Logistic Regression…

and the modeling philosophy

Agenda

- Overview of Machine Learning
- Intro to Classification
- What is Logistic Regression?
- The Inner Mechanics of Logistic Regression
- Logistic Regression in sklearn

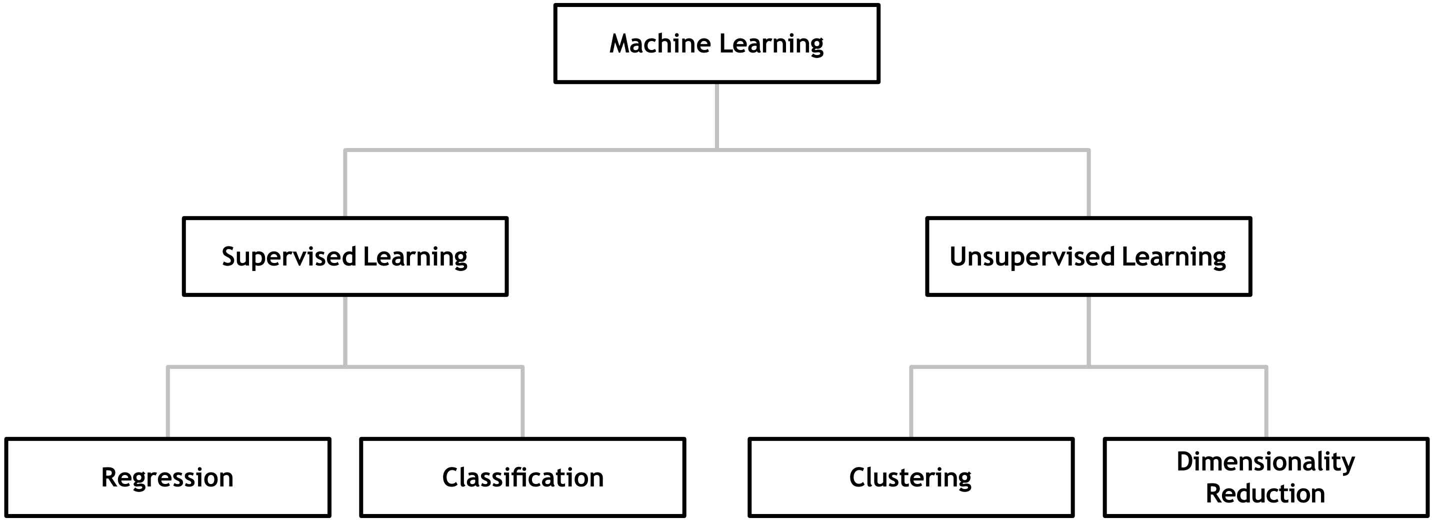
Learning Objectives

- Distinguish between regression and classification problems.
- Understand how logistic regression is similar to and different from linear regression.
- Fit, generate predictions from, and evaluate a logistic regression model in sklearn.
- Understand how to interpret the coefficients of logistic regression.
- Know the benefits of logistic regression as a classifier.

