AIBridge

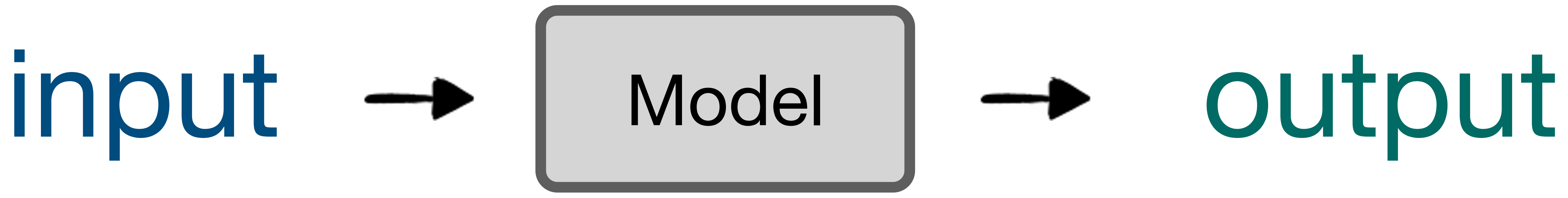
Lecture 4

Our first question:

what *is* Artificial Intelligence?

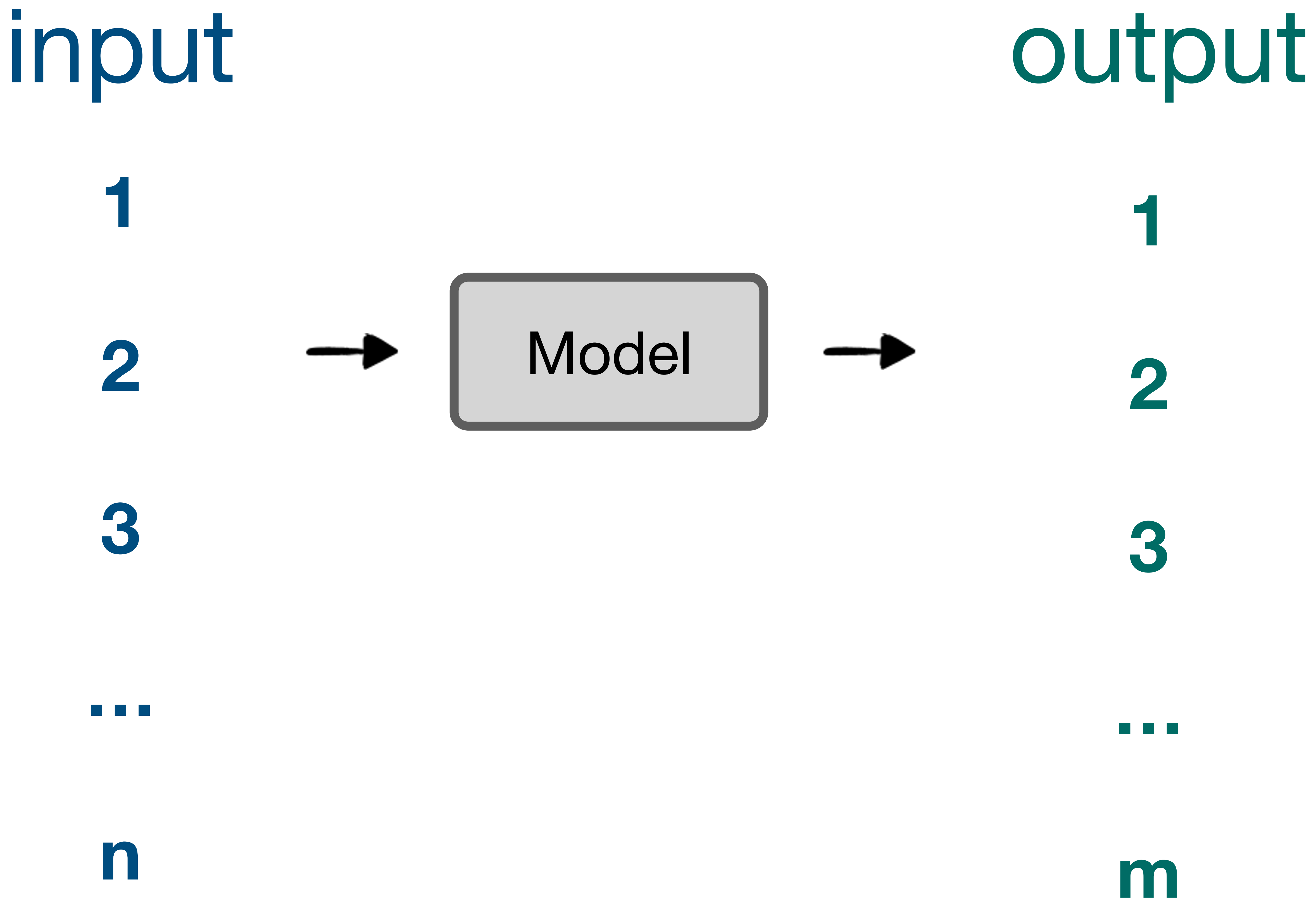
AI Models

- We call different types of AI functions **models**.

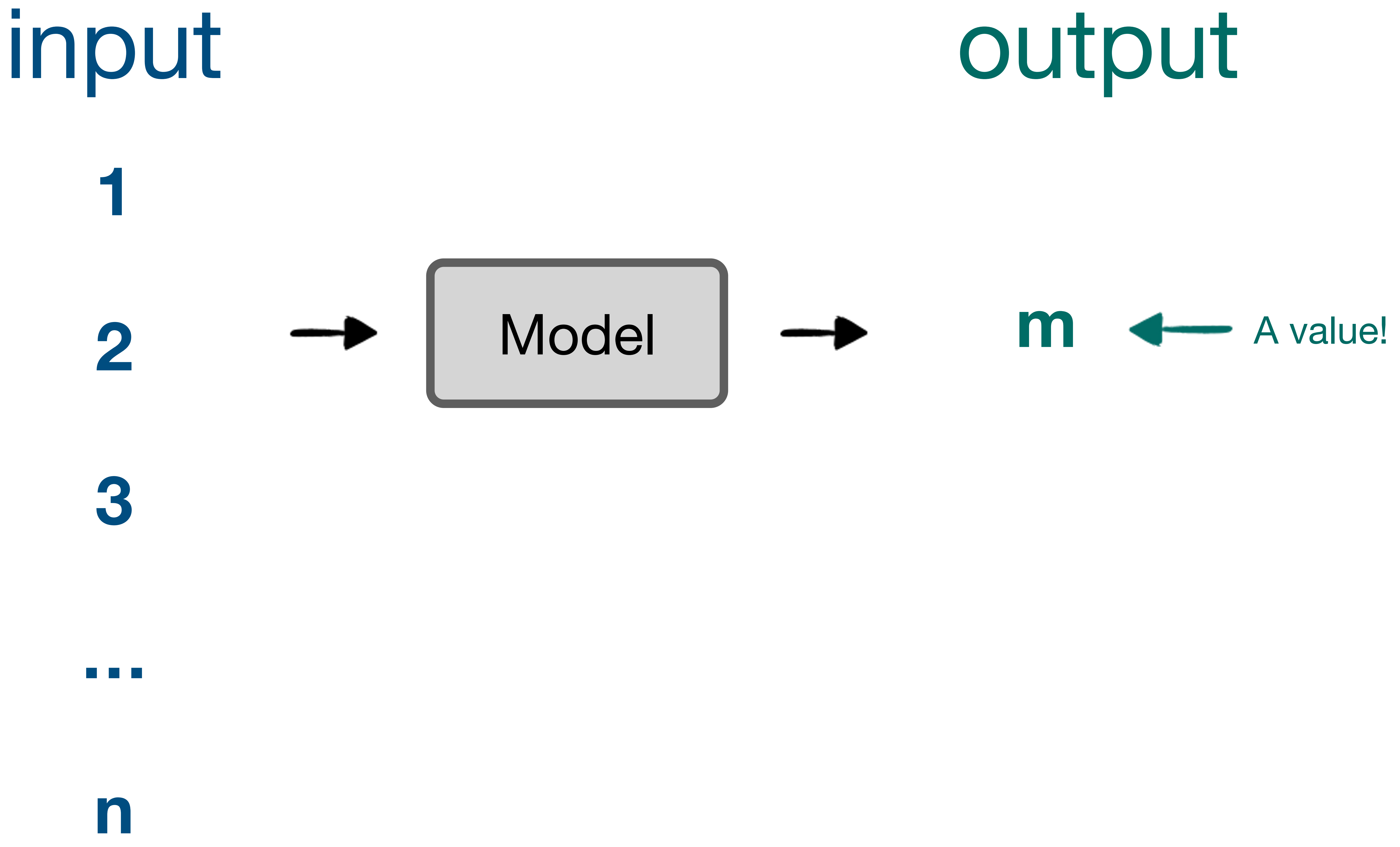


AI Models

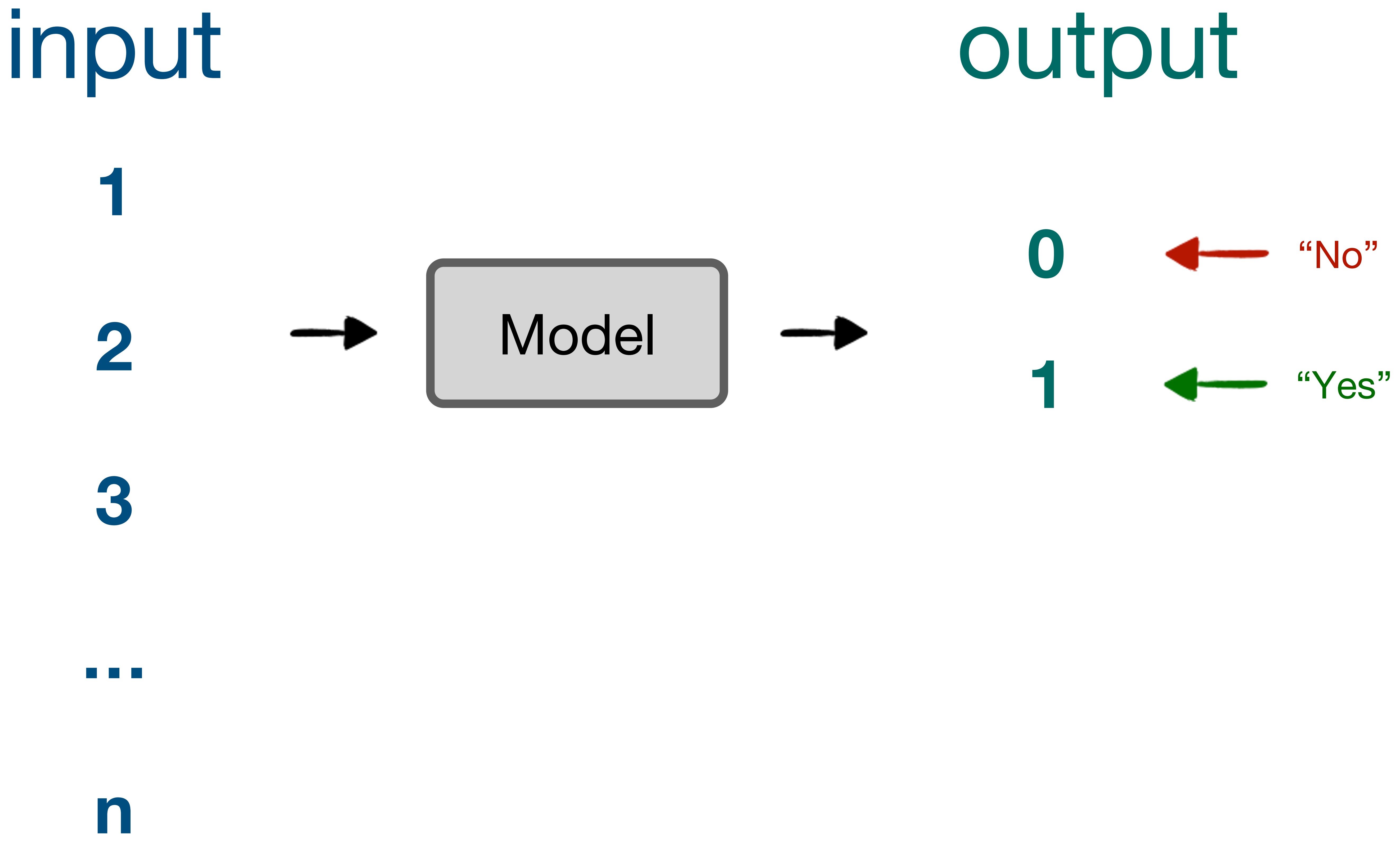
- What are the inputs and outputs?
- R_n to R_m functions



AI Model Examples



AI Model Examples



AI Model Examples

input

output

cat

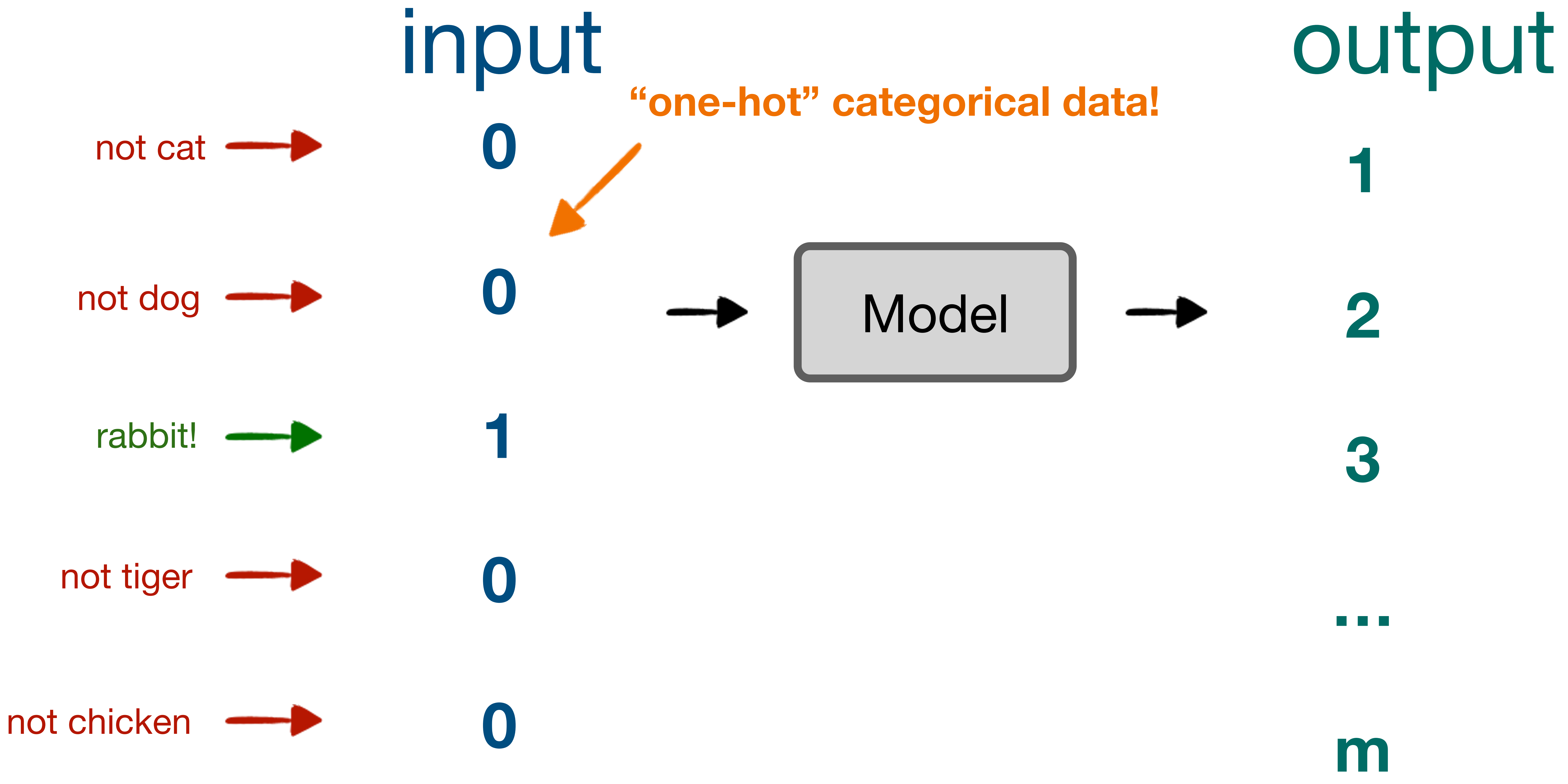
dog

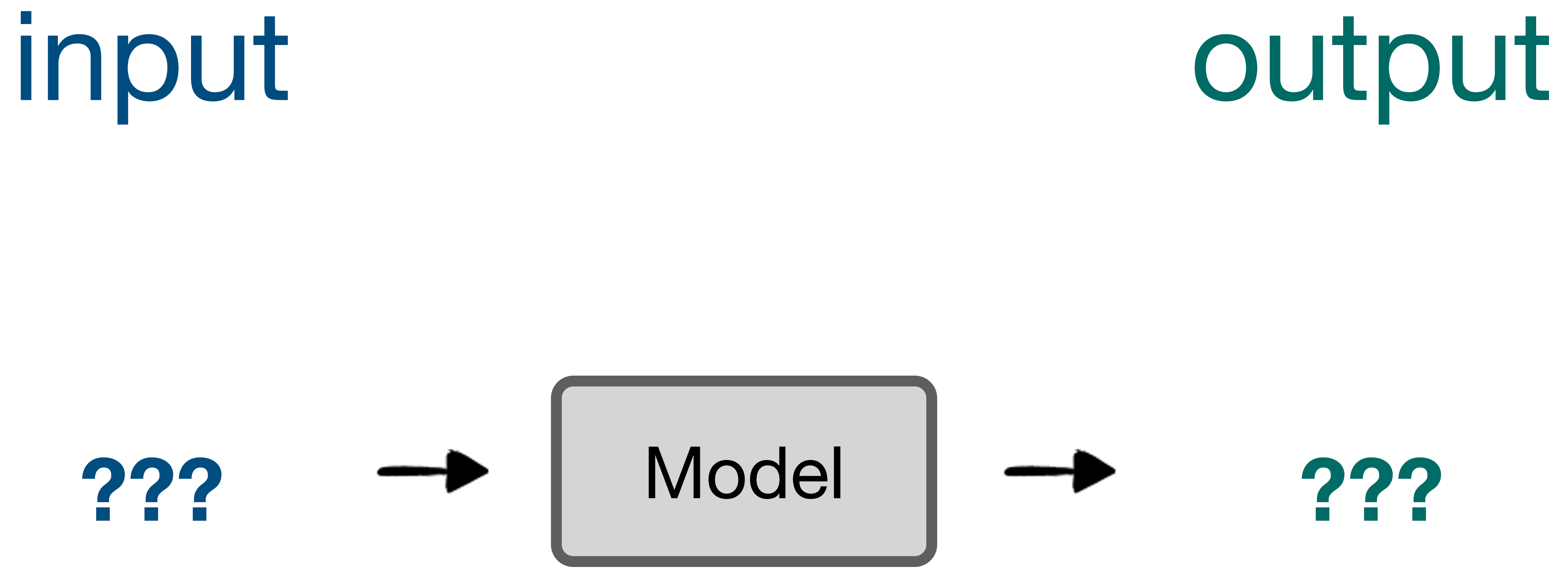
rabbit!

tiger

chicken

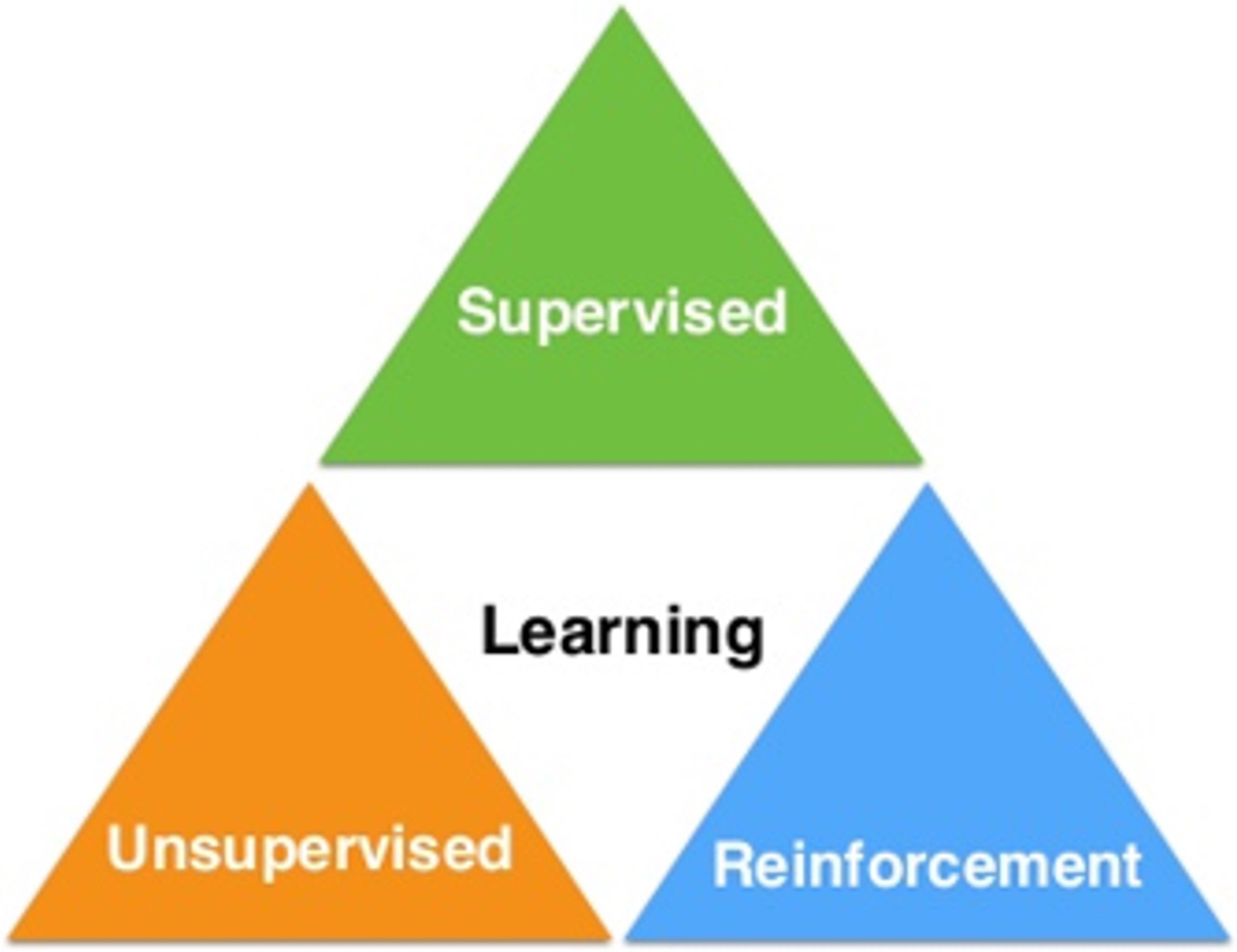
AI Model Examples





- by changing the shapes of input and output, models can represent a lot of different problems

- Labeled data
- Direct feedback
- Predict outcome/future



- No labels
- No feedback
- “Find hidden structure”

- Decision process
- Reward system
- Learn series of actions

SUPERVISED LEARNING

Supervised Learning

- “**Supervised learning** (SL) is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.” – Wikipedia
- Input-output pairs: Features and labels
- Training/learning and inference
- Most widely used ML techniques in real-world applications.

