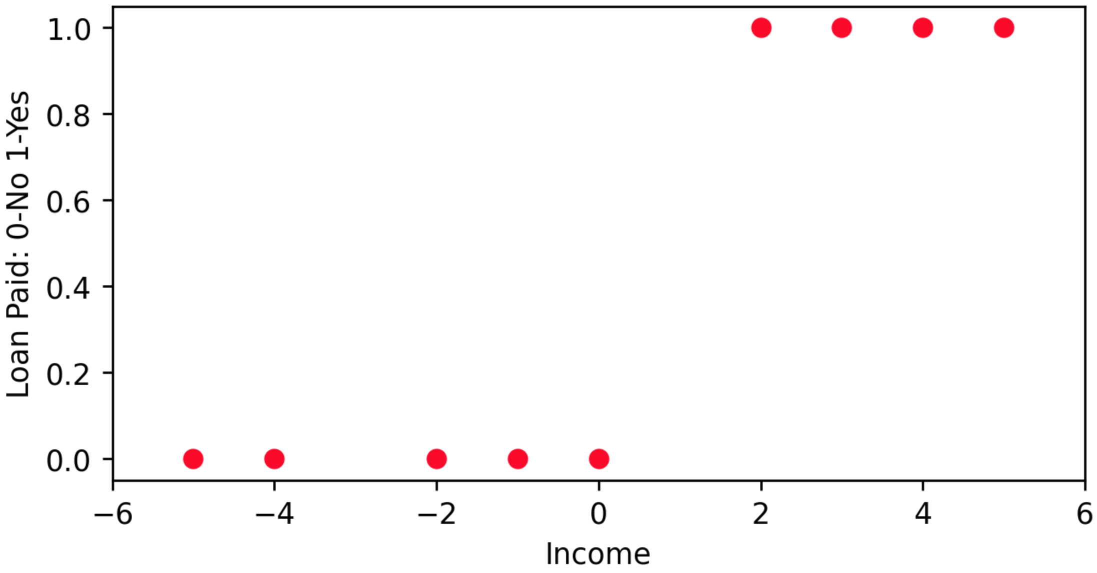
Let's explore how to convert a Linear Regression model used for a regression task into a Logistic Regression model used for a classification task.
Imagine a dataset with a single feature (previous year's income) and a single target label (loan default)

Our data set:

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

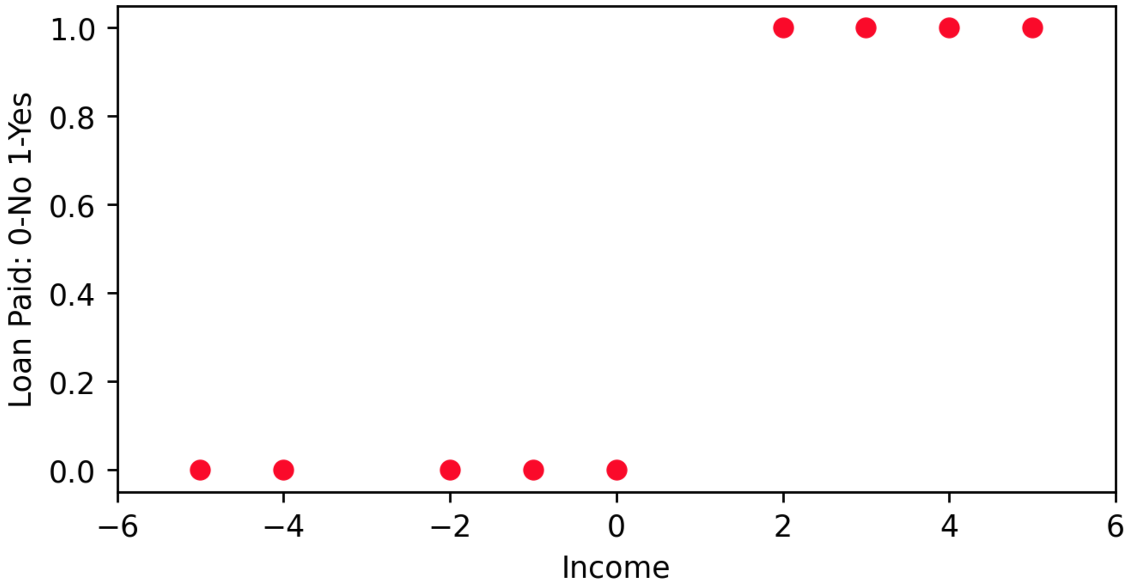
Logistic regression

Let's begin by plotting income versus default:



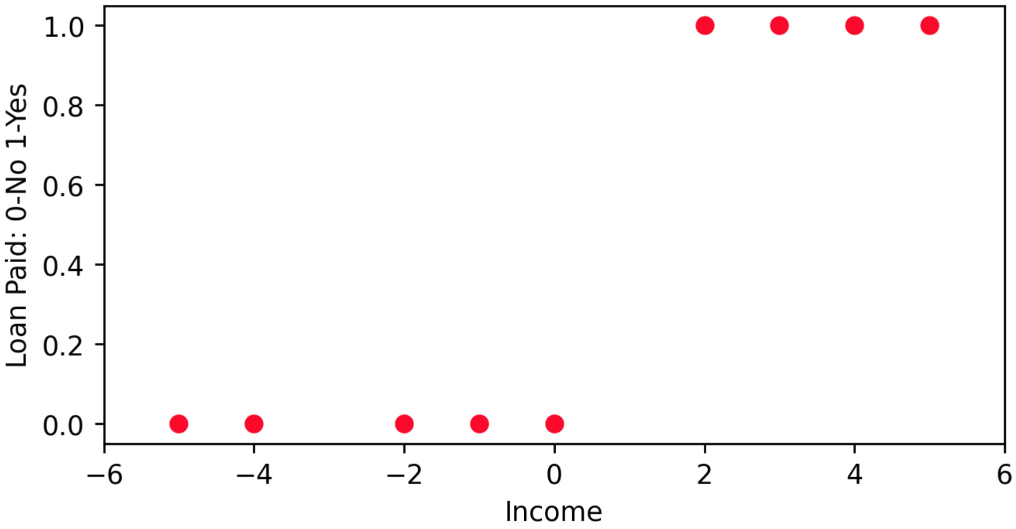
Logistic regression

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



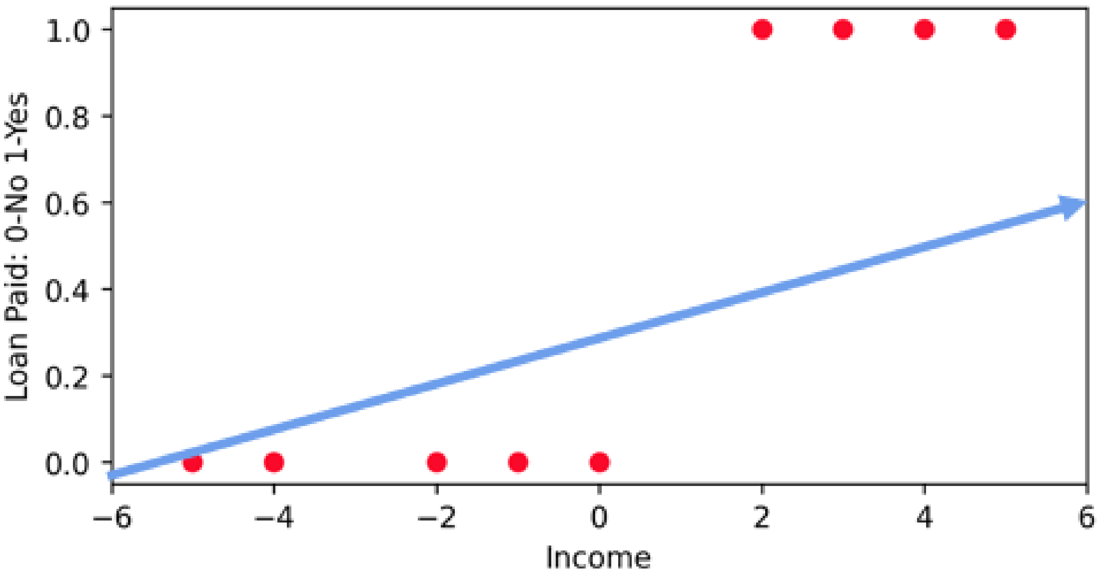
Logistic regression

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



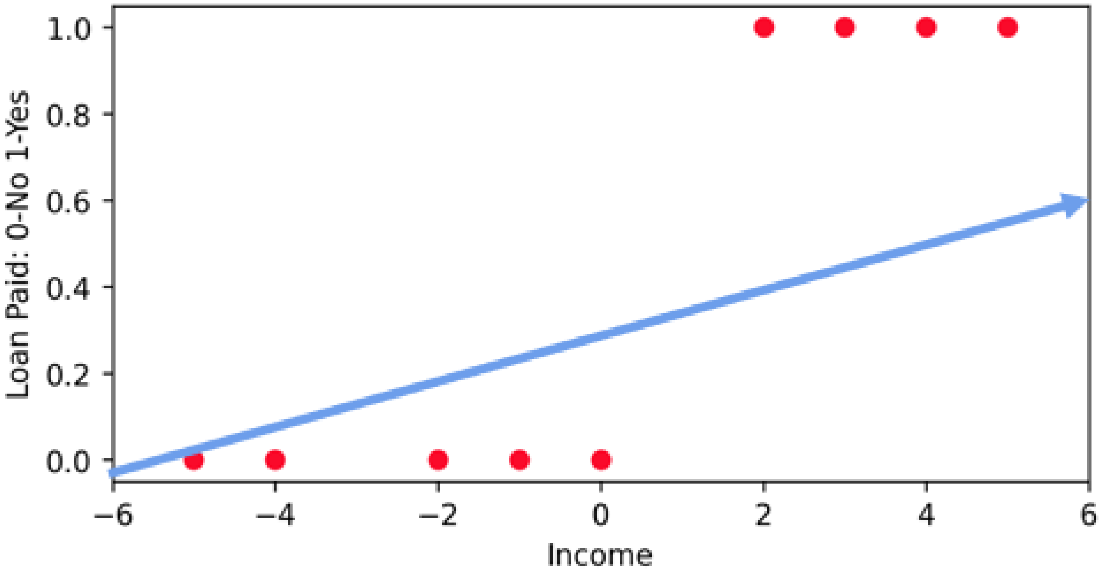
Logistic regression

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



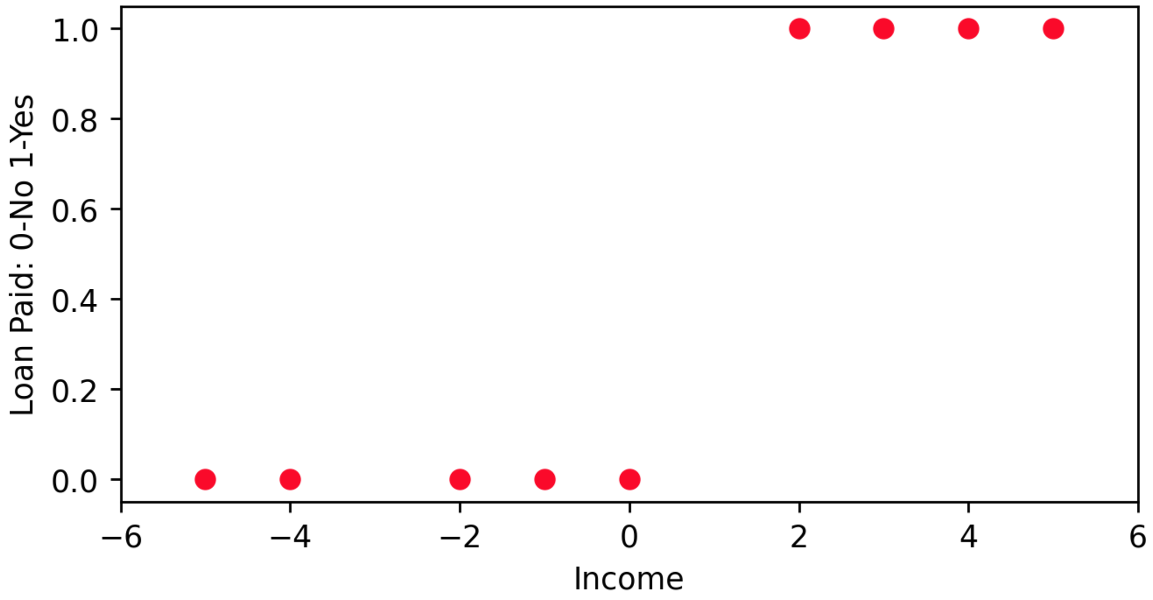
Logistic regression