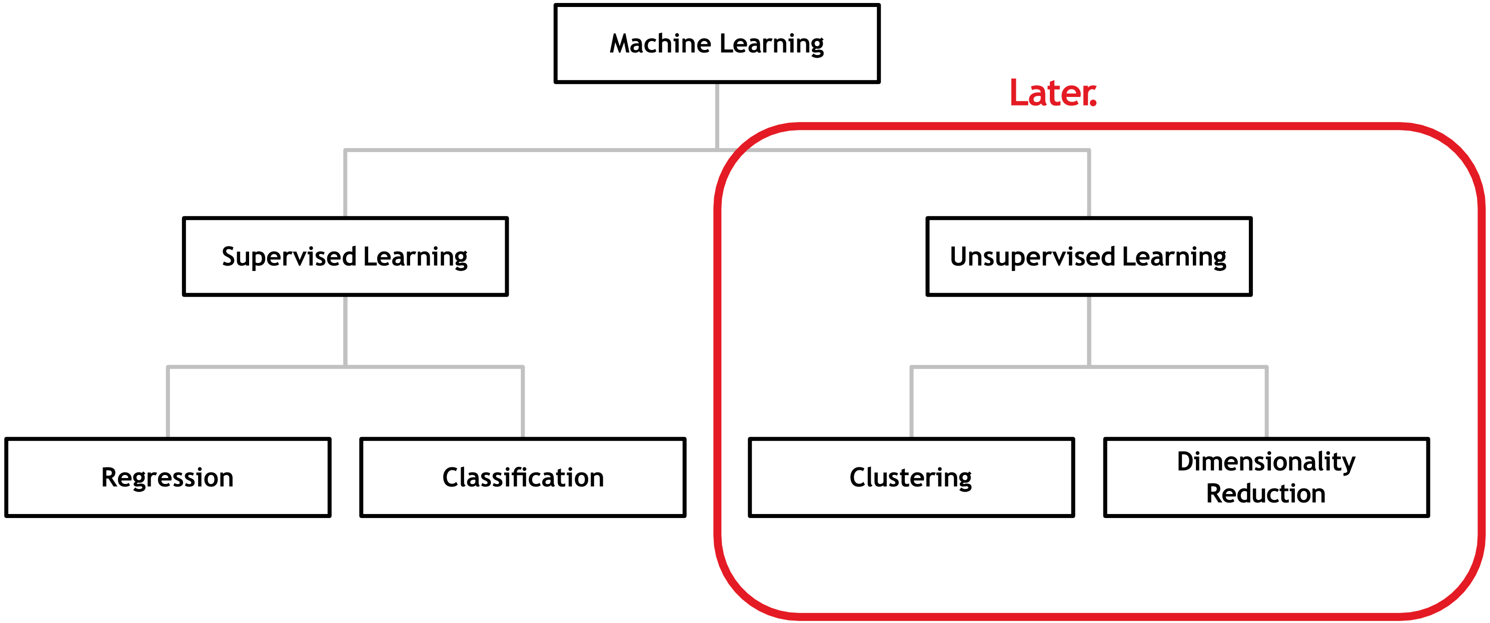
Data Science Process

1. Define problem.
2. Gather data.
3. Explore data.
4. Model with data.
5. Evaluate model.
6. Answer problem.

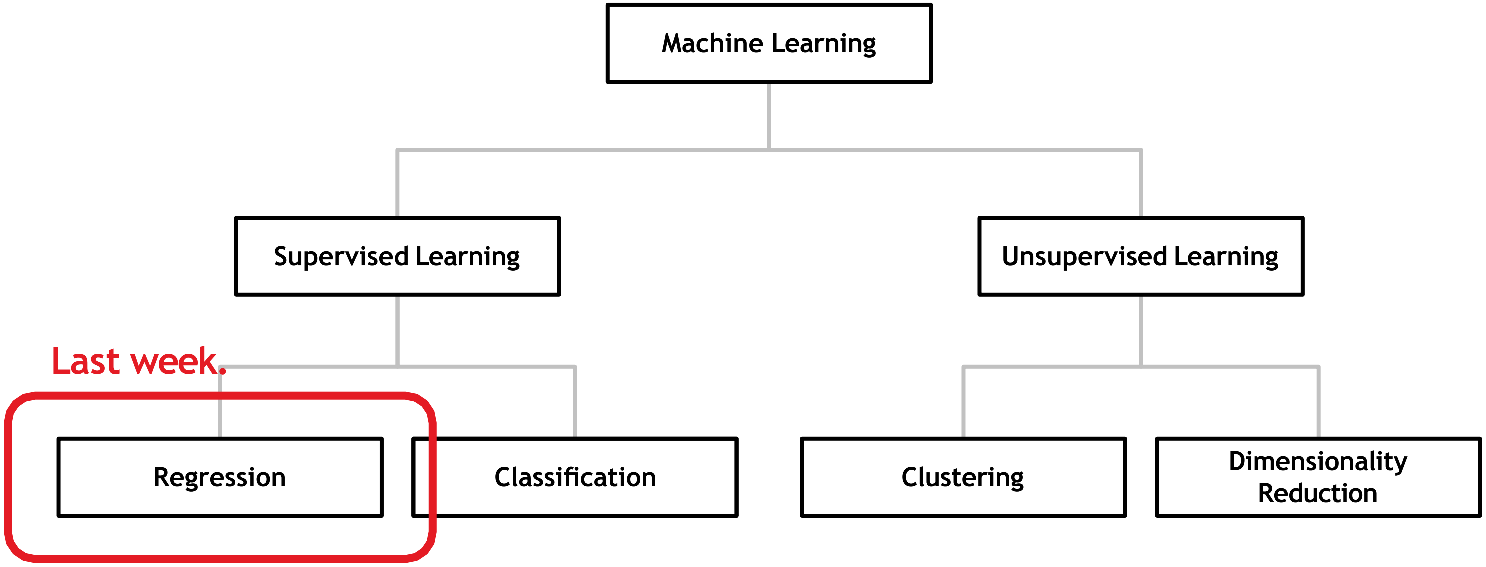
