We have lots of data and the **promess** to give -> **Meaning**

to all this data

“Sufficently advanced techno … Magic”. Arture C. Clark

**Machine Learning** is rather tools and technologies used to answer **questions** with your data

Let me give you a definition of ML using **5 words**

**Using data to answer questions**

We can split this definition into:

Using data -> is often what we refer to as **Training**

While

Answer question -> is what we call making **Prediction** or **Inference**

|  |
| --- |
| **Training** |
| refers to **Using our data** to **create and tune** a **predictive model** |
|  |
| This **predictive model.** |
| Can then be used to **Serve our Prediction** on **previously unseen data** |

Let’s look at an example:

Suppose we’ve been asked to create a system -> That answers the question:

**On** **whether a drink is wine or beer?**

|  |
| --- |
| This **question answering machine** is called a**model** |
| This **model** is created via a process called **Training** |

In machine learning the goal is to create an accurate model

# Choosing a model

With only 2 features: color and Alcohol %

We’ll go for a **small linear model** that will get the job done

*Now we move on to what is often considered the* ***Bulk of ML***

The Training

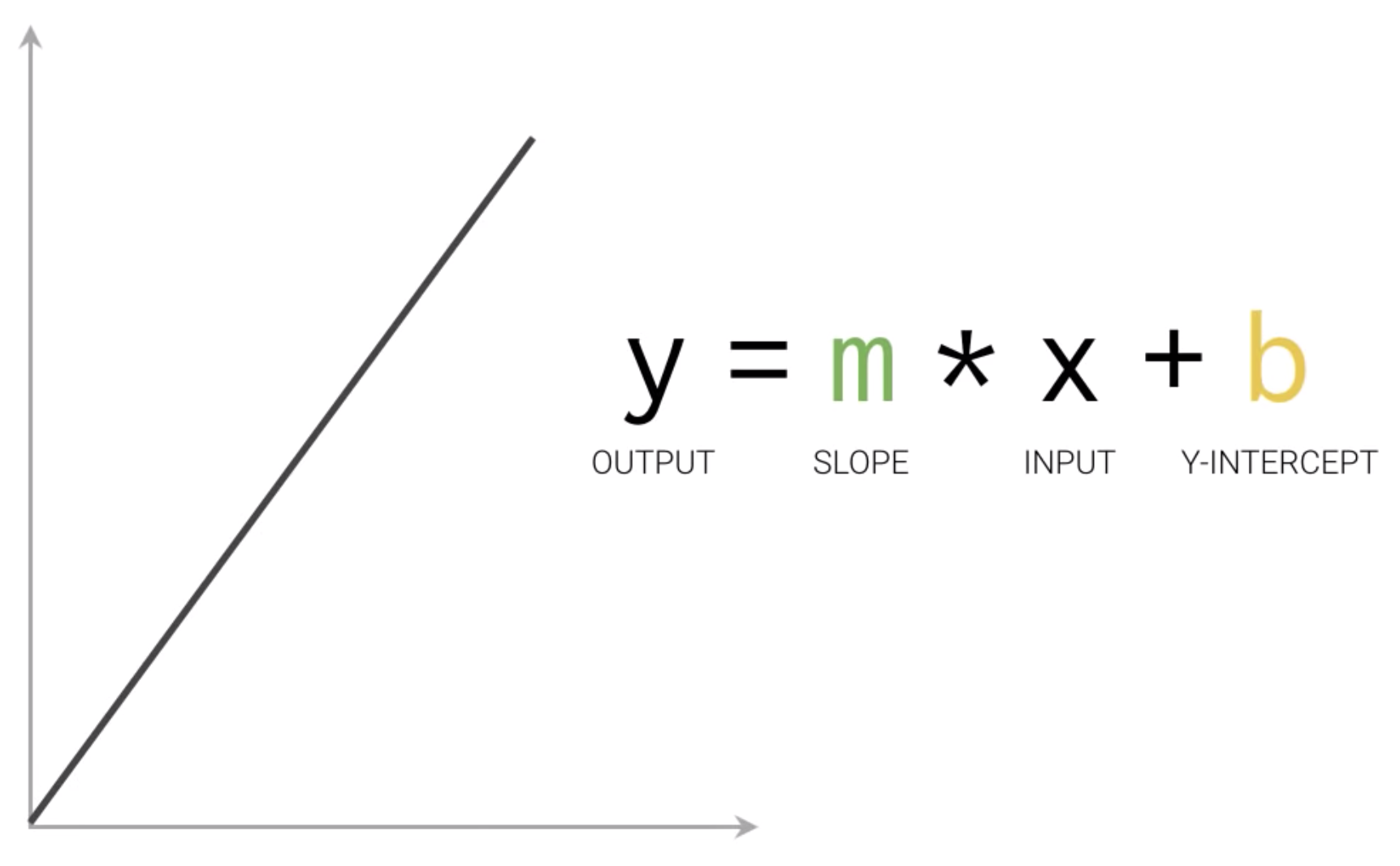
# Training (bulk of machine learning)