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



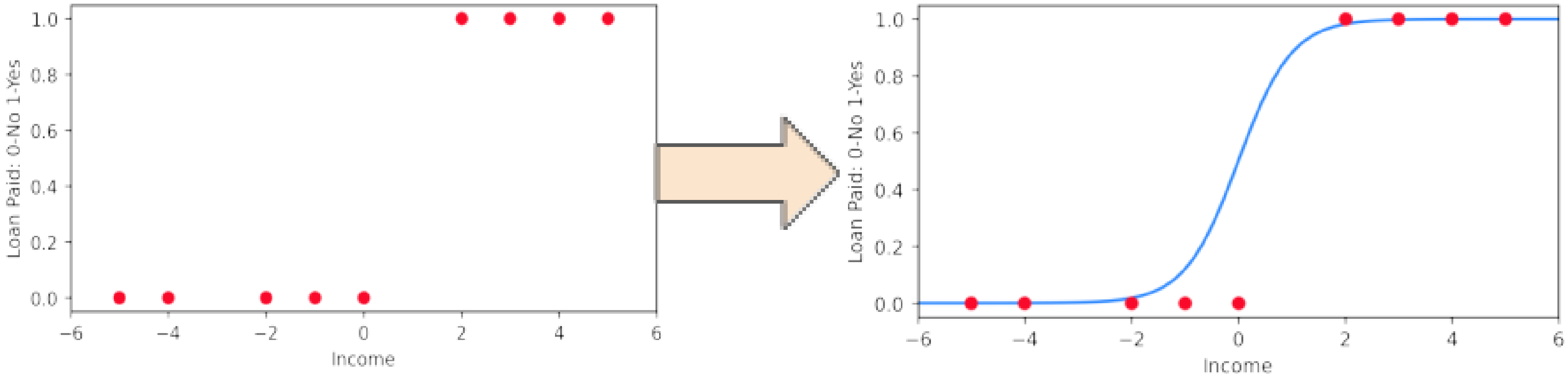
Logistic regression

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



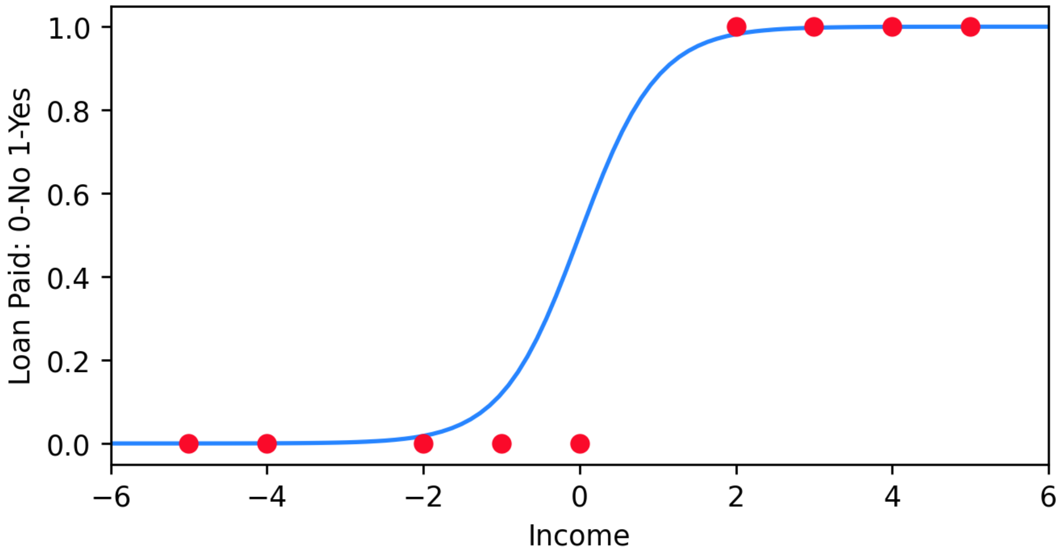
Logistic regression

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



Logistic regression

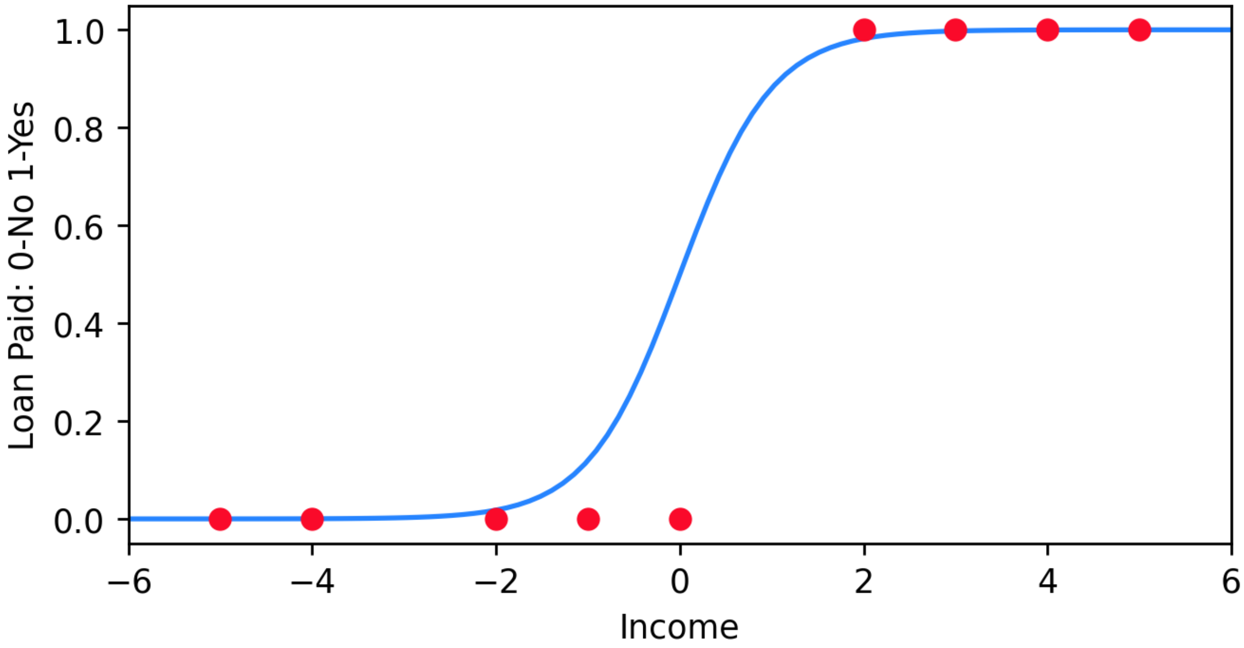
Let’s first focus on what this Logistic Regression would look like.



