AIBridge

Lecture 4

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

- Most widely used ML techniques in real world applications.

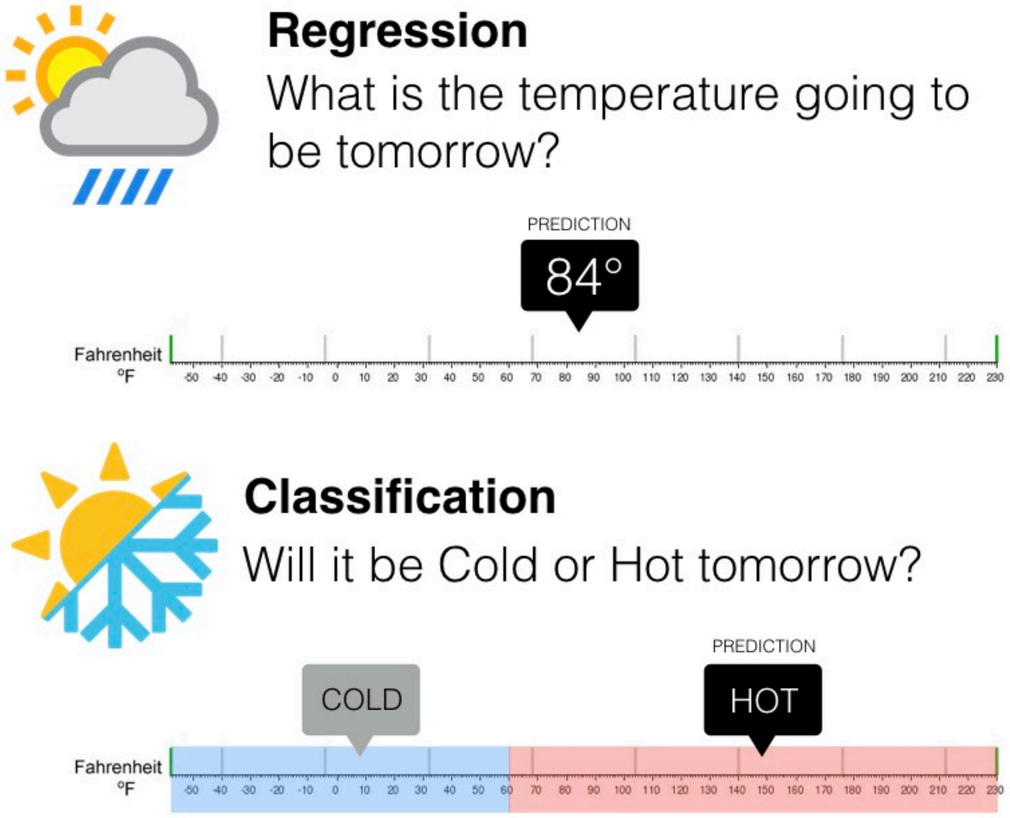
Supervised Learning

- Classification:
  - Predicting a label/class/category
  - Ex: spam or not, cancer or not, cat or dog, red wine vs. white wine
- Regression:
  - Predicting a (continuous) quantity
  - Ex: Survival rate, wine quality, yield prediction

Example

- 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?

Conversion



Credit: https://www.springboard.com/blog/data-science/regression-vs-classification/

Example

- 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?

Ok, so, like…

what *is* Artificial Intelligence?

what *is* Artificial Intelligence

~~what *is* Artificial Intelligence~~

what can Artificial Intelligence do

…not everything

what can Artificial Intelligence do

“Given a thing, tell you a thing."

what can Artificial Intelligence do

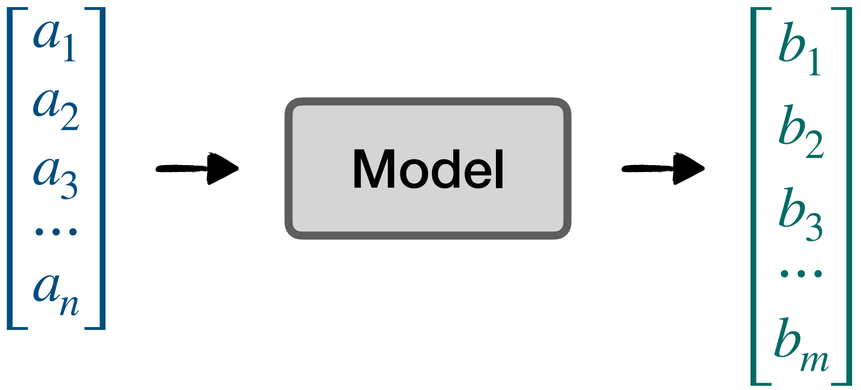
what can Artificial Intelligence do

- **models** map inputs to outputs

what can Artificial Intelligence do

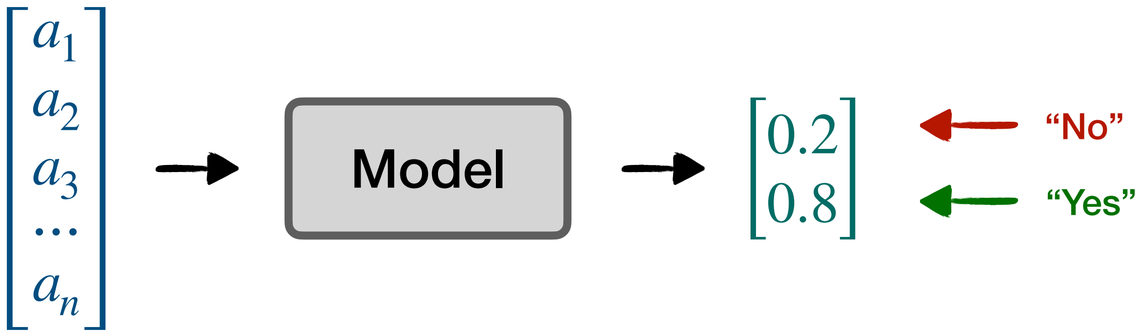
functions
~~models~~ map inputs to outputs

what can Artificial Intelligence do



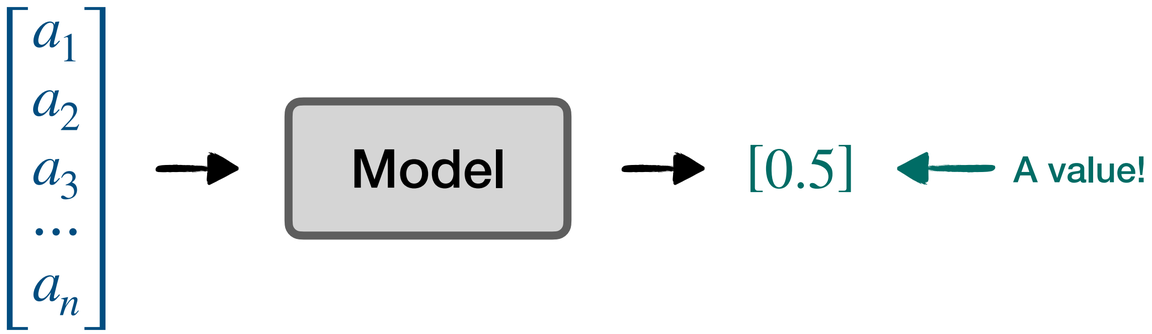
models are $\mathbb{R}^n \rightarrow \mathbb{R}^m$ functions

what can Artificial Intelligence do



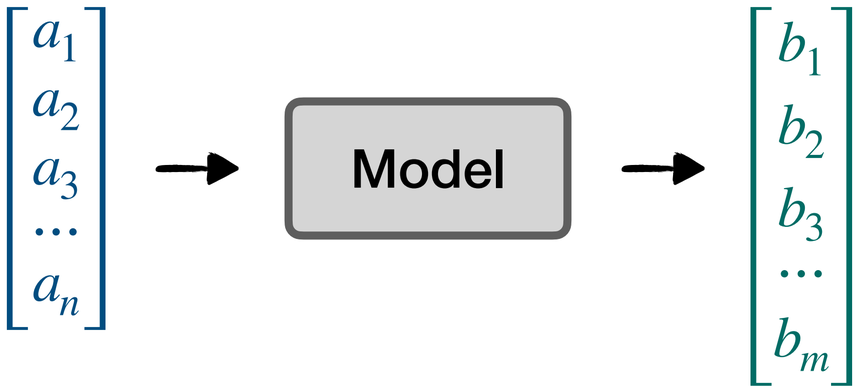
models are $\mathbb{R}^n \rightarrow \mathbb{R}^2$ functions

what can Artificial Intelligence do



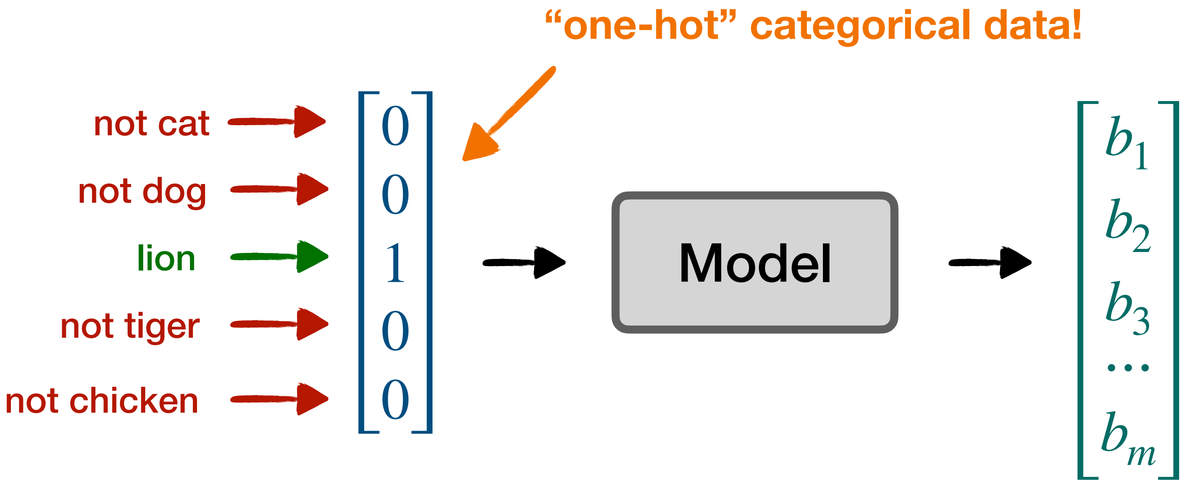
models are $\mathbb{R}^n \rightarrow \mathbb{R}^1$ functions

what can Artificial Intelligence do



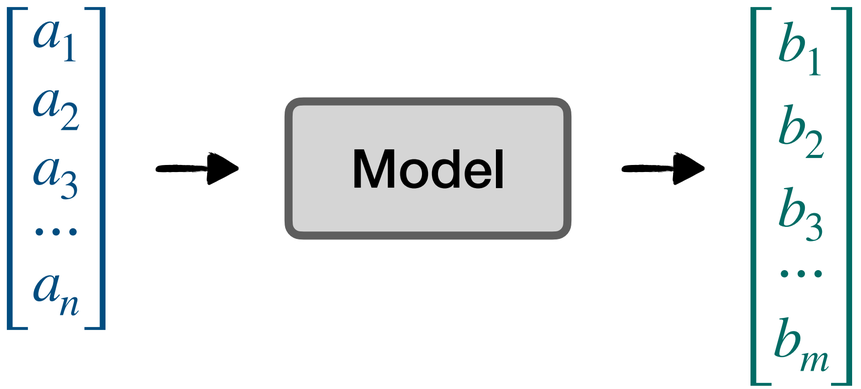
models are $\mathbb{R}^n \rightarrow \mathbb{R}^m$ functions

what can Artificial Intelligence do



models are $\mathbb{R}^n \to \mathbb{R}^m$ functions

what can Artificial Intelligence do



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

we will get to those

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

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact

Search

Repository Web

Google™

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]).