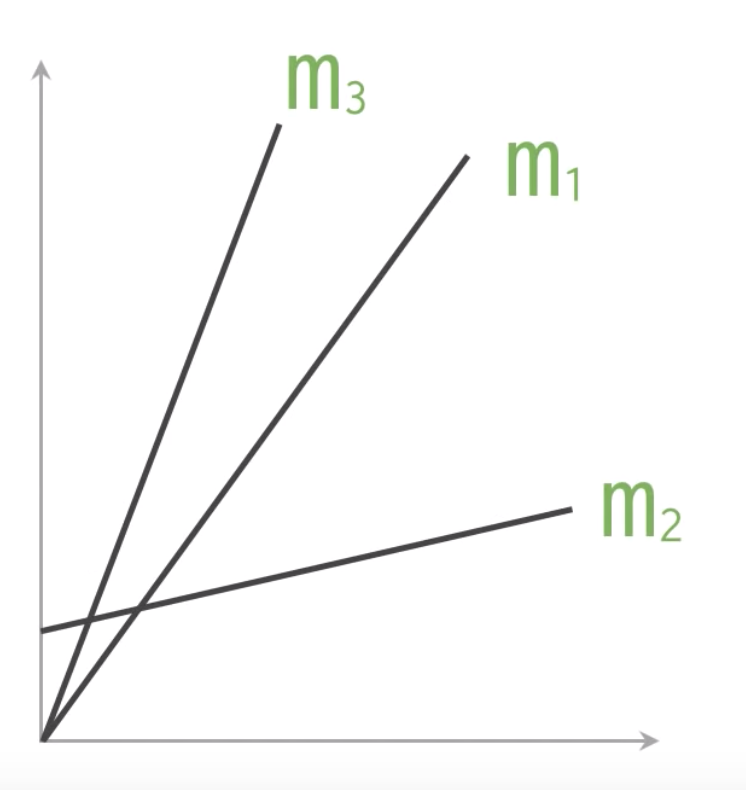
It’s a bit like driving

* You’ve been shown how to use the clutch and wheel
* But you need hours of practice to be able to drive efficiently