Logistic regression

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

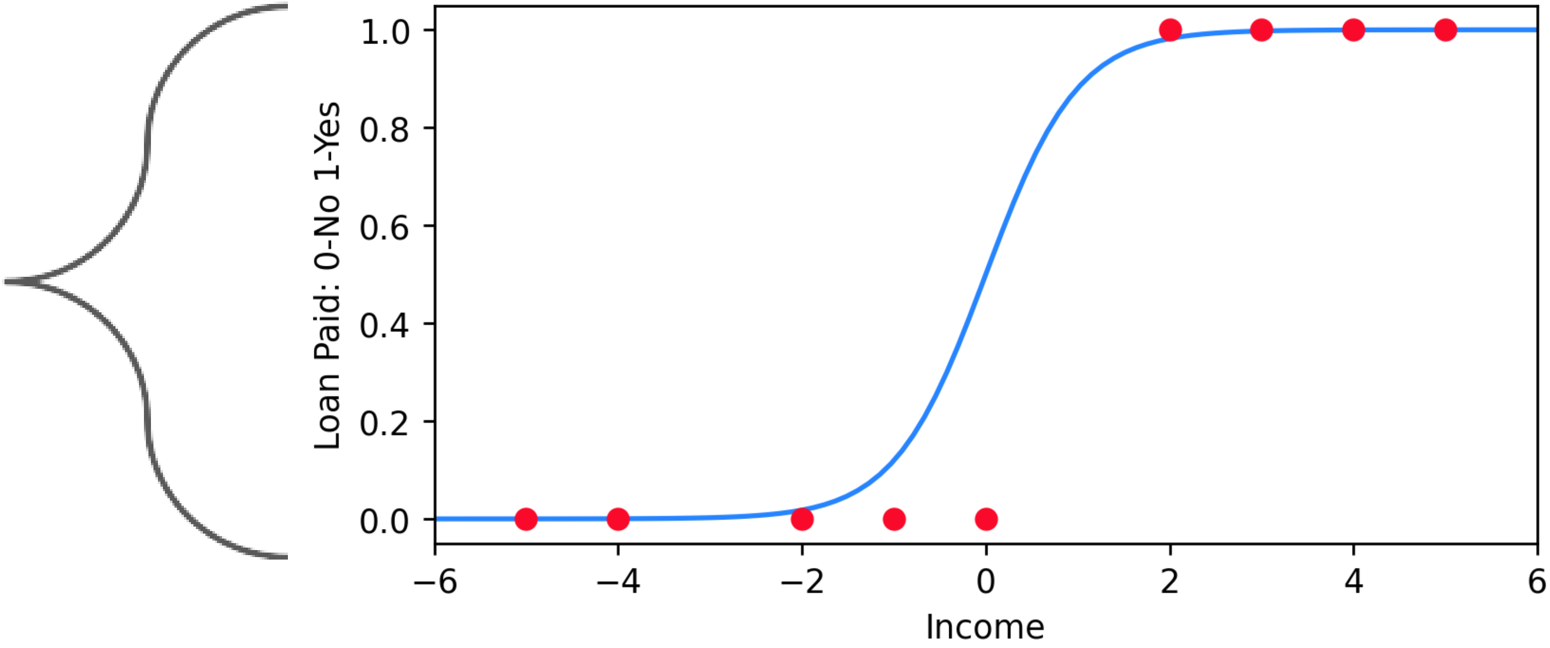
P(y=1)



Logistic regression

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

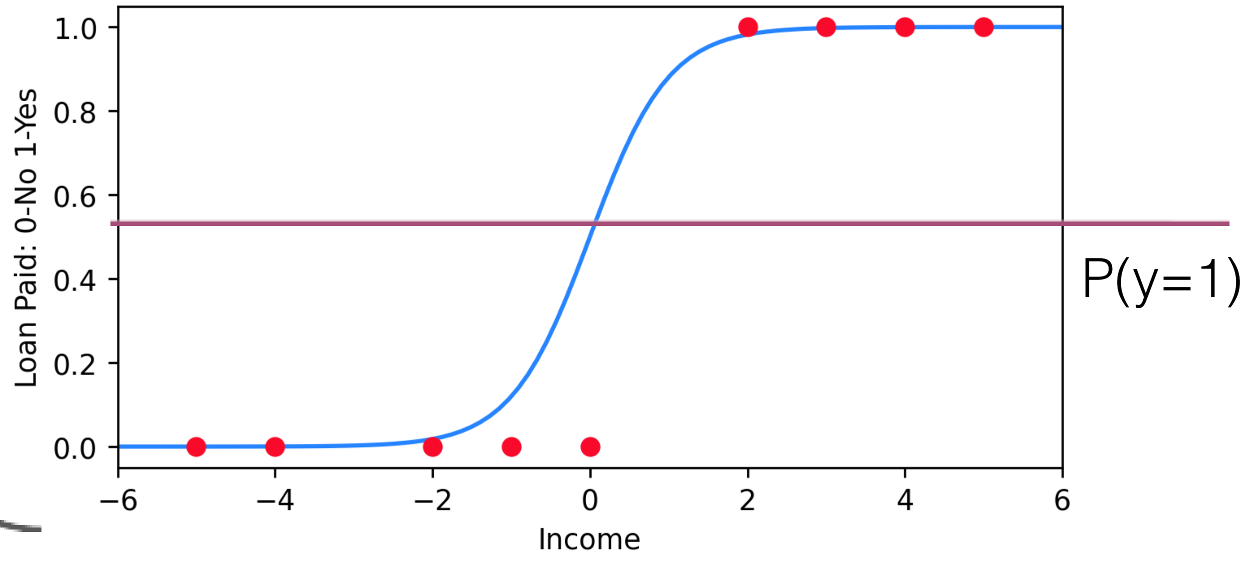
P(y=1)