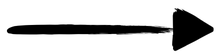
Data Set Characteristics:	Multivariate	**Number of Instances:**	4898	**Area:**	Business
Attribute Characteristics:	Real	**Number of Attributes:**	12	**Date Donated**	2009-10-07
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	1891084

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

1. Fixed acidity
2. Volatile acidity
3. Citric acid
4. Residual sugar
5. Chlorides
6. Free sulfur dioxide
7. Total sulfur dioxide
8. Density
9. pH
10. Sulphates
11. Alcohol
12. White/Red
13. **Quality**

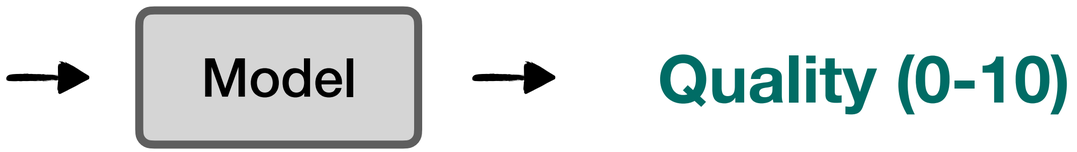
- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red



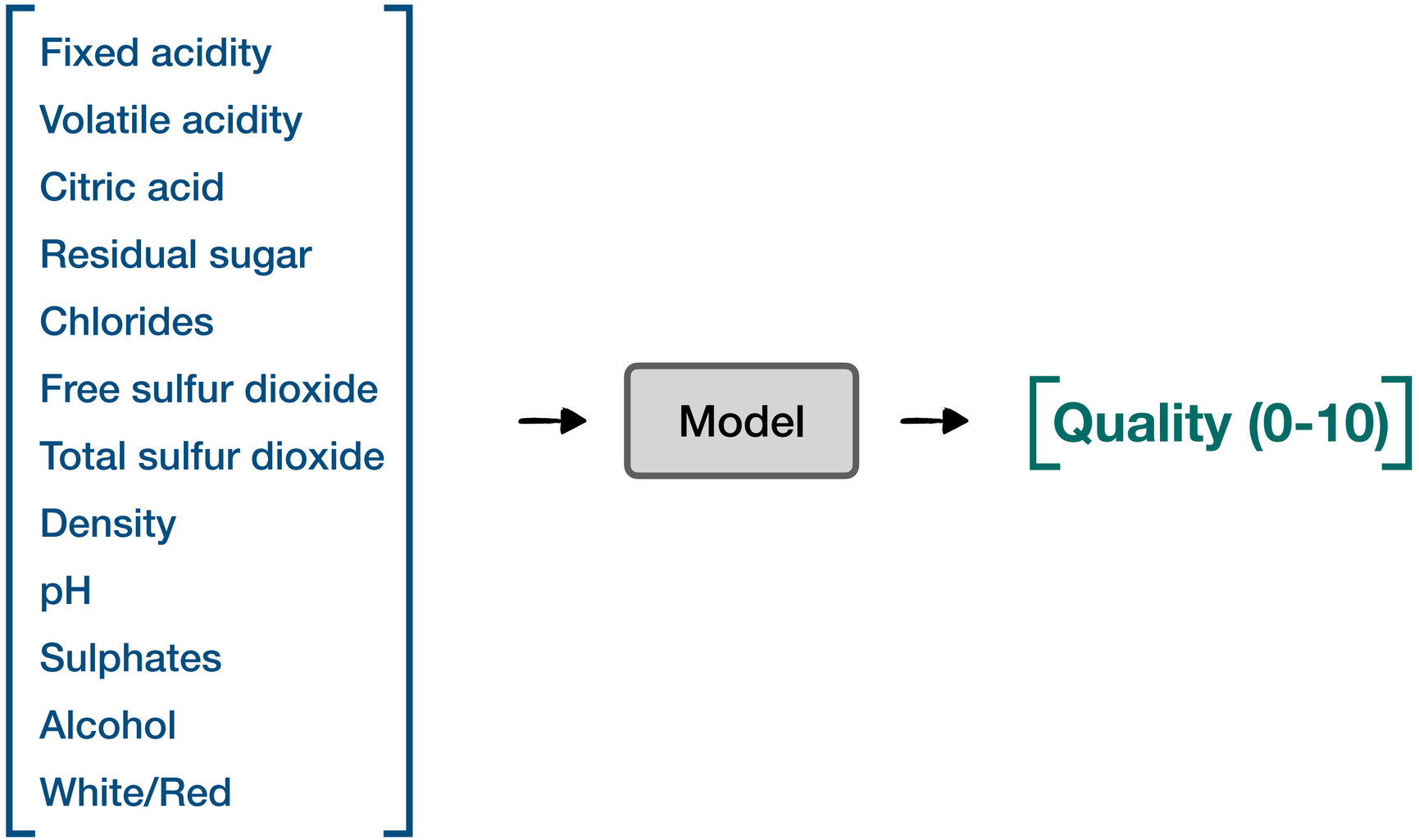
Quality (0-10)

Linear Regression

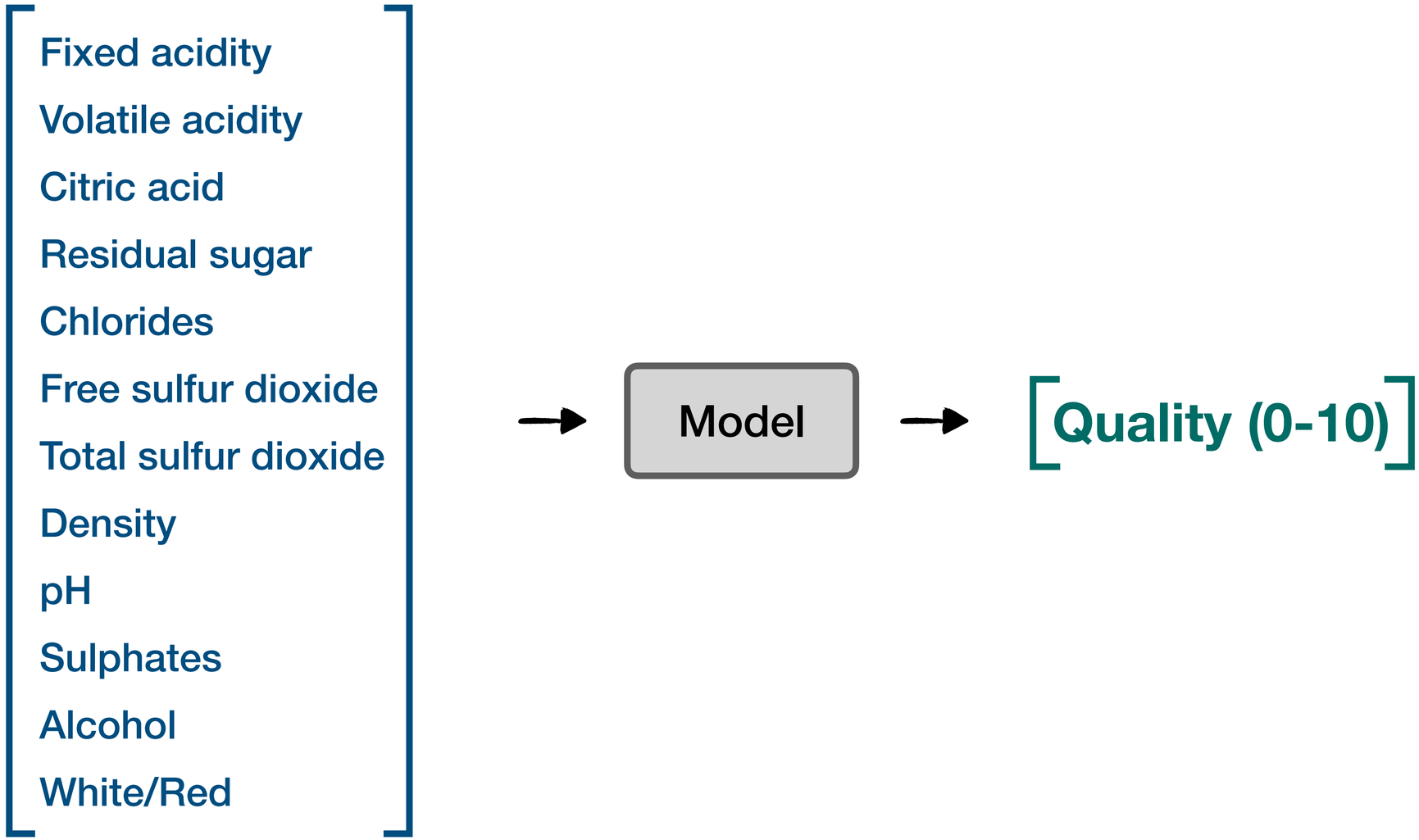
- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red



