

A scenic view of a river flowing through a lush green forest. In the foreground, tall, slender bamboo stalks stand vertically, their green leaves and branches partially obscuring the view. The river, with a brownish, slightly murky water, flows from the background towards the foreground, curving slightly to the right. The surrounding forest is dense with various green trees and foliage, creating a vibrant, natural backdrop. The overall atmosphere is peaceful and serene.

HUMAN LEARNING ABOUT MACHINE LEARNING

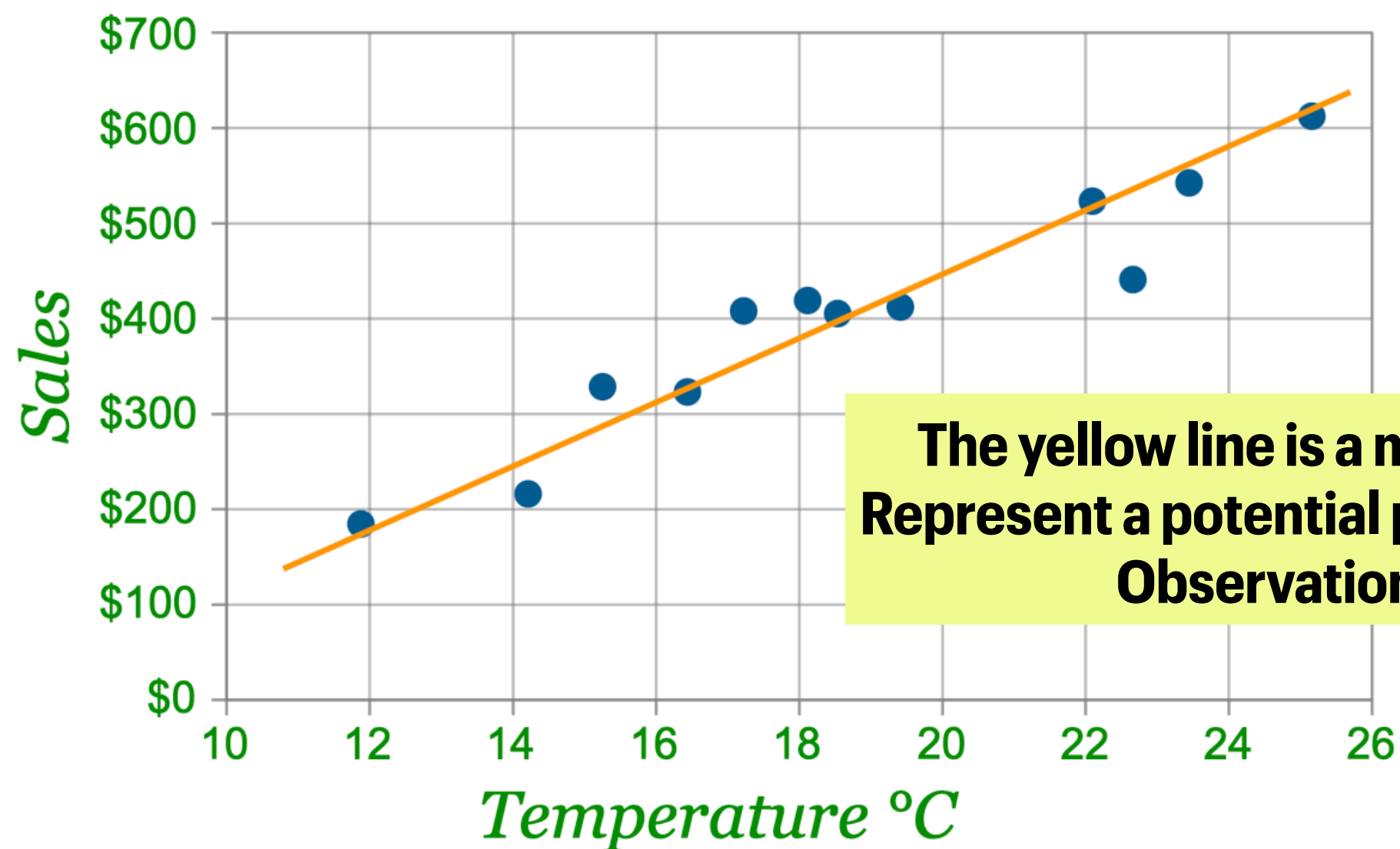
DR PAT RYSER-WELCH

SALES OF ICE-CREAMS IN THE SUMMER



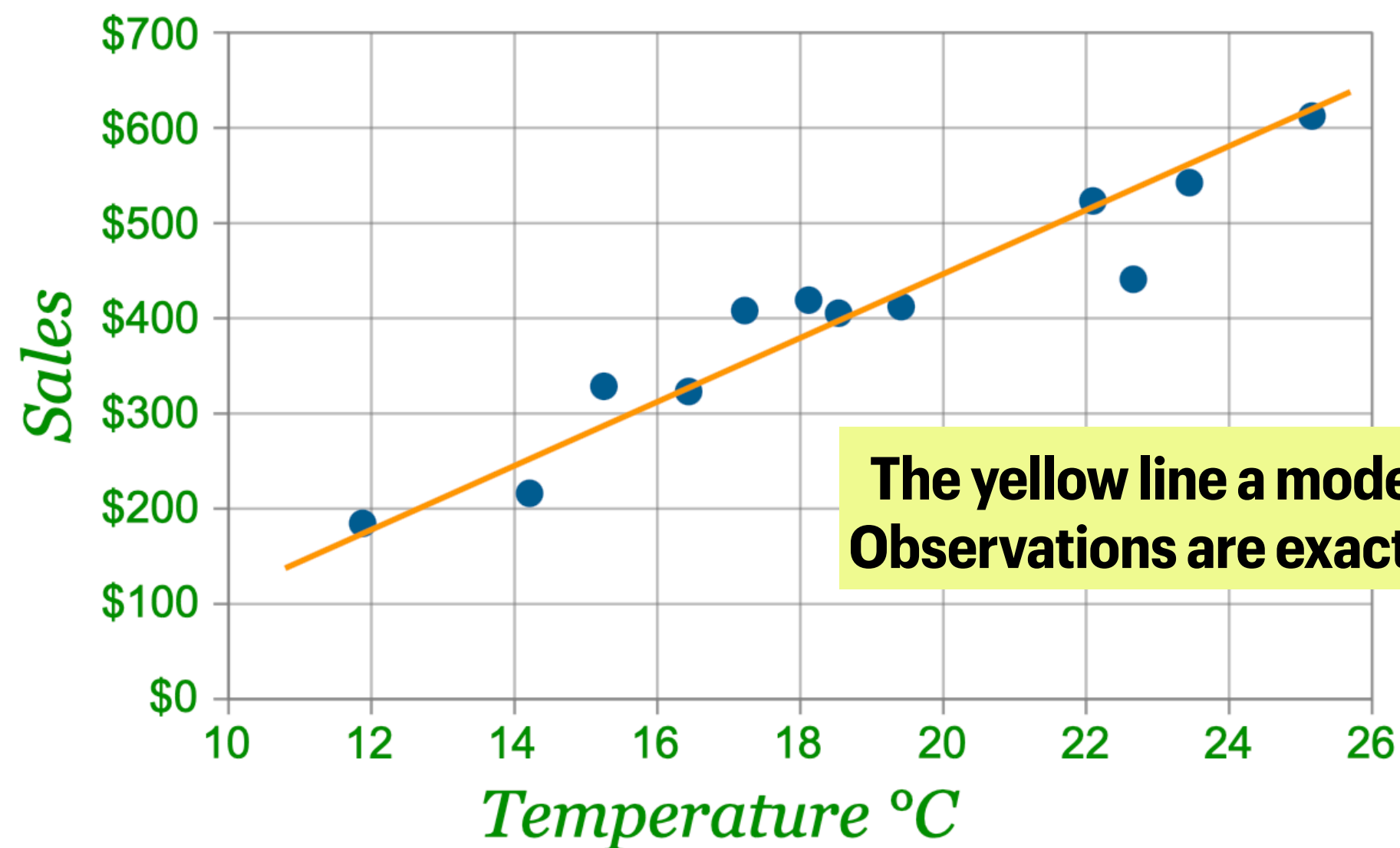
If I have captured the sales of ice-creams and the temperature in degree C. Could I possibly predict the sales of ice-cream?