Terminology alignment

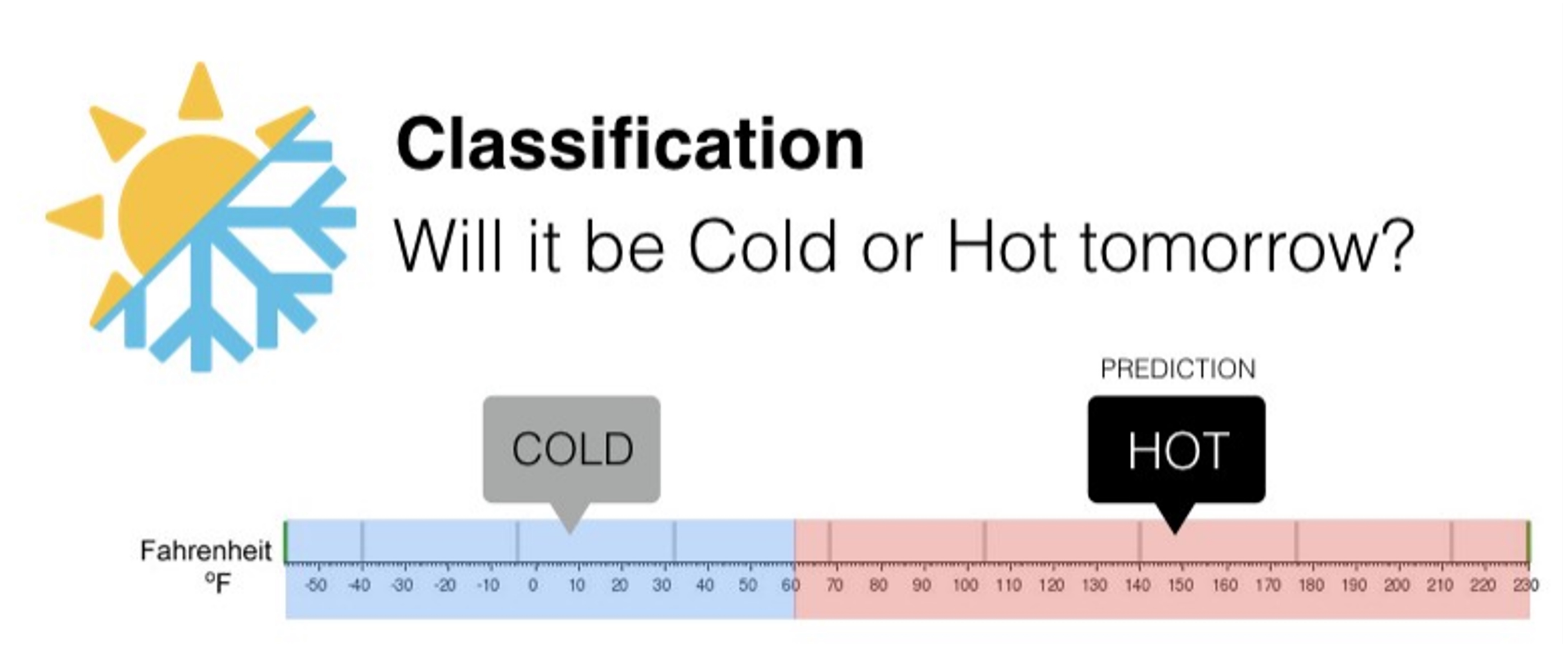
- Sample = (features, label)
- Features: independent variables, attributes, predictors, input variables, input, covariates, explanatory variables, treatment variables,
- Label: dependent variable, outcome, target variable, outcome variable, response variable
- Samples: cases, observations, individuals, participants, data points

- If you have other names for these, please let me know.

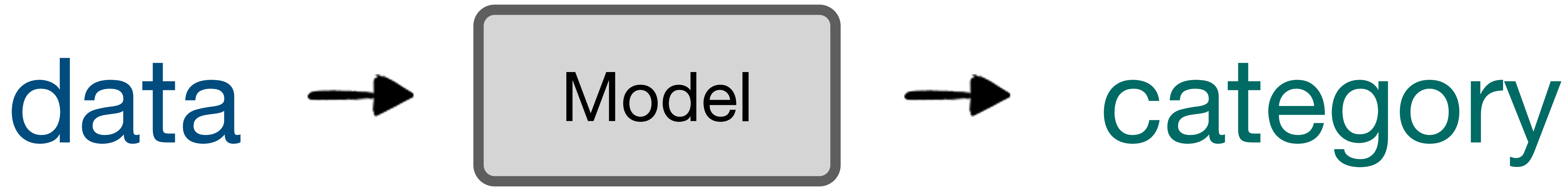
What types are there?

Classification

- Predicting a label/class/category
  - Ex: spam or not, cancer or not, cat or dog, red wine vs. white wine

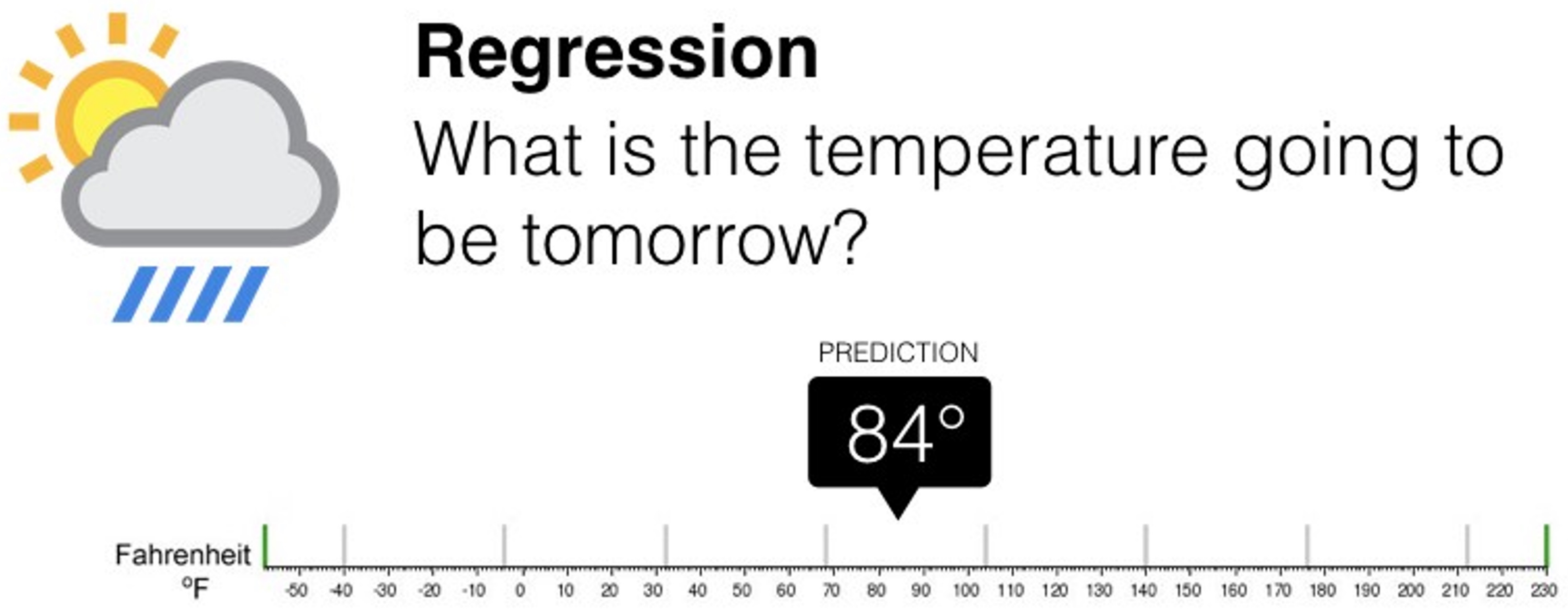


What does classification do?

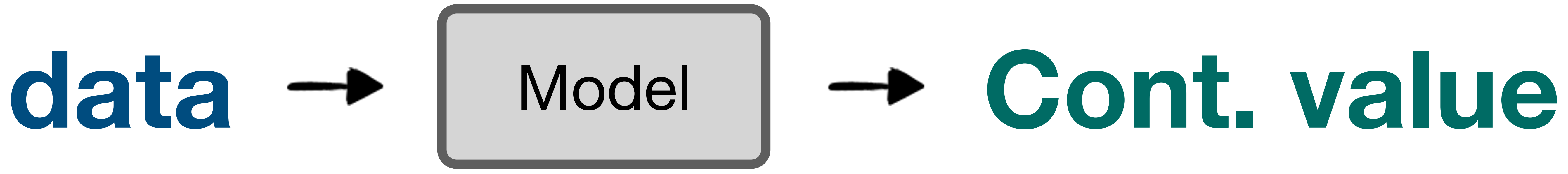


Regression

- Predicting a (continuous) quantity
  - ➢ Ex: Survival rate, wine quality, yield prediction



What does regression do?



Examples

- You're running a company, and you want to develop learning algorithms to address each of two problems.
  - Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
  - Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.
- Are they **classification or regression?**

Supervised Algorithms Practice

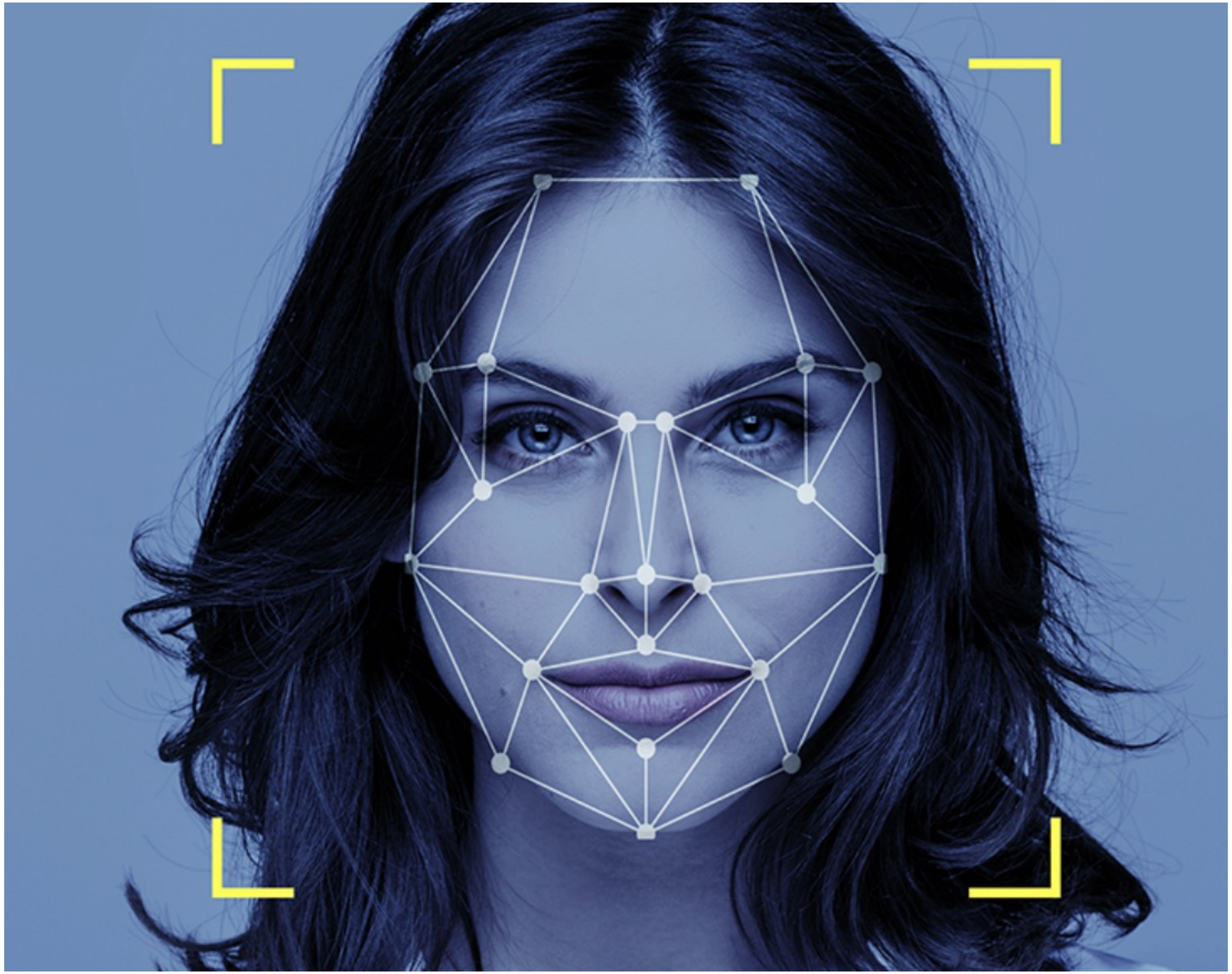
- **You're running a company, and you want to develop learning algorithms to address each of two problems.**
  - Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
- **Can we formulate it as a classification problem?**

We could. Sometimes, we can reformulate.

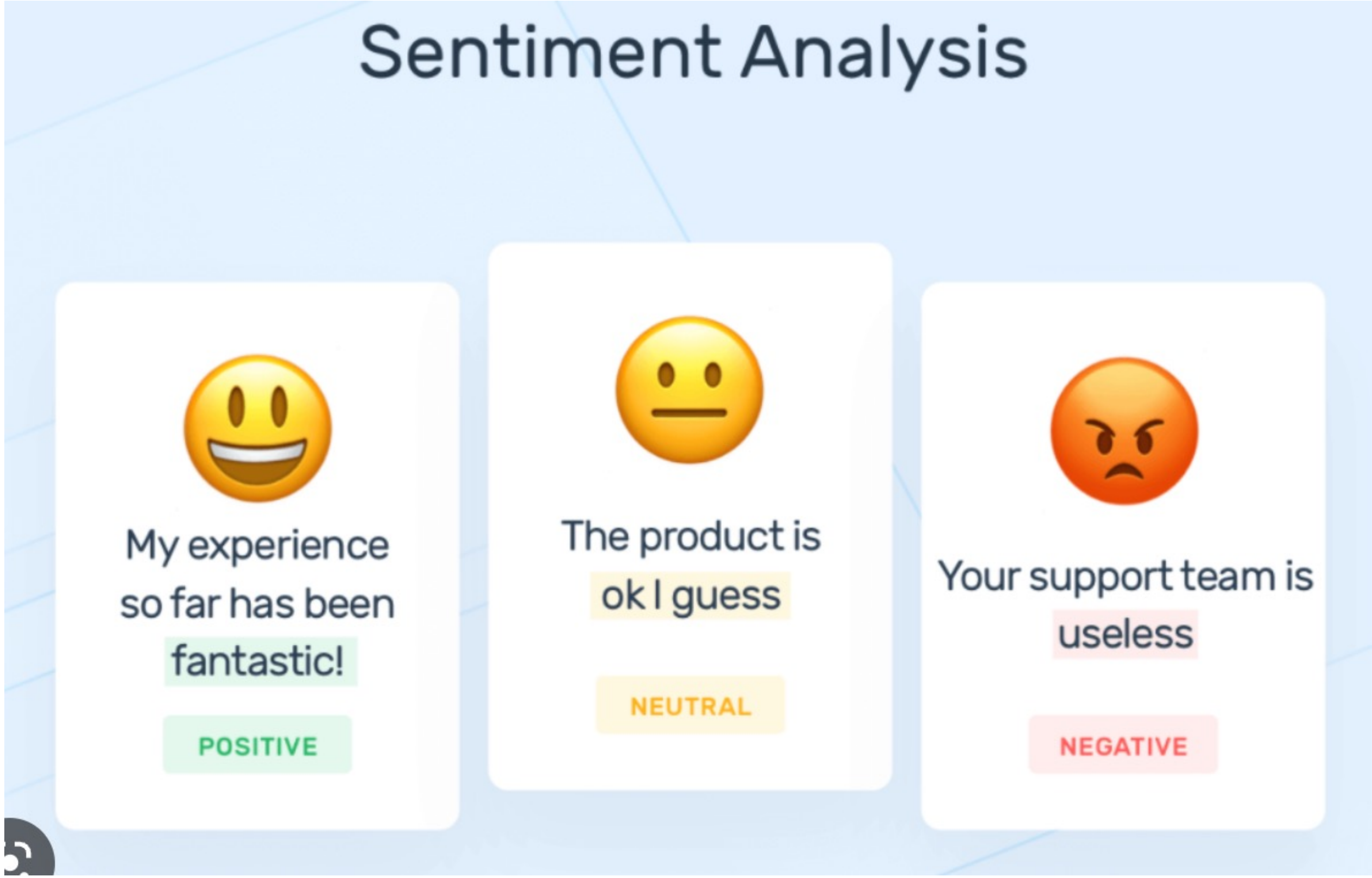
Let's start with regression instead!

Speech Recognition

Face Recognition



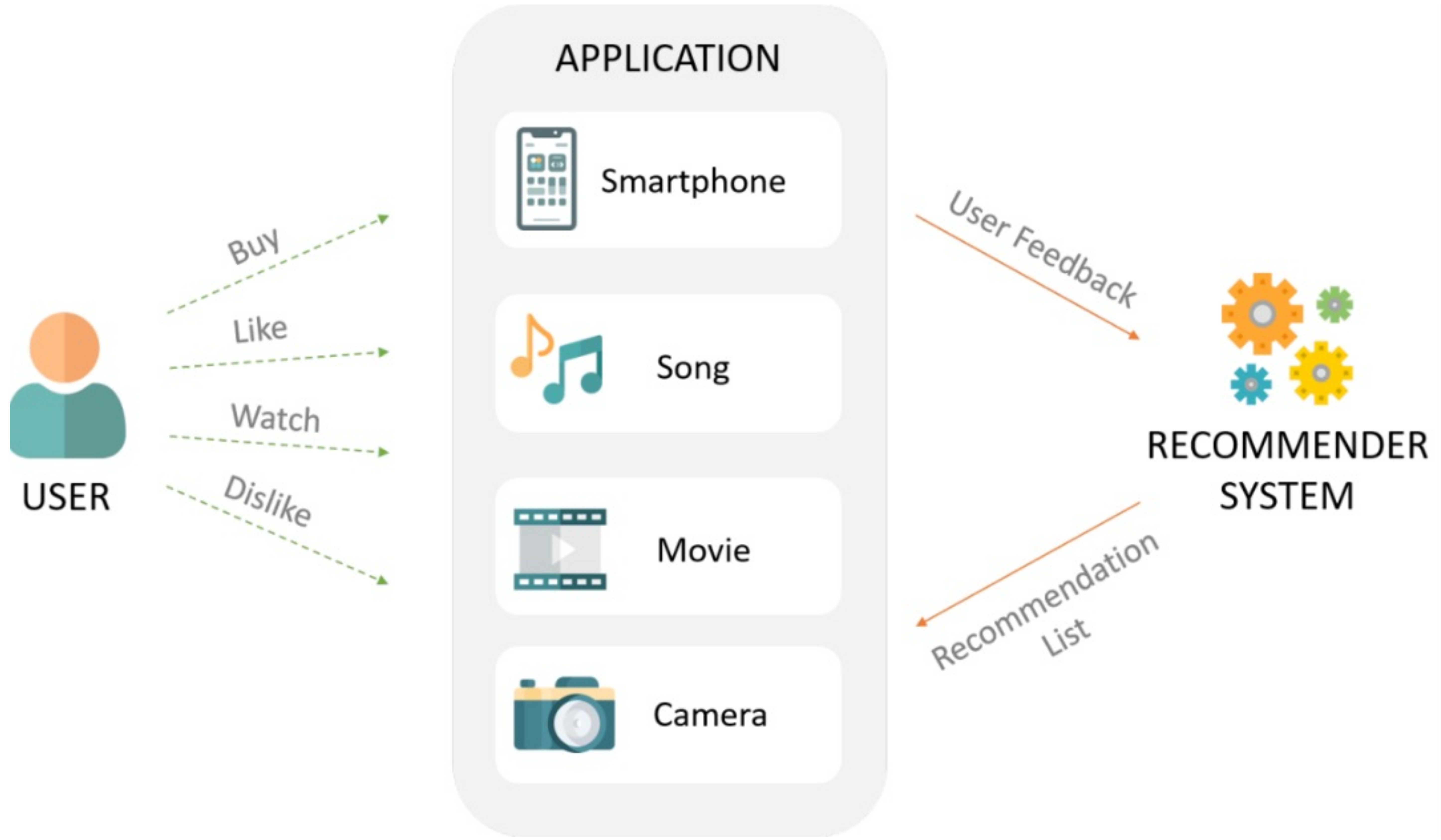
Sentiment Analysis



Spam Filter

Fraud Detection

Recommendation



Stock Price Prediction

Food Quality/Safety Prediction

“All models are wrong, some are useful”

- George Box

Our dataset: wine quality

What does our dataset look like?

About Citation Policy Donate a Data Set Contact

Search

Repository Web

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Wine Quality Data Set

Download: Data Folder, Data Set Description

Abstract: Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], [Web Link]).

Data Set Characteristics:	Multivariate	**Number of Instances:**	4898	**Area:**	Business
Attribute Characteristics:	Real	**Number of Attributes:**	12	**Date Donated**	2009-10-07
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	1891084

Source:

Paulo Cortez, University of Minho, Guimarães, Portugal, http://www3.dsi.uminho.pt/pcortez
A. Cerdeira, F. Almeida, T. Matos and J. Reis, Viticulture Commission of the Vinho Verde Region(CVRVV), Porto, Portugal
@2009

Wow! 12 attributes!

(and quality, which can be counted as a 13th)

Structuring our dataset

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red

a feature of input data

- a **sample** is a collection of **features**

So if we want to predict wine quality…

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red

→ **Quality (0-10)**

…we'll need a model!

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red

→ Model → **Quality (0-10)**

LINEAR REGRESSION

Linear Regression

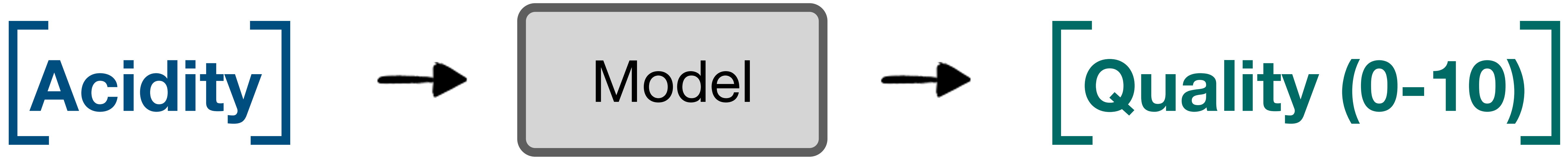
- **What does linear regression represent?**

The features	vs.	The label
Fixed acidity		**Quality (0-10)**
Volatile acidity		
Citric acid		
Residual sugar		
Chlorides		
Free sulfur dioxide		
Total sulfur dioxide		
Density		
pH		
Sulphates		
Alcohol		
White/Red		

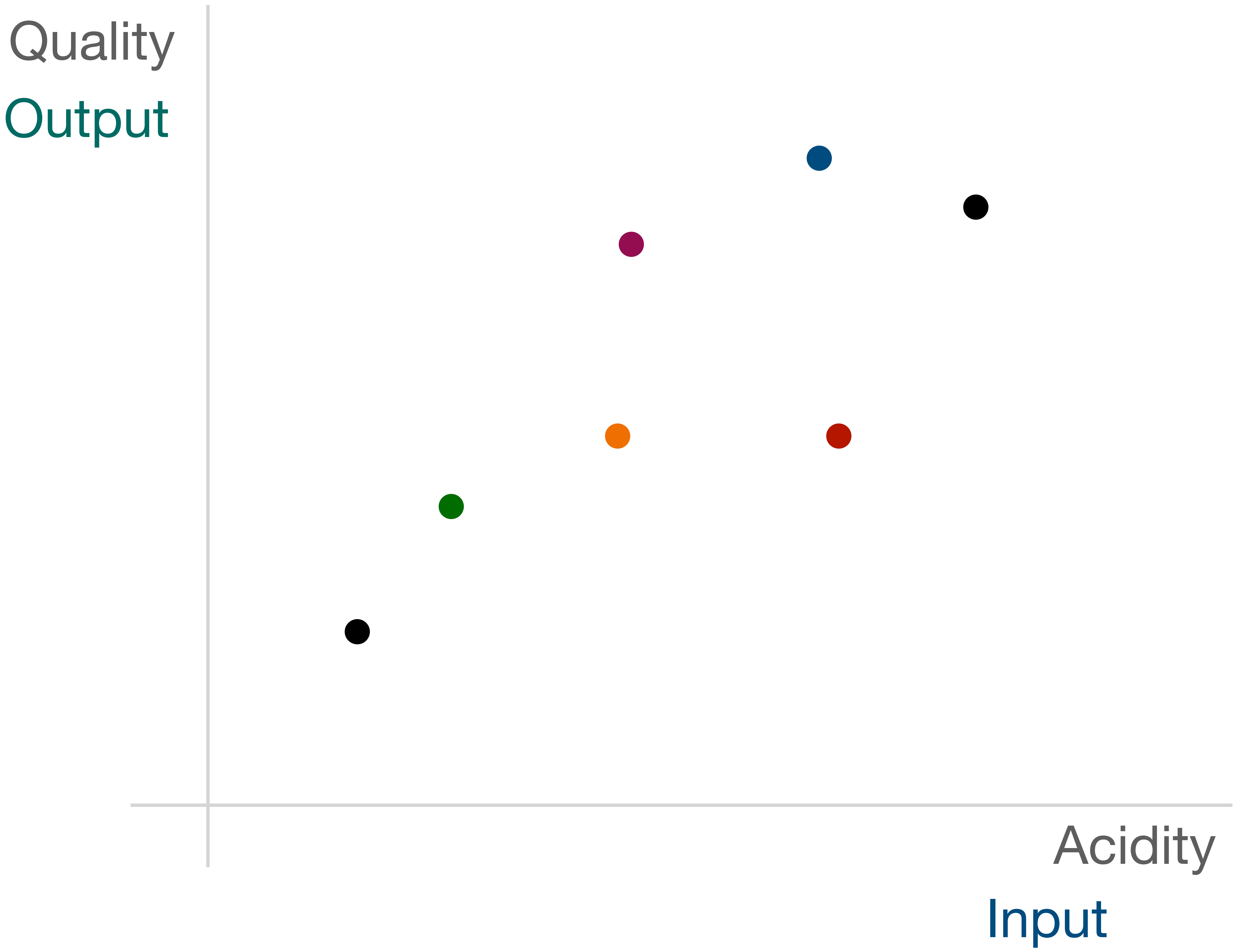
Let’s start with one variable.