Linear Regression

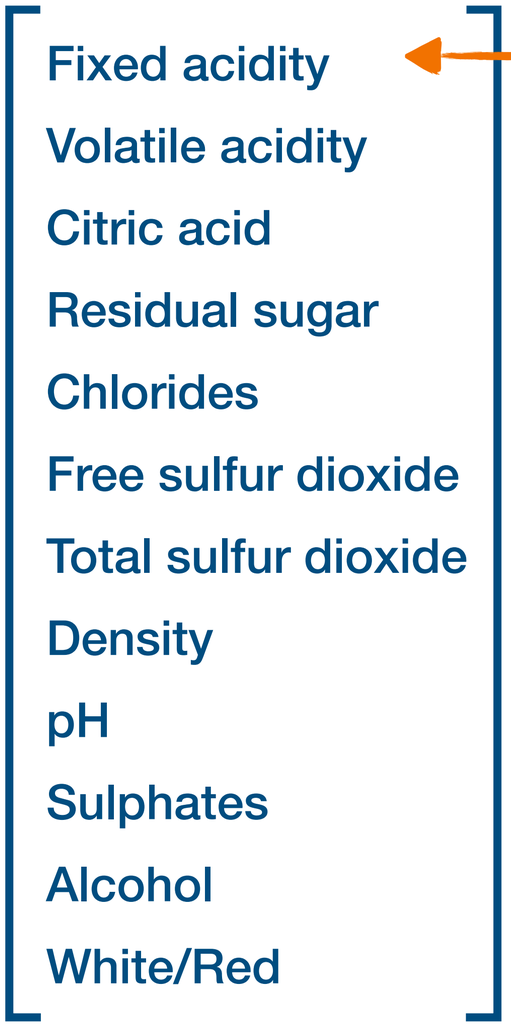


Linear Regression



this model could be a $\mathbb{R}^{12} \rightarrow \mathbb{R}^{1}$ function

Linear Regression

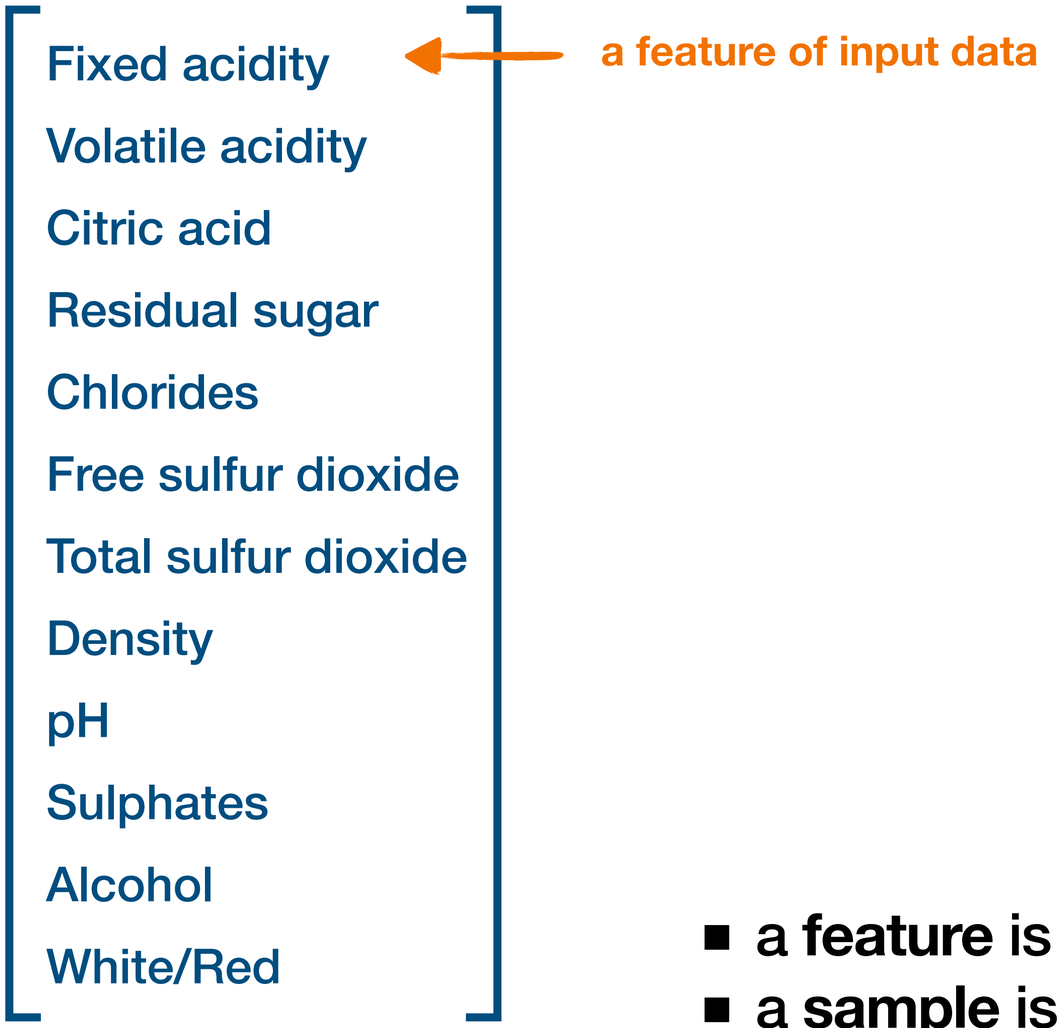


a feature of input data

- a **feature** is a facet of input data

hint: “this model could be a $\mathbb{R}^{12} \rightarrow \mathbb{R}^{1}$ function”

Linear Regression

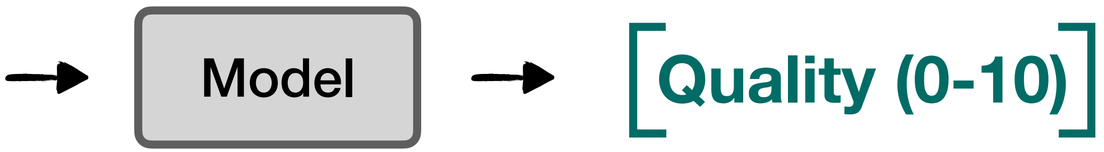


- a **feature** is a facet of input data
- a **sample** is a collection of **features**

Linear Regression

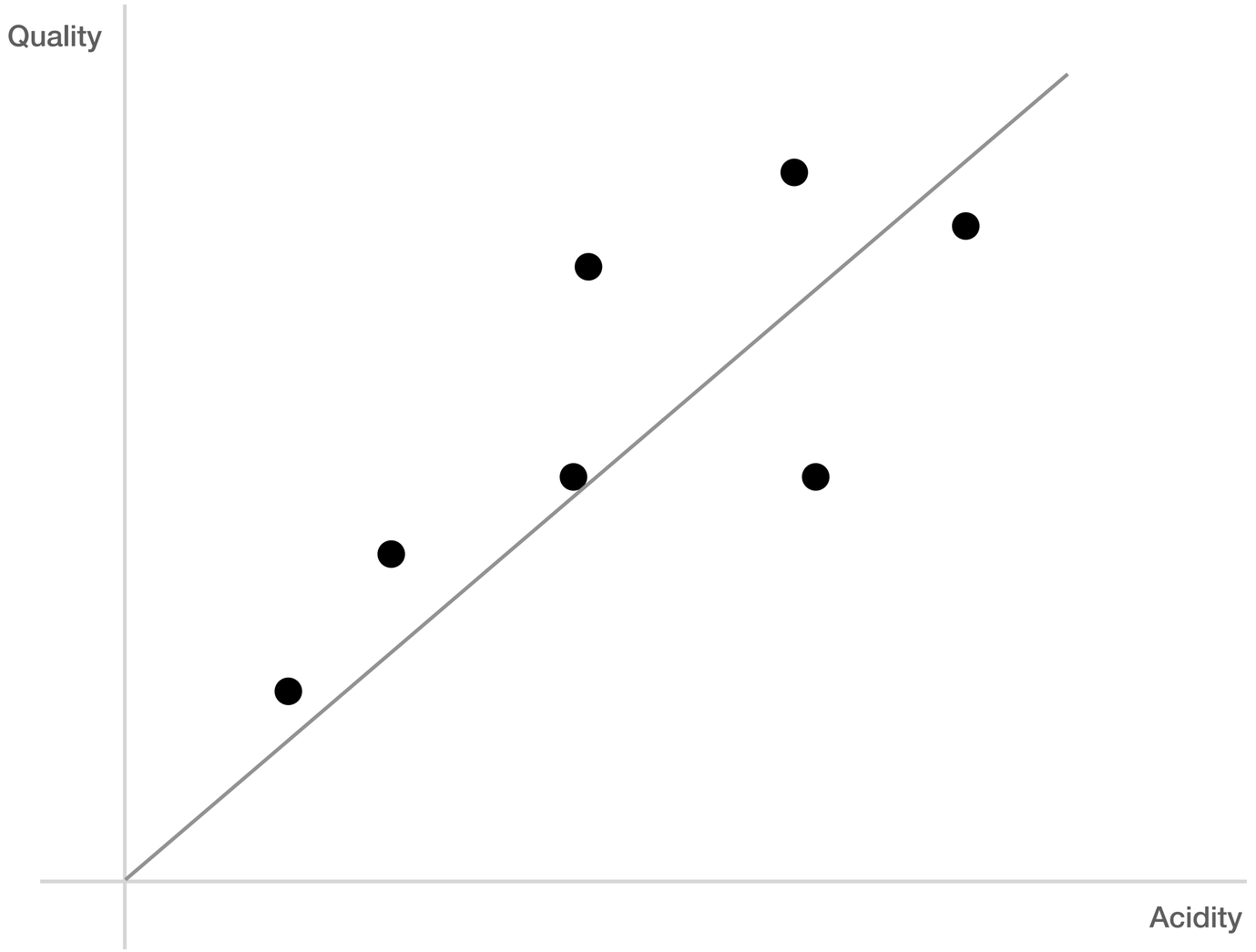
- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- White/Red

ok, cool, but what's here



Linear Regression

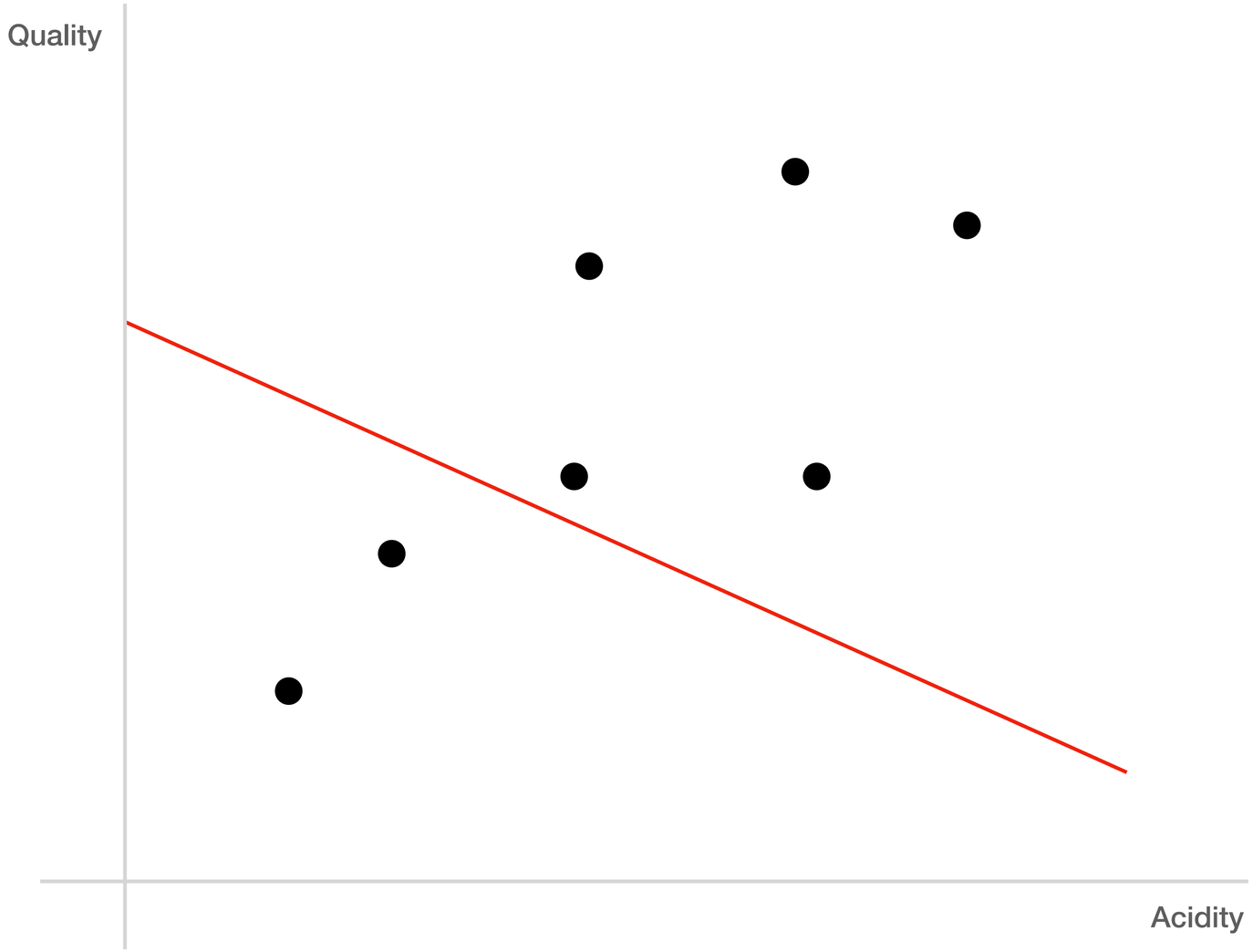
STATISTICS!