**The blue dots match some observed sales
against the sales**



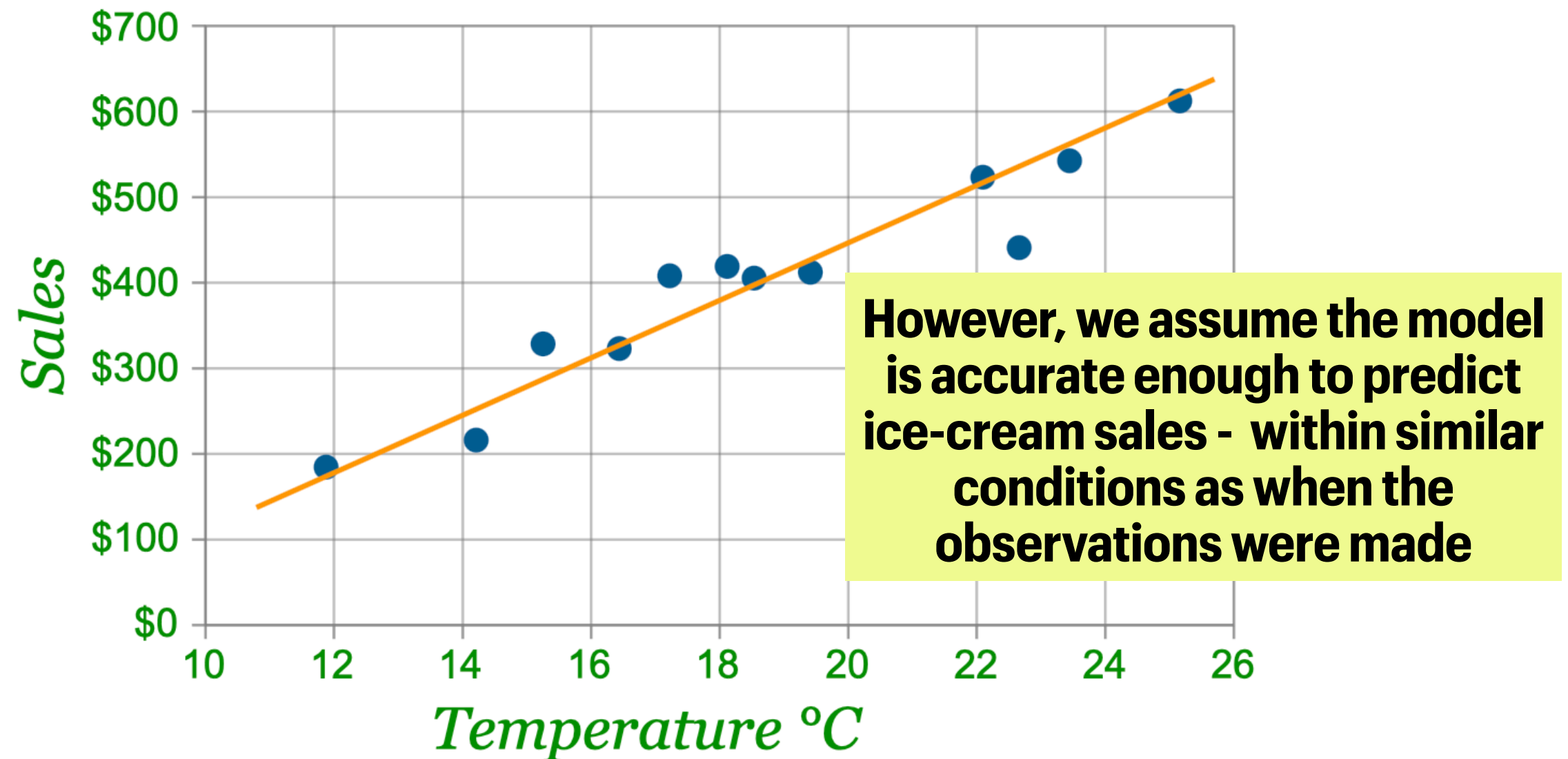
This is a fictitious data

**The blue dots match some observed sales
against the sales**



This is a fictitious data

**The blue dots match some observed sales
against the sales**



This is a fictitious data

$$\mathbf{DATA = MODEL + ERROR}$$

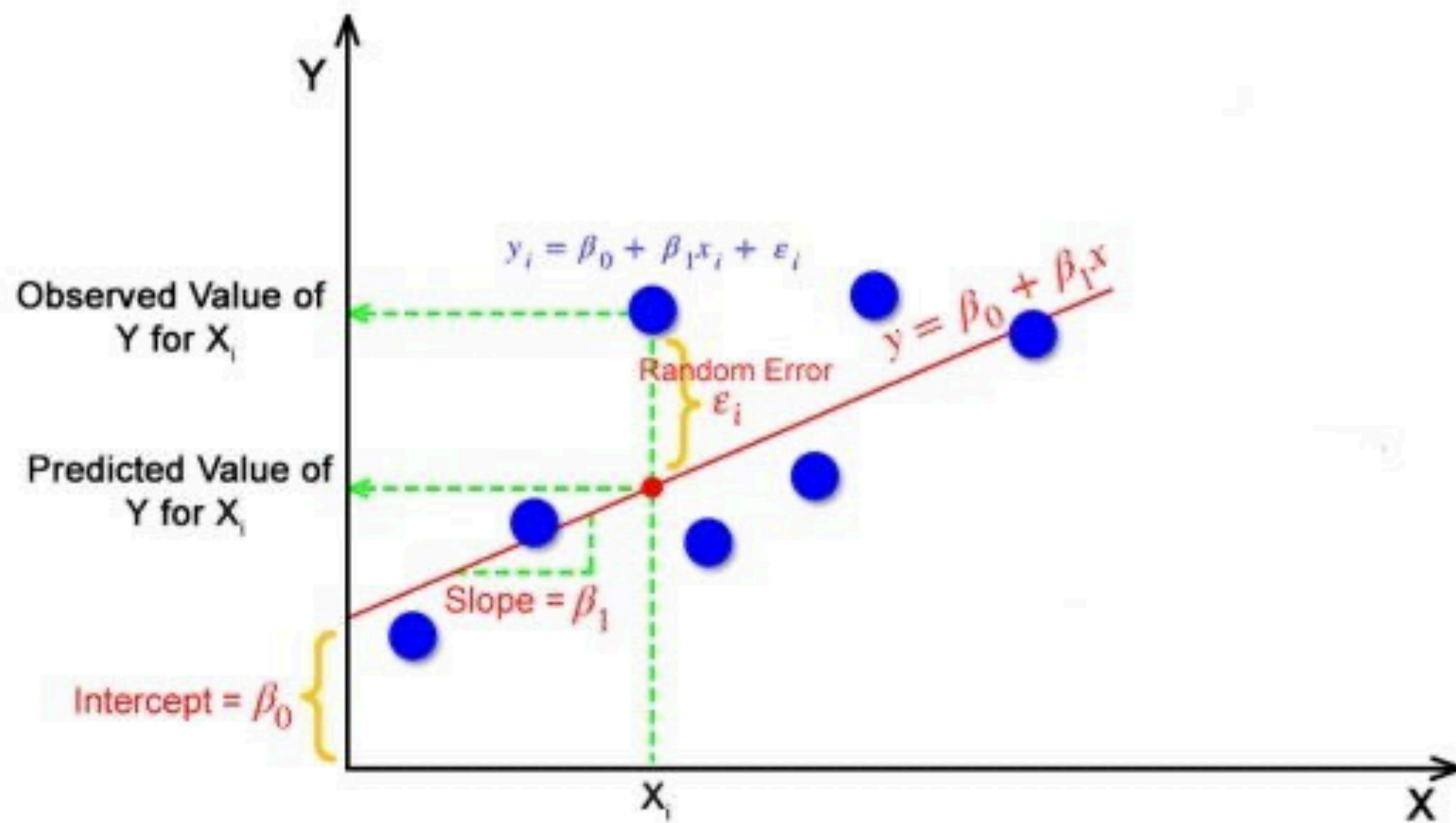
Models

Models capture the nature of some data as simply as possible. The basic structure of a statistical model considers data as the sum of a model and some errors.

$$data = model + error$$

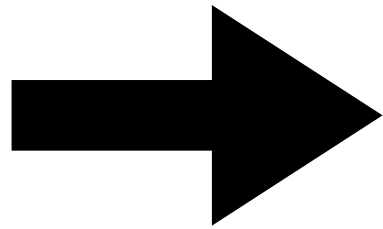
- The model expresses the values we expect the data to be take given our knowledge.
- The error reflects the differences between the model's prediction and the observed data.