Roadmap of ML



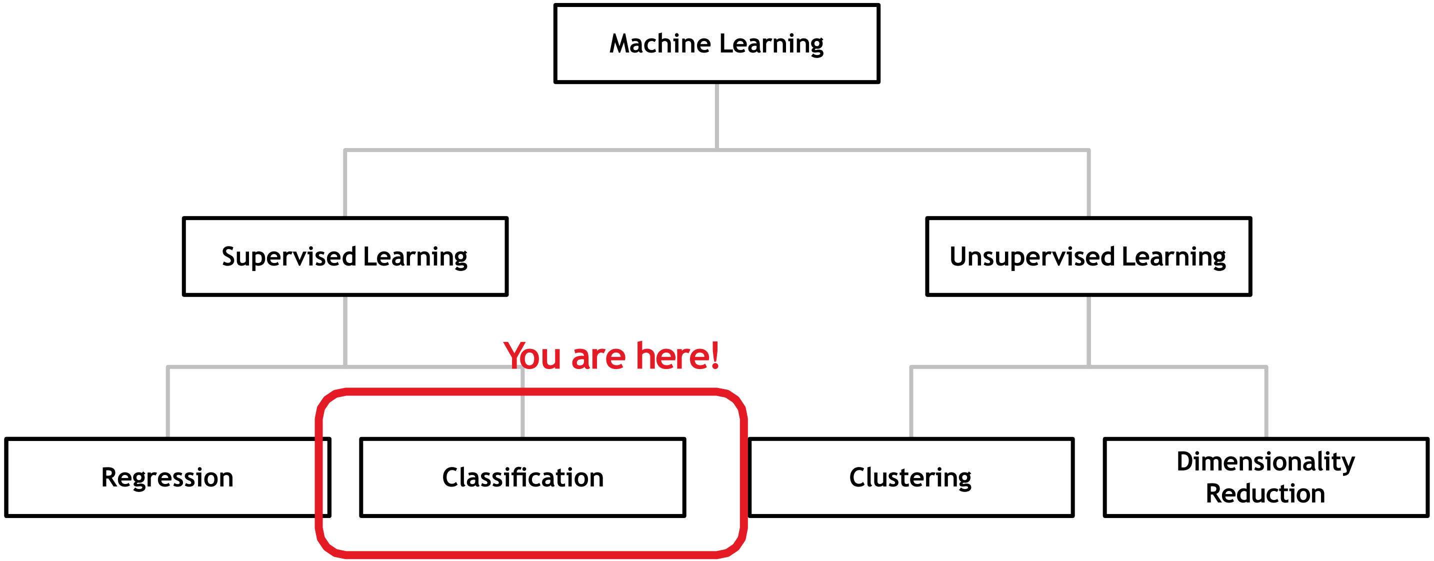
Roadmap of ML



Roadmap of ML



Roadmap of ML



Reminder: Regression

Reminder: Regression

Regression - this is when our *y*-variable is numeric.