this model could be a $\mathbb{R}^1 \to \mathbb{R}^1$ function

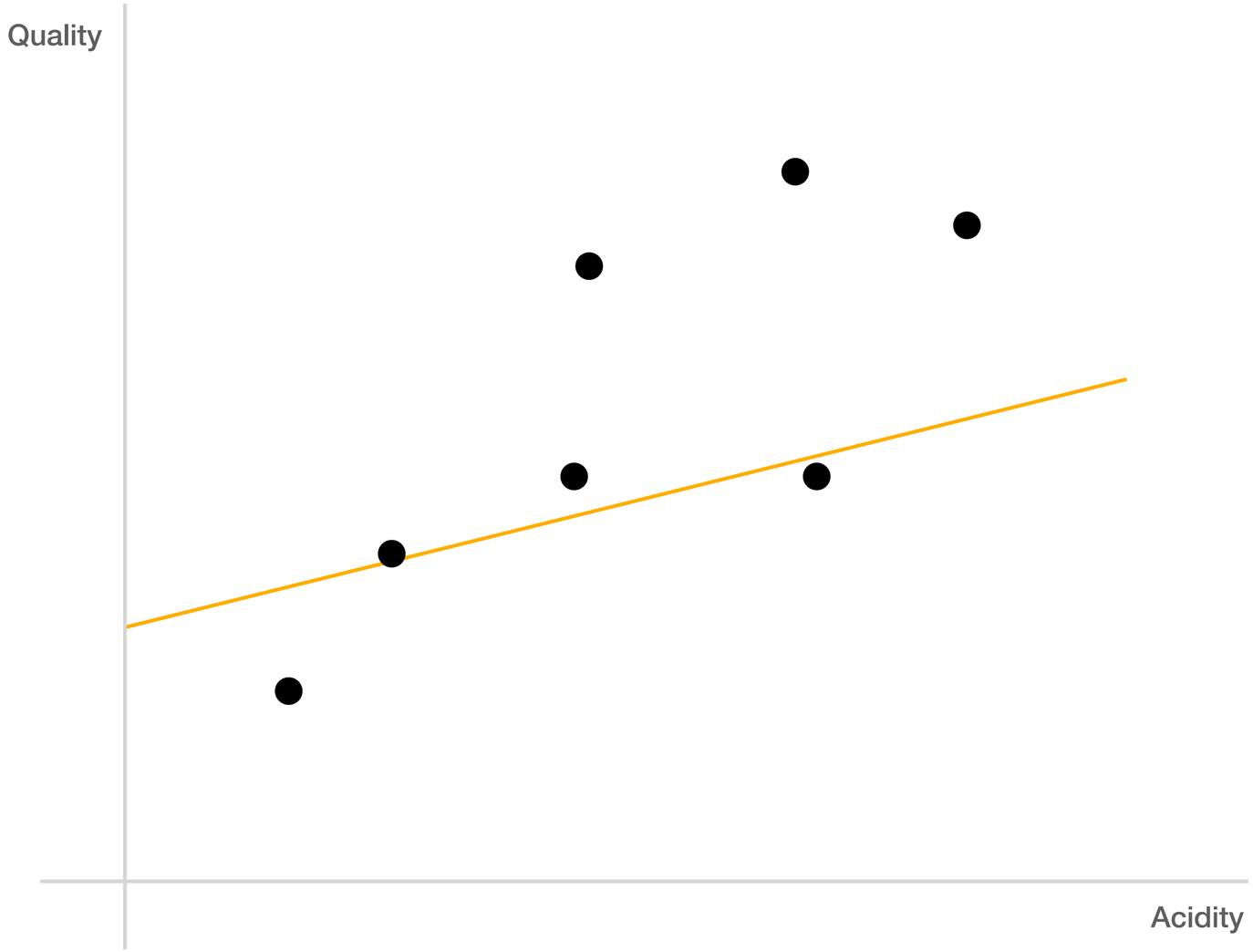
Linear Regression

STATISTICS!



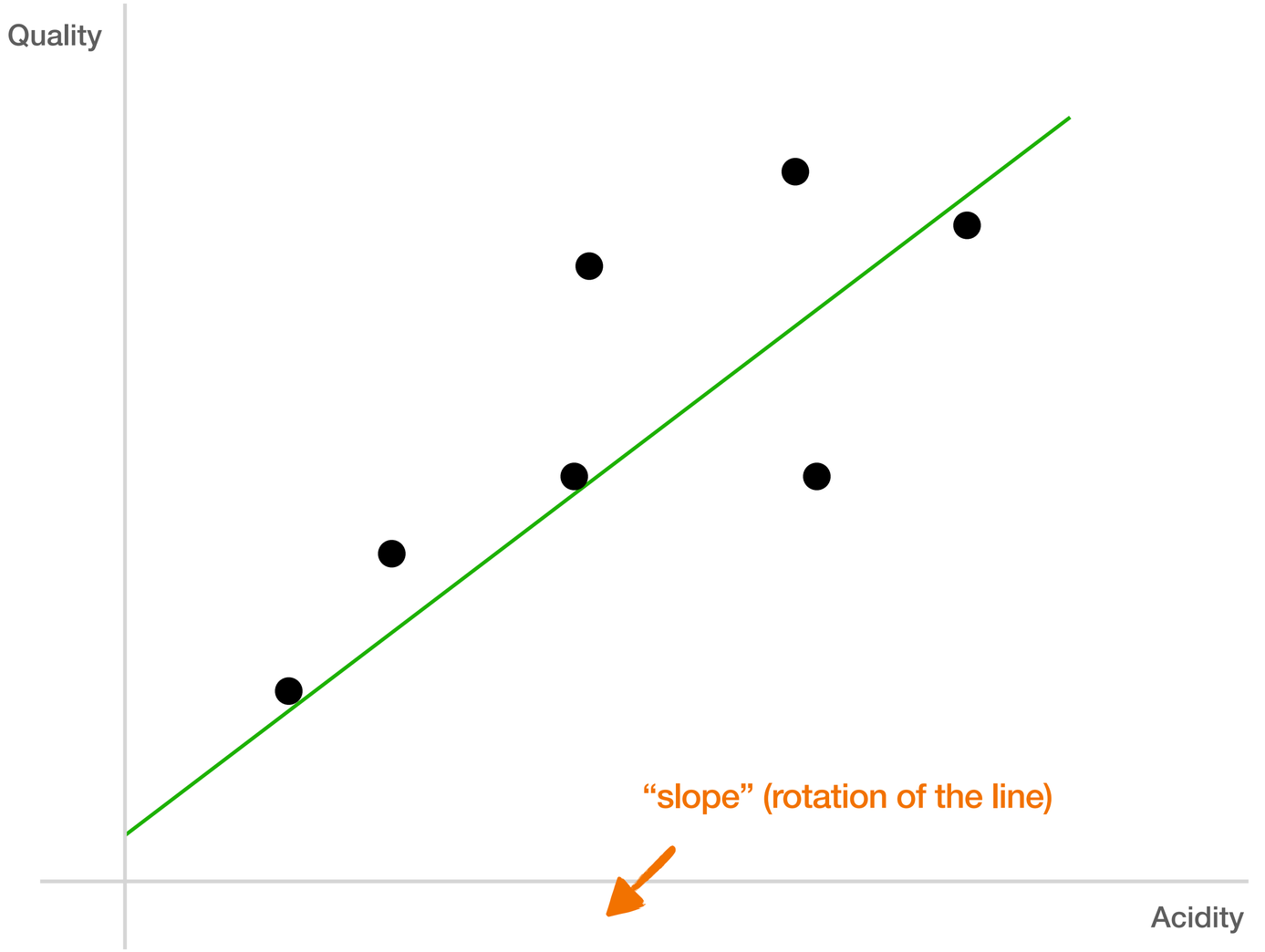
Linear Regression

STATISTICS!



Linear Regression

STATISTICS!

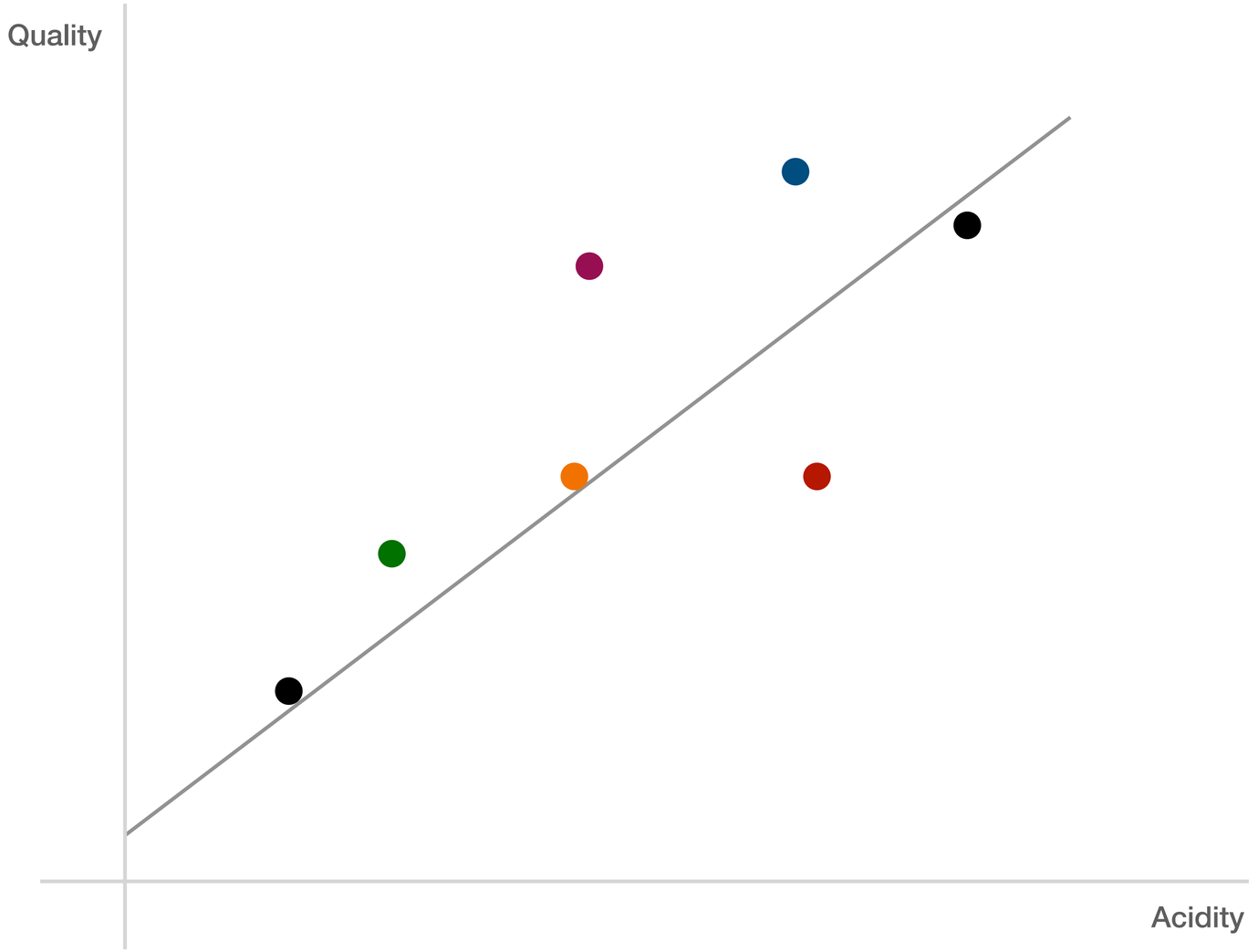


$$y_i = mx_i + b$$

“intercept” (shift of the line)