Linear Regression

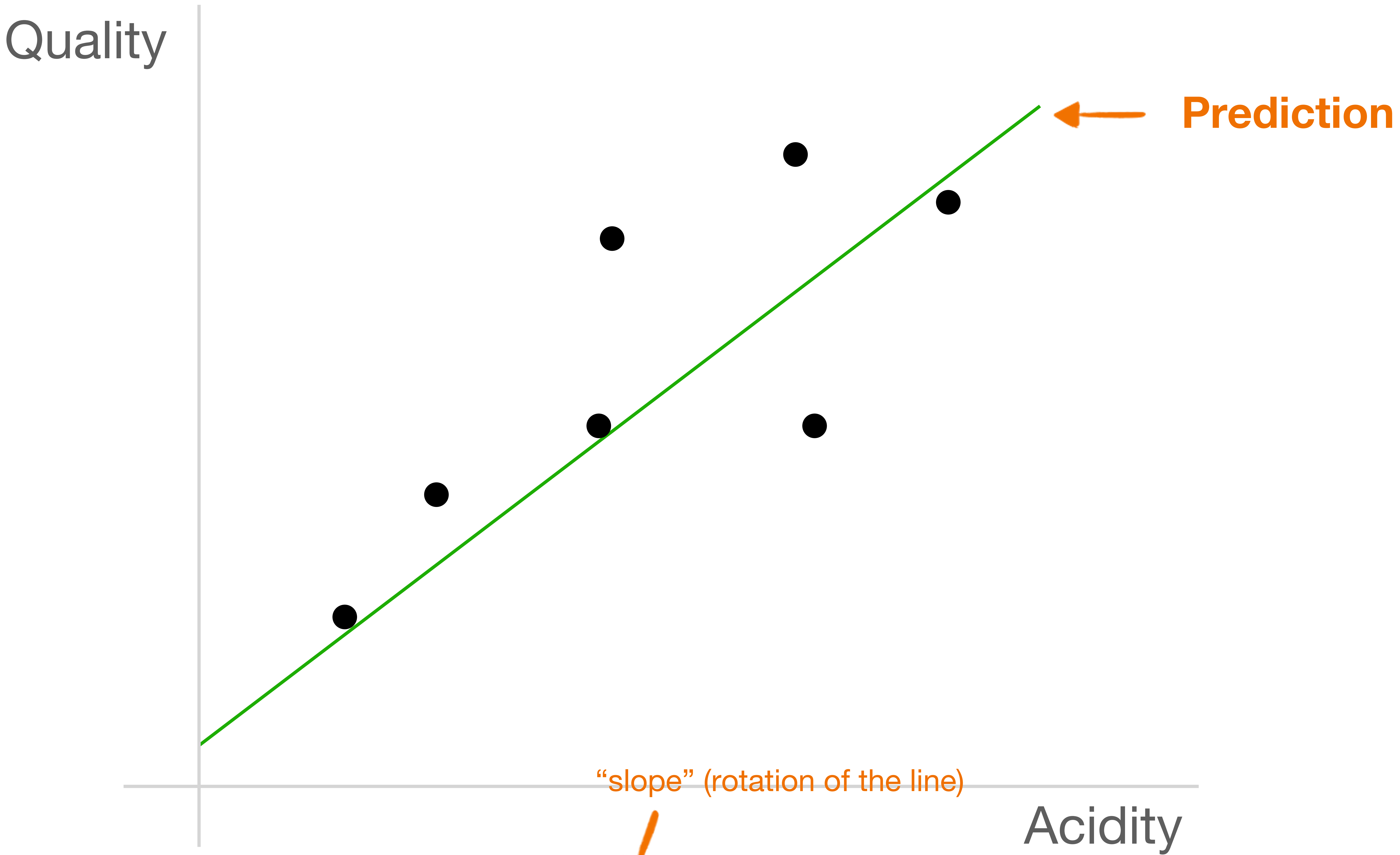
- We'll just consider **one feature** of our **sample**.



Linear Regression

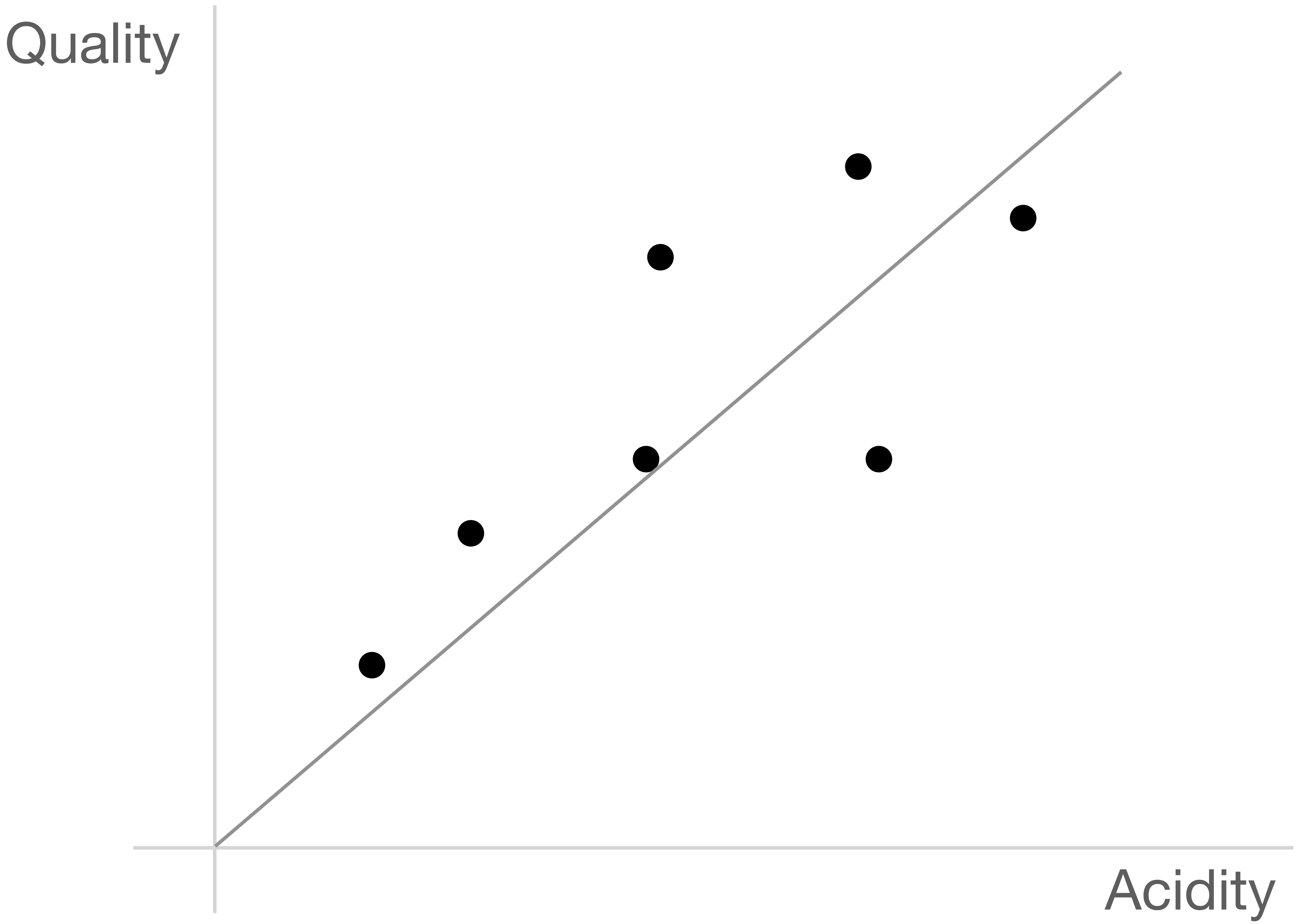


Linear Regression

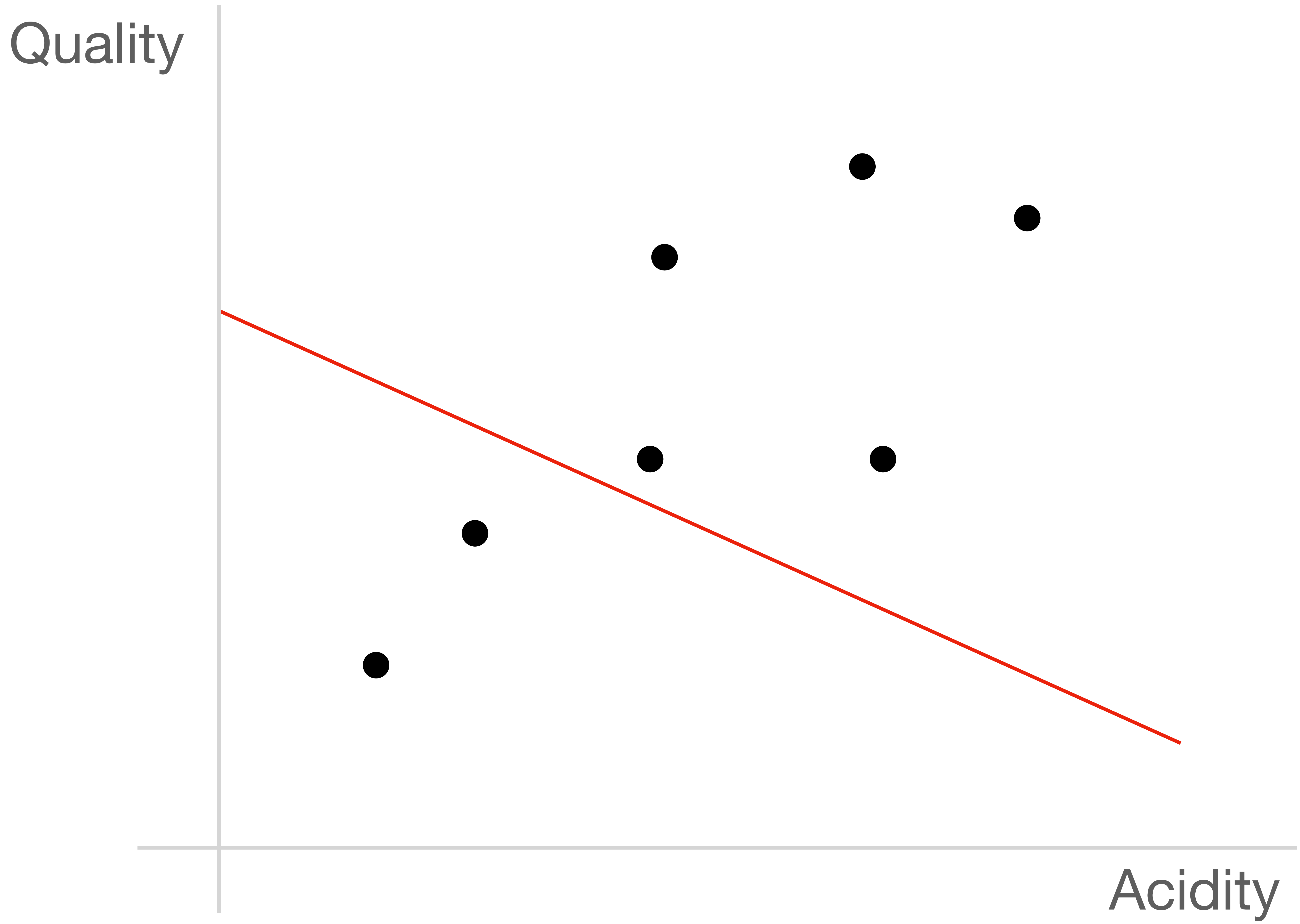


$$y_i = mx_i + b$$

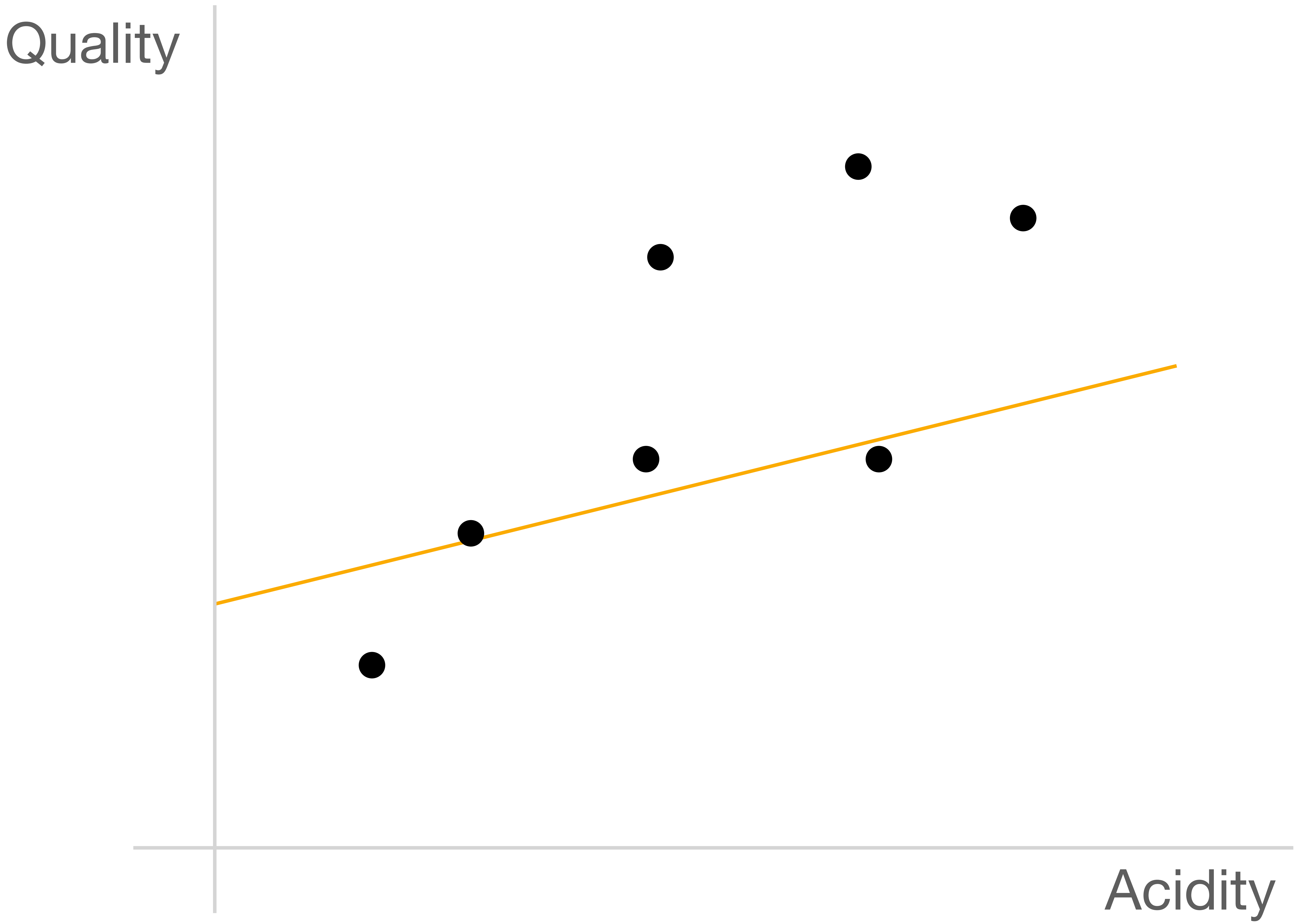
Linear Regression



Linear Regression



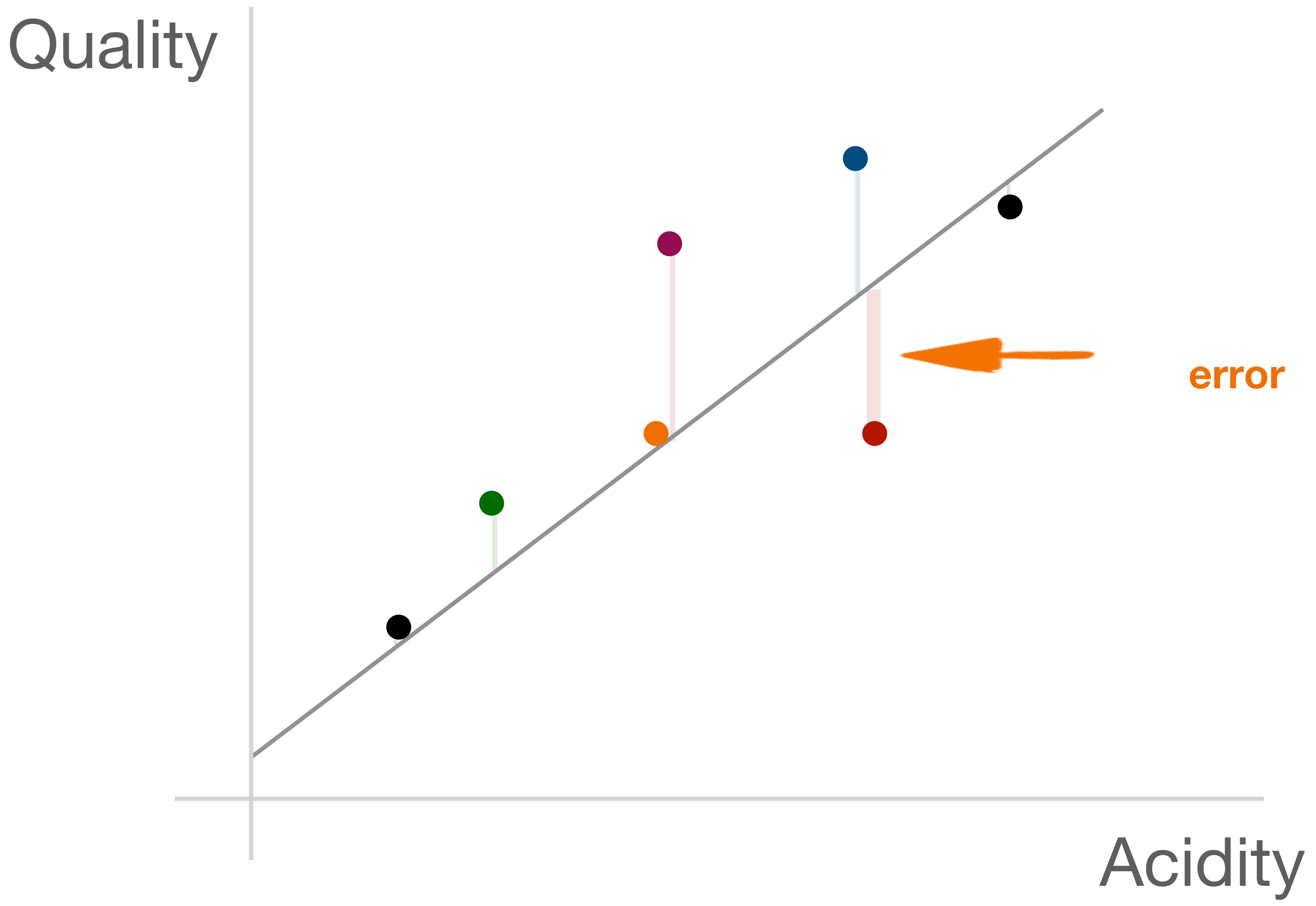
Linear Regression



How do we evaluate these models objectively?