Logistic regression

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

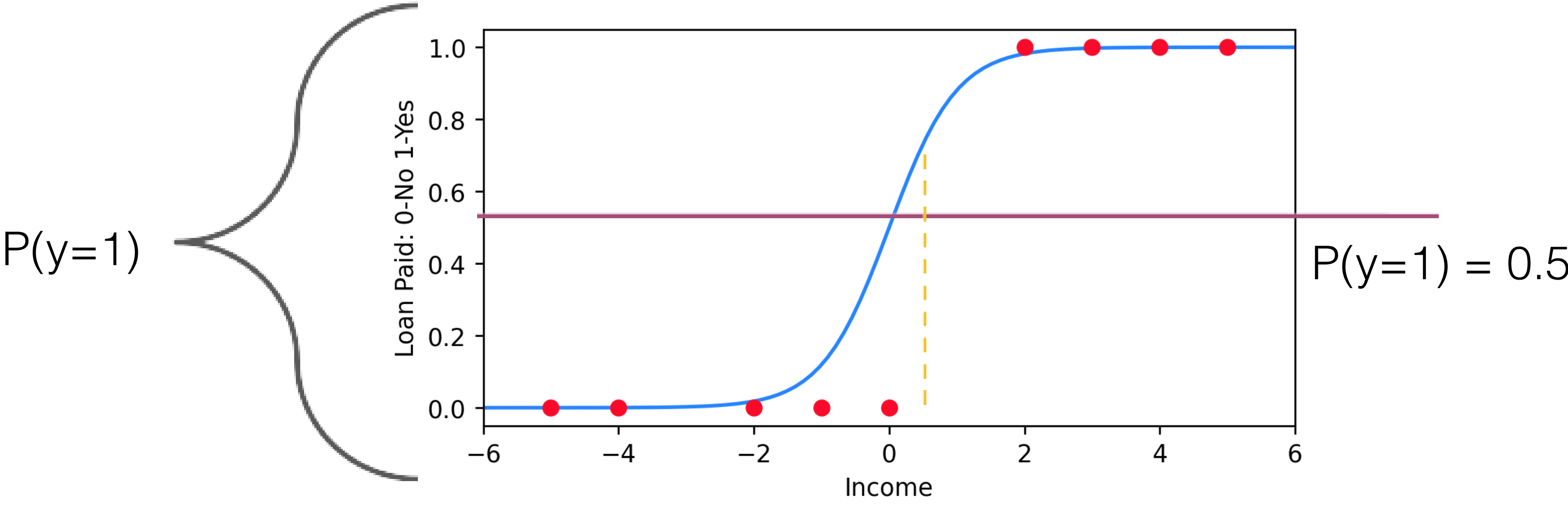
P(y=1)