AN ILLUSTRATION

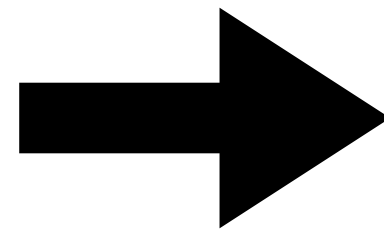


**HOW CAN WE LEARN A
MODEL?**

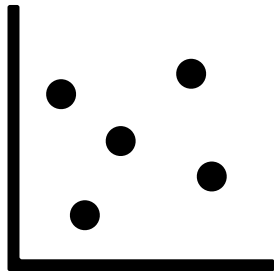
**DATA OR
SAMPLE**



LEARNING ALGORITHM



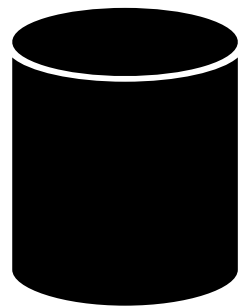
**MODEL
AND
ERROR**



Neural AI
Regression

Neural network

Decision Tree



Supervised learning

Model fitting

Regression line

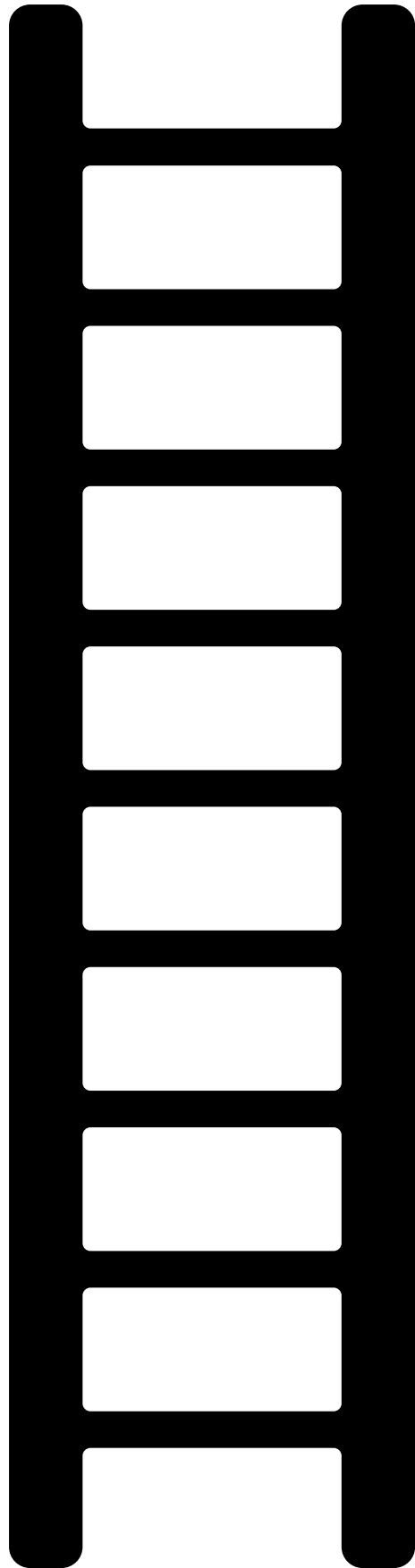
Non-deterministic
Methodologies

General linear model



Gradient-descent

Probabilistic model
Statistical model



STEP 4 - LEARNING OR MODEL FITTING

STEP 3 - PREPARE DATA FOR LEARNING

STEP 2- CLEAN AND PREPARE DATA

STEP 1 - EXPLORE THE DATA

**HOW CAN MEASURE THE
QUALITY OF A MODEL**

1. Apply the model on a dataset

**2 . Compare the predicted values
against known expected values.**

3 . Compute some metrics

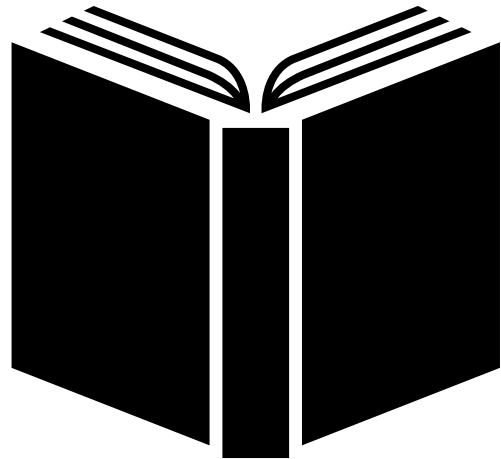
EXPECTED VALUE



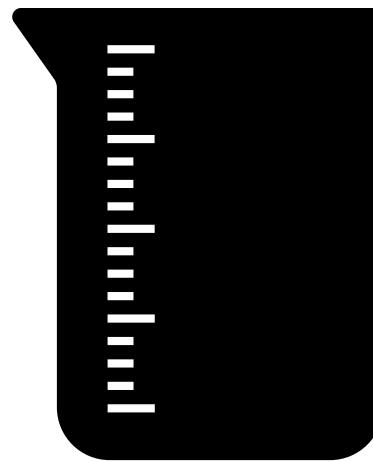
PREDICTED VALUE



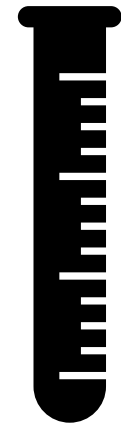
THREE DATASETS:



TRAINING



VALIDATION



TEST

Measures of quality

The measures of quality for a **predictive** model are expressed using two performance metric referred as *precision*, *recall*, and *accuracy*. These metrics are probabilities computed using the following mathematical formulae.

Accuracy describes how the model perform across *all* classes (True and False). It adds the diagonal values of a confusion matrix and divide by the total of possible outcomes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

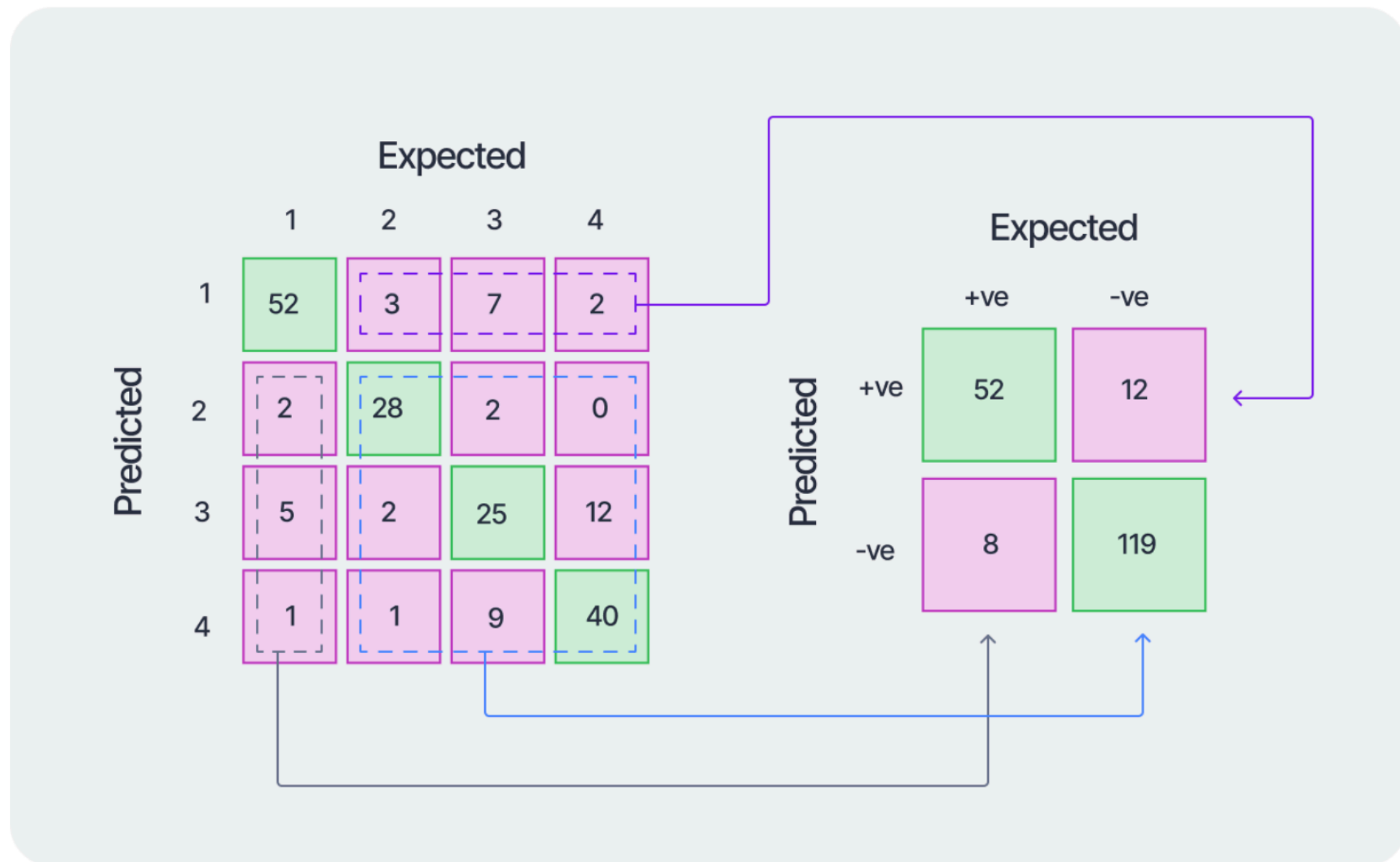
Precision measures the model's accuracy in **prediction** the True class as a sample. It reflects how reliable the model is in classifying samples as positive. It is a probability based on the class columns the confusion matrix (green cells). The precision can be computed for each class.