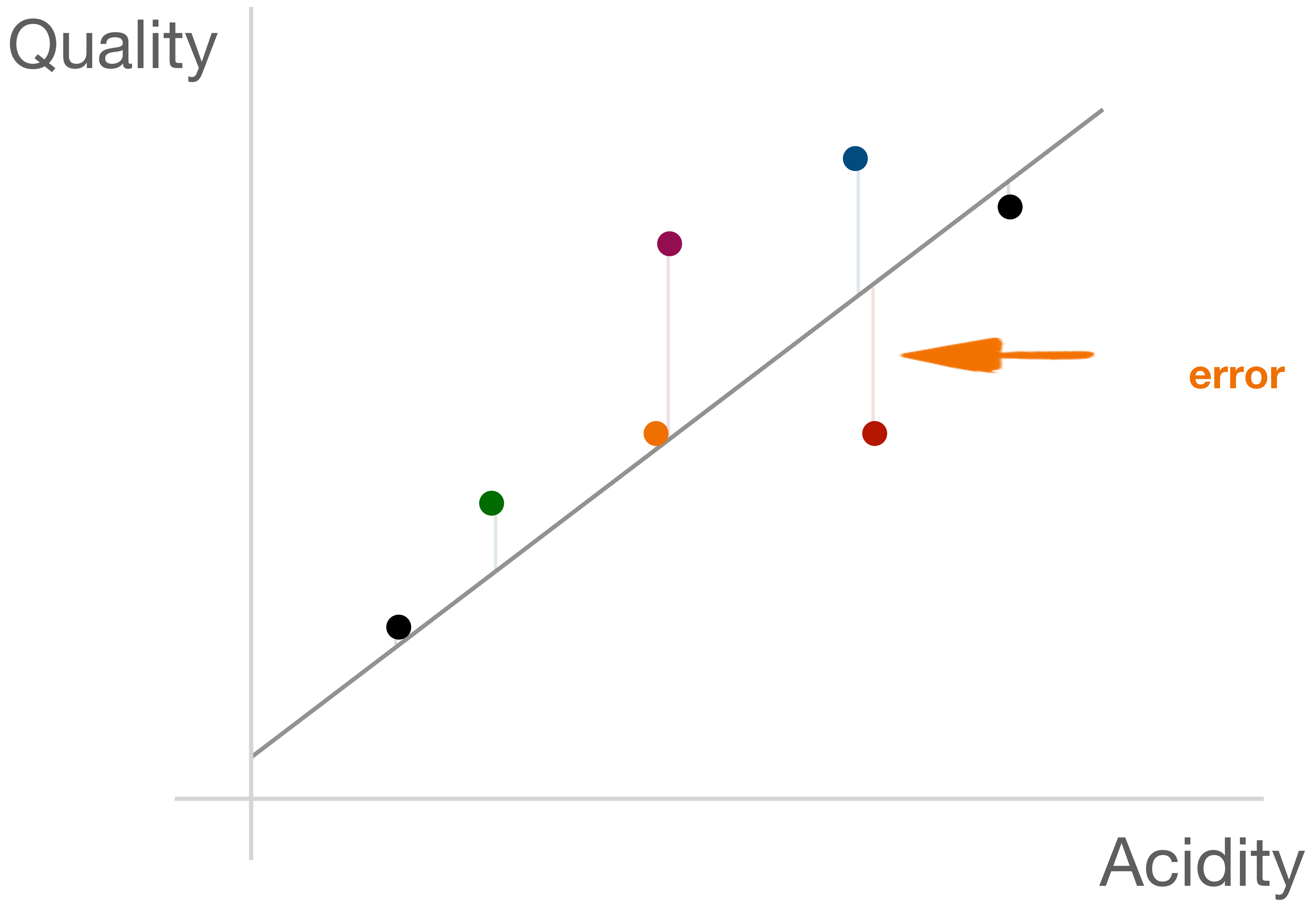
Calculating Error

- **error** is a measure of the “incorrectness” of a line



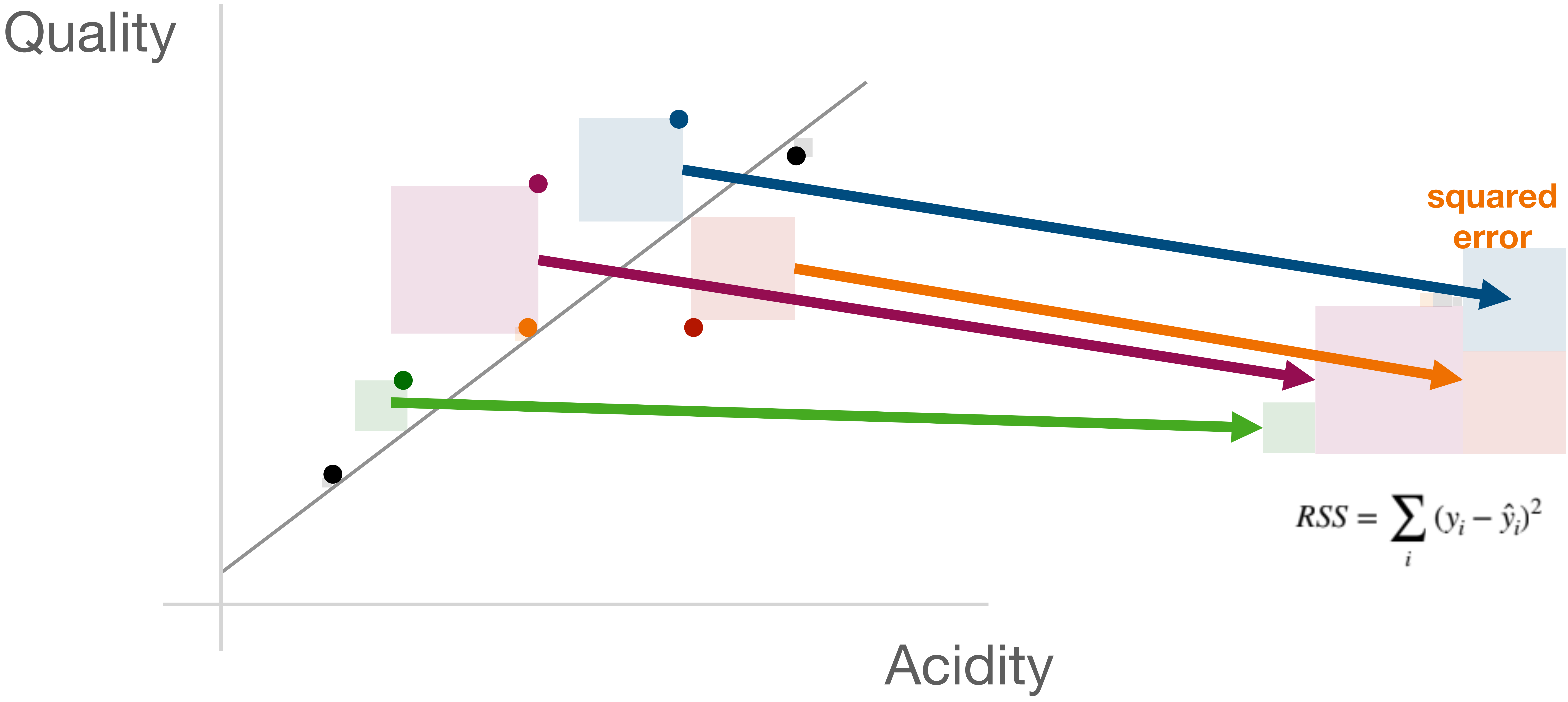
Calculating Error

- **simple error:** difference between the predicted value and the actual value



Calculating Error

- **sum-of-squares error**: sum of the squared difference between predicted and actual values



$$RSS = \sum_i (y_i - \hat{y}_i)^2$$

Calculating Error

- **How do we minimize error?** Cost function

Sum of all of the squared differences

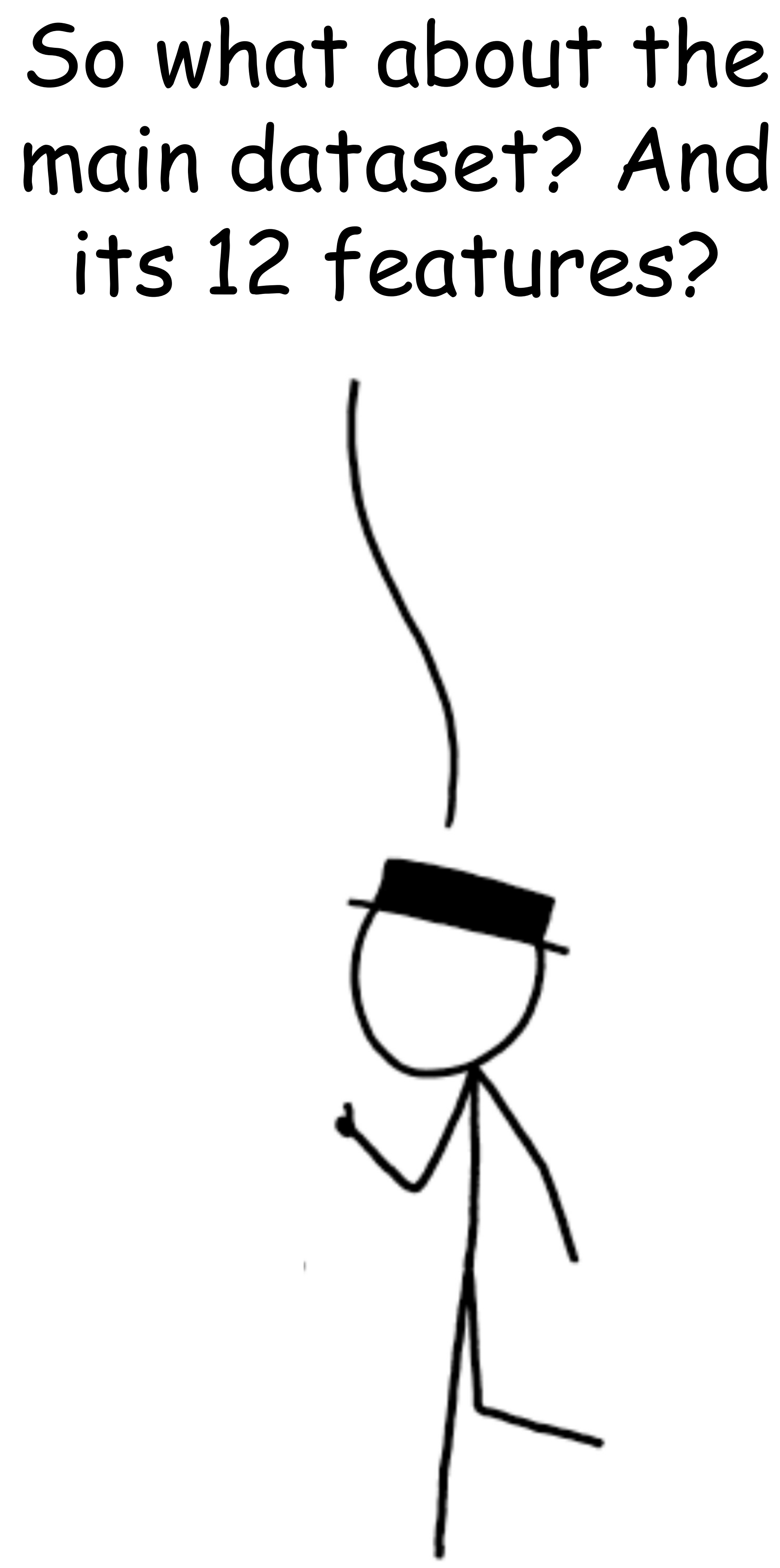
Square it!

$$\text{COST} = \frac{1}{n}\sum_{i=0}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

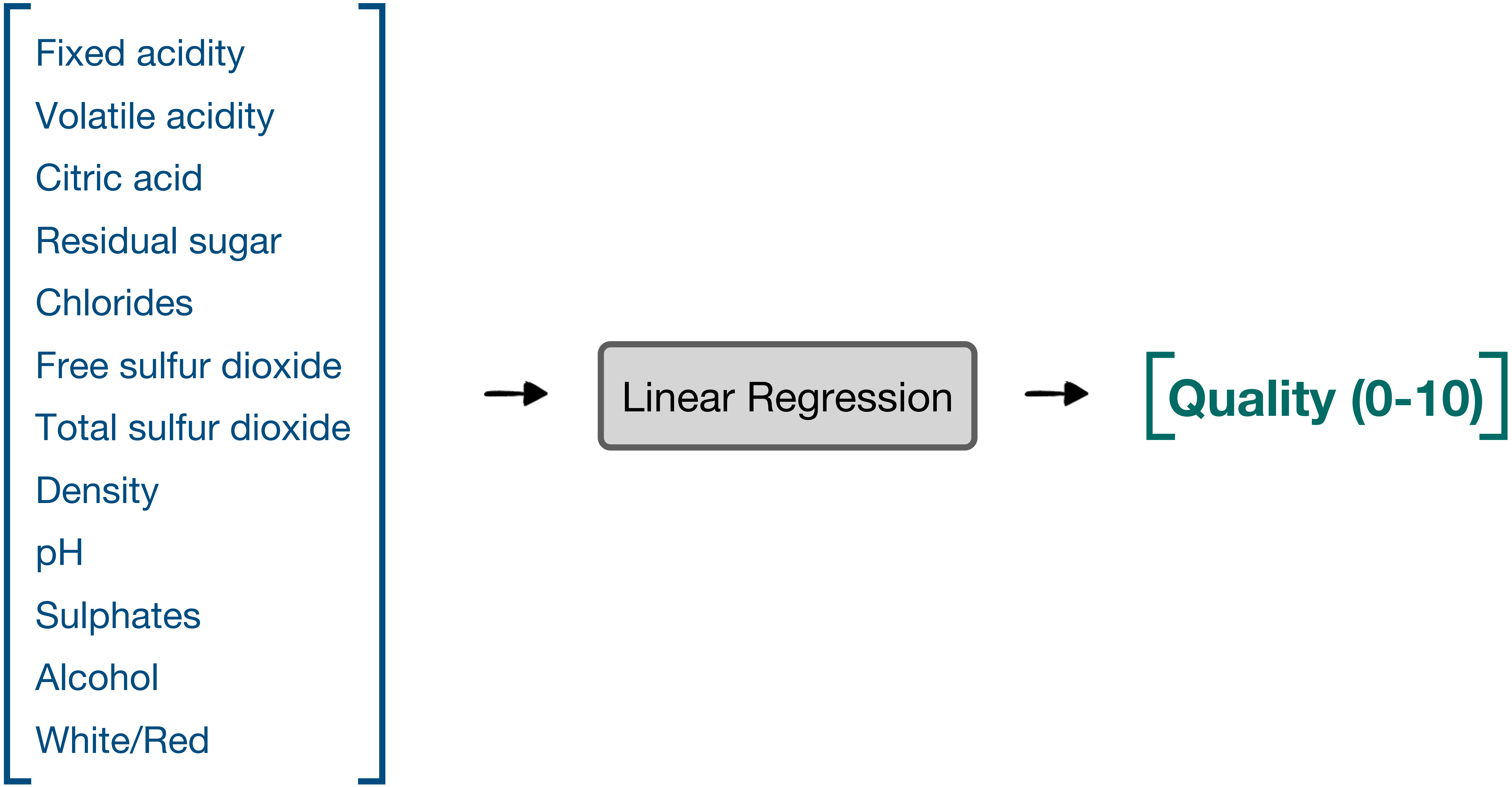
Divide by number of samples: mean error, not total error

Difference between predicted and actual

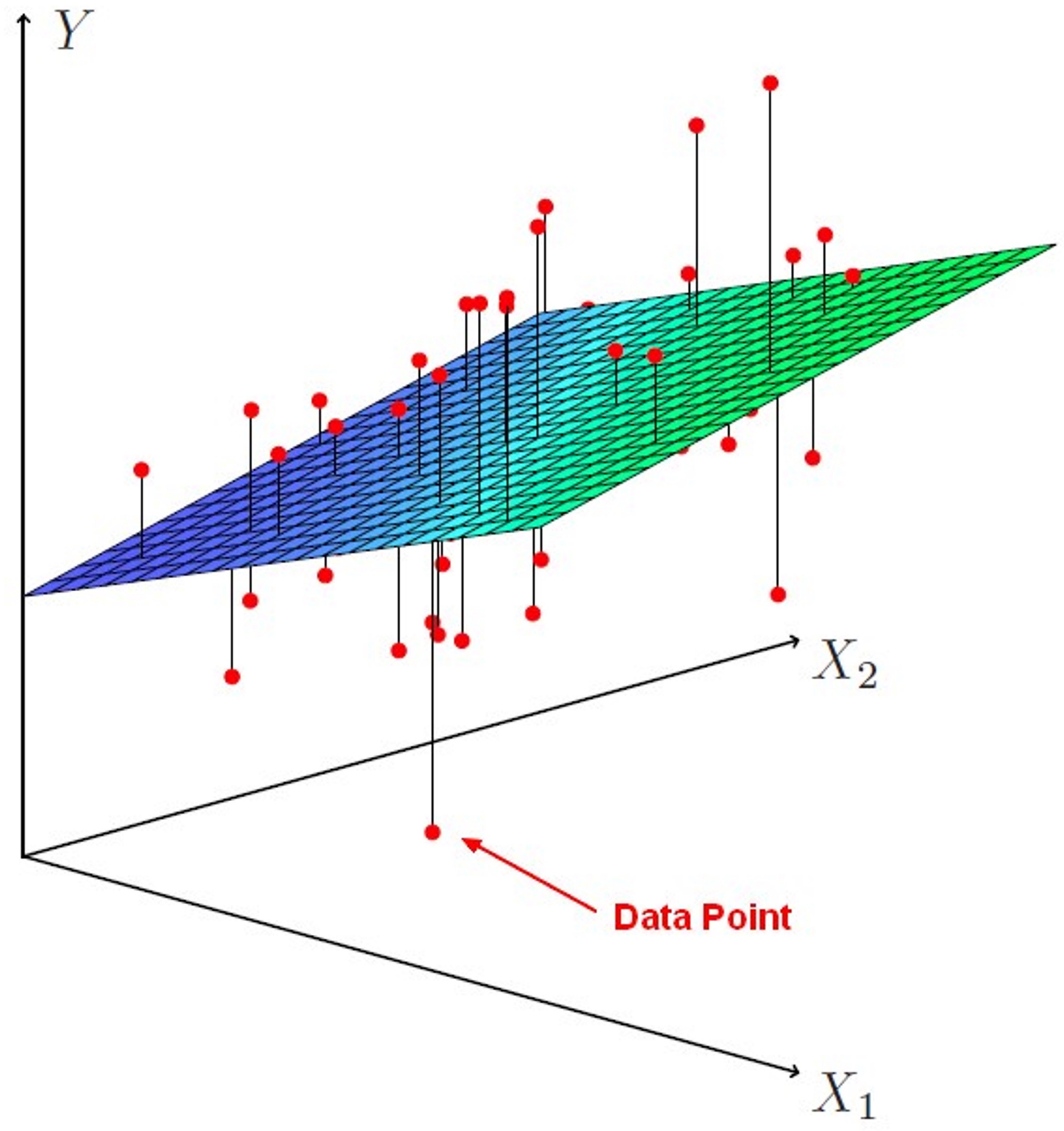
Higher-dimensional linear regression



Higher-dimensional linear regression



Higher-dimensional linear regression



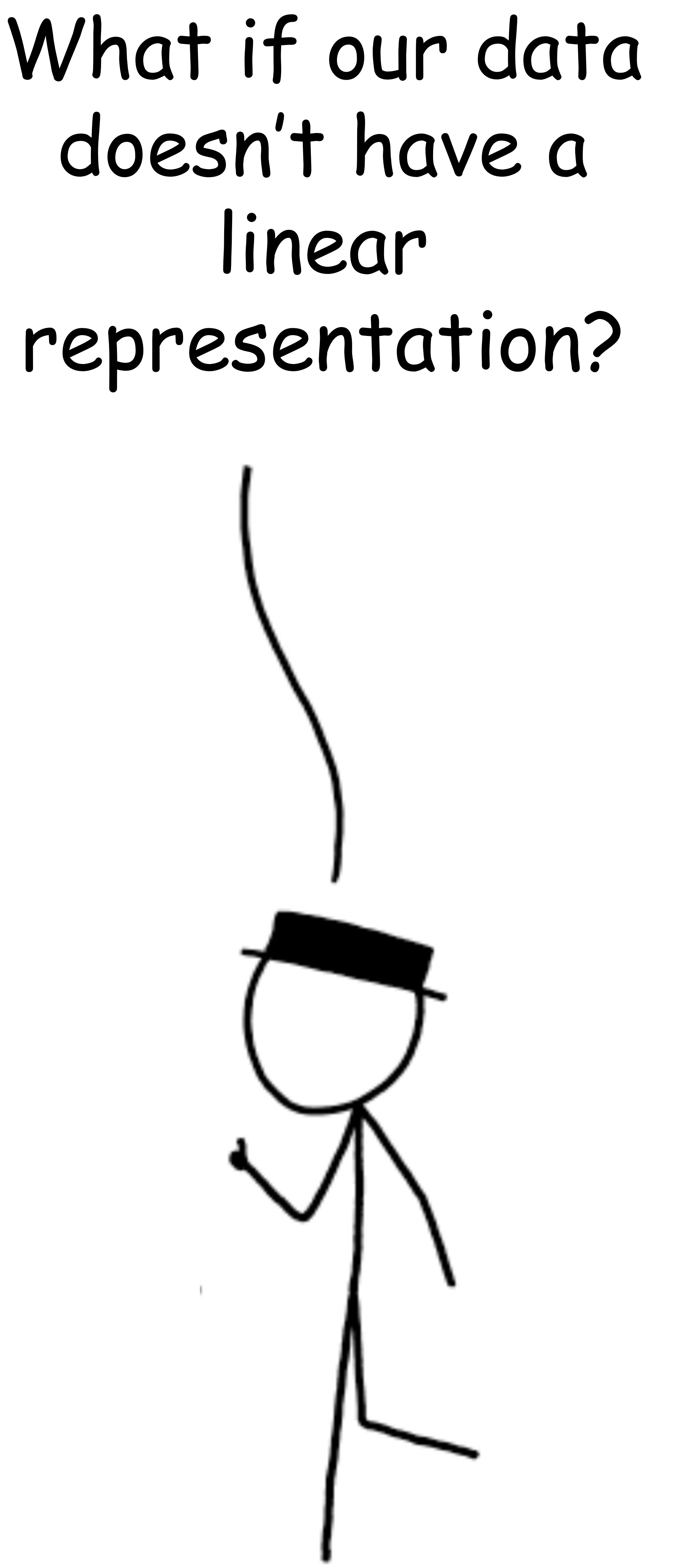
$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n$$

Hyperplanes!

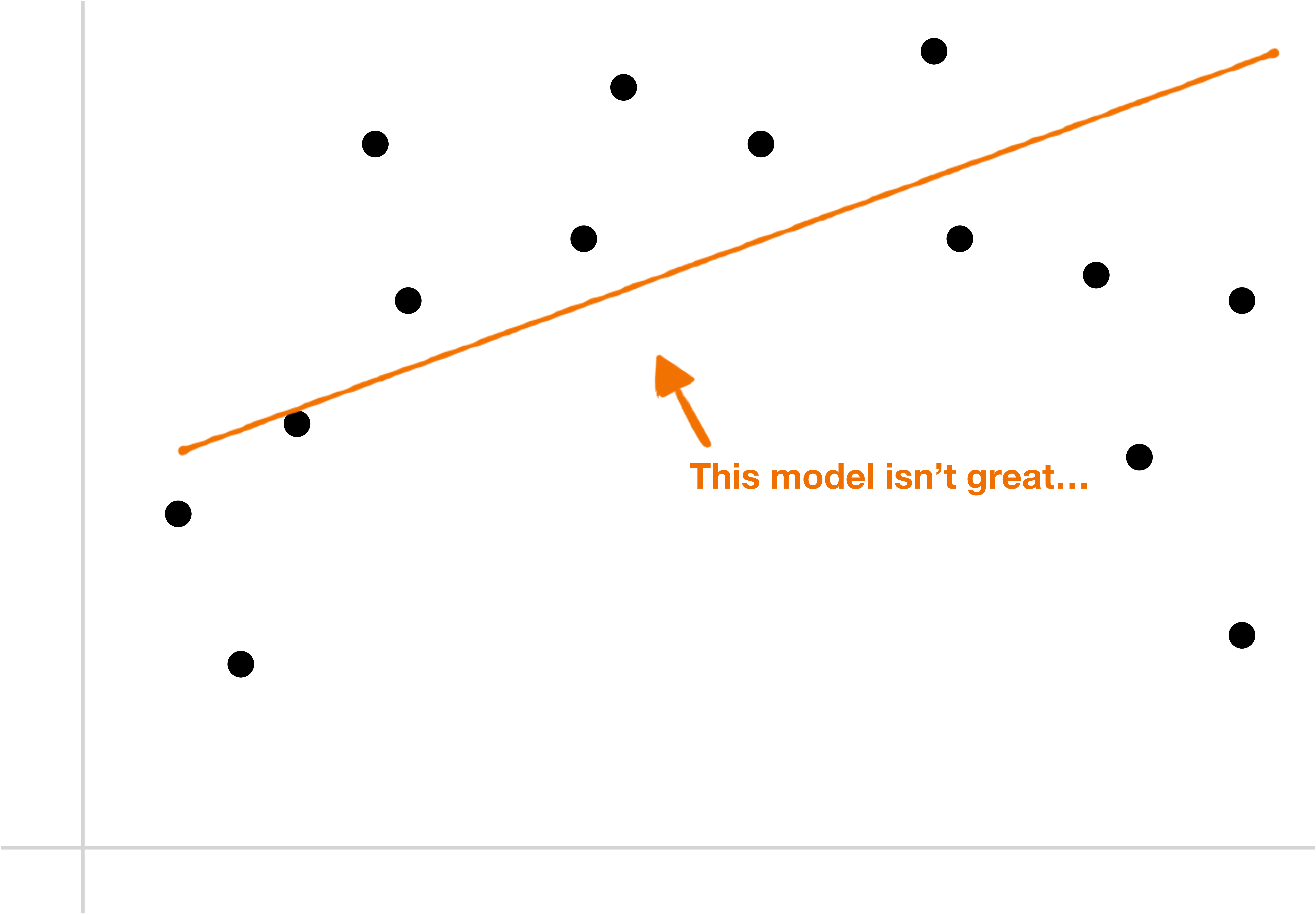
More dimensions, similar math.

One key assumption we make

- dataset **linearity**


What if our data doesn't have a linear representation?

The assumption fails…



- model is unable to capture relationship between variables

We need to make our model more powerful!