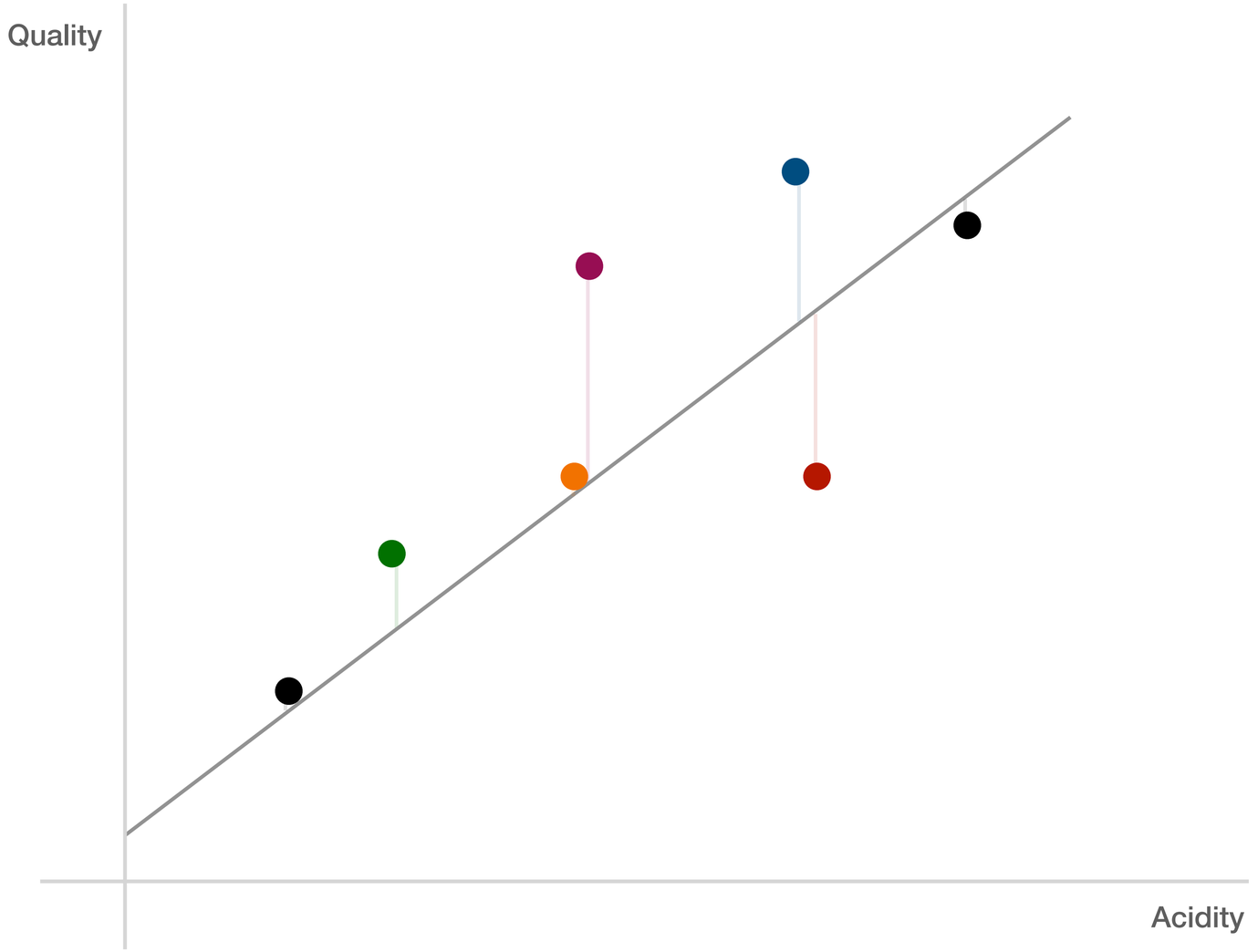
Linear Regression

STATISTICS!



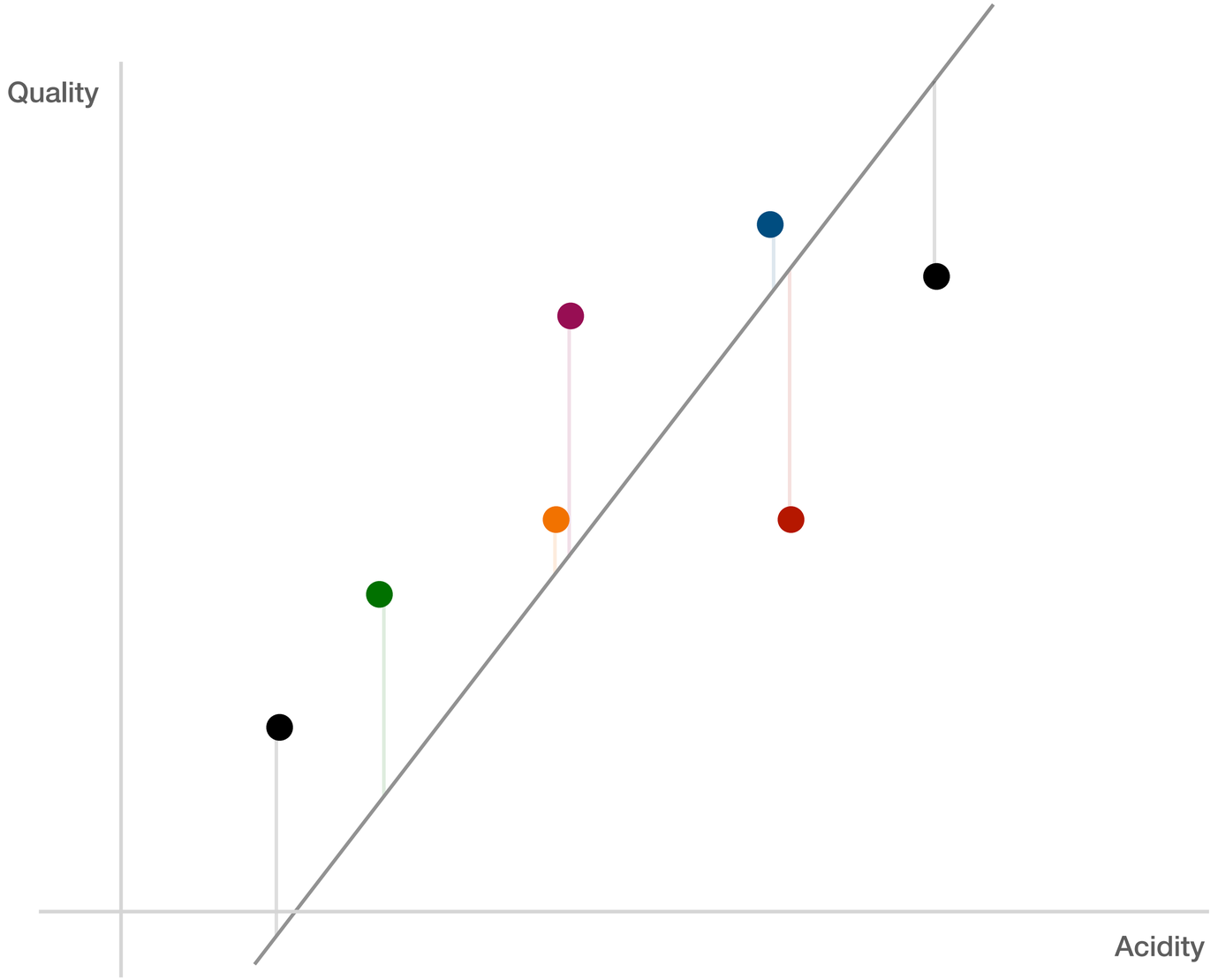
Linear Regression

STATISTICS!



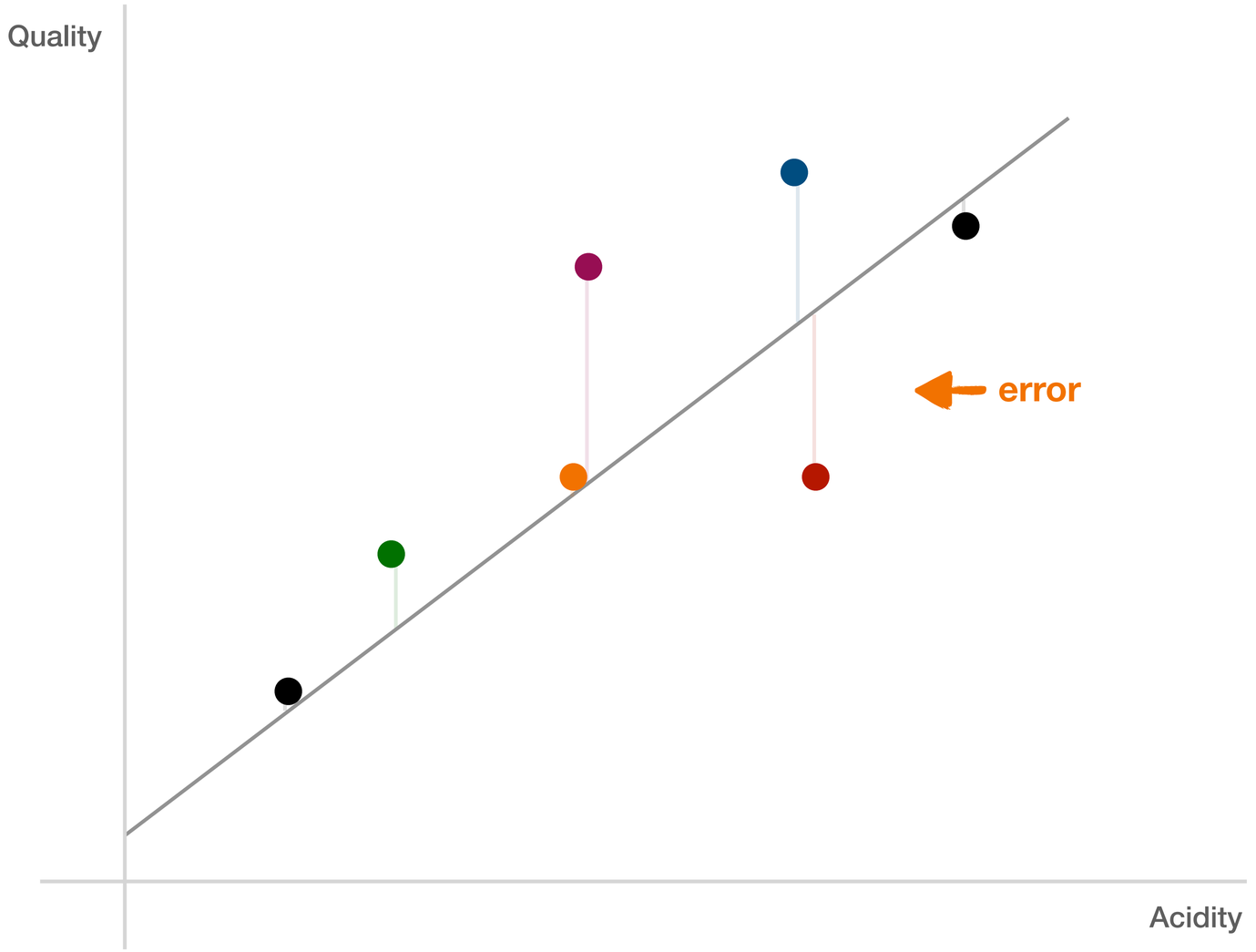
Linear Regression

STATISTICS!