## In our case

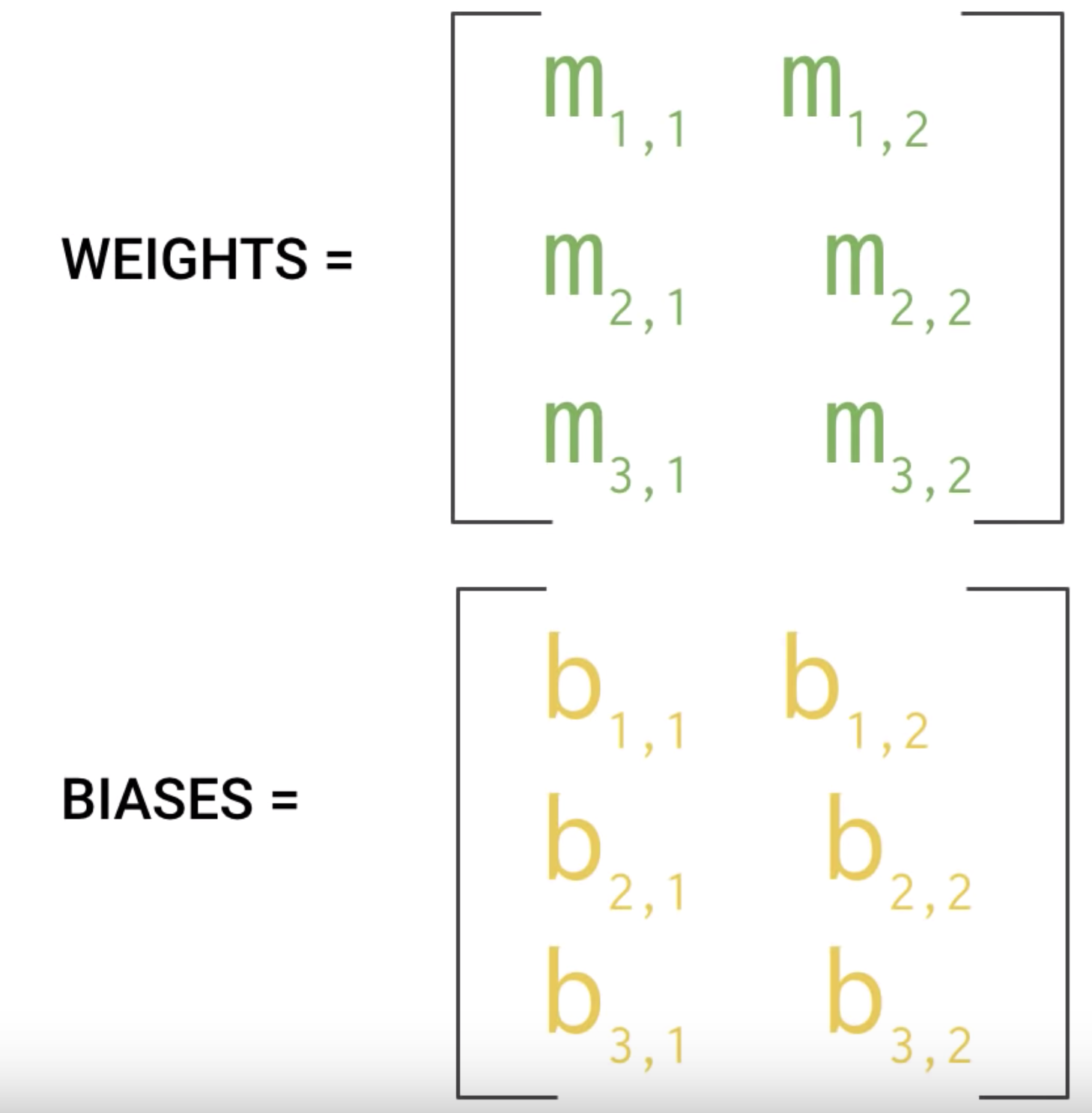
Linear model therefore equation of a straight line is



In machine learning there are many M’s as there can be many features

The collection of these values is usaly formed into a matrix that is called Weight

Similarly, for B’s we arrange them together and create a matrix called the Biases



So we use some random data from our training dataset and apply our model with w and B’s.

It does poorly at first

What we do is : comparing the outcome of the model with the results that the model should have produced.