POLYNOMIAL REGRESSION

Polynomial Regression

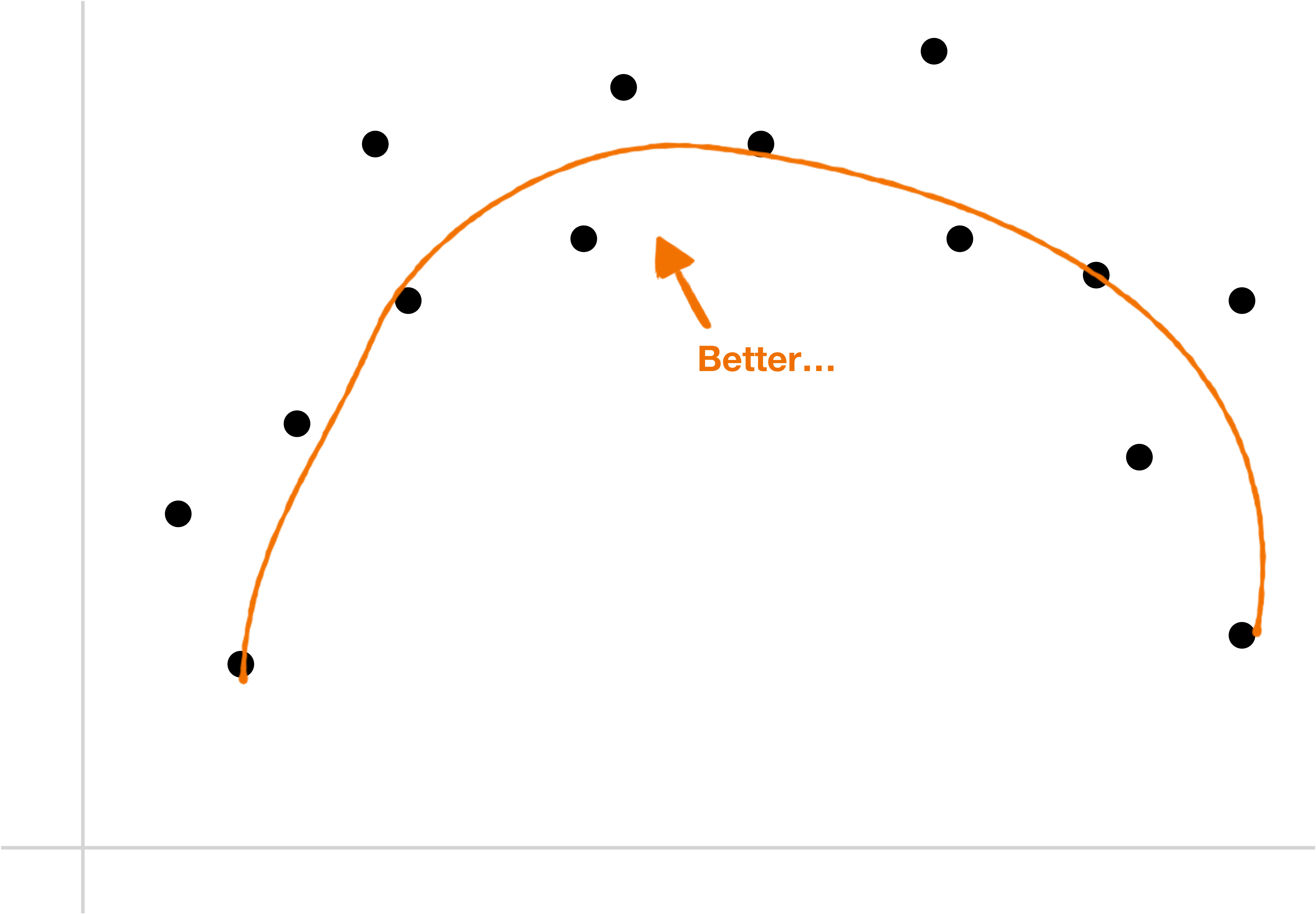
- introducing **higher-dimensional terms** to add **curvature**

nth degree term

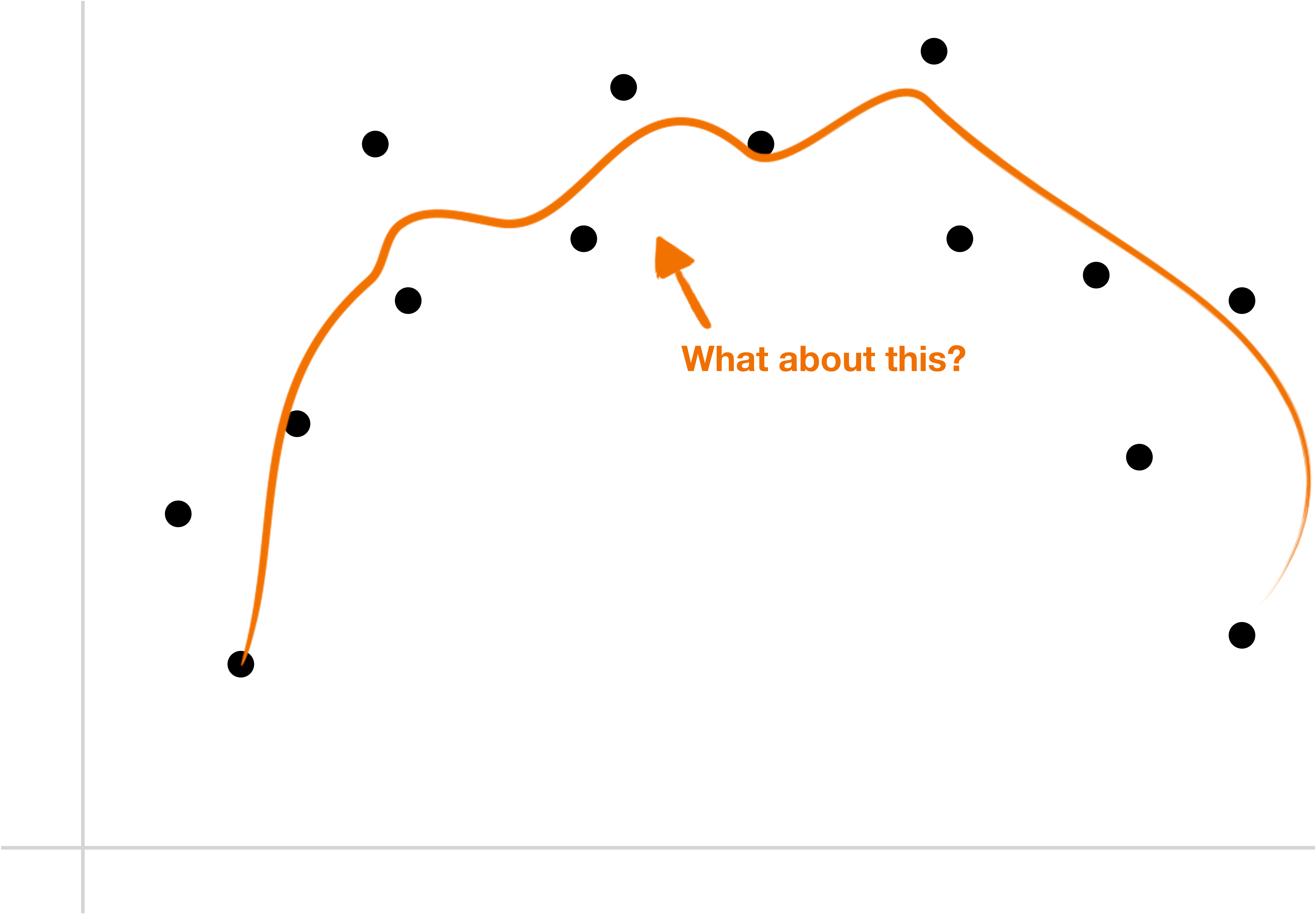
$$\hat{y} = w_n x^n + w_{n-1} x^{n-1} + w_{n-2} x^{n-2} + \cdots + w_1 x + w_0$$

nth term coefficient

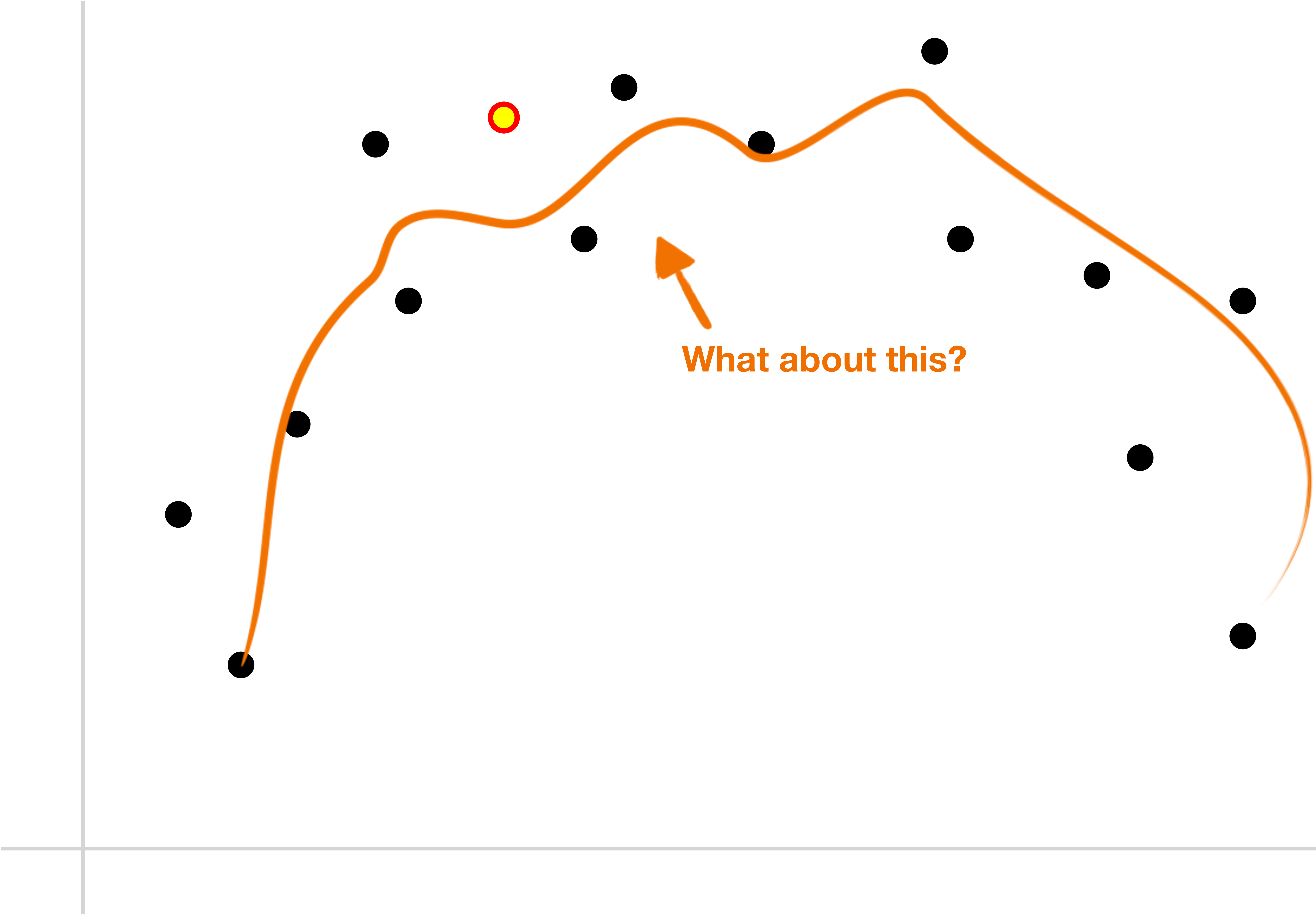
Polynomial Regression



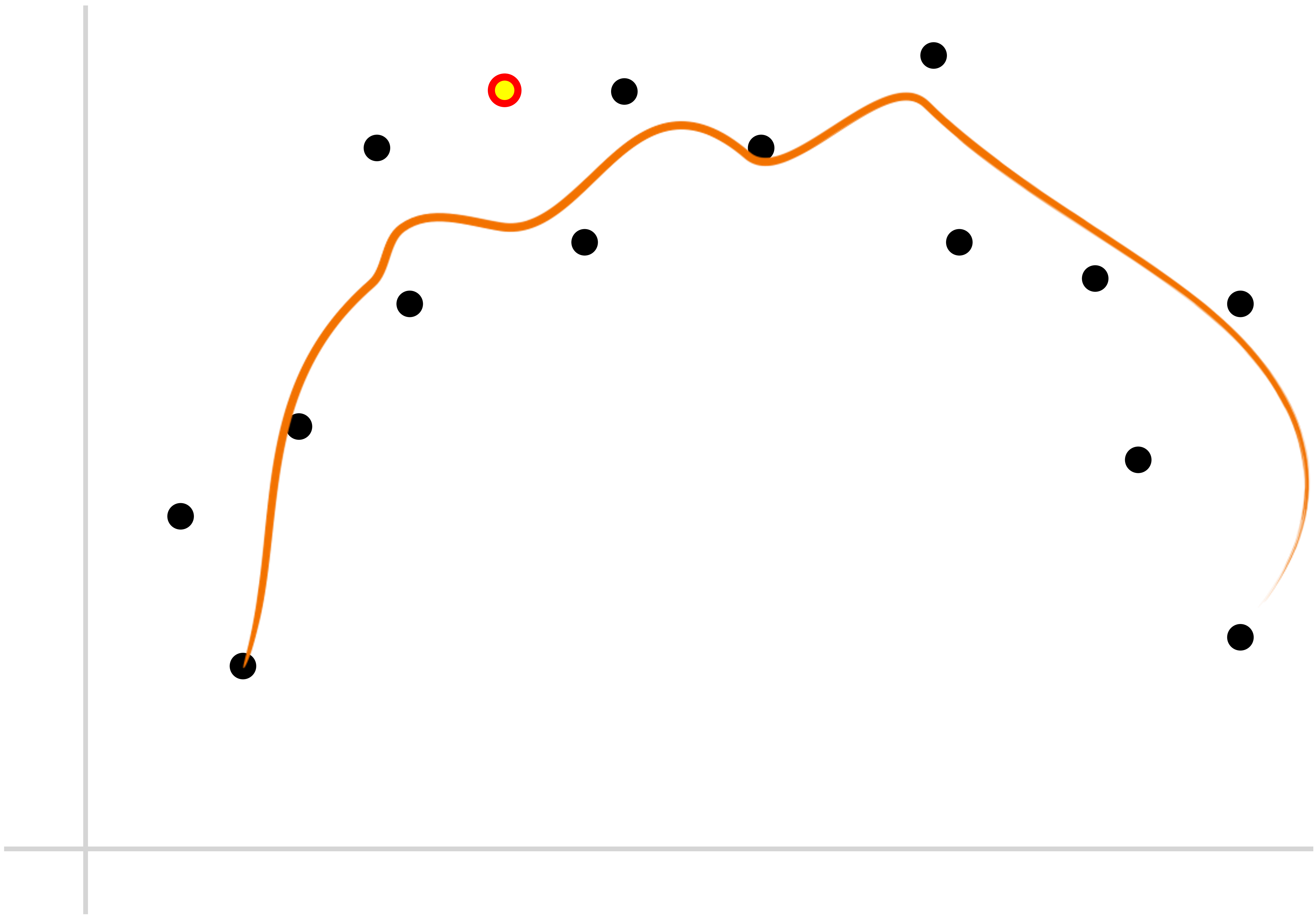
Polynomial Regression

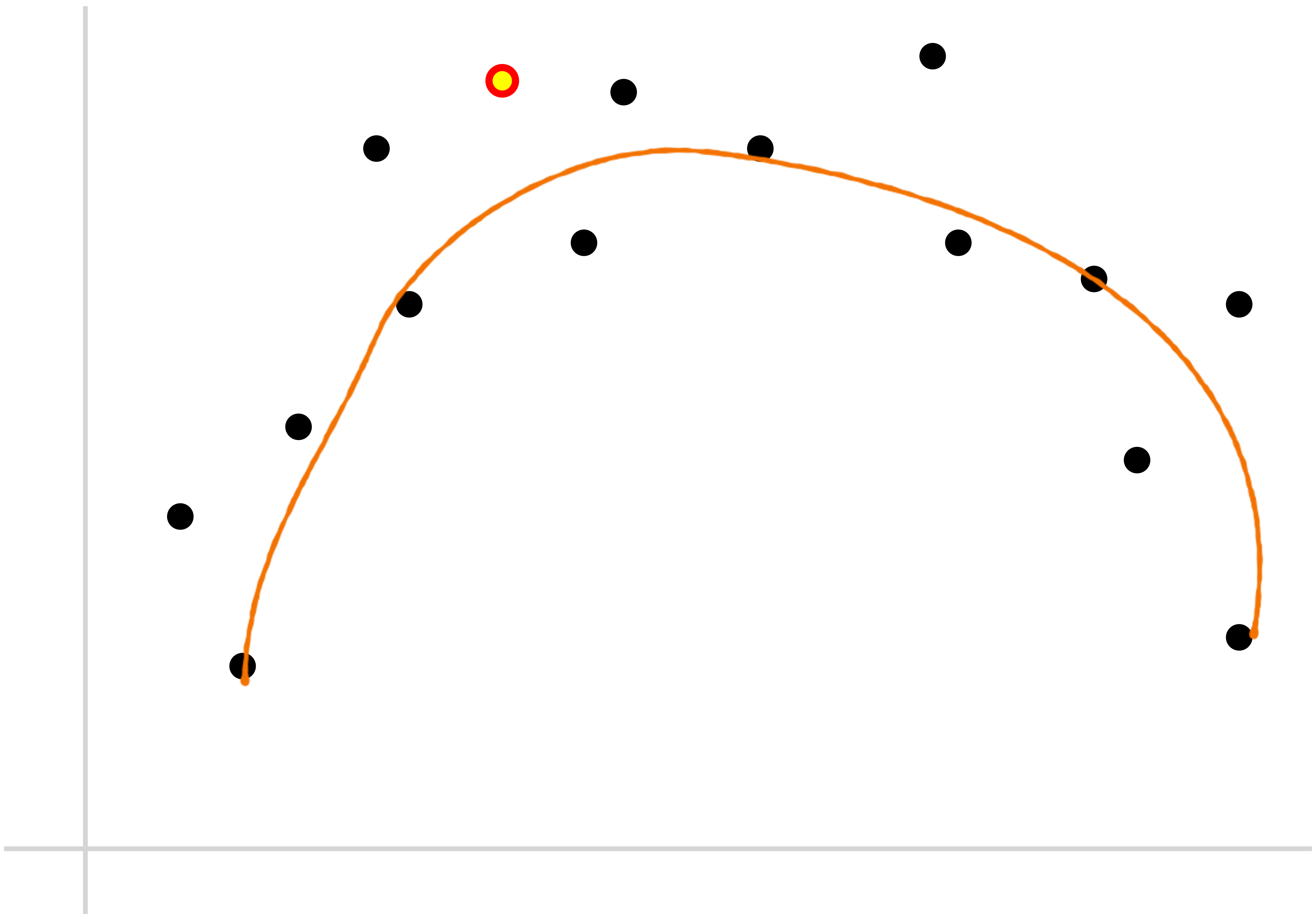


Polynomial Regression

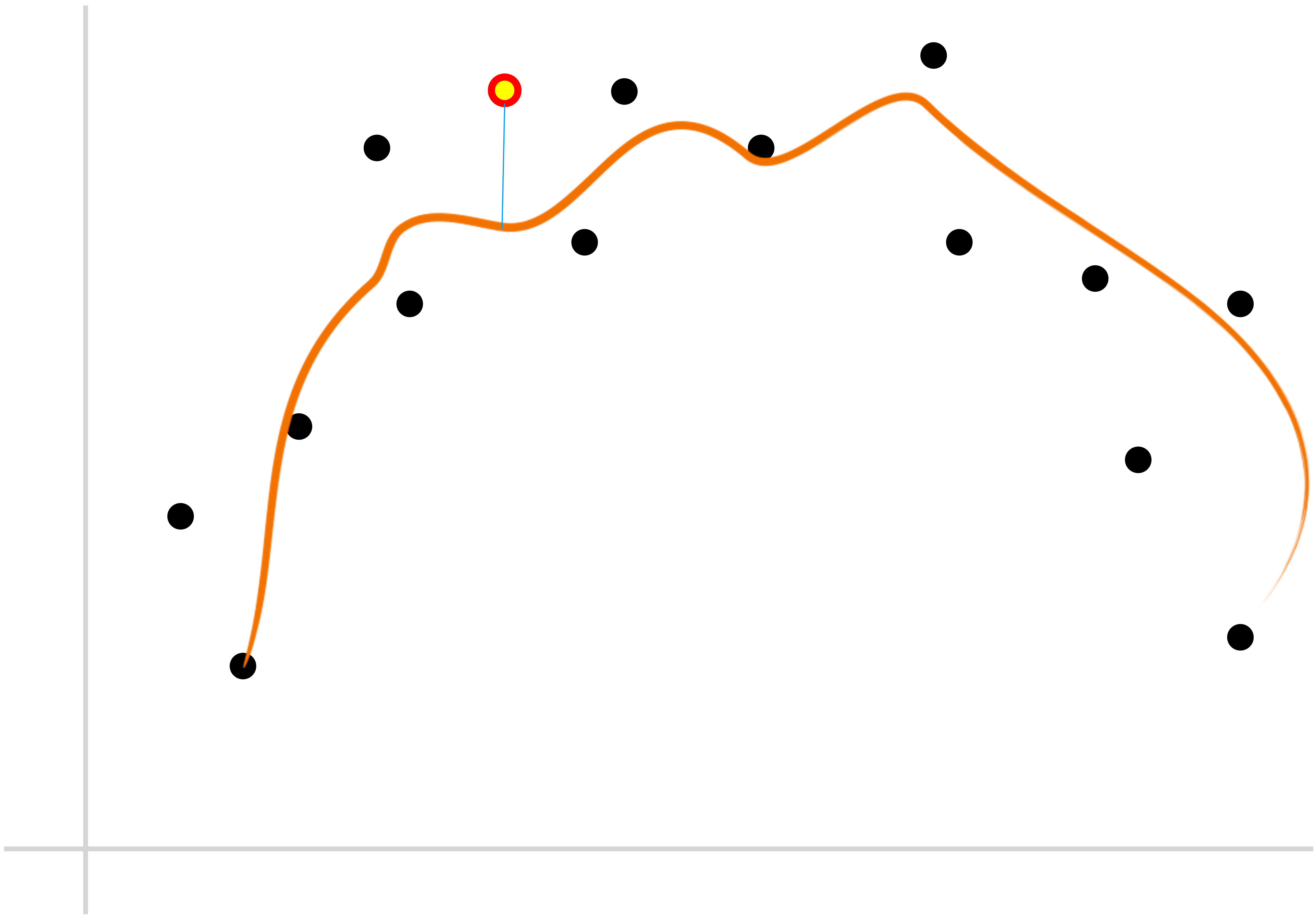


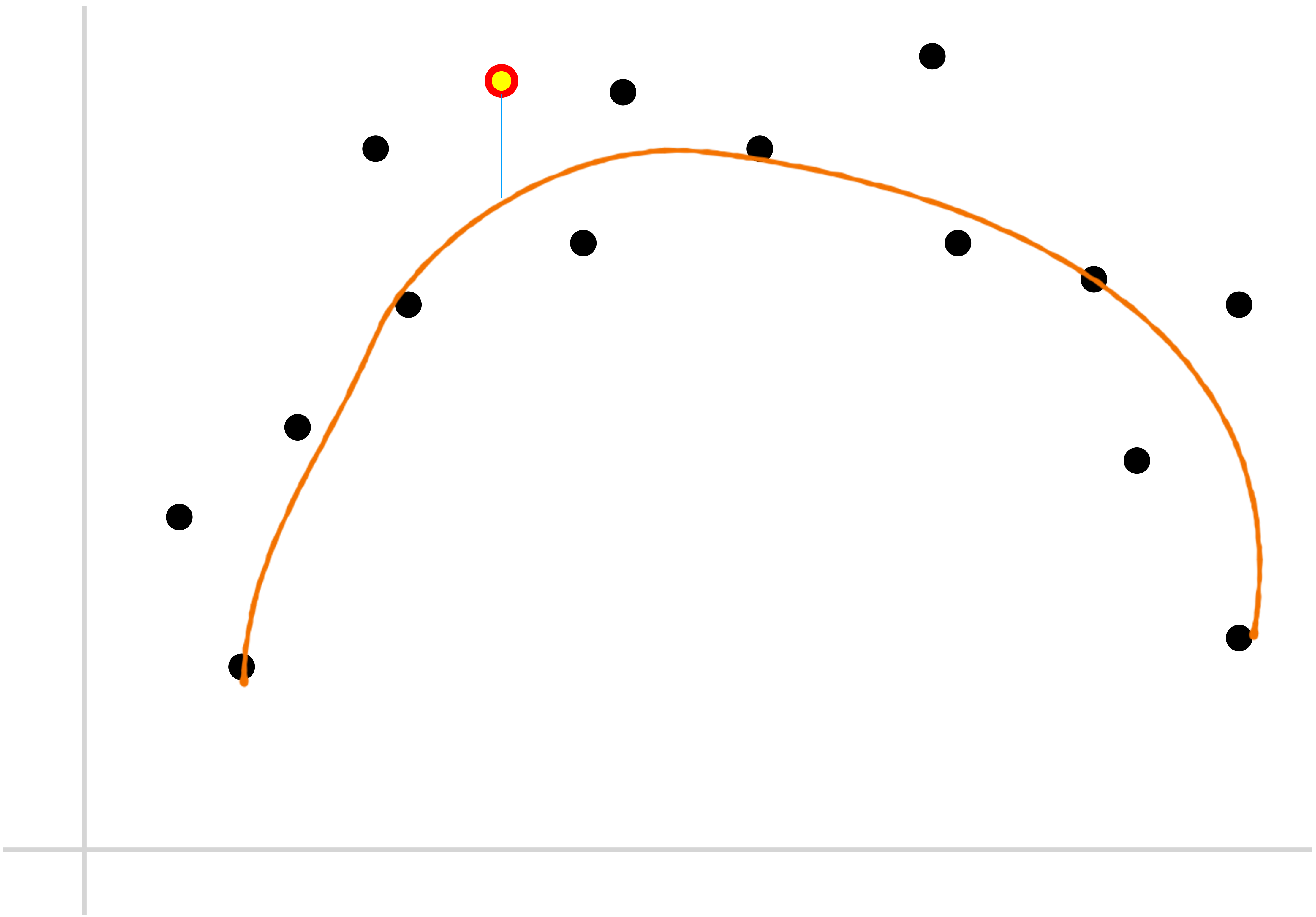
Polynomial regression - Which one is better?