Linear Regression

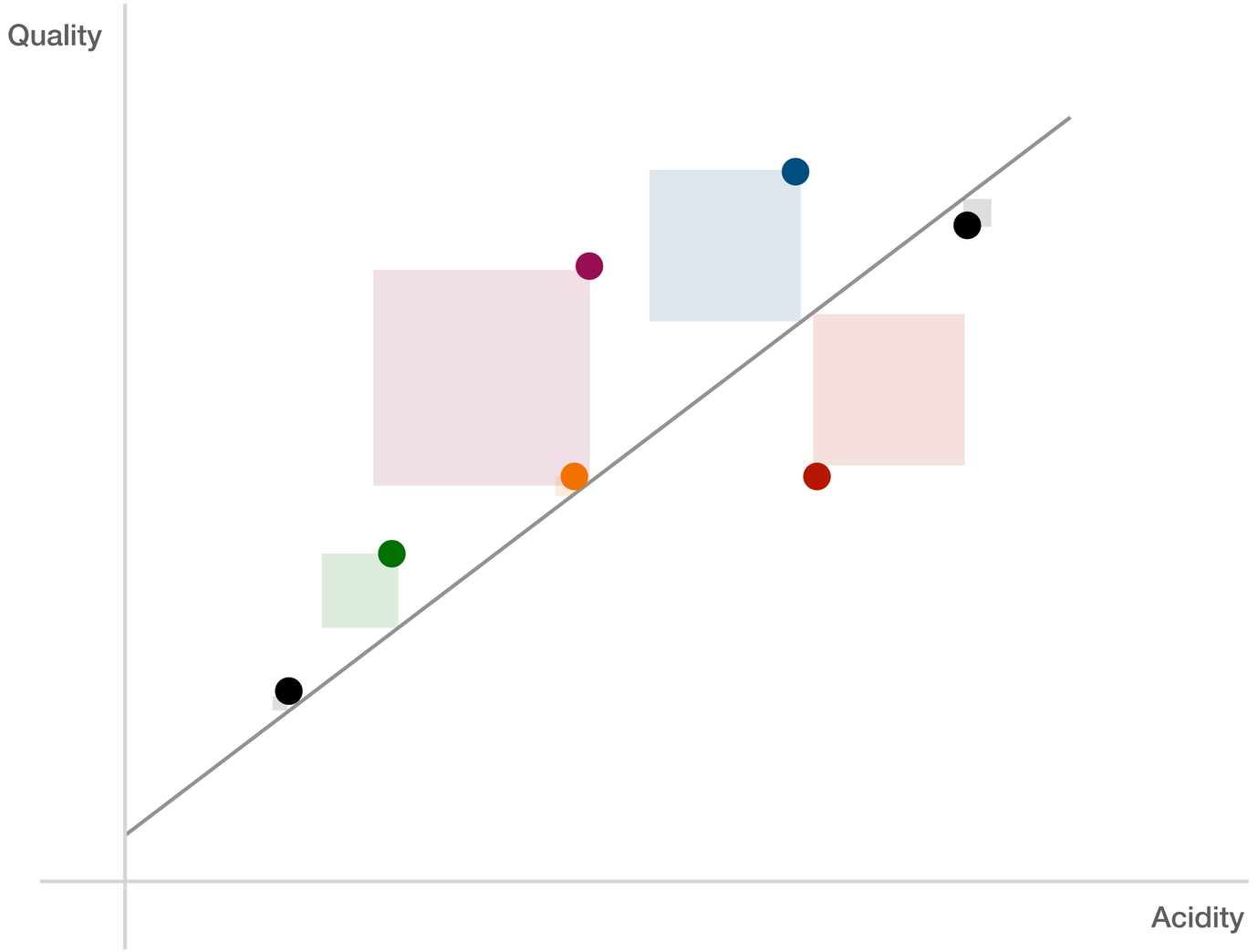
STATISTICS!



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

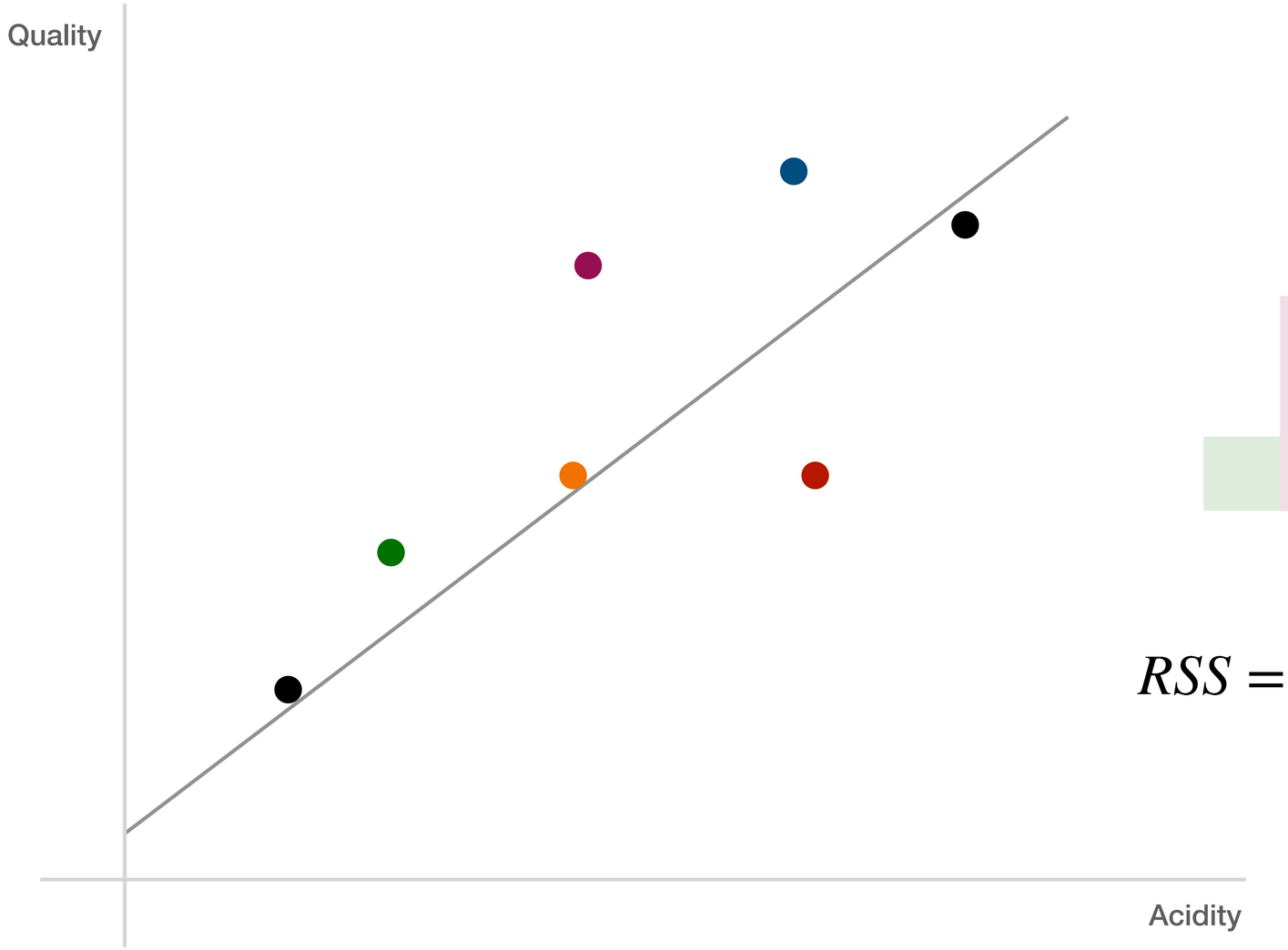
Linear Regression

STATISTICS!



Linear Regression

STATISTICS!