P(y=1) = 0.5

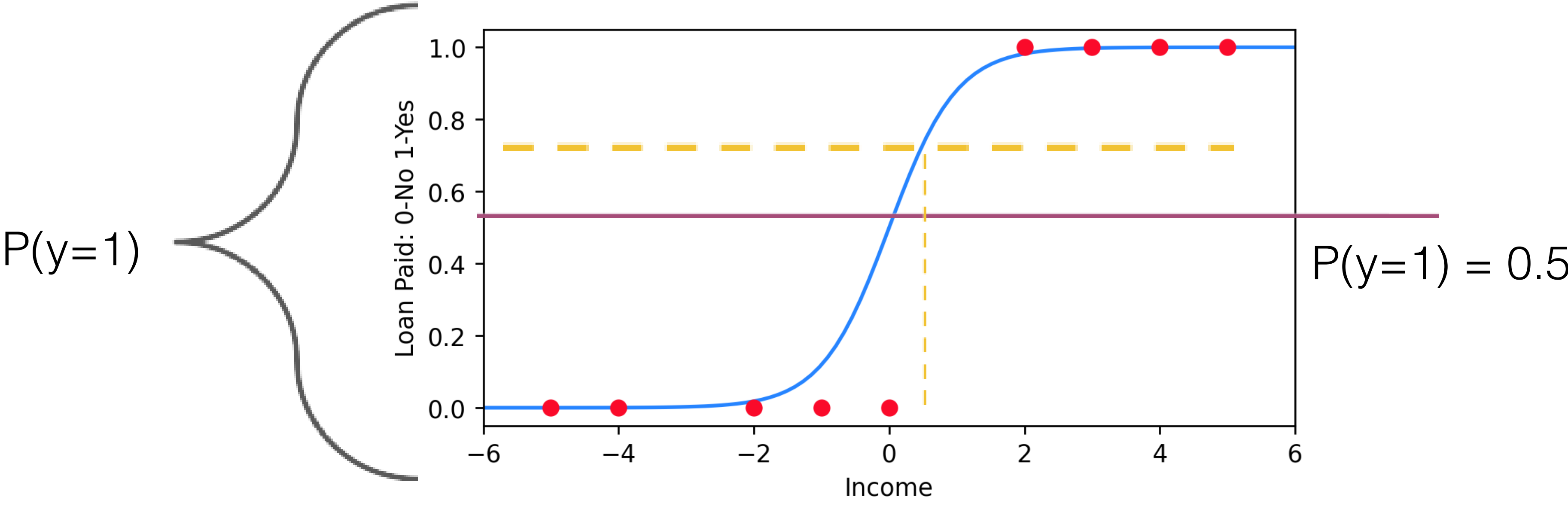
Logistic regression

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



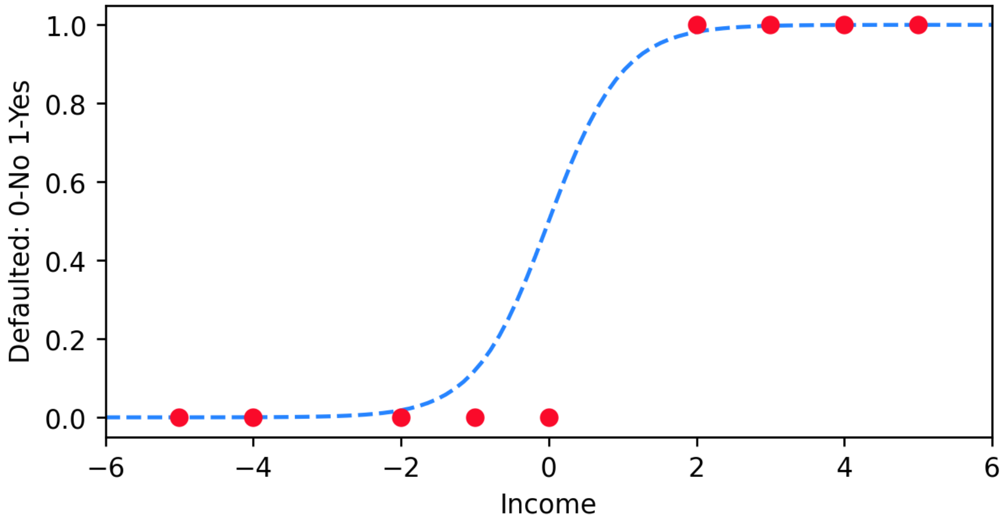
Logistic regression

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



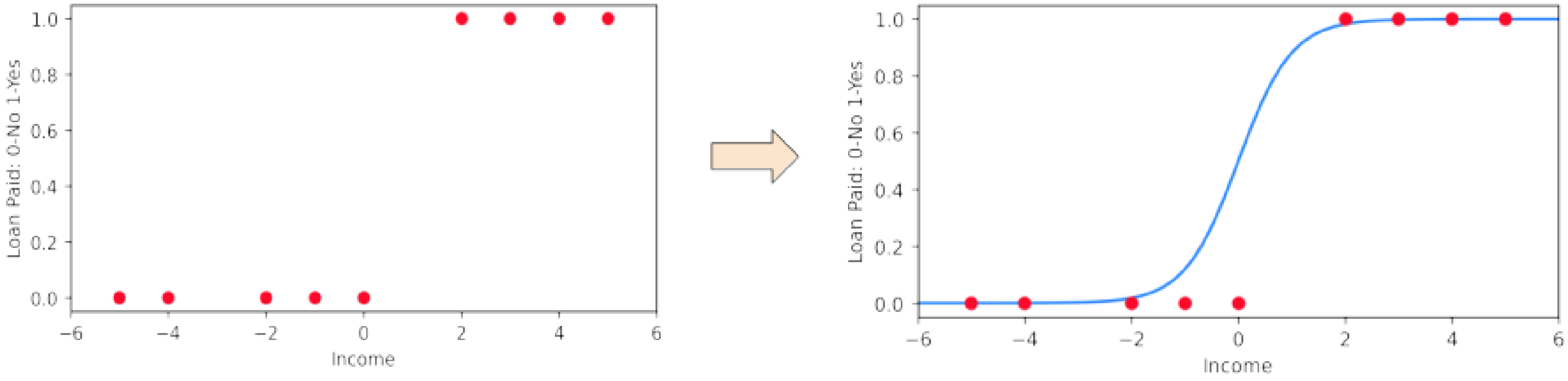
Logistic regression

But how do we actually create this line?



Logistic regression

Fortunately, the mathematics of the conversion are quite simple!



Logistic regression

We already know the Linear Regression equation:

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

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

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

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

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

$$\sigma(x) = \frac{1}{1 + e^{-x}} \Rightarrow \hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

Logistic regression

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

After some mathematical magic, we can make it look familiar again

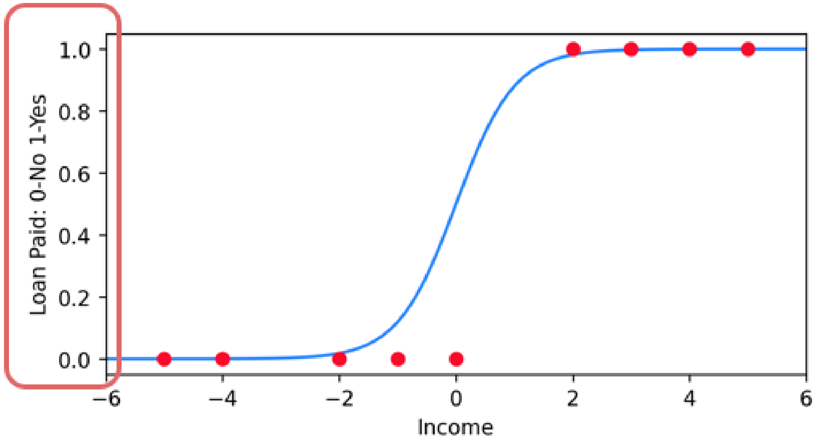
$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

Looks like the formula for linear regression

Logistic regression

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

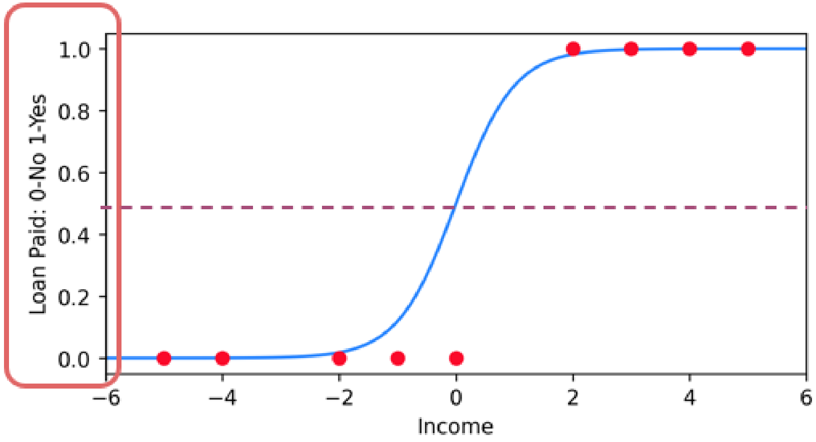
$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) \Leftarrow \hat{y}$$