A significant difference!





Polynomial Regression

We overfit the dataset…

Calculating Error

- The **same** cost function!

Sum of all of the squared differences

Square it!

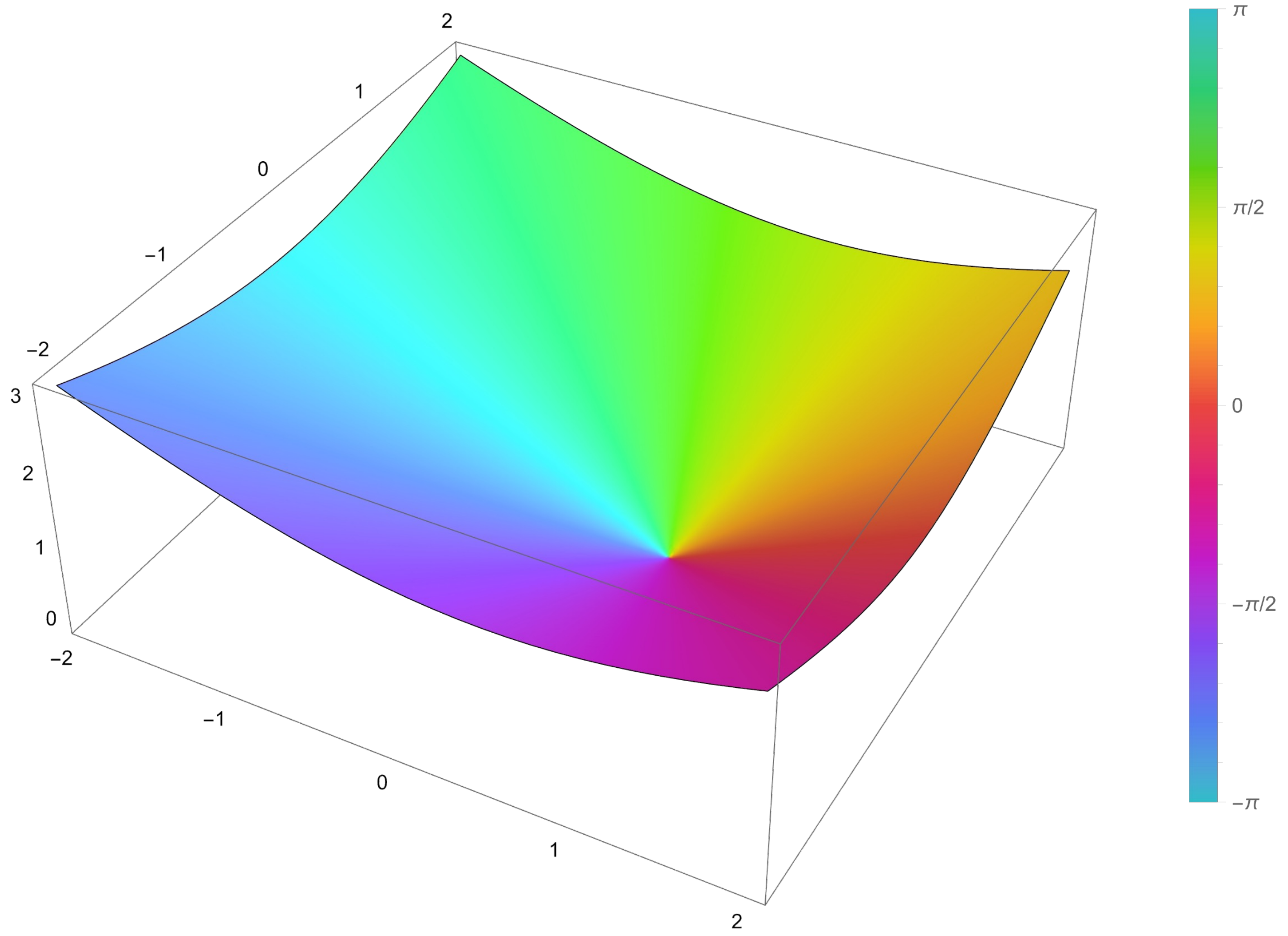
$$\text{COST} = \frac{1}{n}\sum_{i=0}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

Divide by number of samples: mean error, not total error

Difference between predicted and actual

Higher-dimensional polynomial regression

- **More powerful** (and complex) models!



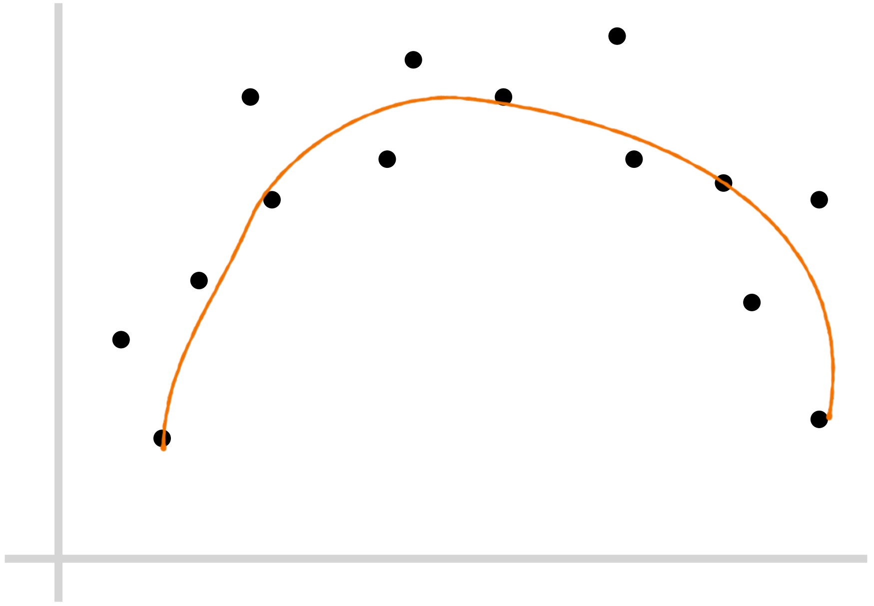
Higher-dimensional polynomial regression

- **More powerful** (and complex) models!

$$\hat{y} = w_1 x_1 + w_2 x_2^2 + w_3 x_3^3 + \cdots + w_n x_n^n + \cdots + w_i x_1 x_2 + \cdots$$

Higher-dimensional polynomial regression

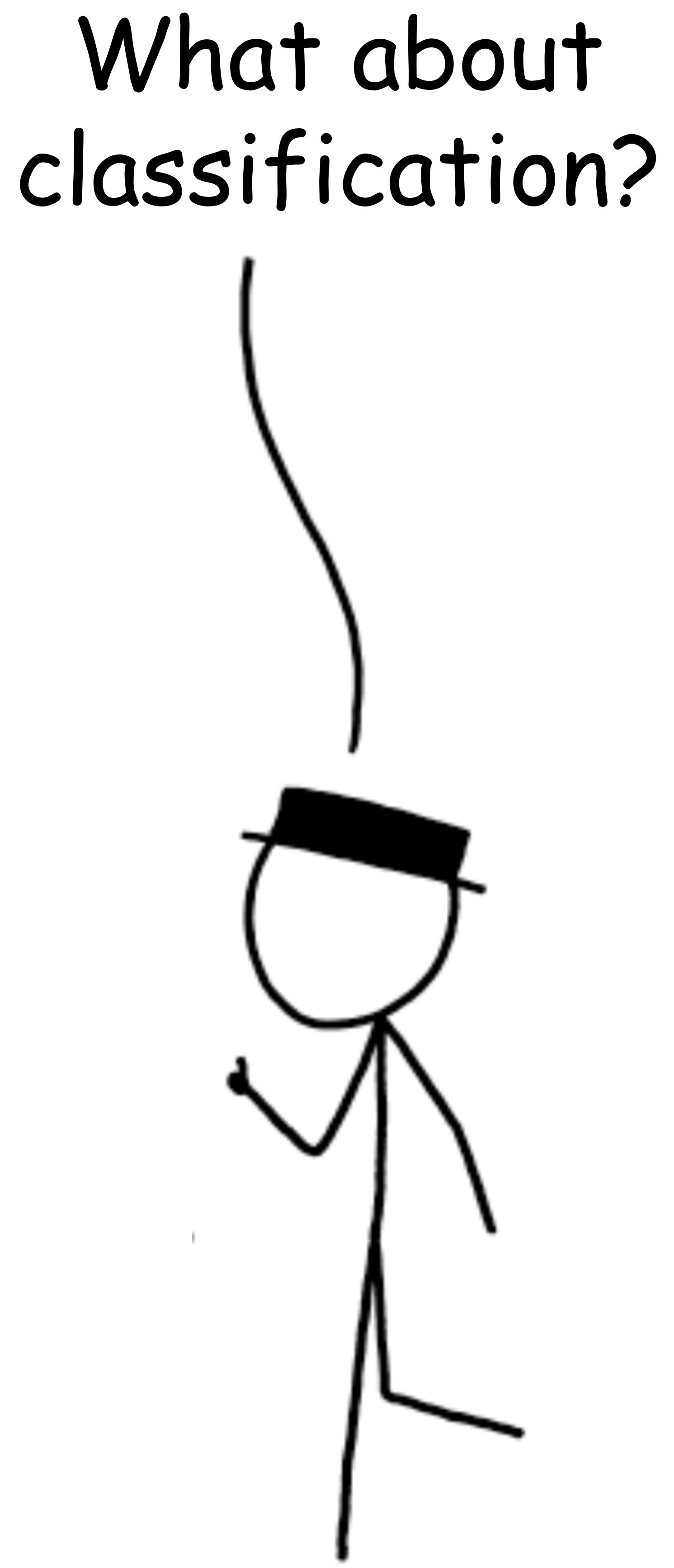
- Don’t worry, the math stays the same:



$$\hat{y} = w_n x^n + w_{n-1} x^{n-1} + w_{n-2} x^{n-2} + \cdots + w_1 x + w_0$$

Build a polynomial regression model → **Evaluate the cost function** → **Minimize the cost function and optimize our model**

$$\text{COST} = \frac{1}{n} \sum_{i=0}^{n} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$$


What about classification?

LOGISTIC REGRESSION

Logistic Regression

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol

→

a class

0

1

- categorical label outputs are named "**classes**"

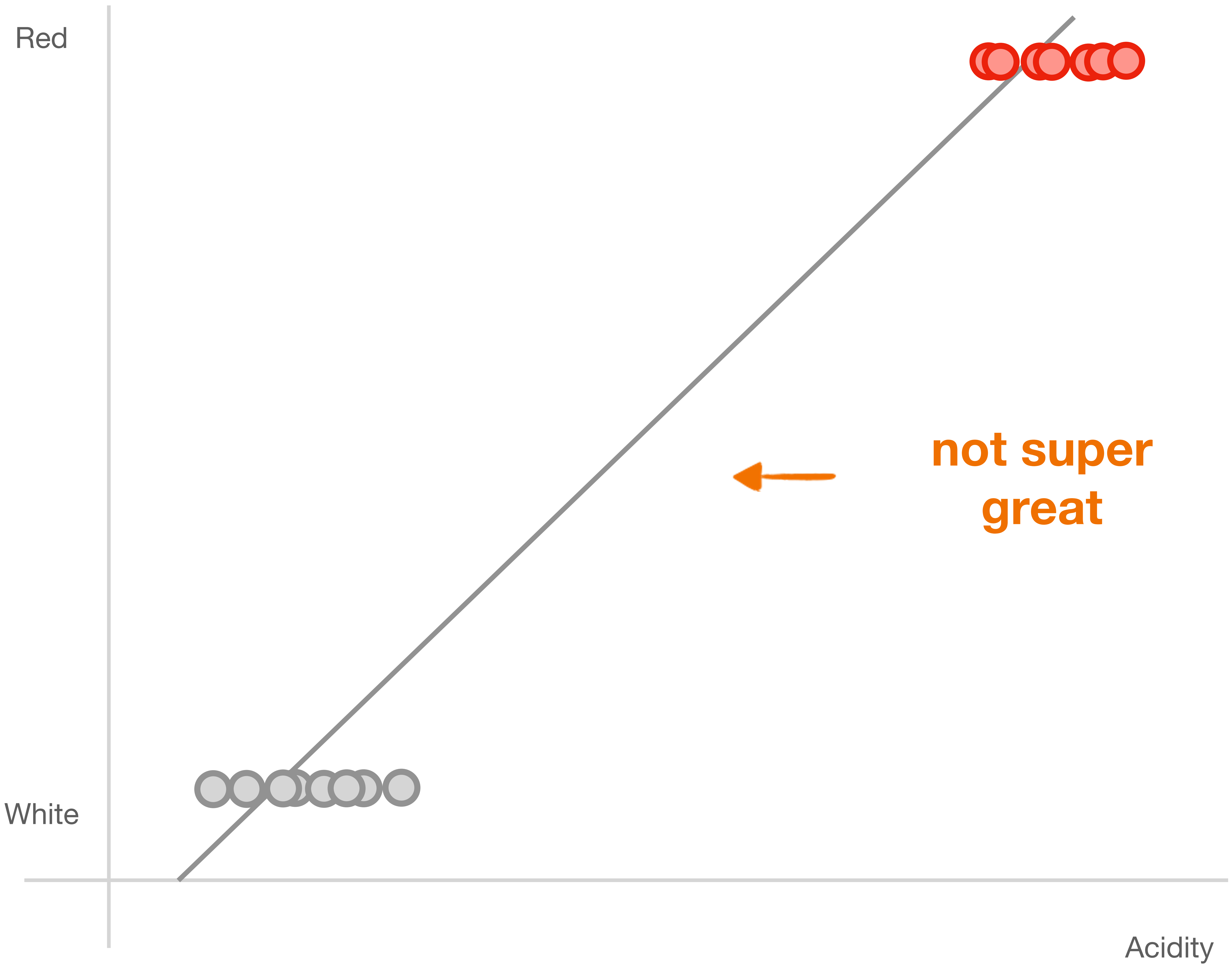
Logistic Regression

White = 0

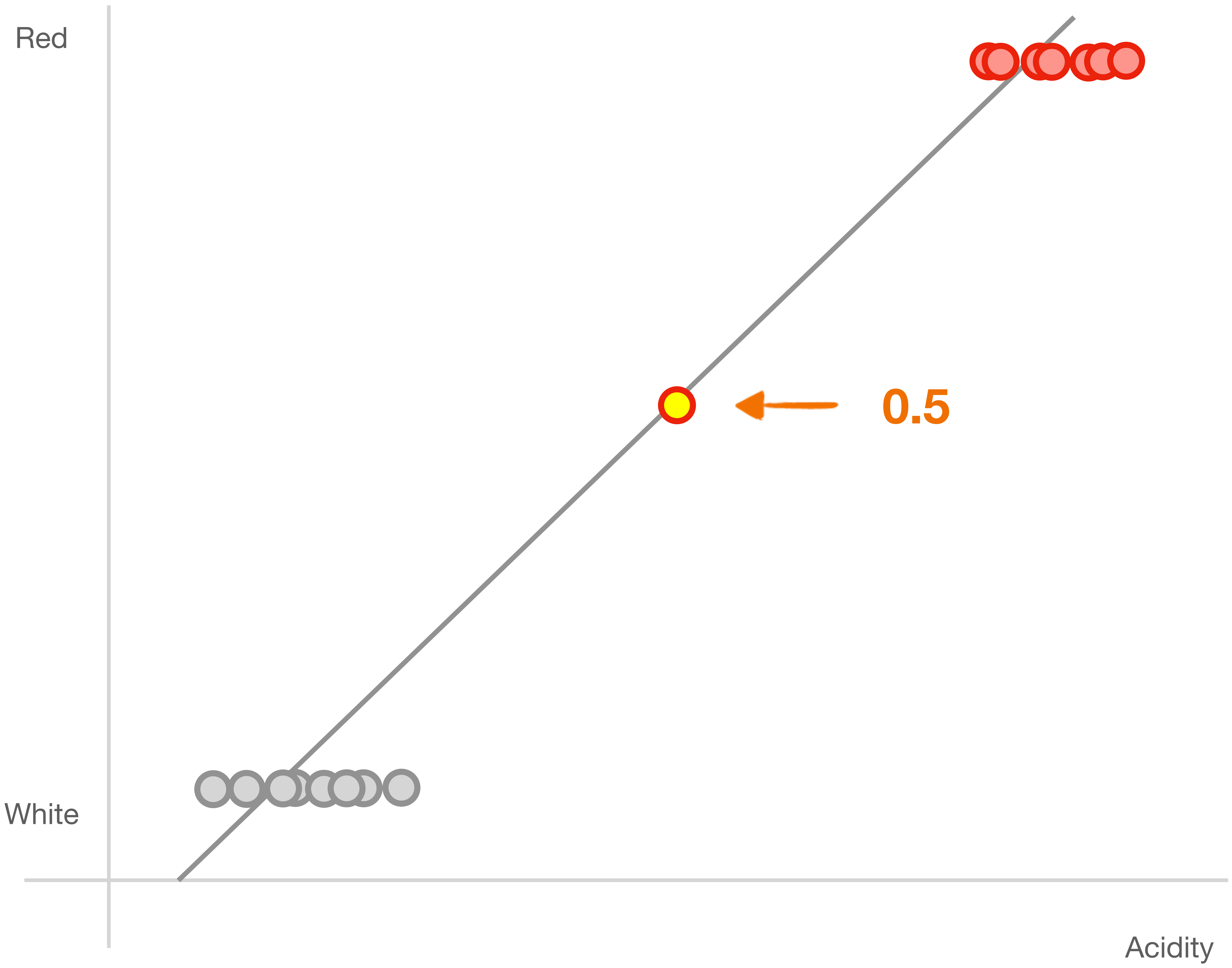
Red = 1

Why can’t we just use linear or polynomial regression?