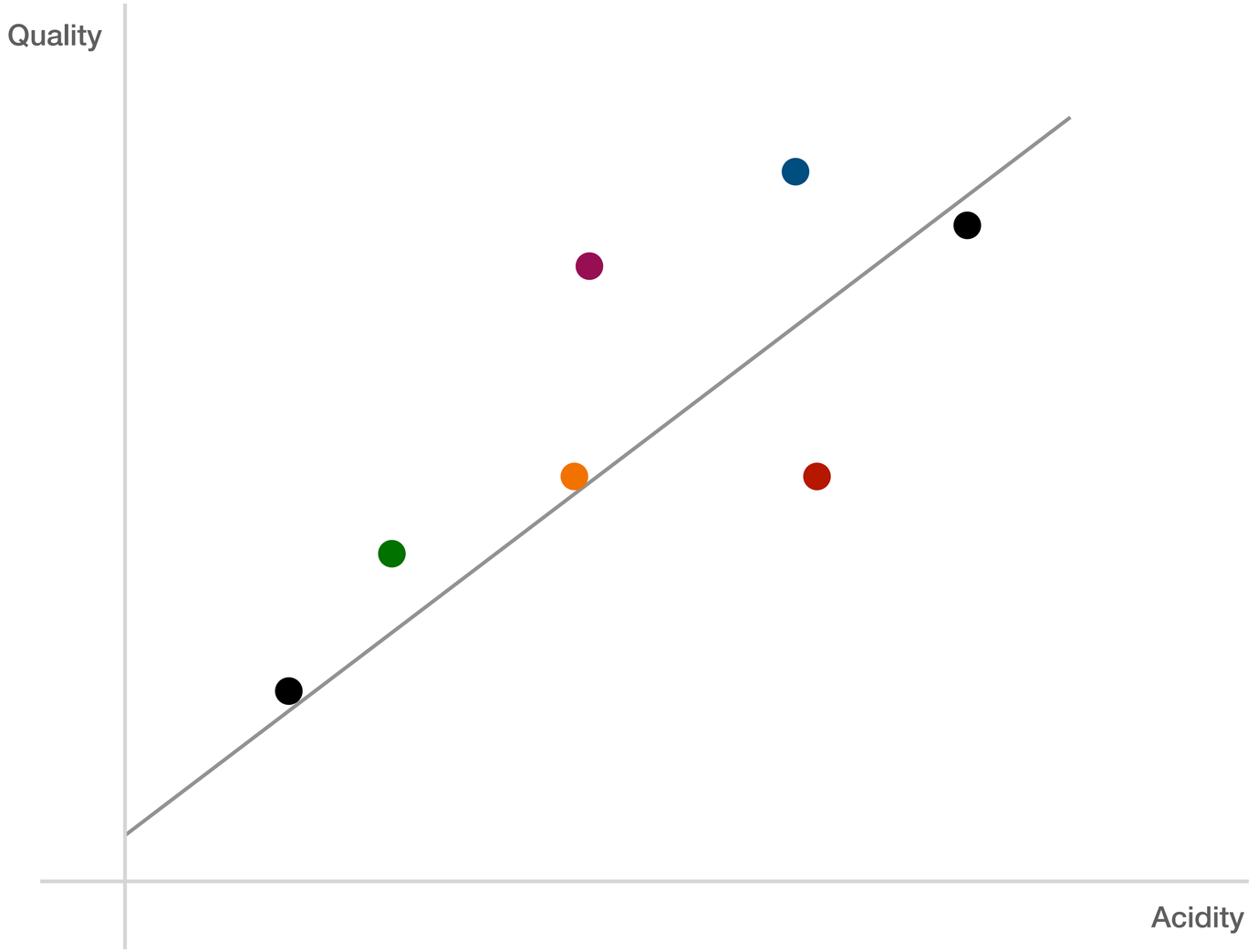
squared error

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

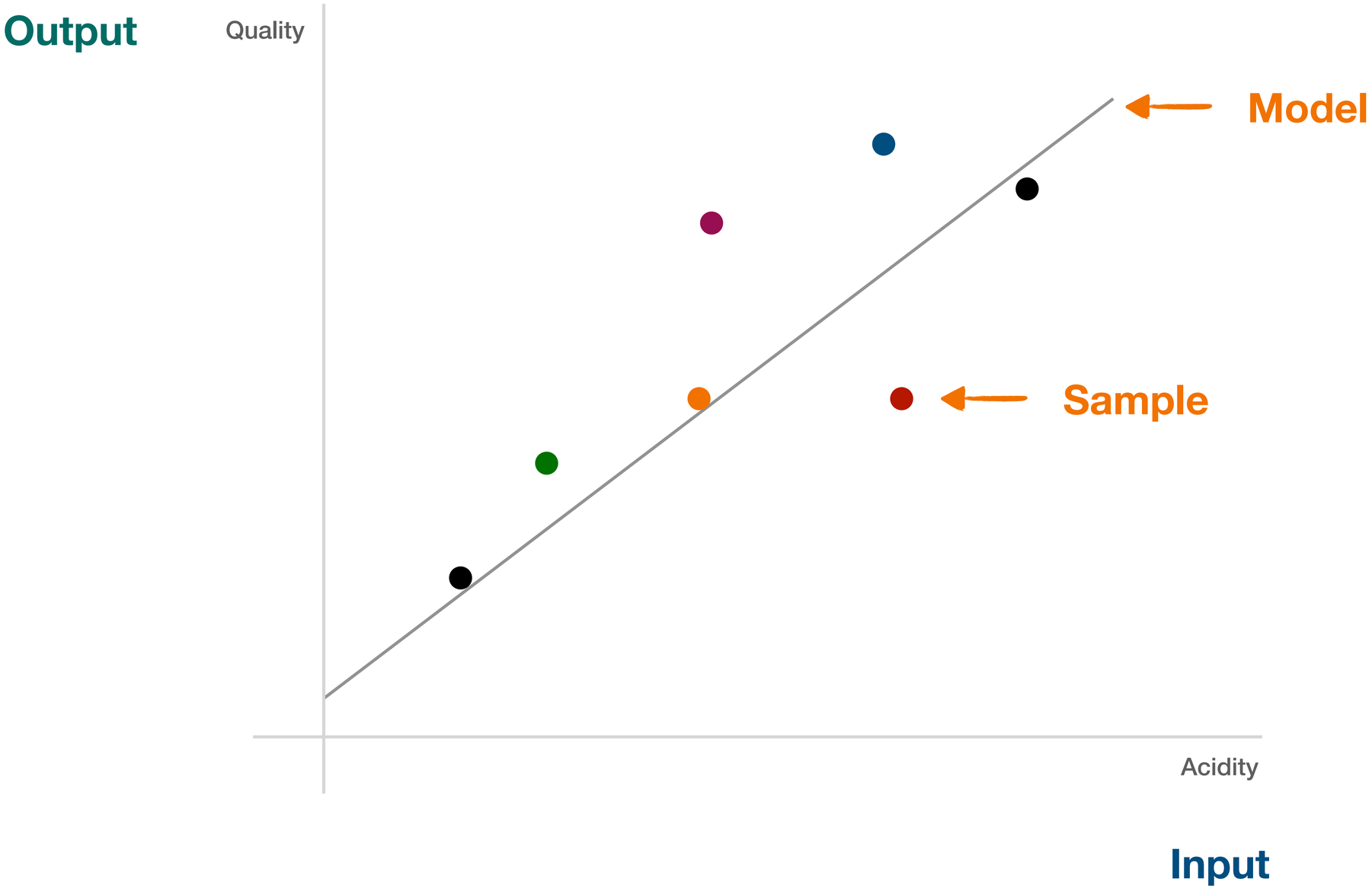
- **sum-of-squared** error is a common error metric for linear regression
- **sum-of-squared** error is also known as “L2 Penalty”

Linear Regression

STATISTICS!



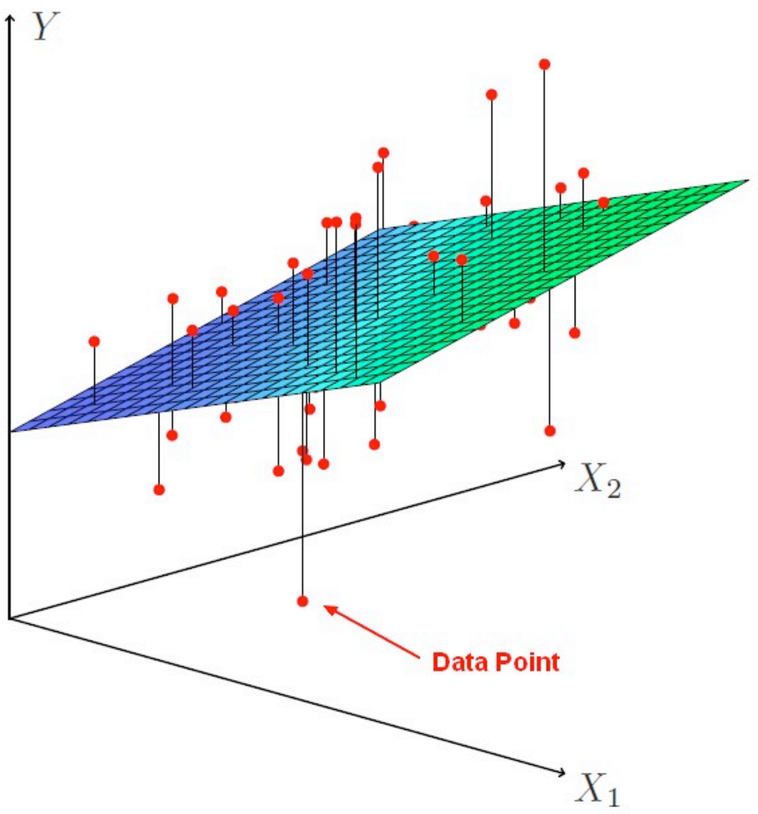
Linear Regression



Linear Regression

What if we have a high-dimensional vector for samples?

Linear Regression

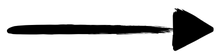


$$y_i = m_0 w_i + m_1 x_i + \ldots + b$$

Hyperplanes!

(details are about the same)

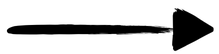
- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol
- Red or White



Quality (0-10)

Logistic Regression

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



White or Red

Logistic Regression

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

White or Red

how do we turn this into numbers?