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 ∞

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

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

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

As p goes to 0 then log odds becomes -∞

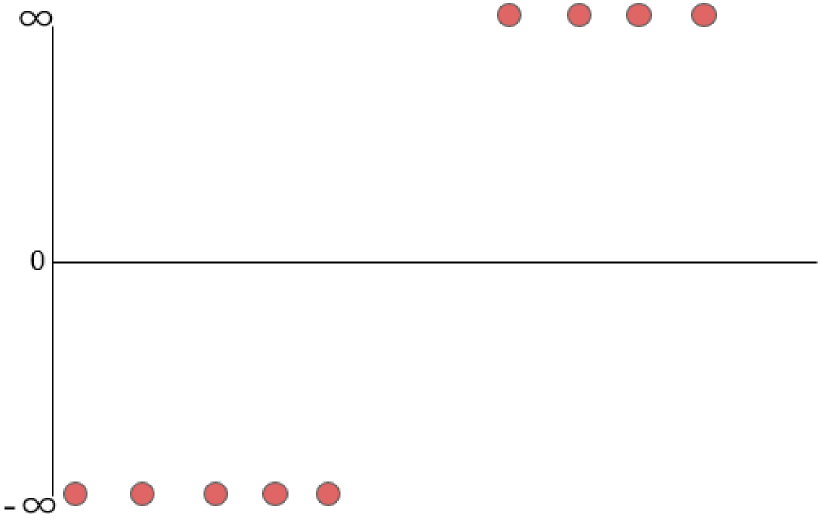
Logistic regression

Class points now at infinity

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

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

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



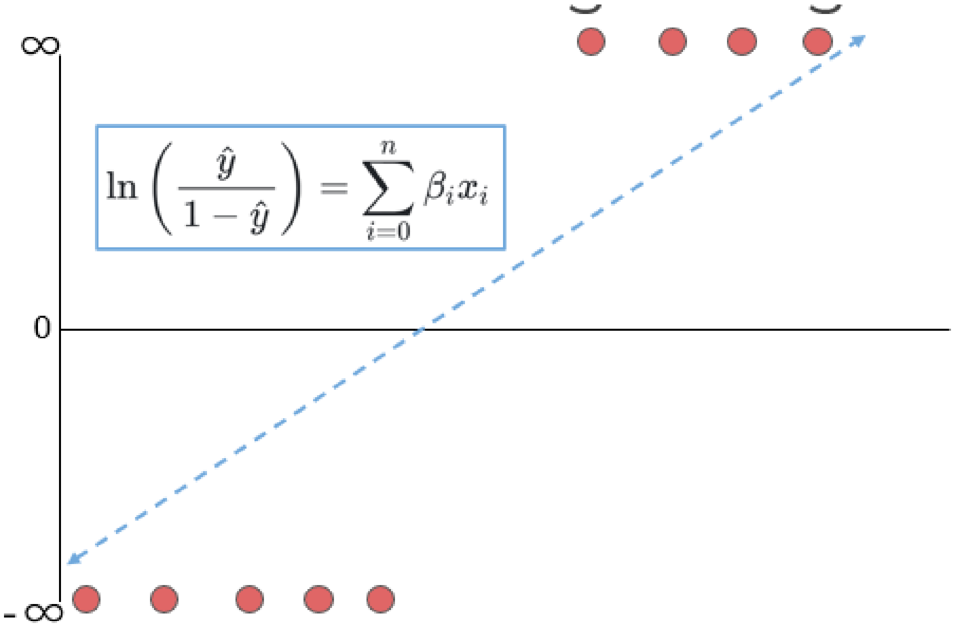
Logistic regression

- On log scale logistic function is straight line
- Coefficients in terms of change in log odds.

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

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

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



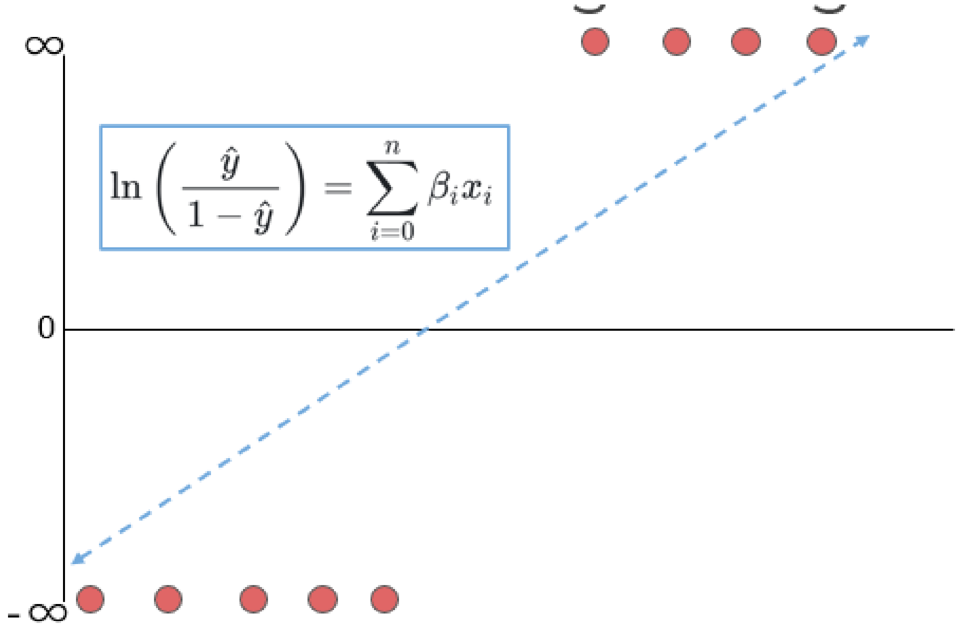
Logistic regression

- On log scale logistic function is straight line
- Coefficients in terms of change in log odds.

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

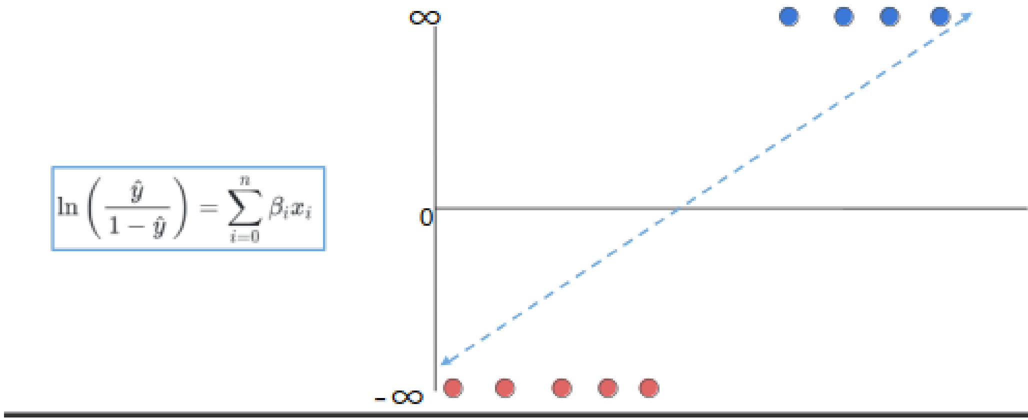
$$ln(\frac{0.5}{1-0.5}) = 0$$

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



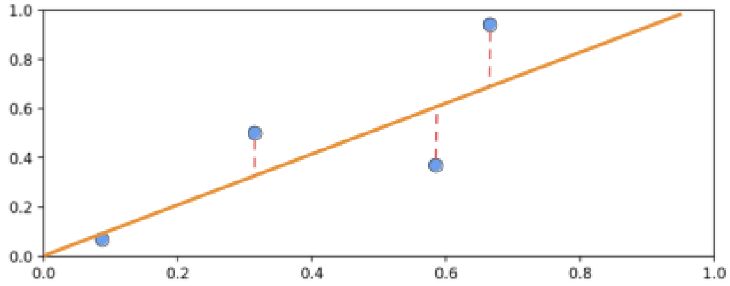
Logistic regression - Finding the Best Fit