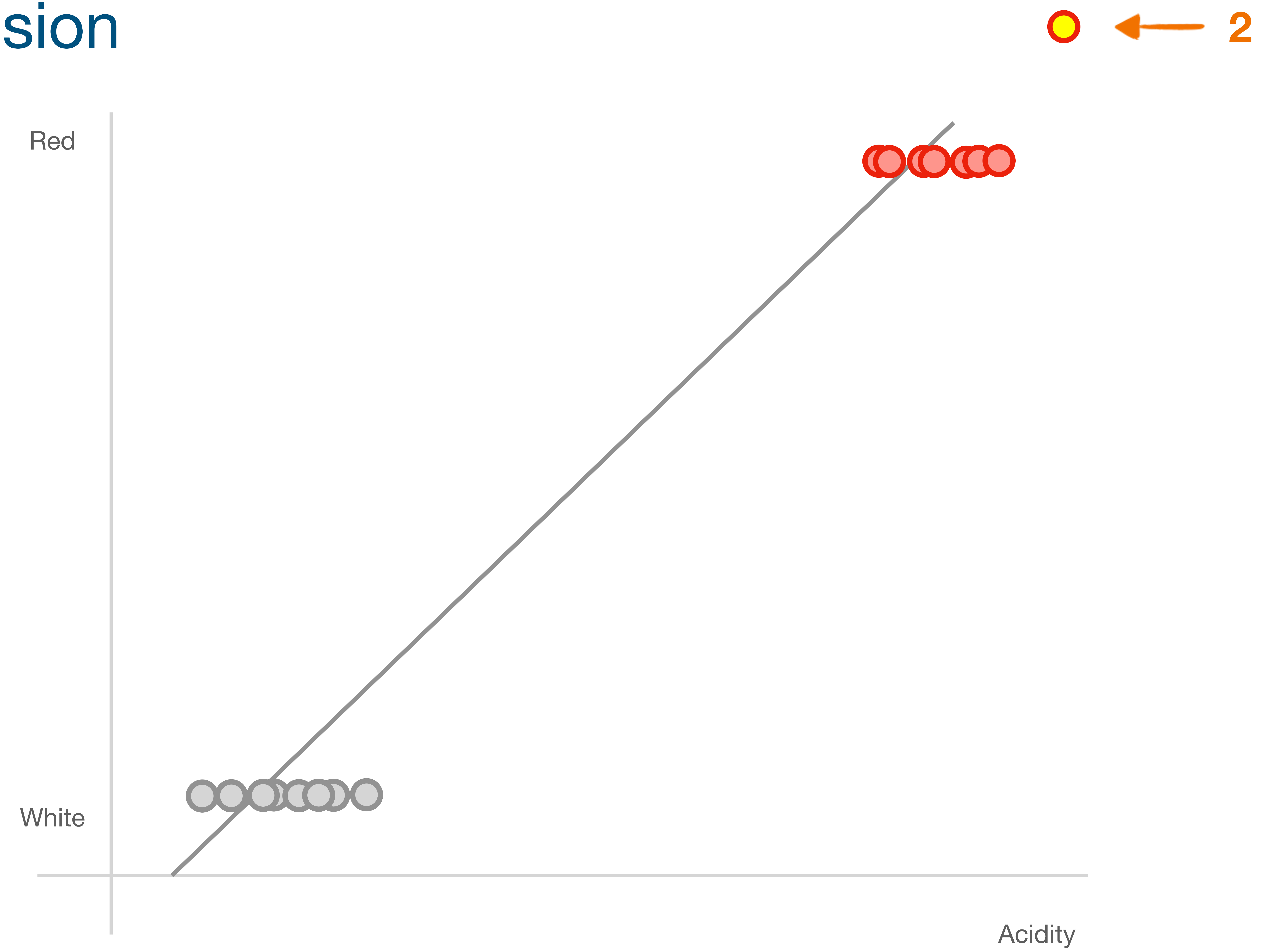
Logistic Regression



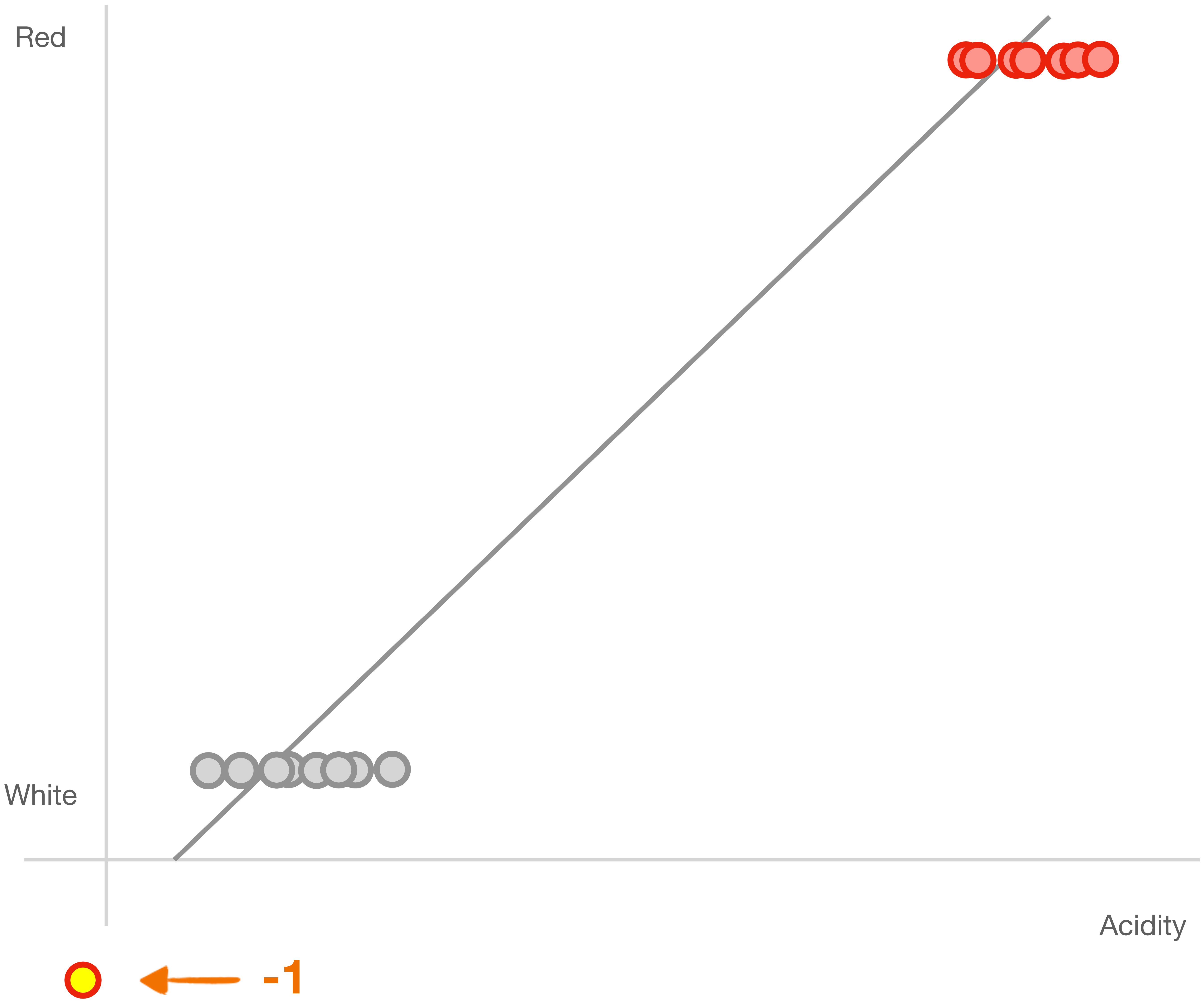
Logistic Regression



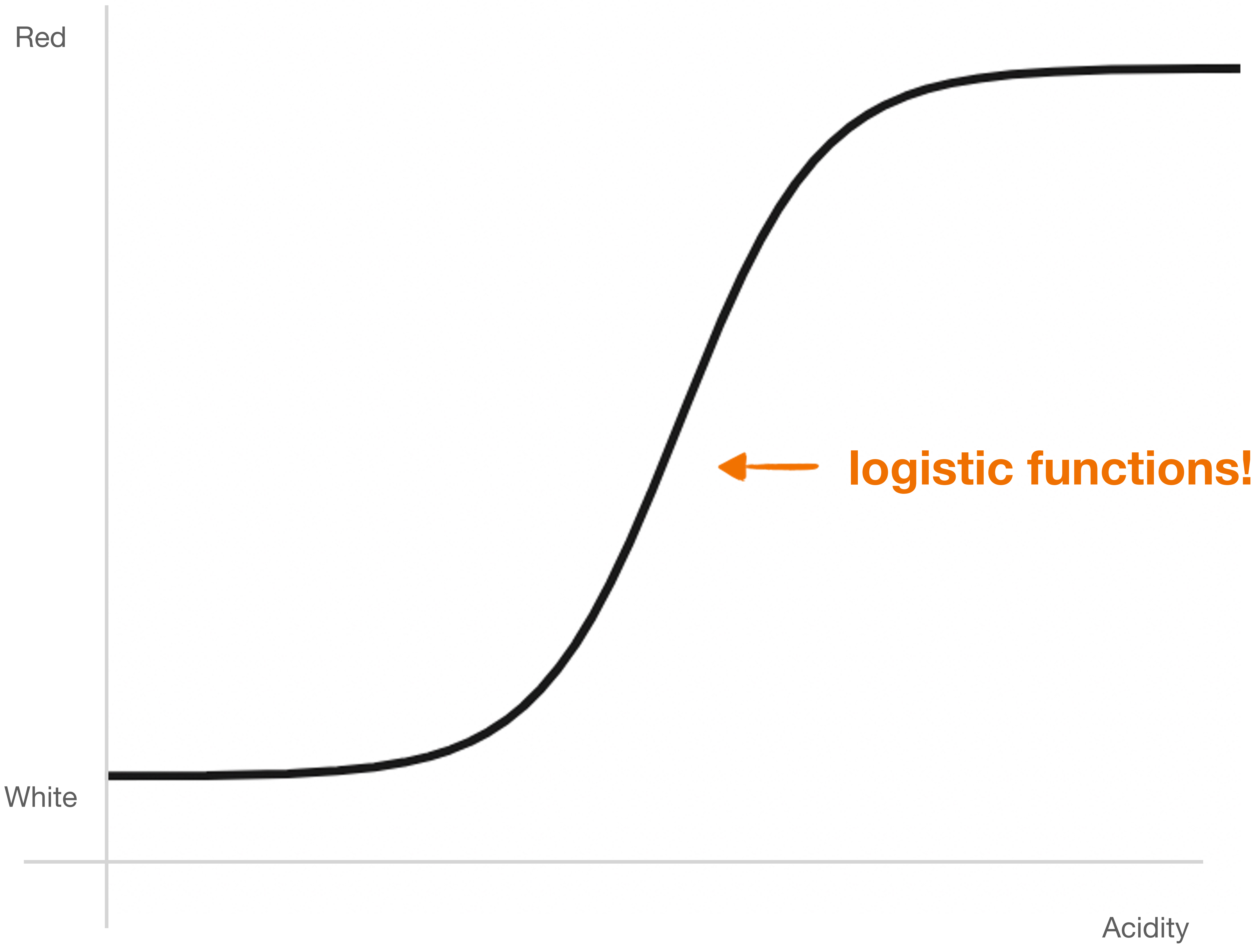
Logistic Regression



Logistic Regression



Logistic Regression



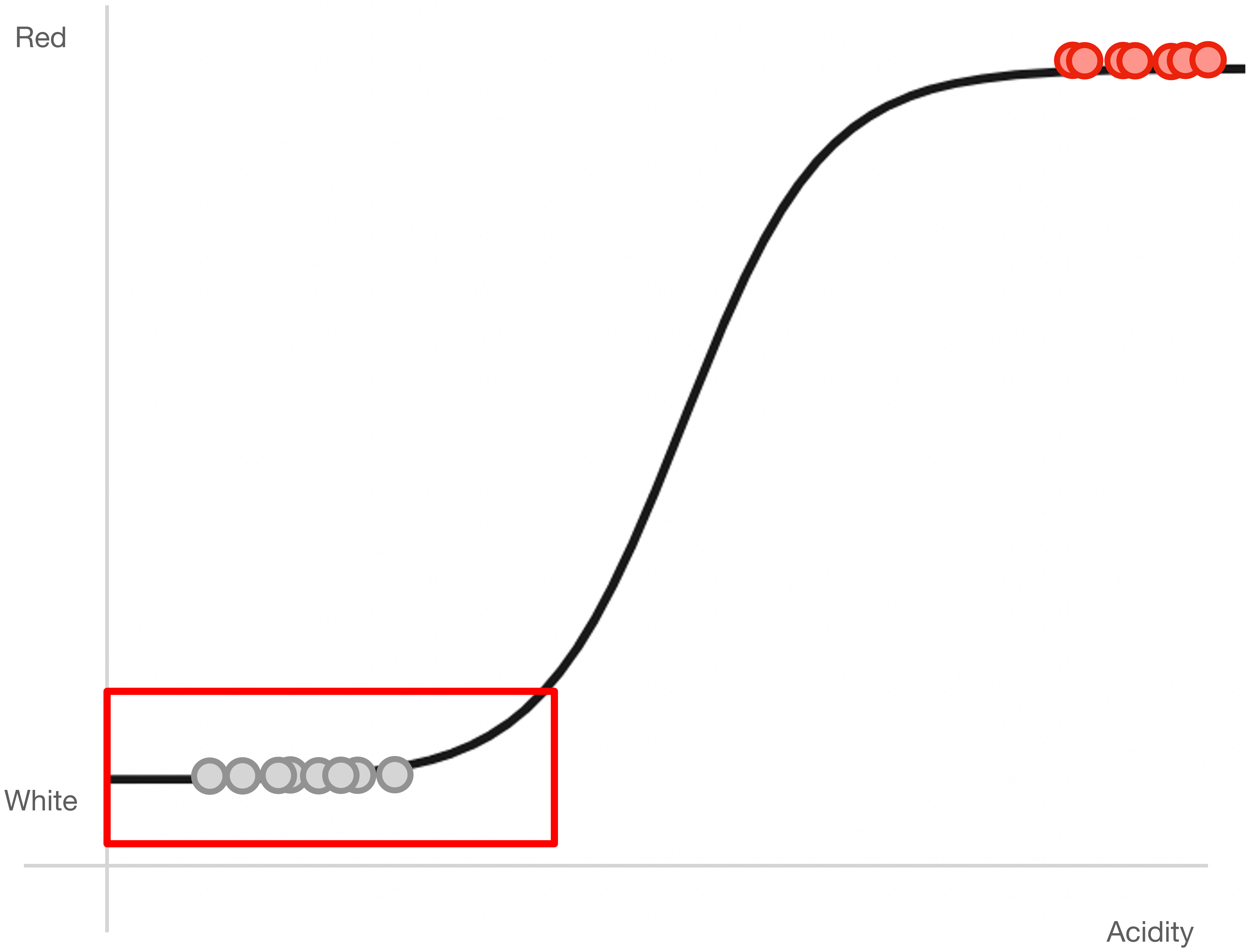
Logistic Regression

A logistic function:

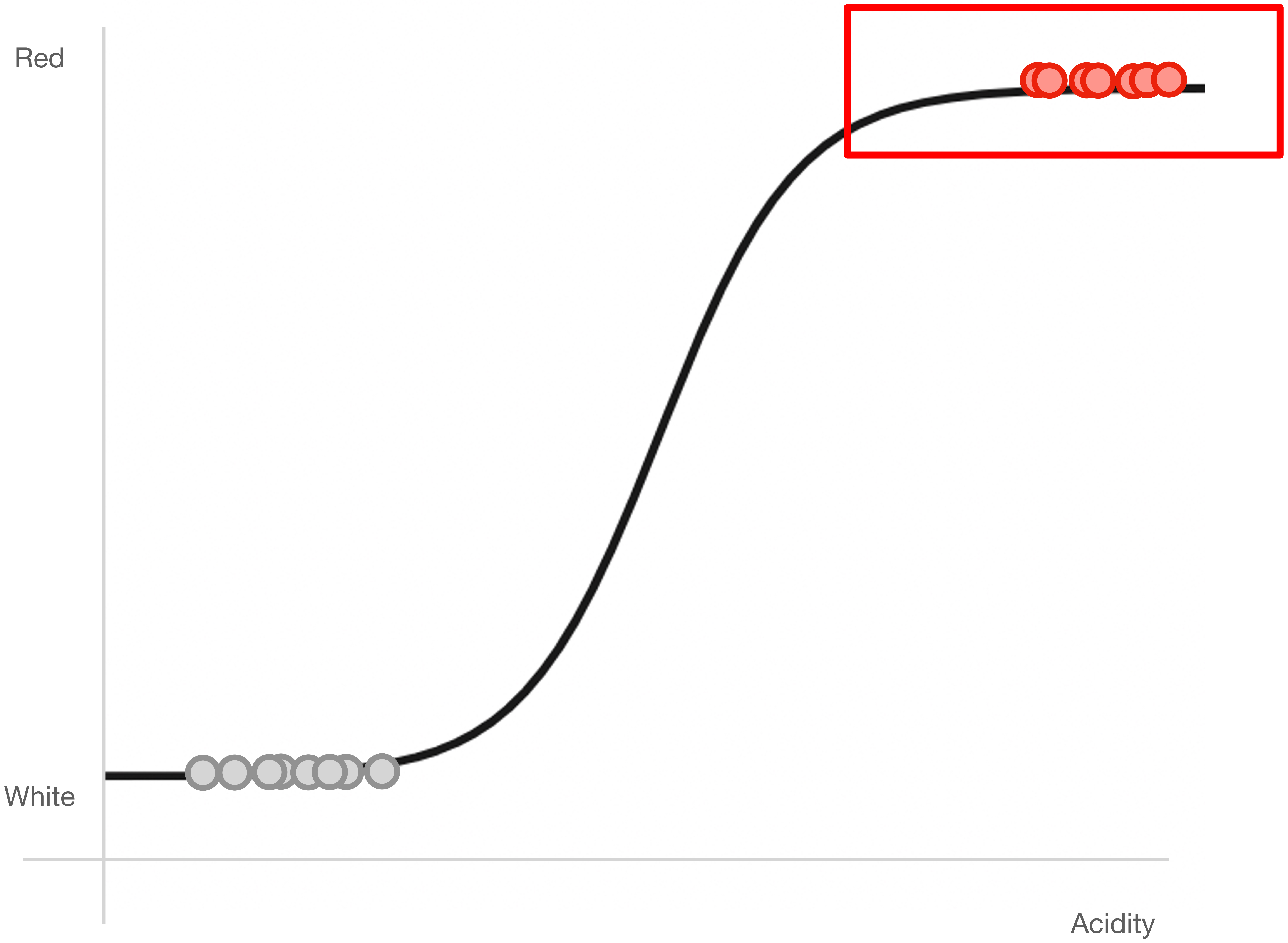
$$\hat{y} = \frac{1}{1 + e^{-w_0 + w_1 x_1 + \cdots + w_n x_n}}$$

our input features are here

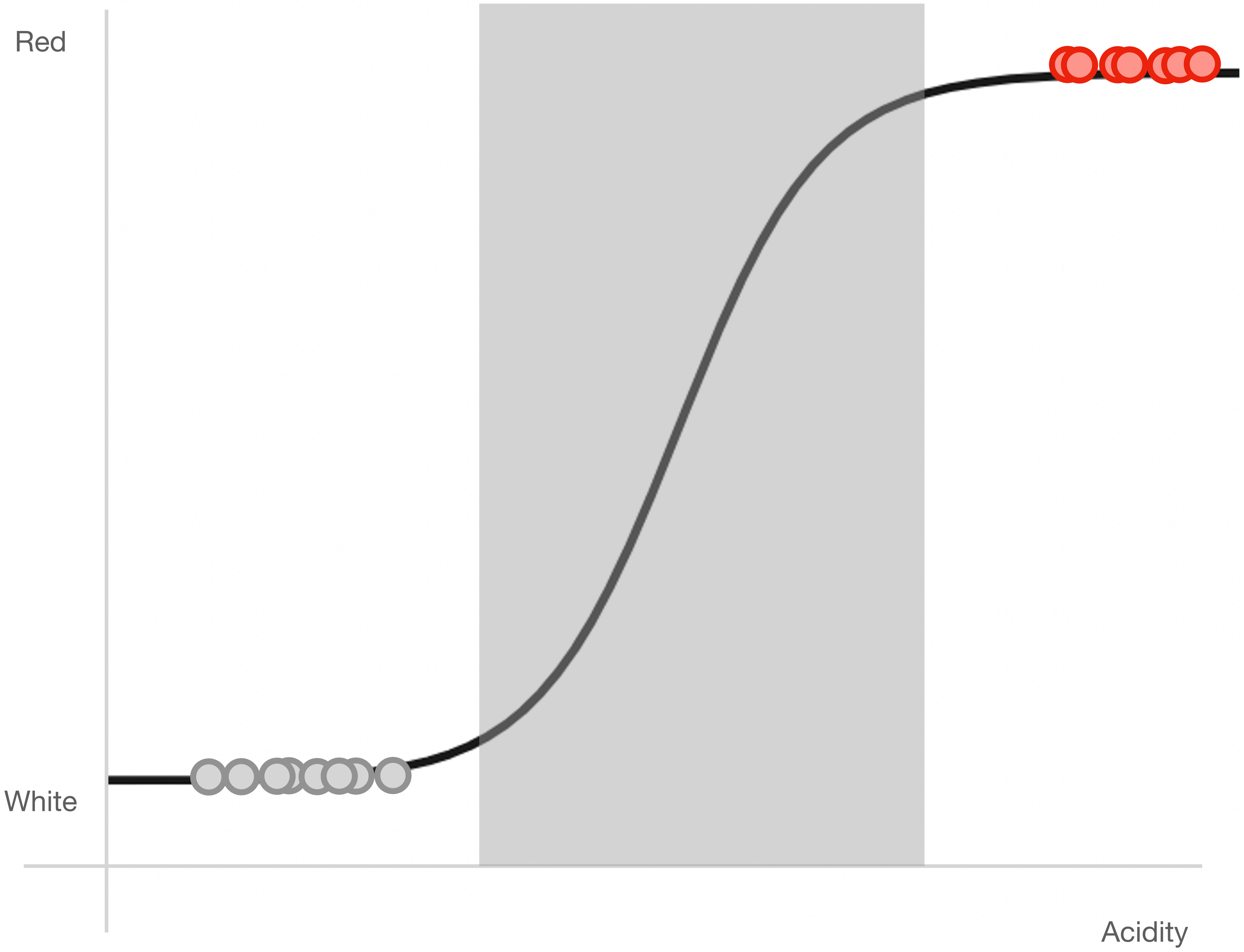
Logistic Regression



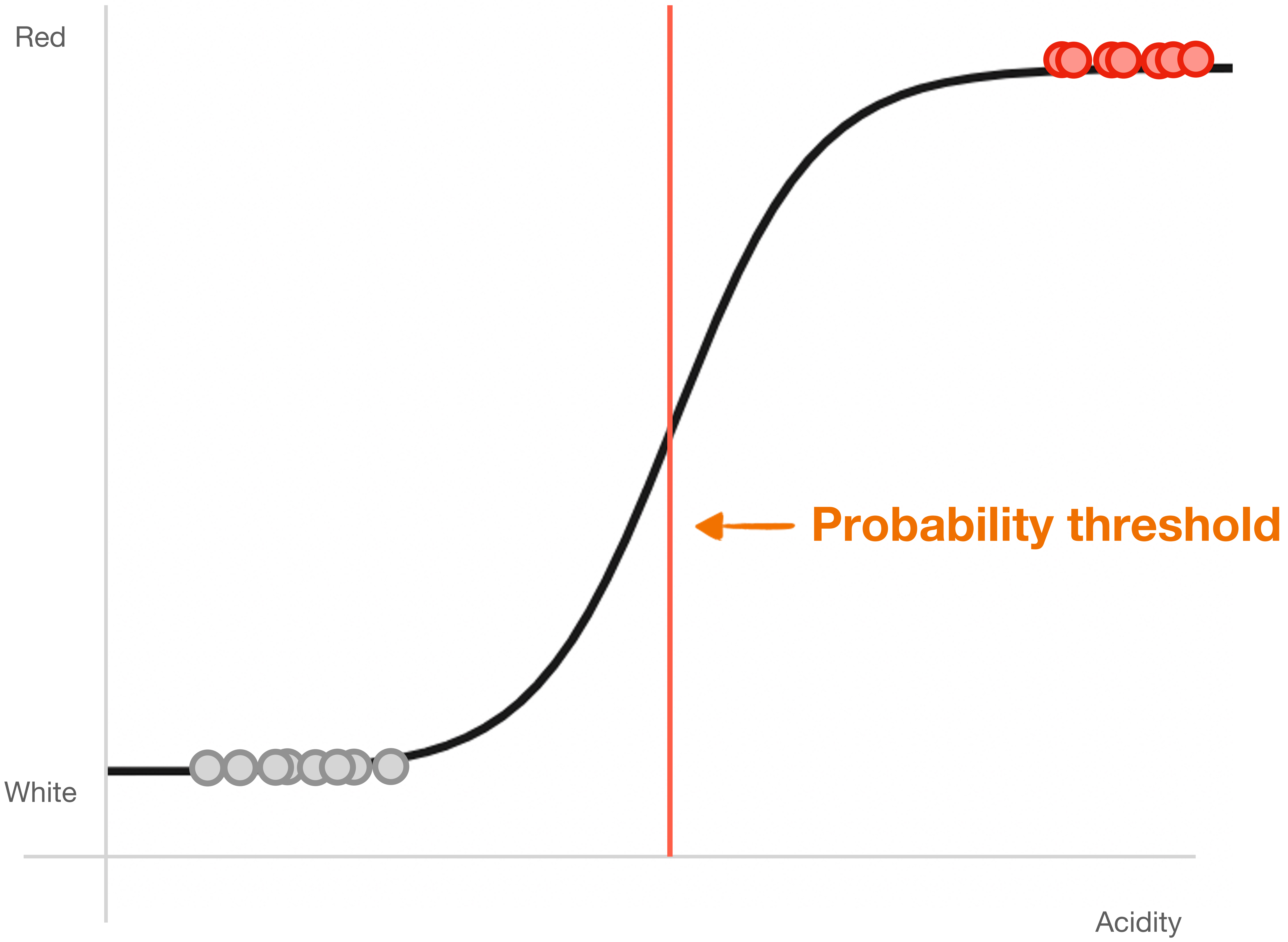
Logistic Regression



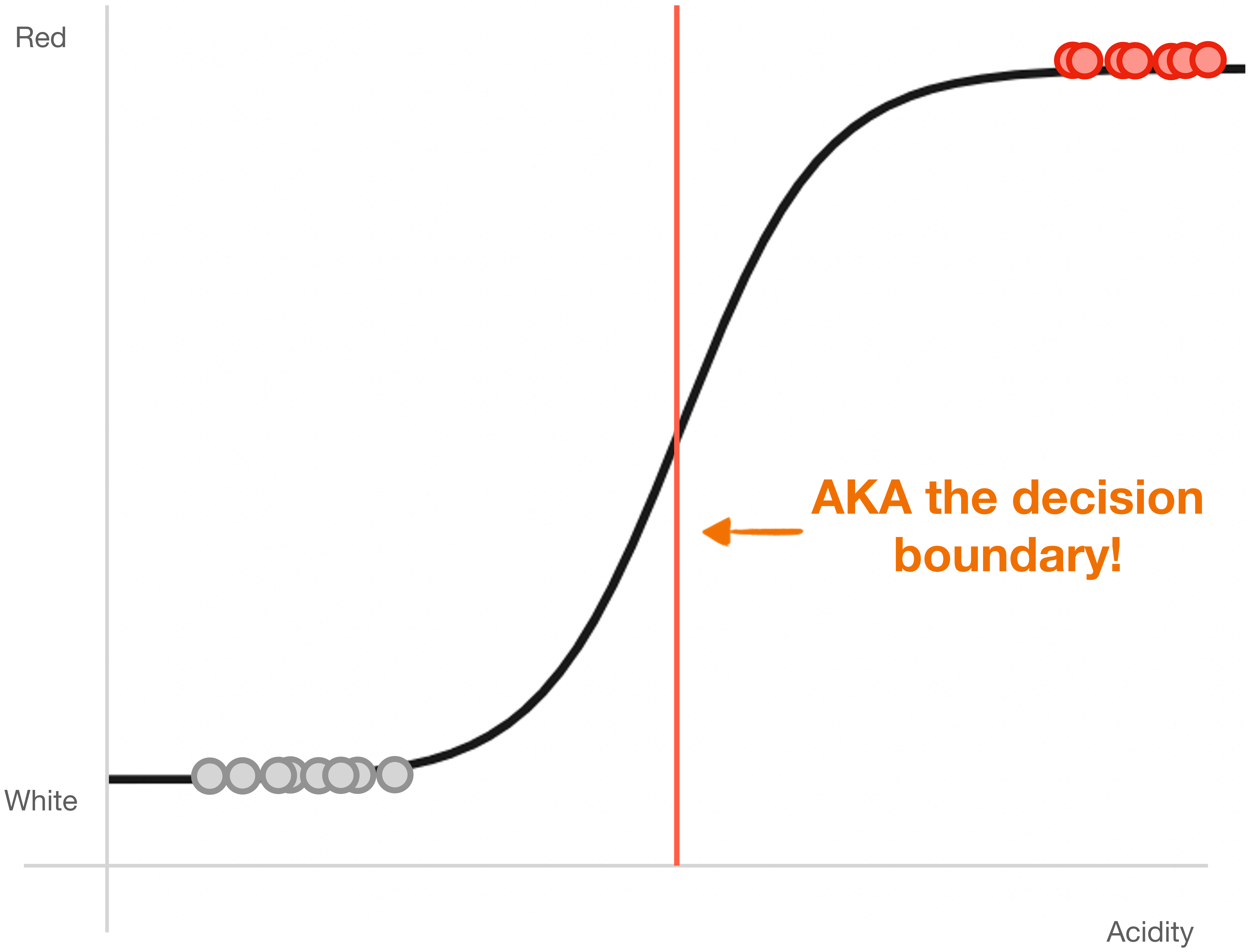
Logistic Regression



Logistic Regression



Logistic Regression



Logistic Regression

A new cost function…

$$\text{COST} = -\frac{1}{n}\sum_{i=0}^{n}\boxed{\left(y^{(i)}\log(\hat{y}^{(i)}\right)} + \left(1 - y^{(i)}\right)\log(1 - \hat{y}^{(i)})$$