Logistic Regression

a possible solution…

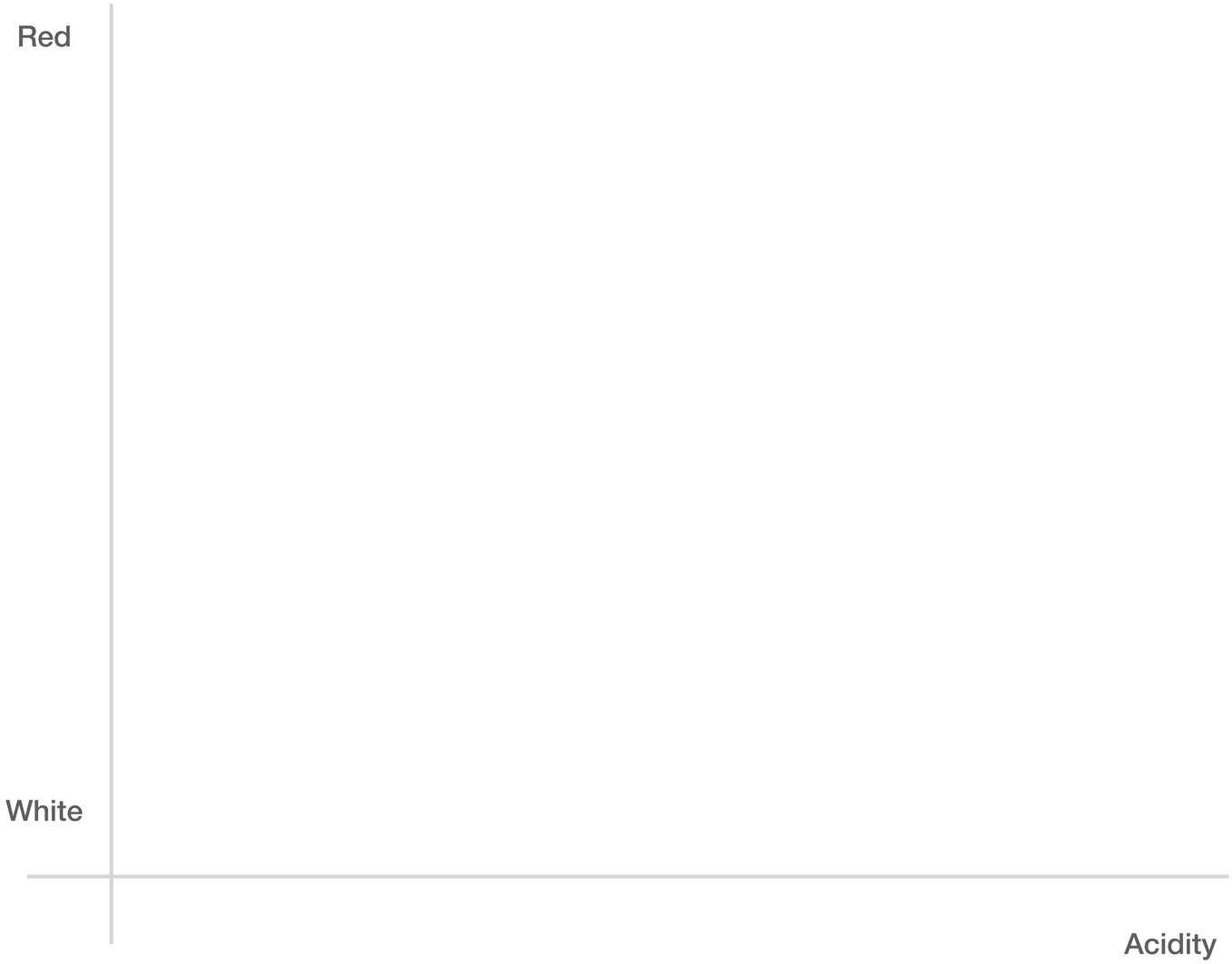
a class

White = 0

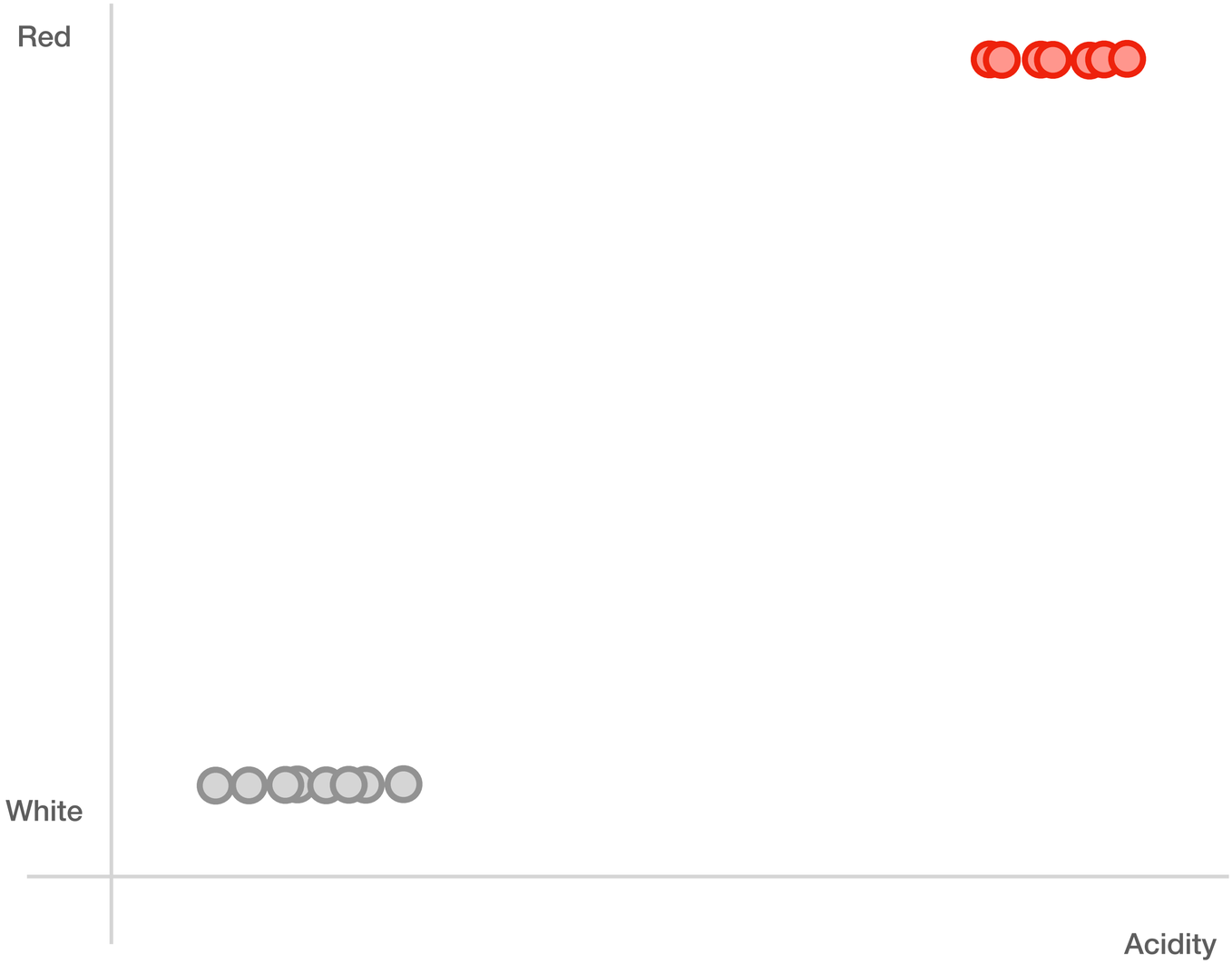
Red = 1

- categorical label outputs **are named “classes”**

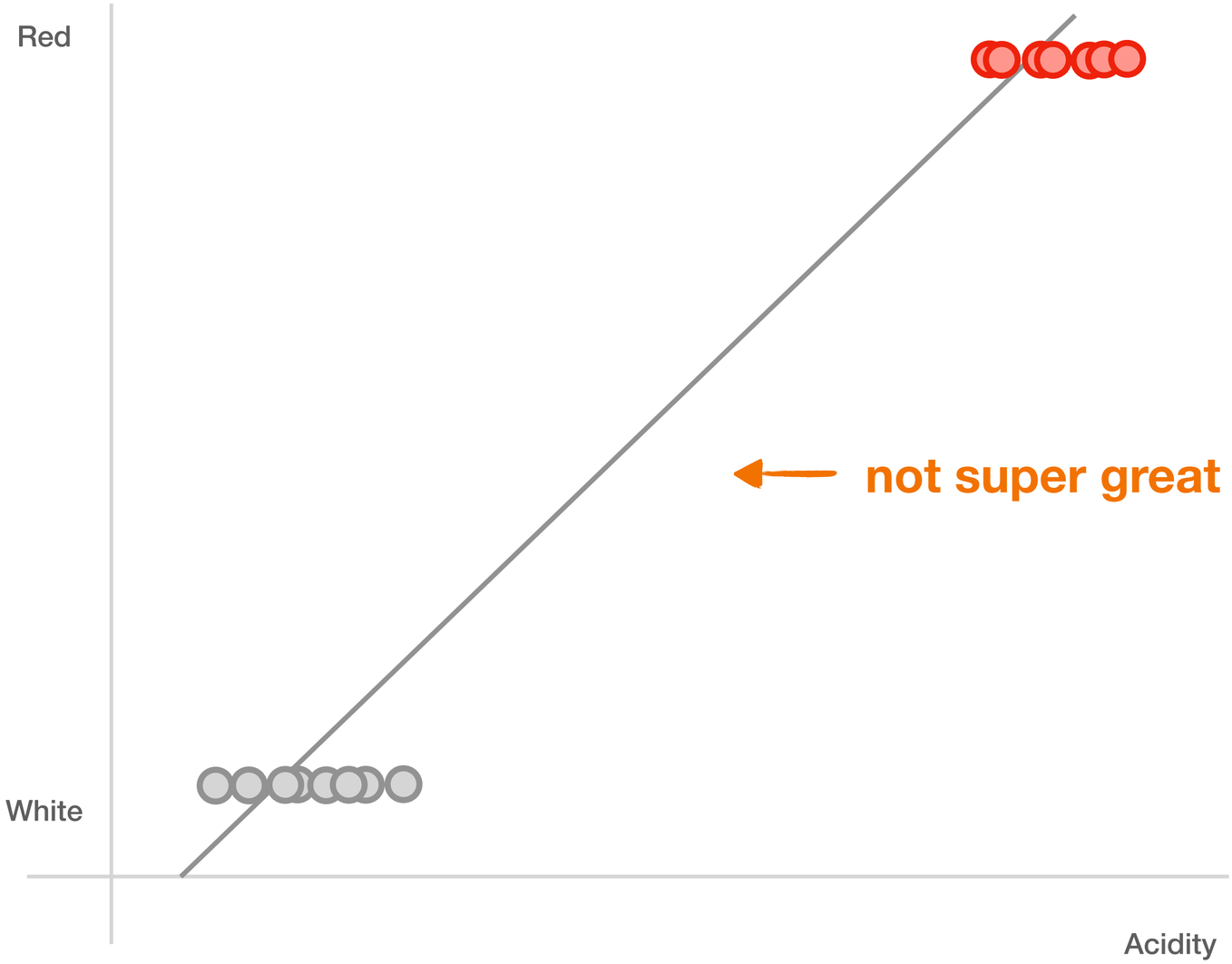
Logistic Regression



Logistic Regression



Logistic Regression



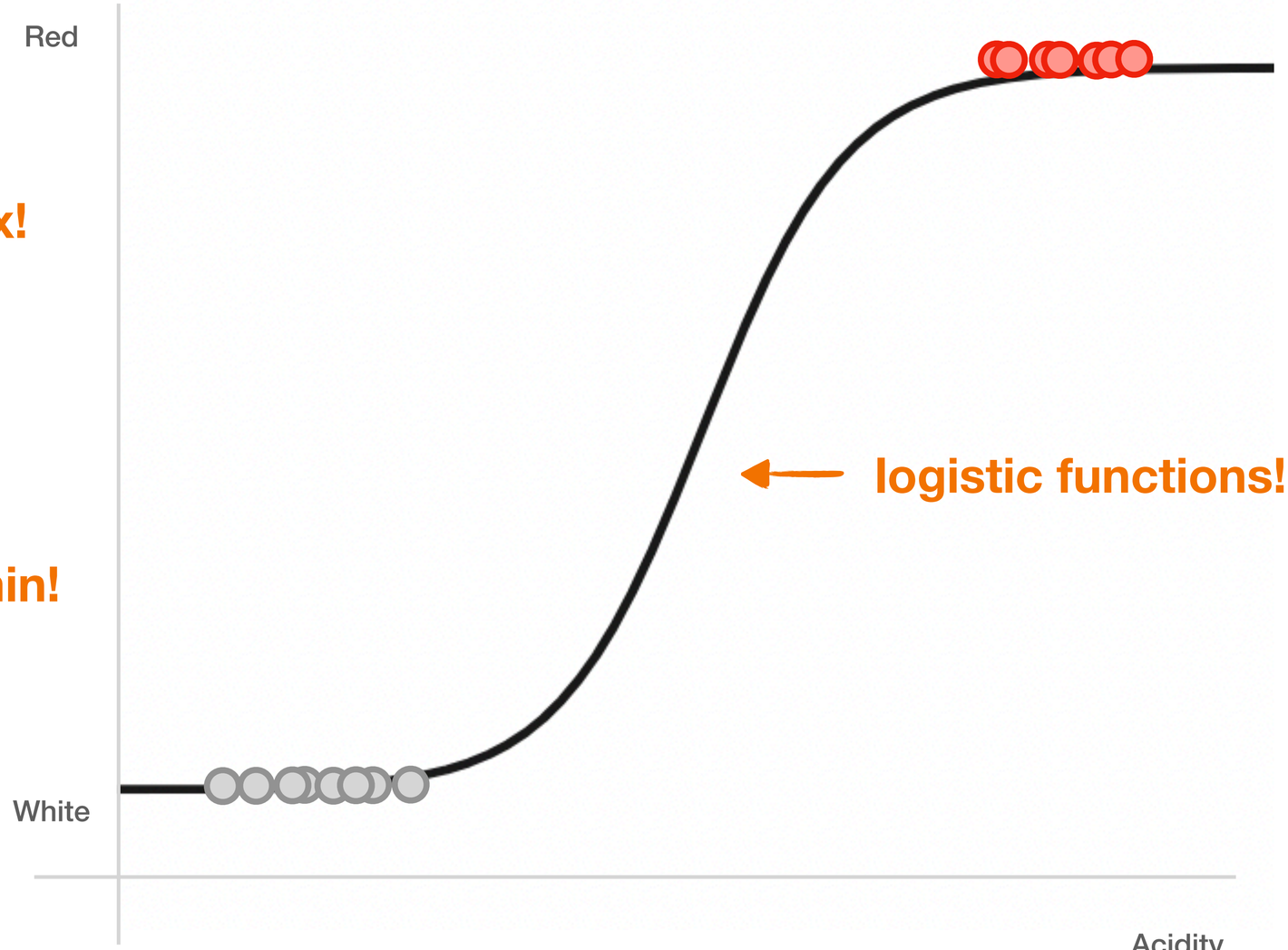
Logistic Regression

$$y_i = \frac{L}{1 + e^{-k(x_i - x_0)}}$$

max!

steepness!

min!



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

- **logistic functions** are a good way of making predictions on binary/categorical data
- **sum-of-squared** error is *still a* common error metric for logistic regression

Logistic Regression

White Red Champagne

pair them up!

Logistic Regression

White Red

White Champagne

Red Champagne

Logistic Regression

White White Red

Red Champagne Champagne

Logistic Regression

White **White** **Red**

Red Champagne Champagne

Logistic Regression

White	**White**	Red
Red	Champagne	**Champagne**

Logistic Regression

more pairs voted for white!

White **White** Red

Red Champagne **Champagne**

Its White!

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

Logistic Regression

White Red

White Champagne

Red Champagne

Logistic Regression

White

Red Champagne

Red

Champagne White

Champagne

Red White

Logistic Regression

White

Red Champagne

Red

Champagne White

Champagne

Red White

pick the answer with highest probability

- **One-vs-all** multiclass classification uses the most highest probability paring result among all combinations

Ok. Time for doing stuff.