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



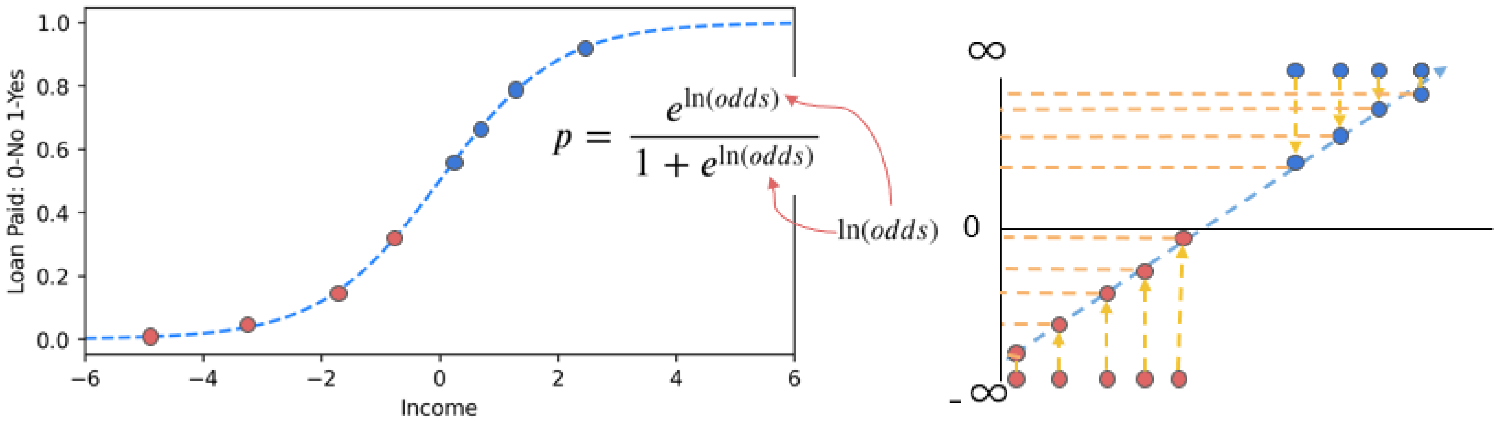
Logistic regression - Finding the Best Fit

- Logistic Regression uses Maximum Likelihood to find the best fitting model.
- We'll also then display the cost function and gradient descent that is solved for by the computer.
- Recall in Linear Regression we seek to minimize the Residual Sum of Squares (RSS).



Logistic regression - Finding the Best Fit

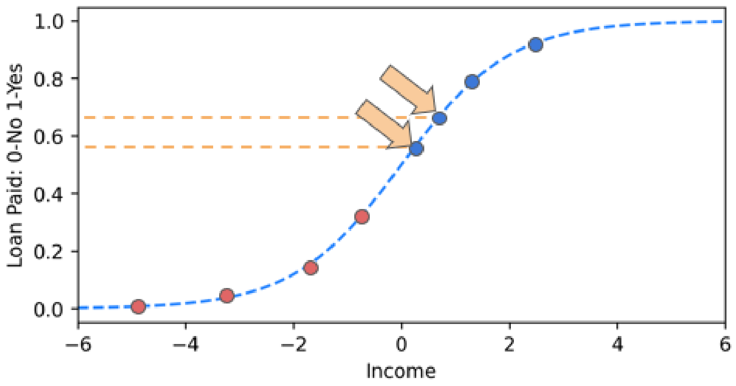
We are able to convert ln(odds) into a probability.



Logistic regression - Finding the Best Fit

How do we find the most optimized function? Maximizing the **likelihood**

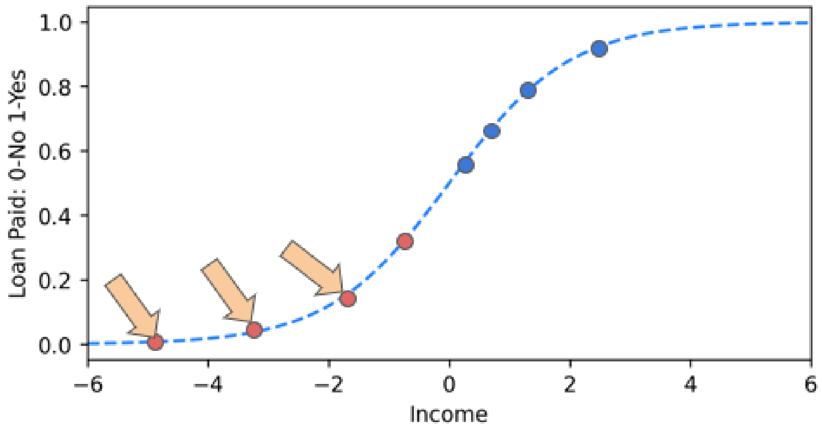
Likelihood = 0.9 × 0.8 × 0.65 × 0.55 × ...



Logistic regression - Finding the Best Fit

How do we find the most optimized function? Maximizing the **likelihood**

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

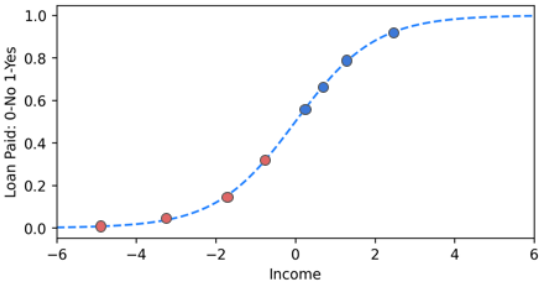


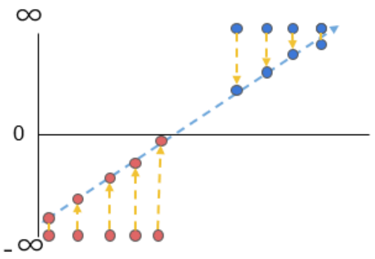
Likelhood = 0.129

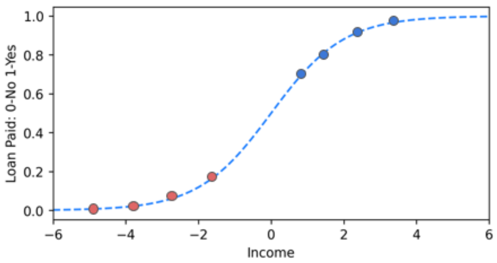
Logistic regression - Finding the Best Fit

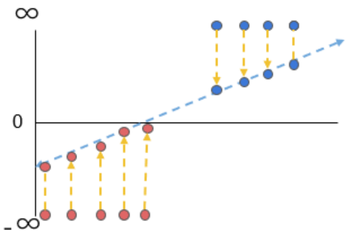
!Note in practice we actually maximize the **log** of the likelihoods. (e.g. ln(0.9)×ln(0.8)×…)

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









Fitting the line that maximizes the log of likelihoods, is (just like with linear regression) done by performing the gradient descent algortithm

Logistic regression - Practice

Let's check the notebooks!