Logistic Regression

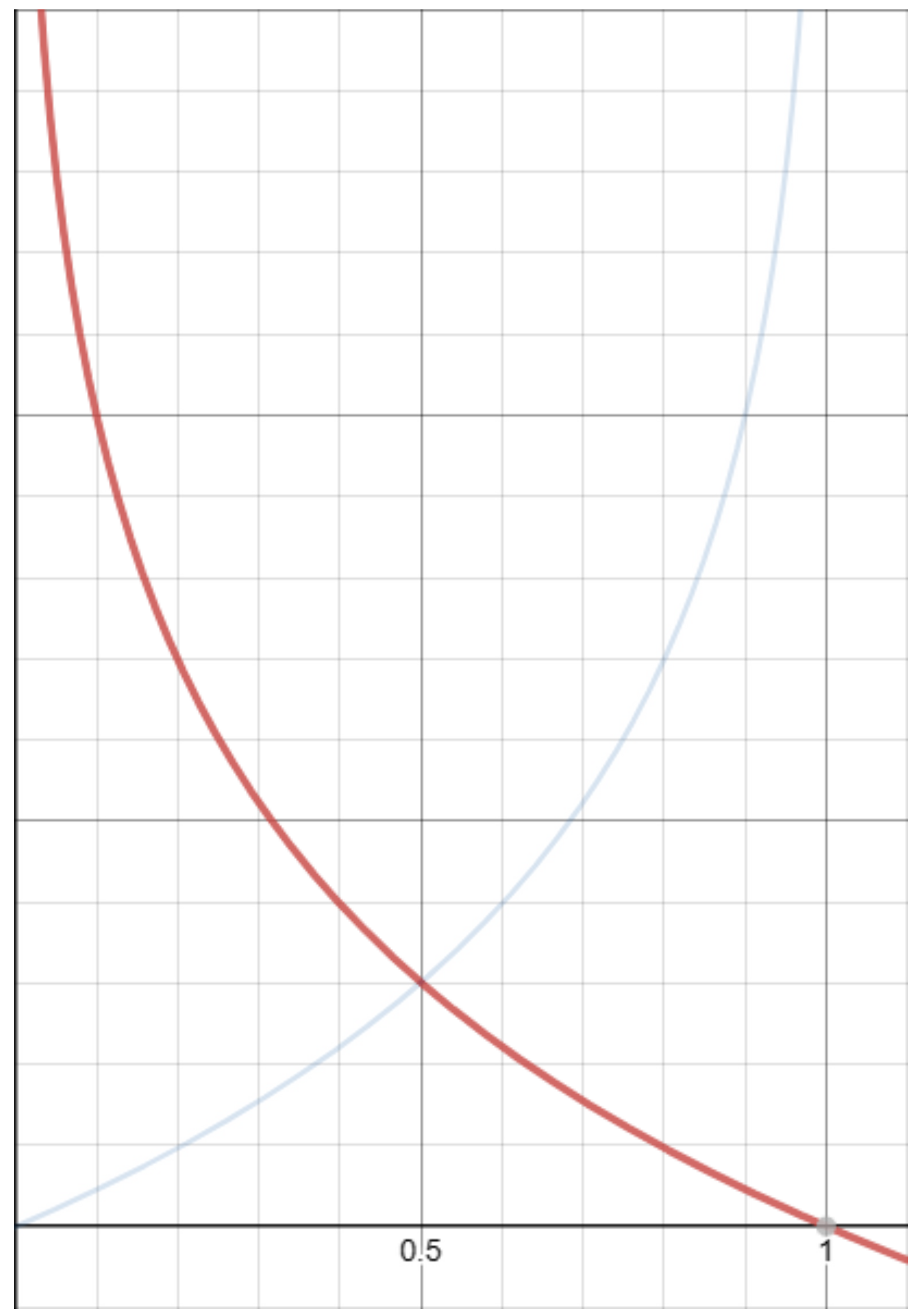
A new cost function…

$$\text{COST} = -\frac{1}{n}\sum_{i=0}^{n}(y^{(i)}\log(\hat{y}^{(i)}) + \boxed{\left(1-y^{(i)}\right)\log(1-\hat{y}^{(i)})}$$

Logistic Regression

A new cost function…

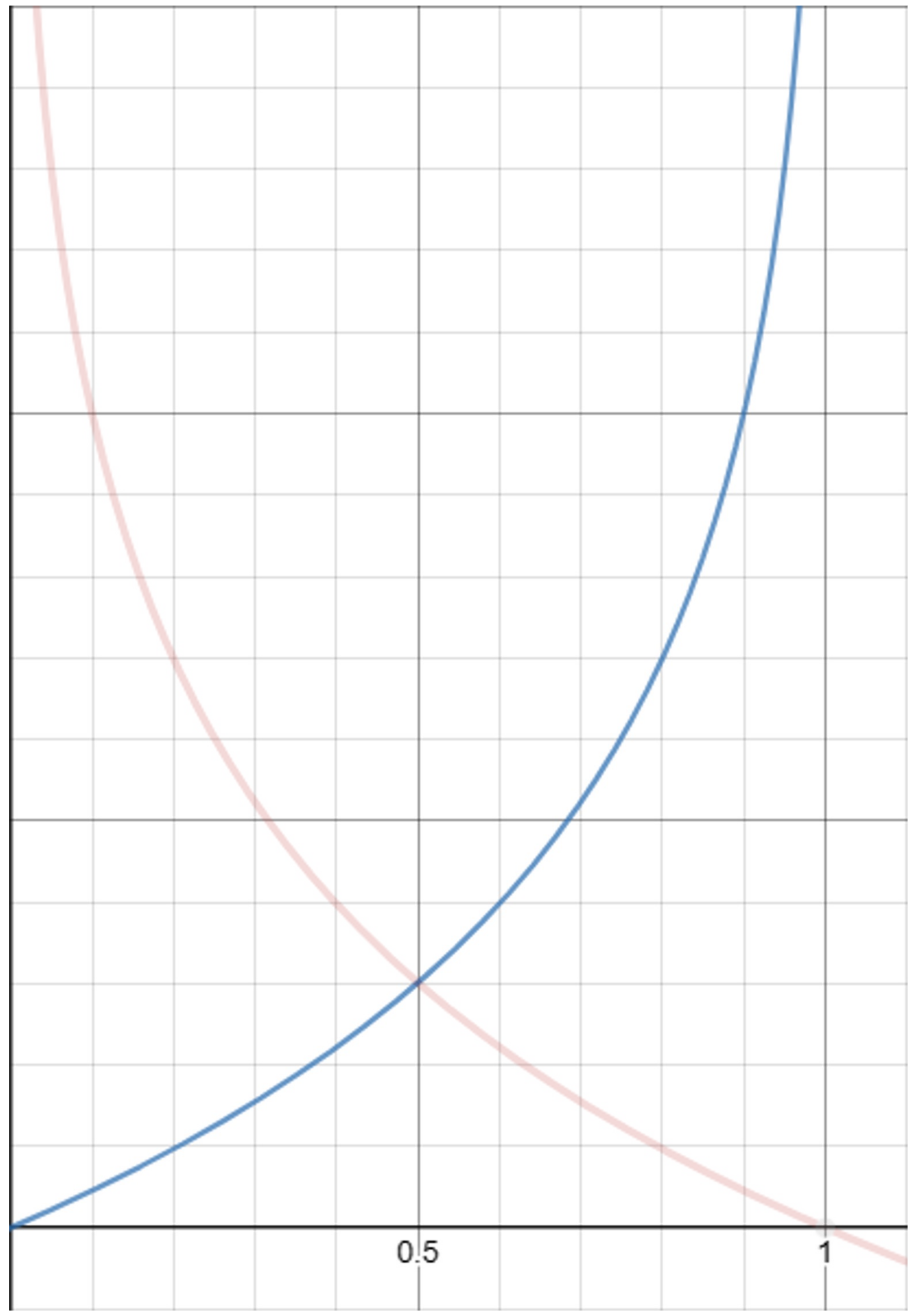
$$\text{COST} = -\frac{1}{n}\sum_{i=0}^{n}\left(y^{(i)}\log(\hat{y}^{(i)}) + \left(1 - y^{(i)}\right)\log(1 - \hat{y}^{(i)})\right)$$



Logistic Regression

A new cost function…

$$\text{COST} = -\frac{1}{n}\sum_{i=0}^{n}\left(y^{(i)}\log(\hat{y}^{(i)}) + \left(1 - y^{(i)}\right)\log(1 - \hat{y}^{(i)})\right)$$



But what if we have
more than two classes
for output?

Multi-class classification

White **Red** **Champagne**

One vs. One and One vs. All Classification

One vs. One Classification

One vs. One Classification

White

Red

One vs. One Classification

White	White
Red	Champagne

One vs. One Classification

White	White	Red
Red	Champagne	Champagne

One vs. One Classification

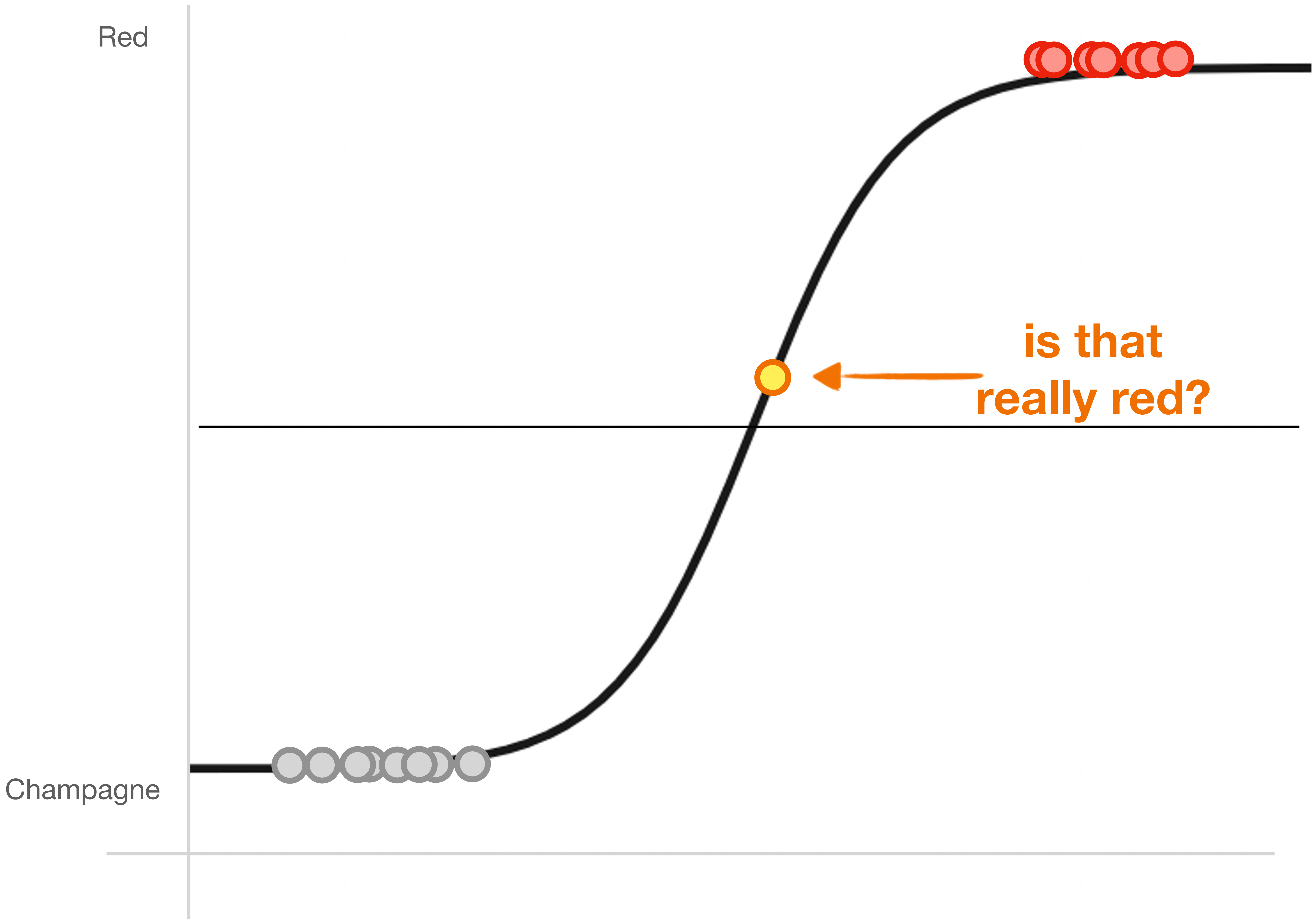
more pairs
voted for white!

White	White	Red
Red	Champagne	Champagne

It's white!

- **One-vs-one** multiclass classification uses the most "voted for" class among paired models

Problems with One vs. One Classification



One vs. All Classification

White	**Red**	**Champagne**
Not white	Not red	Not champagne

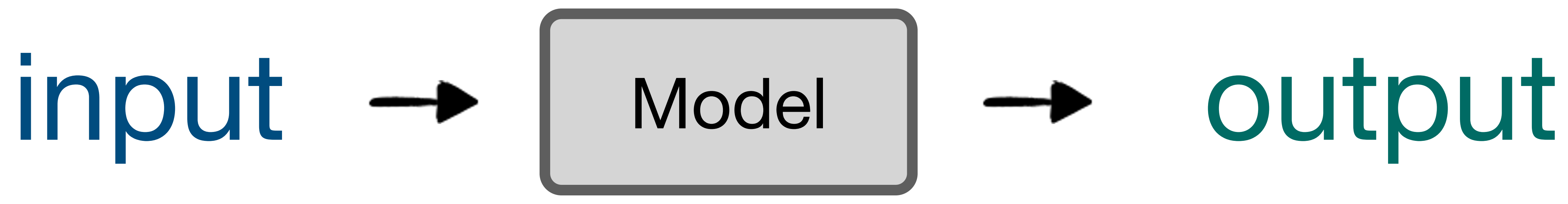
One vs. All Classification

White	**Red**	**Champagne**
Not white	Not red	Not champagne

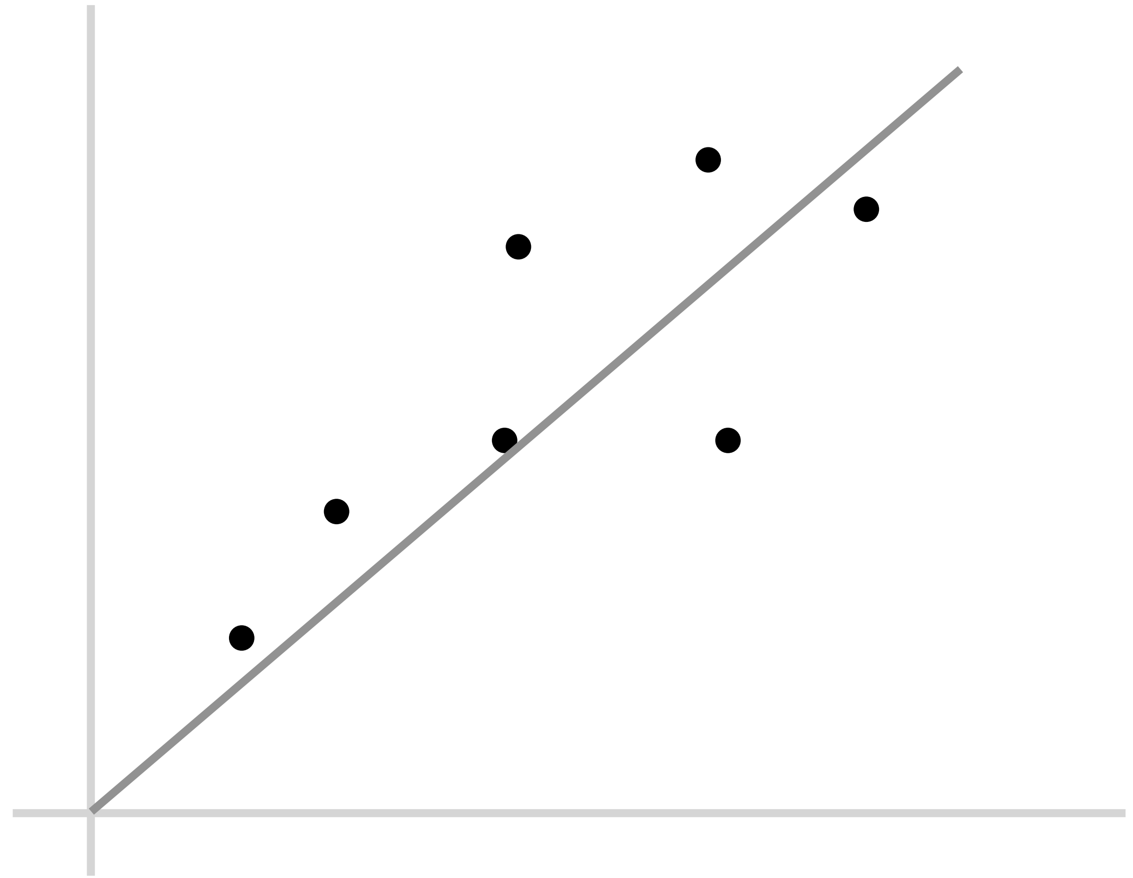
pick the answer with highest probability

A quick summary…

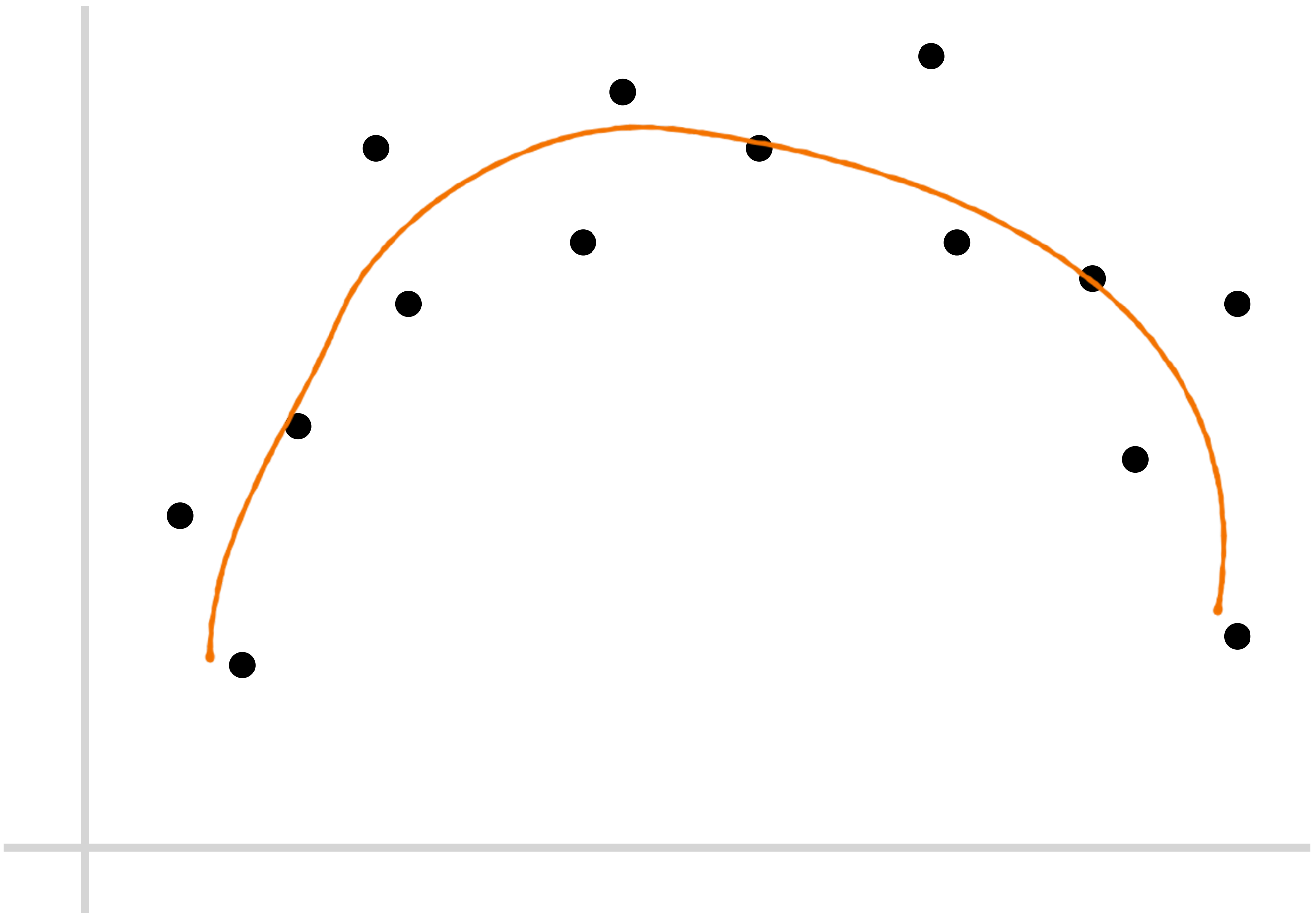
what *is* Artificial Intelligence?



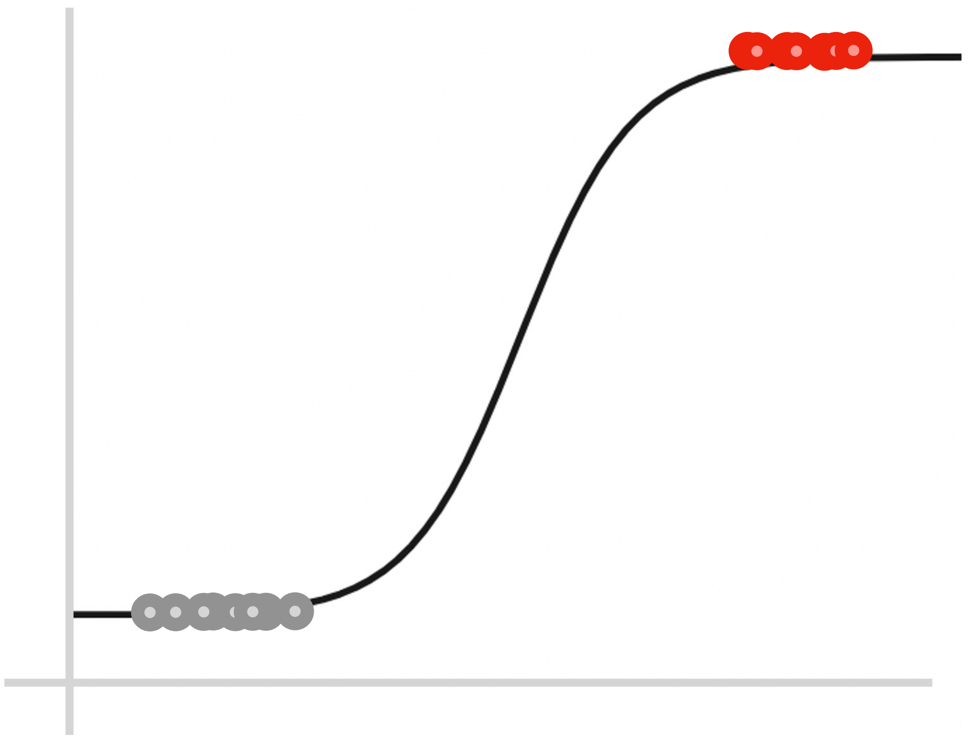
Supervised Learning



Linear Regression



Polynomial Regression



Logistic Regression

Lab time!