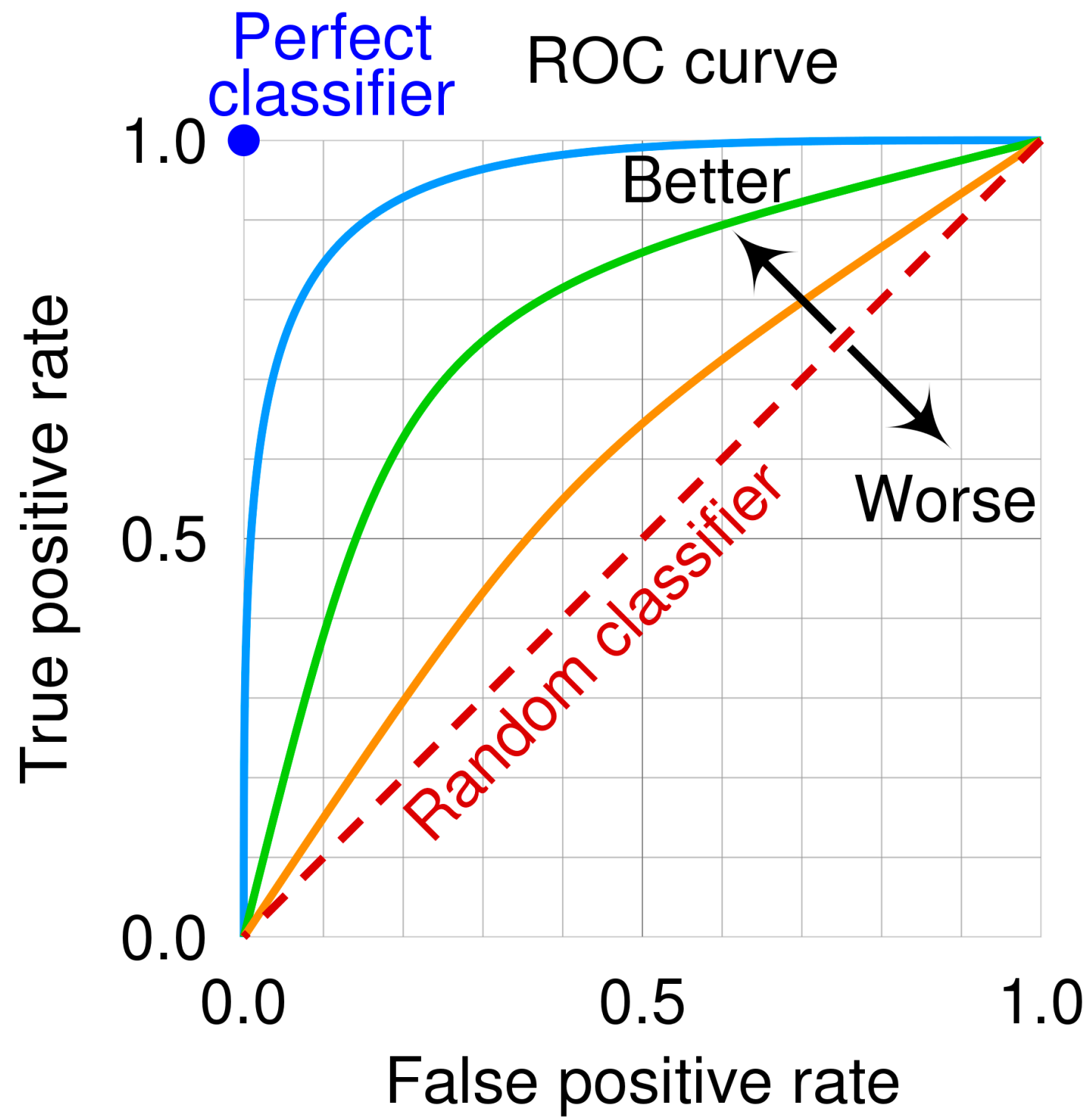
$$Precision = \frac{TP}{TP + FP}$$

Recall is the probability to detect positive classes. It can be calculated for each class. It is a probability is obtained by dividing the true positive by the sum of the true positive and false negative; i.e., the class row of the confusion matrix.

$$Recall = \frac{TP}{TP + FN}$$

CONFUSION MATRIX





**DATA AND LARGE AMOUNT OF IT
SUPPORTS MACHINE LEARNING**

TWO EXAMPLES

Predicting inspection outcome - Chicago food inspection dataset

The data has some repeated patterns

The data appears to have some dependent statistical variables.

Many of the data may have some clusters or strong relationship between them.

We achieve 100 percent accuracy.

[Find the notebook](#)

Predicting surviving the Titanic disaster

The data no clear repeated patterns.

The data appears to have some observations with a lot of complexity.

The data is quite small.

It is hard to achieve more than 80% in models, with machine learning.

[Find the notebook](#)