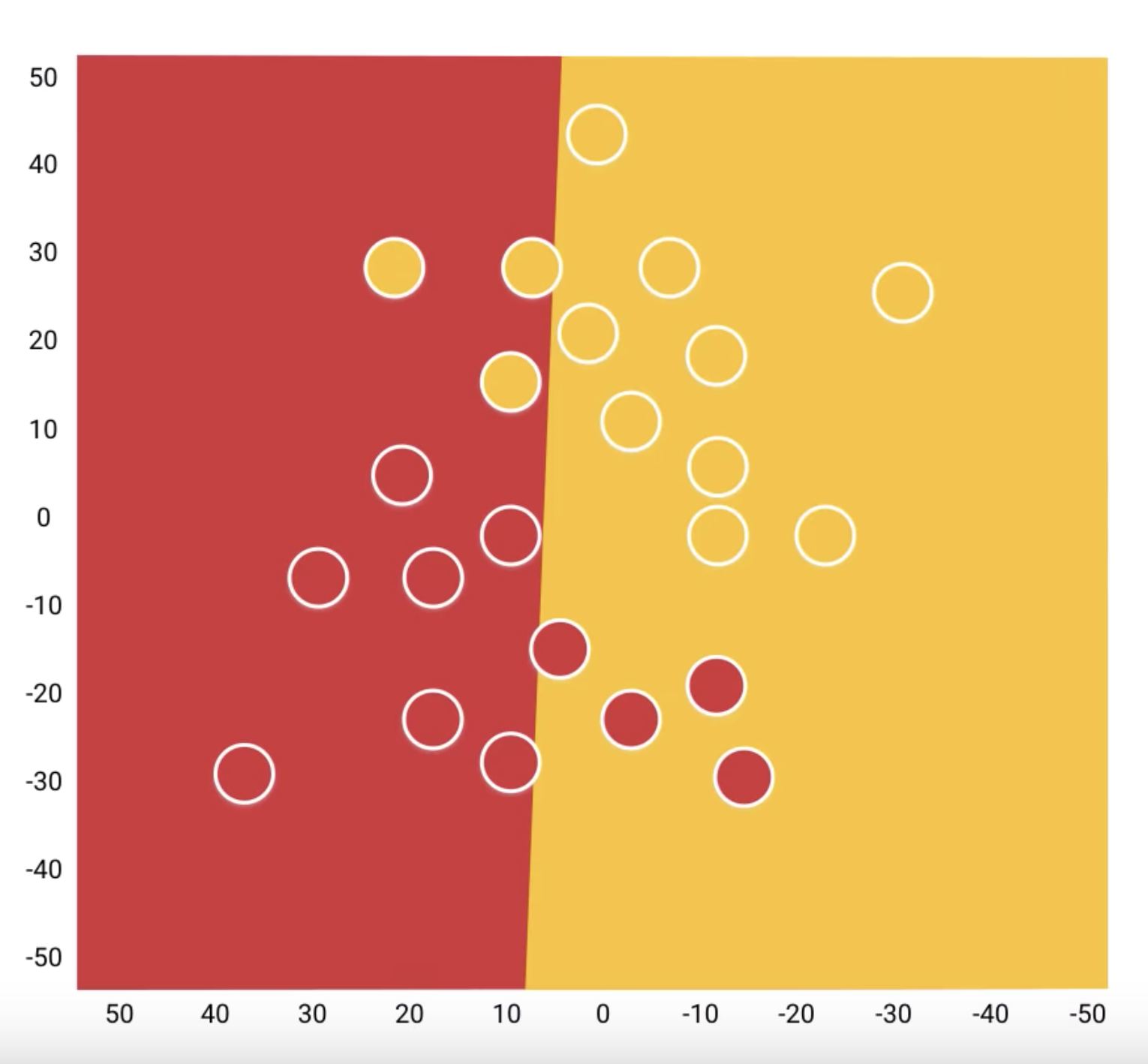
This process then repeats until it gives the results that we expect.