- *"Given the past values of the stock price of Apple, what will tomorrow's closing price be?"*
- *"Given the annual precipitation, average temperature, and soil pH, what will this year's harvest yield be?"*
- *"Given the square footage, number of bedrooms, number of bathrooms, and quality of school district, what will the price of this home be?"*

Classification

Classification - this is when our *y*-variable is a category. If it's a 0/1 yes/no kind of variable, we often call it **binary classification**. Otherwise, **multiclass classification.**

- *"Given this person's demographic information, how many tabs they have open, and where they live, will they make a purchase on my site?"*
- *"Given radar readouts, past weather, and almanac data, will it rain tomorrow?"*
- *"Given how many hours you study, how many hours you sleep, and your course load, will you pass the final exam?"*

Logistic Regression

Supervised, white-box, classification

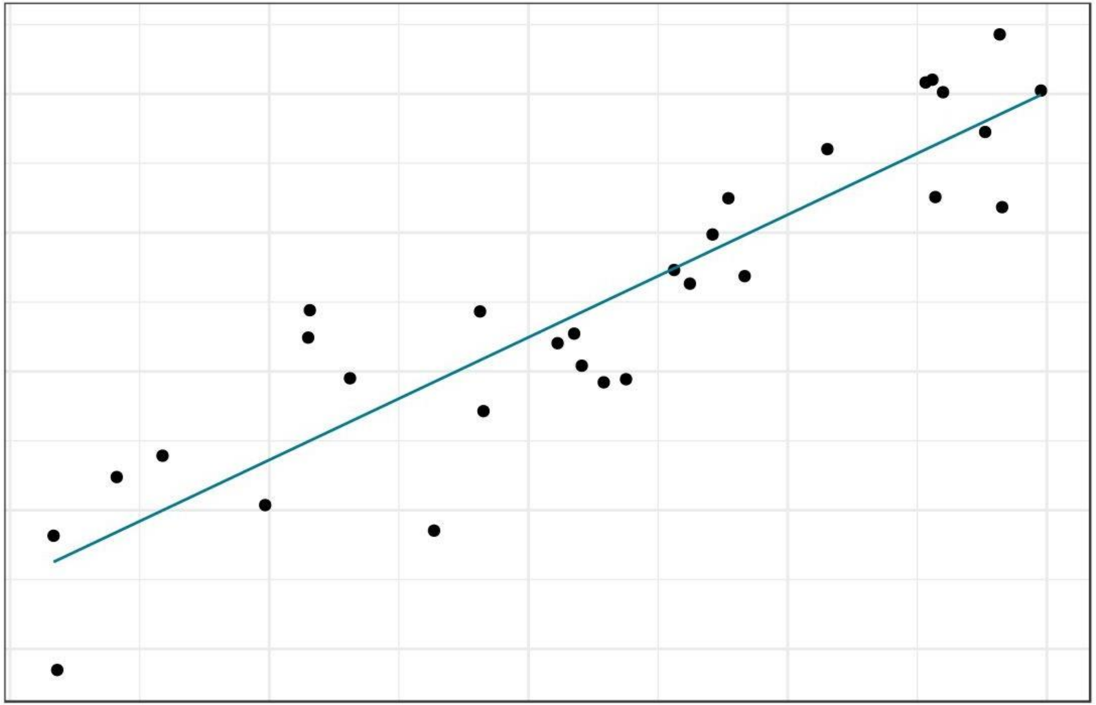
Reminder: Linear Regression

In ordinary least squares linear regression (often just referred to as **OLS**), we try to predict some response variable (*y*) from at least one independent variable (*x*). We believe there is a **linear** relationship between the two:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

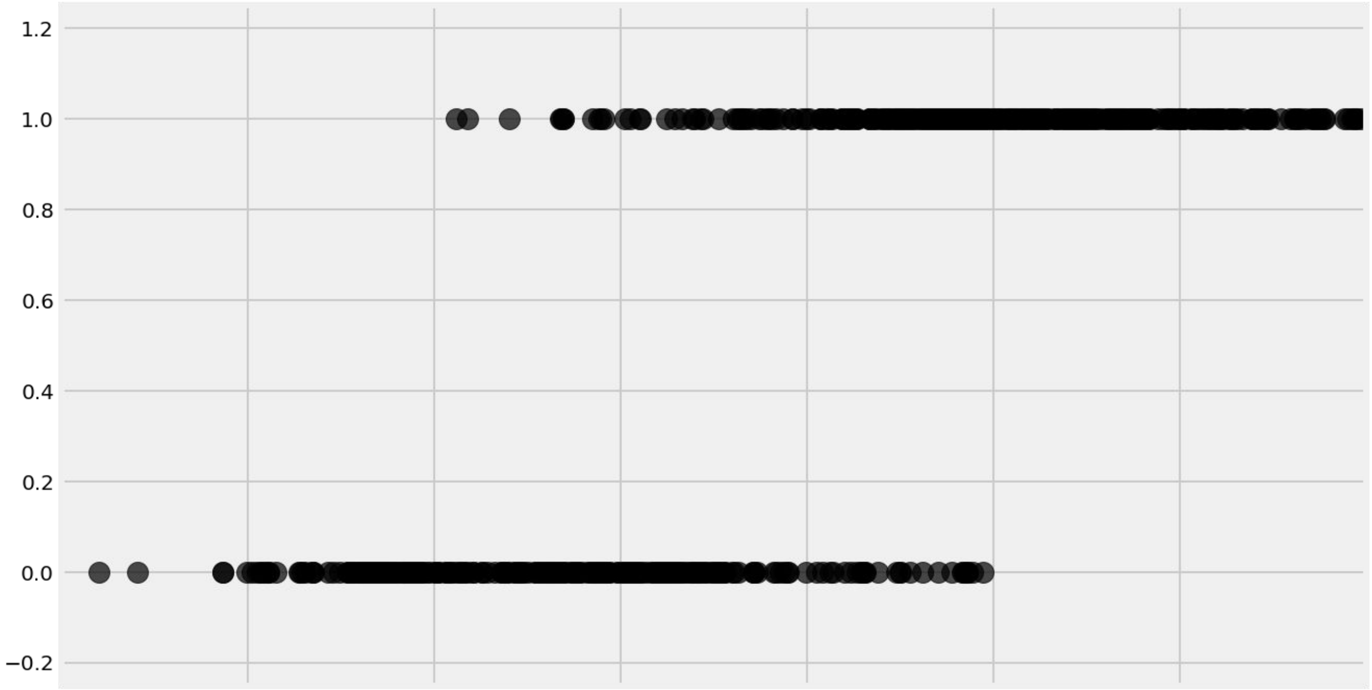
Graphically, this is:

Numeric y-variable!



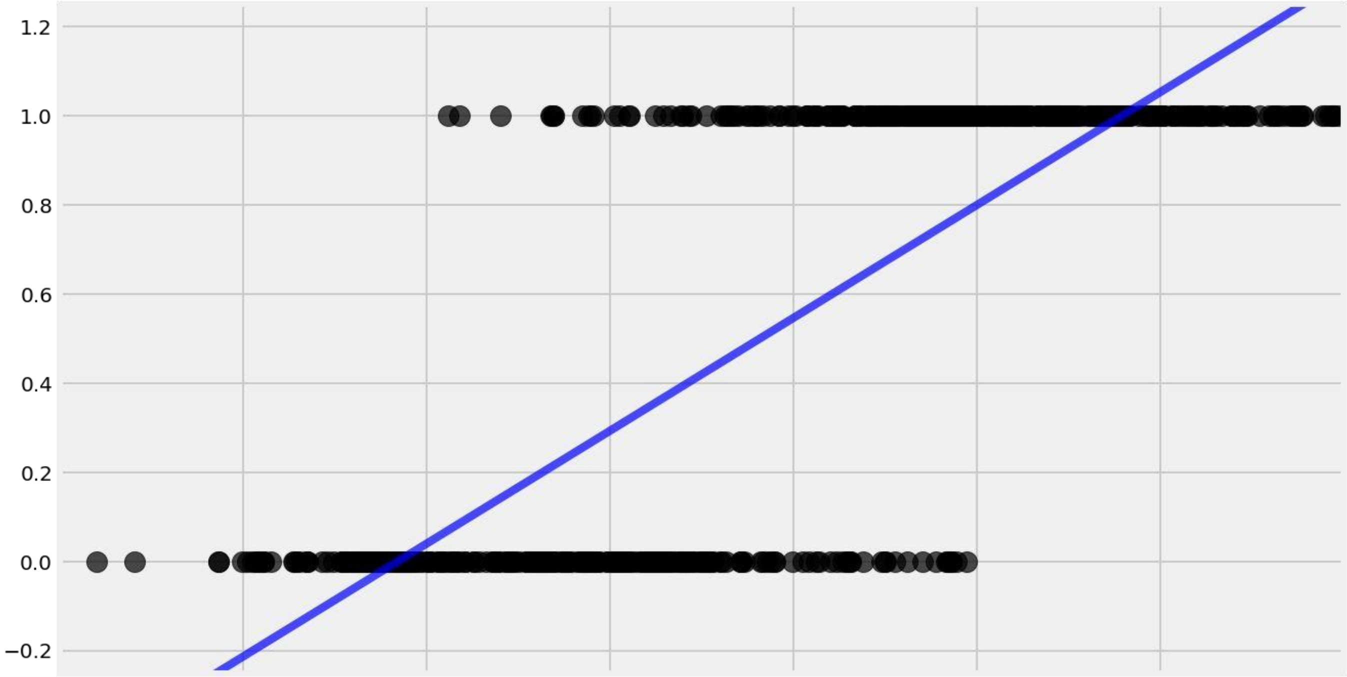
Now, what if our y-variable is a category:

Categorical y-variable!



Now, what if our y-variable is a category:

Categorical y-variable!



Logistic Regression

We will use something called a link function to effectively "bend" our line of best fit so that it is a curve of best fit that matches the range or set of values in which we're interested. For logistic regression, that specific link function that transforms ("bends") our line is known as the logit link.

Warning: math ahead.

Logistic Regression

We will use something called a link function to effectively "bend" our line of best fit so that it is a curve of best fit that matches the range or set of values in which we're interested. For logistic regression, that specific link function that transforms ("bends") our line is known as the logit link.

Reminder: Linear Regression

$$y = \beta_0 + \beta_1 X_1 + \epsilon$$

Logistic Regression

We will use something called a link function to effectively "bend" our line of best fit so that it is a curve of best fit that matches the range or set of values in which we're interested. For logistic regression, that specific link function that transforms ("bends") our line is known as the logit link.

Reminder: Linear Regression → $y = \beta_0 + \beta_1 X_1 + \epsilon$

Apply the Logit Link Function → $\operatorname{logit}(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \epsilon$

Logistic Regression

We will use something called a link function to effectively "bend" our line of best fit so that it is a curve of best fit that matches the range or set of values in which we're interested. For logistic regression, that specific link function that transforms ("bends") our line is known as the logit link.

Reminder: Linear Regression → $$y = \beta_0 + \beta_1 X_1 + \epsilon$$

Apply the Logit Link Function → $$\operatorname{logit}(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \epsilon$$

Which is the same as predicting the log-odds of success → $$\log\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = \beta_0 + \beta_1 X_1 + \epsilon$$

Briefly: Odds

Probabilities and odds represent the same thing in different ways. The odds for probability *p* is defined as:

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

Probability of something happening

Probability of something NOT happening

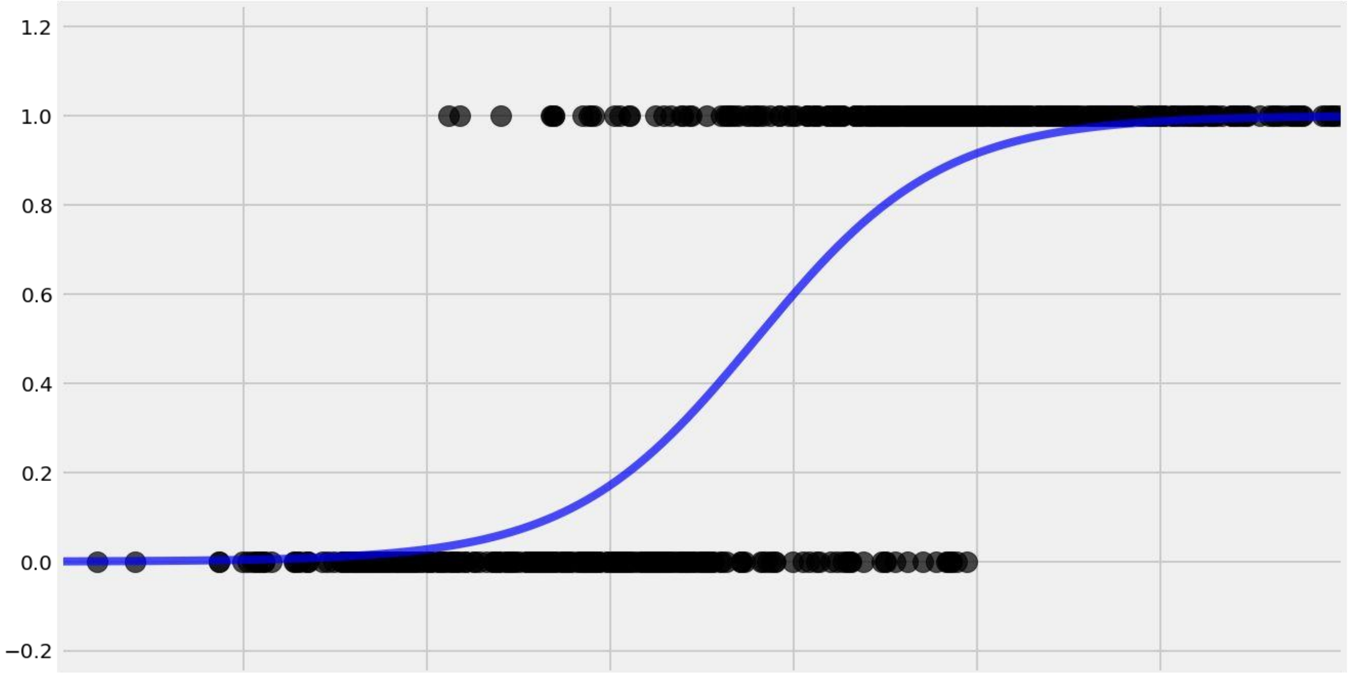
We will come back to this later when talking about how to interpret our coefficients.

Logistic Function

Now we are predicting the logarithm of the odds of success. This is… not very intuitive and can be any value (-∞ to +∞).

So, the logistic function is applied to the model to get values between 0 and 1 (now probabilities!).

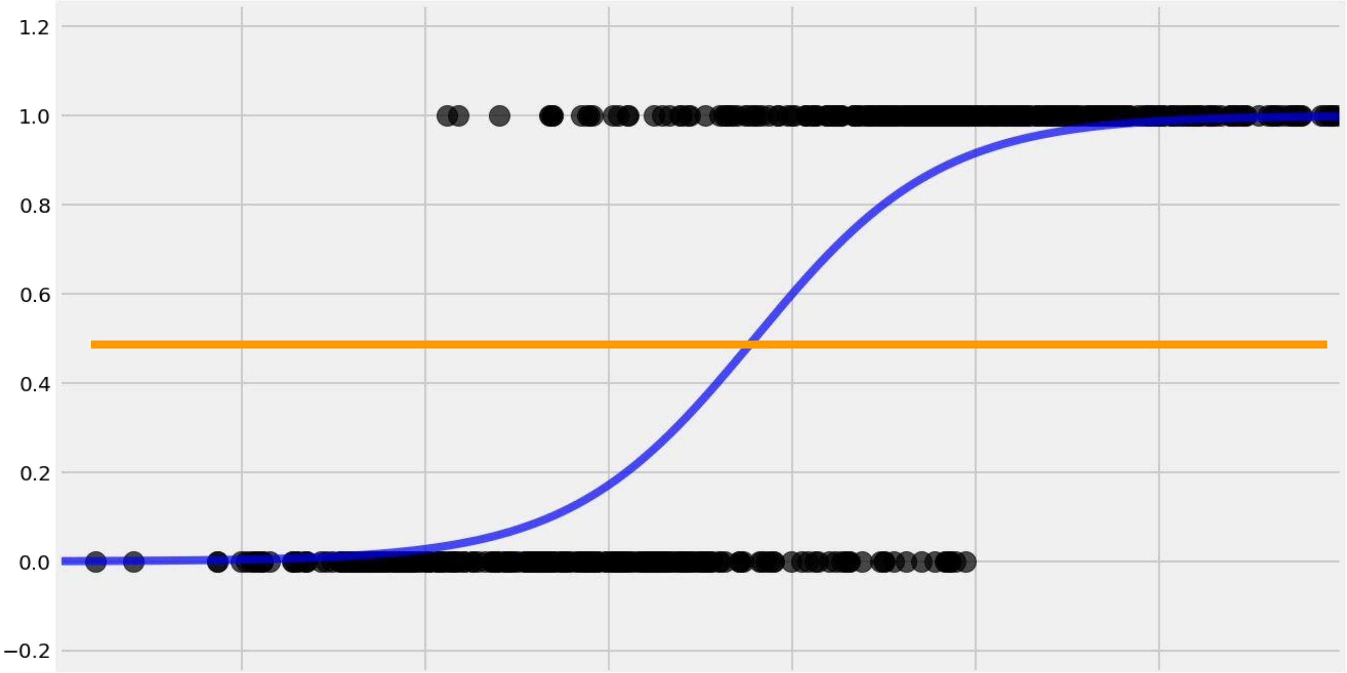
Graphically, this is:



Graphically, this is:

But, our predictions are still off, you say?

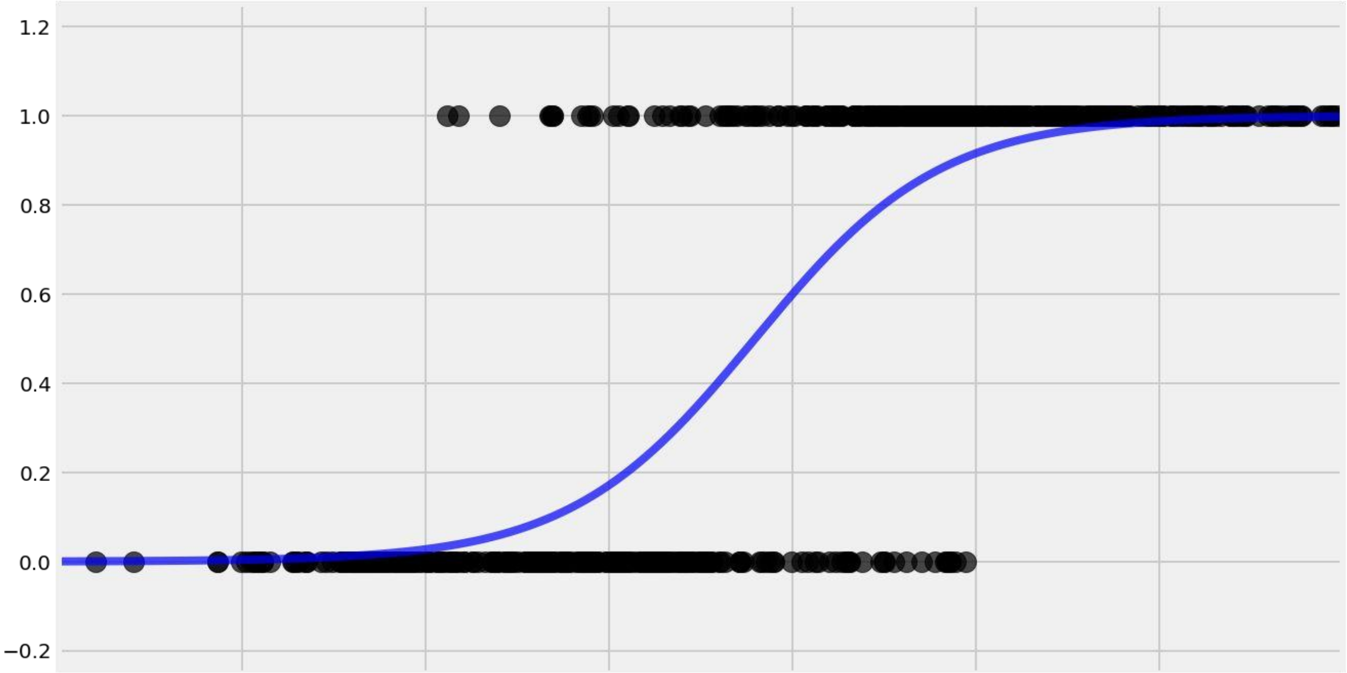
We choose a threshold to decide whether an observation should be predicted as 0 or 1.



So, why is it called logistic *regression* when it is used for *classification*...

We are predicting the probability of success! This is a continuous output between 0 and 1.

But we then transform that into a category to use for classification!



Logistic Regression

To recap:

- Logistic regression gives us the probabilities of being in each class
- We can use these probabilities to choose which class we predict
- The interpretation of coefficients is in terms of the log-odds (more on this later)