Each iteration of updating weights and biases is called: **One training step**

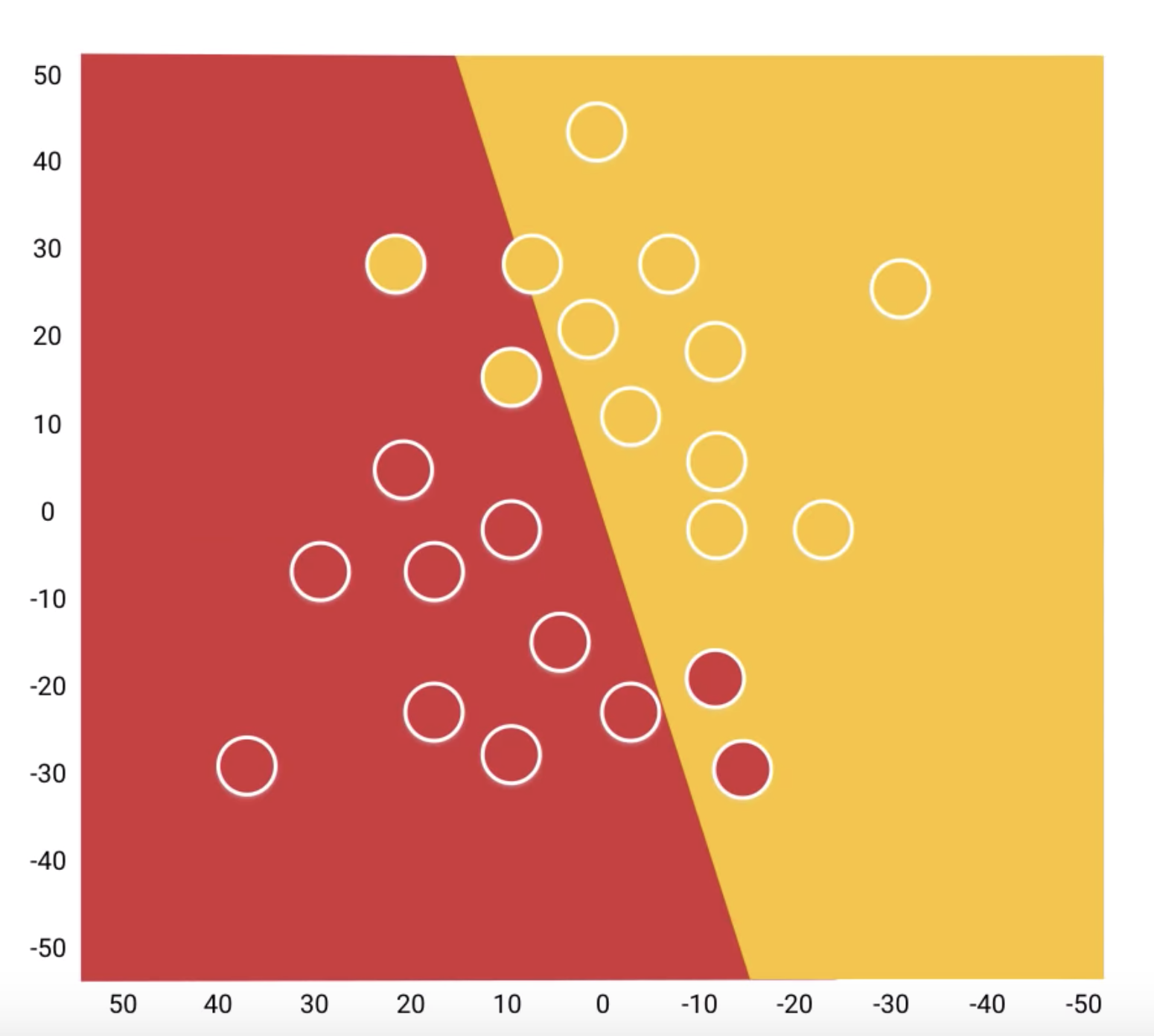
More concretely for our data set

At first our model might perform a bit poorly

Like we drew a line randomly through the data



Then as each steps of the training progresses the line moves step by step



Closer to the ideal separation of the wine and beer

One that training is complete it’s time for evaluation

A good split proportion of the data is around 20% 80 % (depends on the initial data volume)

One thing we could do, is Fit a logistic regression with

just a simple hypothesis like this,

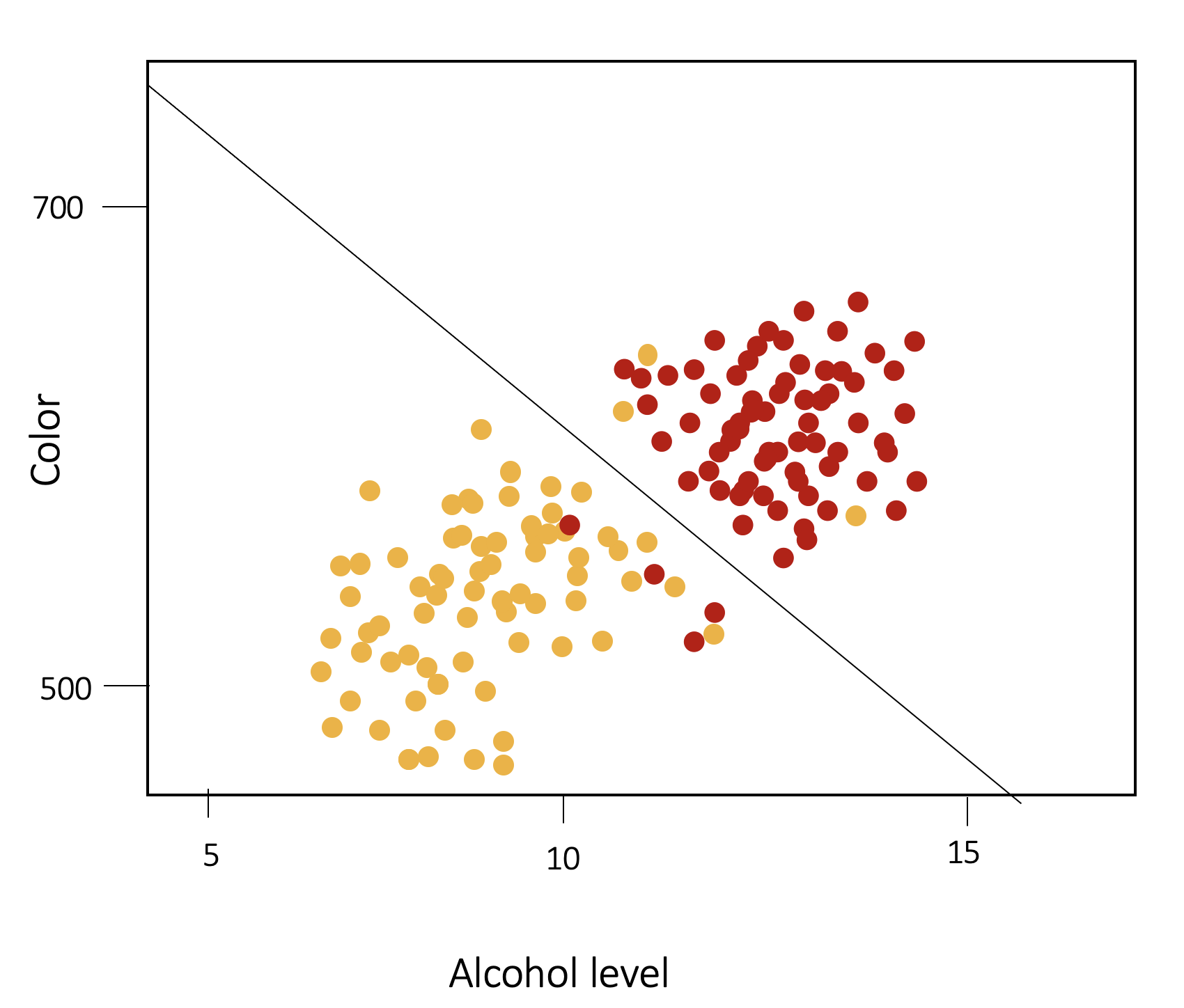
where, as usual, G is my sigmoid function.

And if you do that, you end up

with a hypothesis, trying to

use, maybe, just a straight

line to separate the positive and the negative examples.



Straight line is a logistic regression Classifier

When we work with linear models the basis of our model is a linear equation of type: