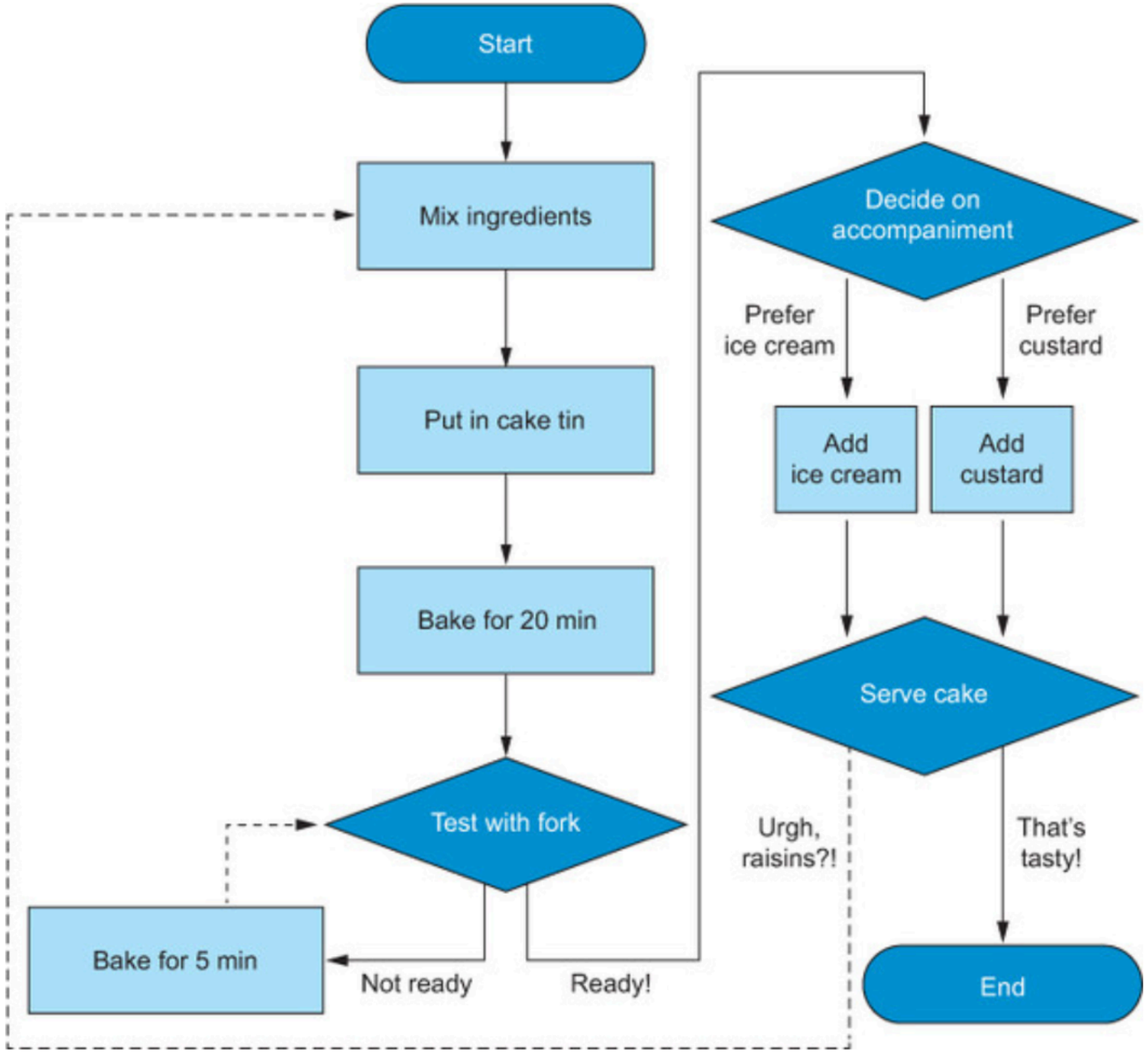
Machine learning: Regression

Alex Di Genova

18/06/2024

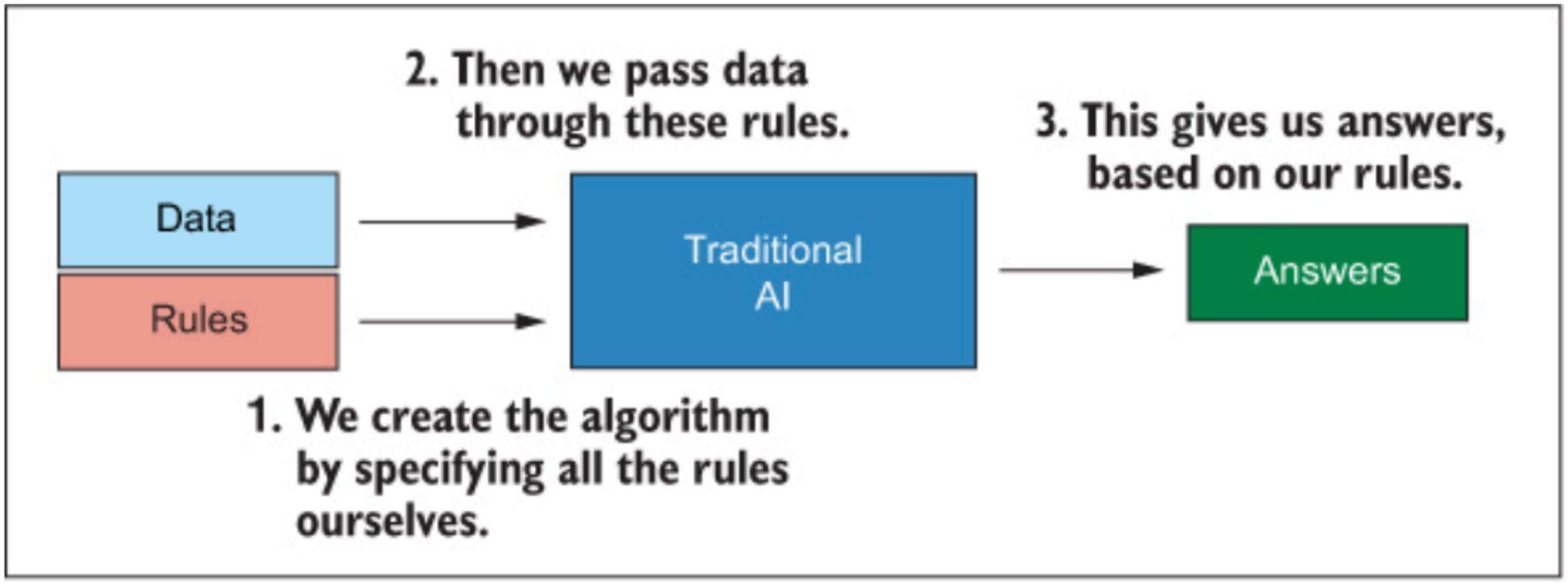
What is Machine learning?

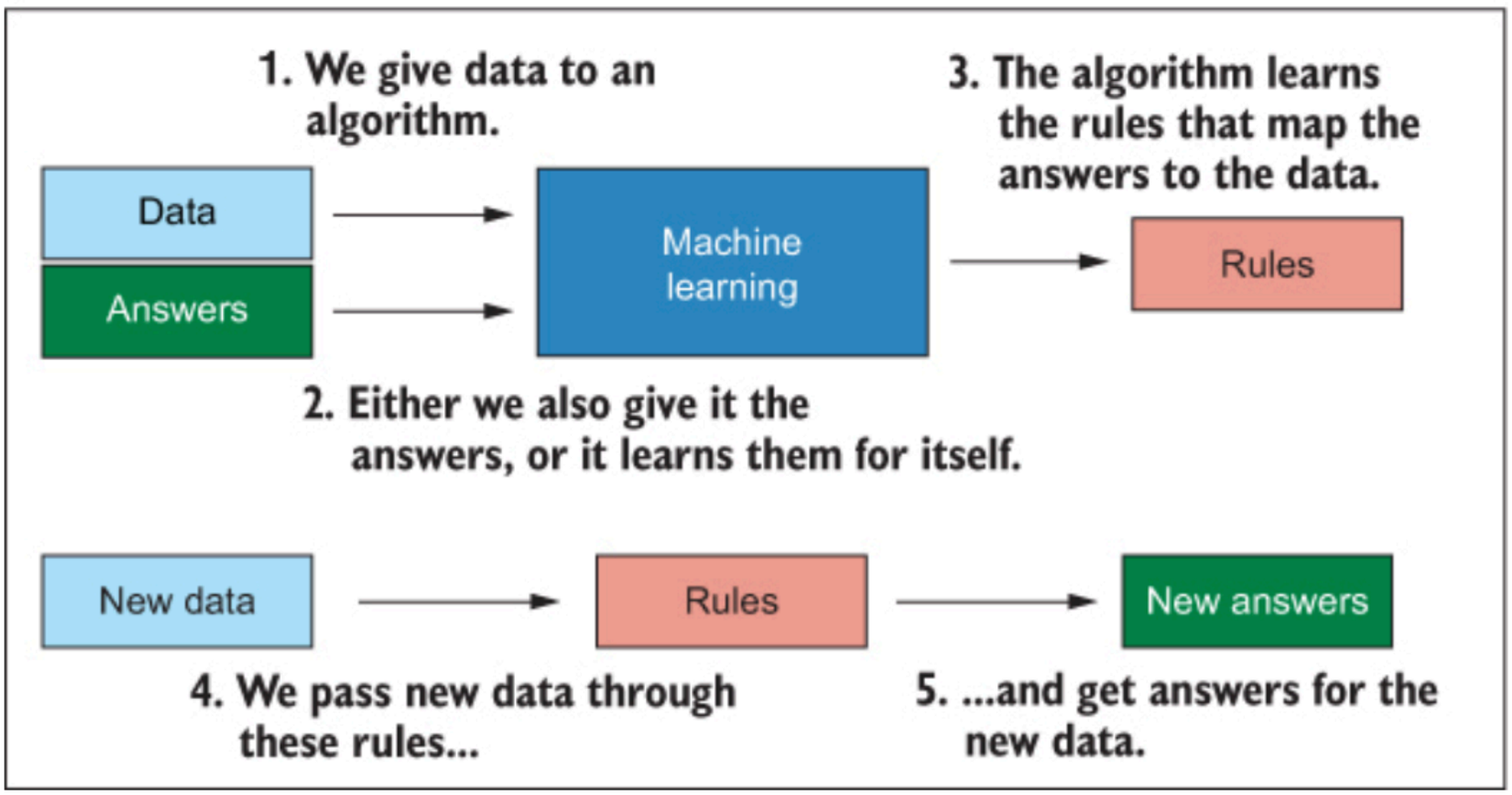
- Machine learning, sometimes referred to as statistical learning, is a subfield of artificial intelligence (AI) whereby algorithms “learn” patterns in data to perform specific tasks.



Recipe

What is Machine learning?

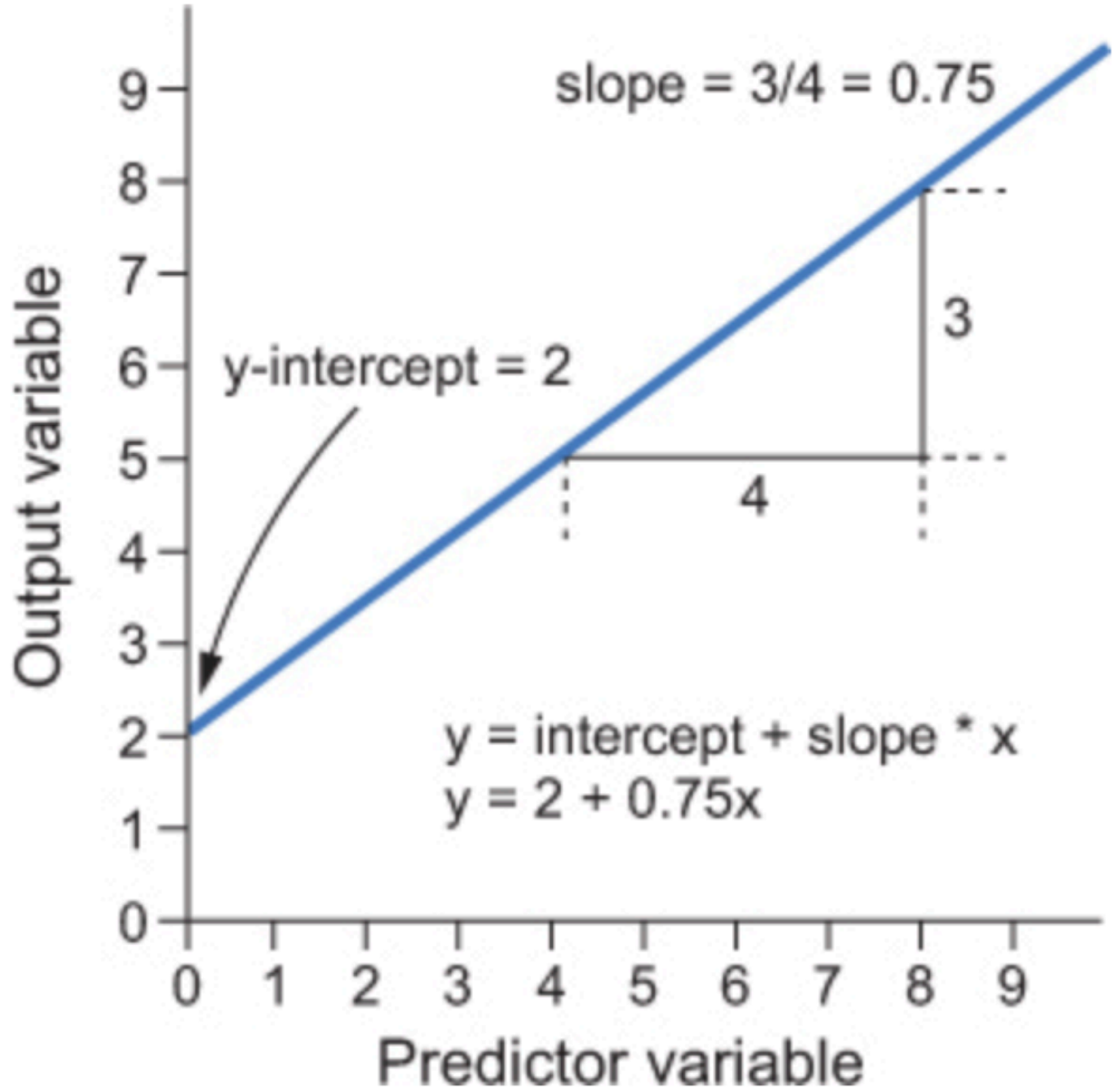




Machine learning

Model and algorithm

- A set of rules that a machine learning algorithm learns is a model.
- Once the model has been learned/trained, we can give it new observations, and it will output its predictions for the new data.
- Algorithm is the process by which a model learn.

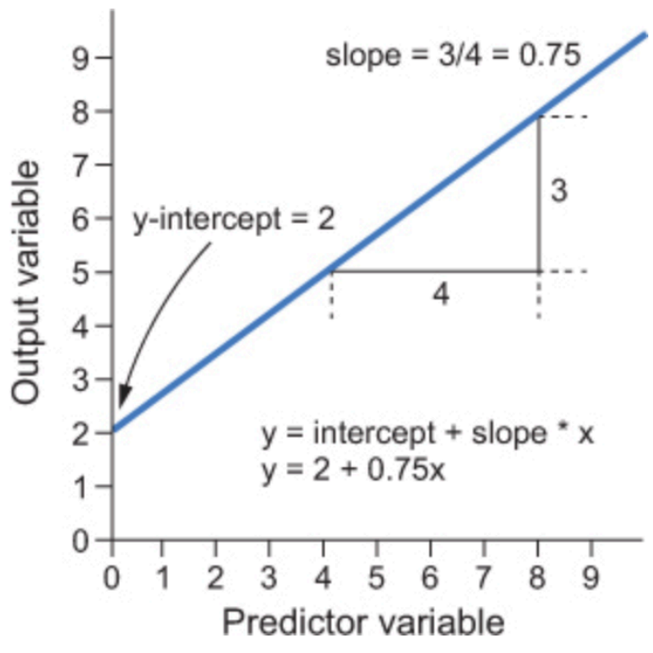


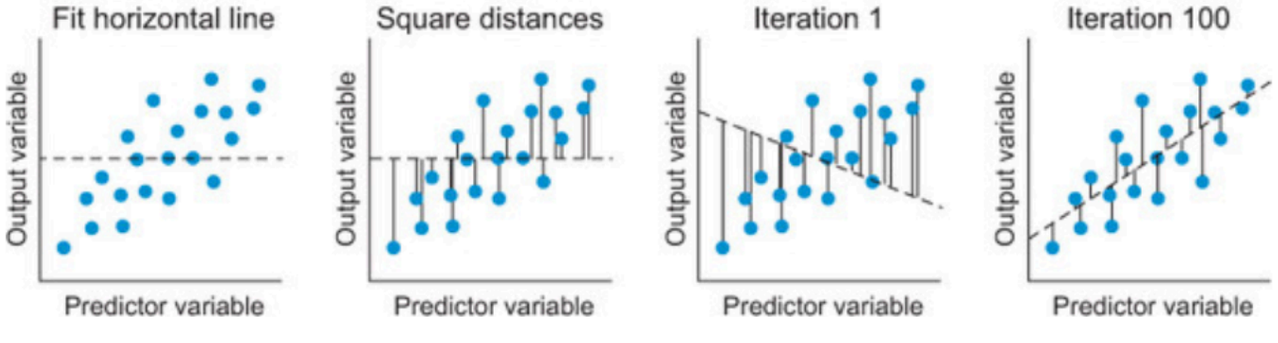
Y = intercept + slope X (parameters)

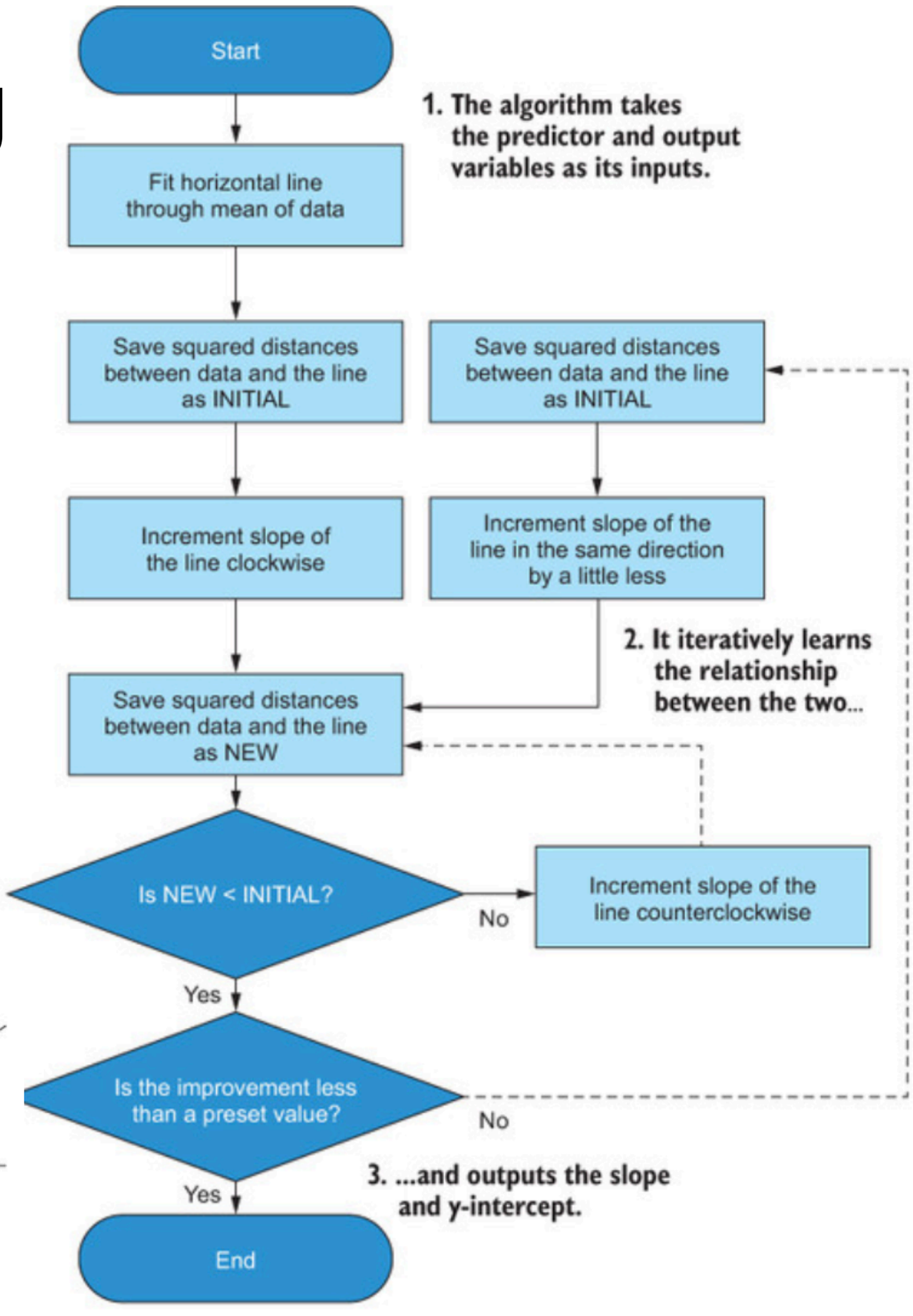
Machine learning

Model and algorithm

Y = intercept + slope X (parameters)



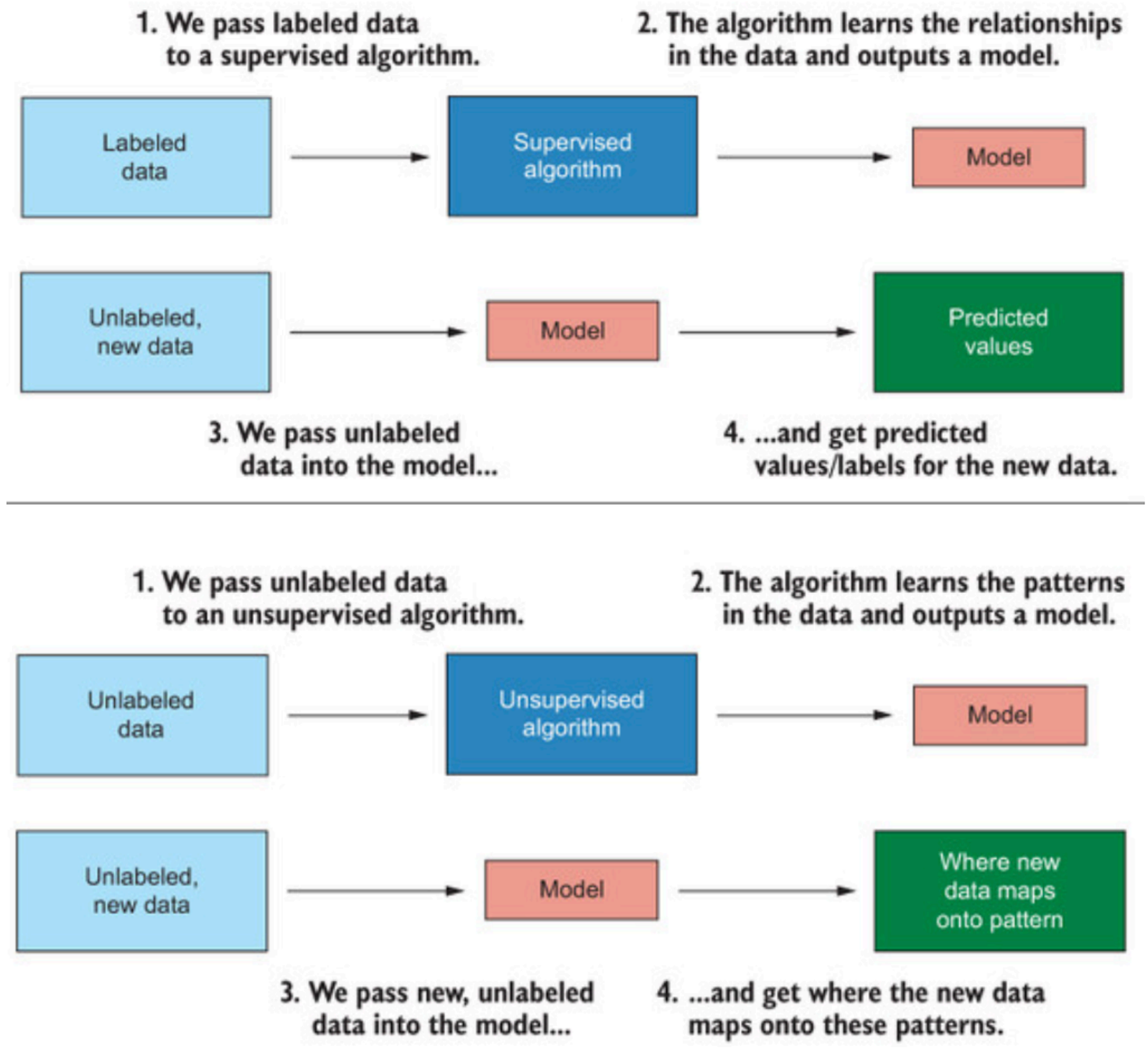




Machine learning algorithms

Classes

- Supervised
  - Classification
  - Regression
- Unsupervised
  - Dimension Reduction
  - Clustering
- Semi-supervised



Machine learning algorithms

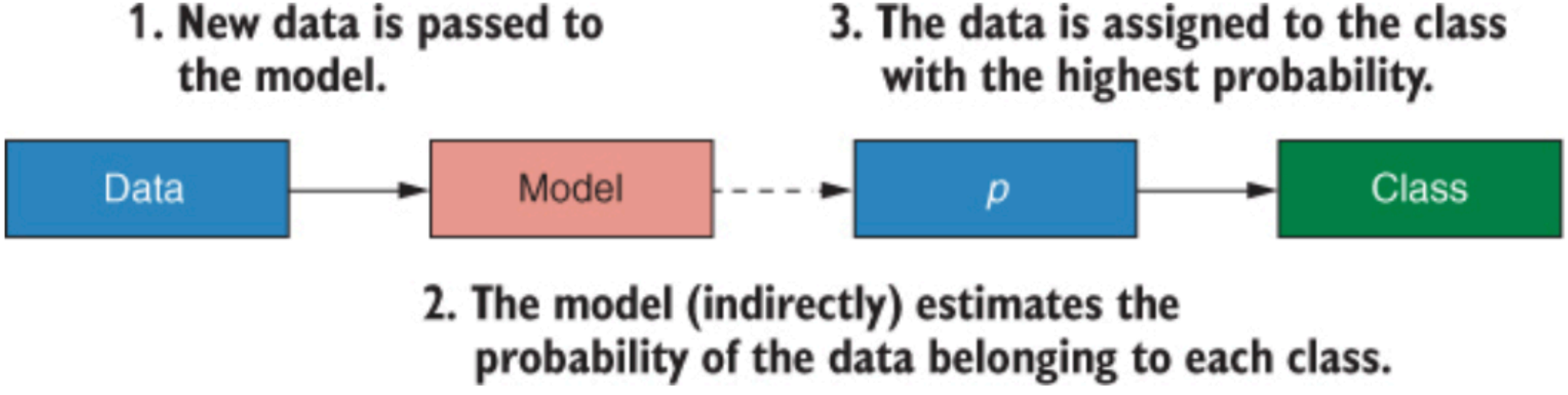
Classes

- Supervised
    - Classification: take labeled data and learn patterns in the data that can be used to predict a categorical output variable.
    - Regression: take labeled data and learn patterns in the data that can be used to predict a continuous output variable.
- Unsupervised
    - Dimension Reduction: take unlabeled and high-dimensional data (data with many variables) and learn a way of representing it in a lower number of dimensions (exploratory).
    - Clustering: take unlabeled data and learn patterns of clustering in the data. A cluster is a collection of observations that are more similar to each other than to data points in other clusters.

Machine learning

Logistic Regression

- Logistic regression predicts the probability that a given input belongs to a particular class.



Machine Learning

Logistic regression

- **Log-Odds**
    - Model the probability that a given input belongs to a particular class
    - Odds = p/1-p —- the ratio of the probability of the event occurring to the probability of it not occurring.
    - Log-Odds = log (p/1-p)
        - log-odds are modeled as a linear combination of the input variables:
            - $\log(1/1-p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$
- **Sigmoid Function** (Logistic function)
    - Maps the log-odds to a probability value between 0 and 1
    - $\sigma(z) = 1 / (1 + e^{-z})$ where $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$
- **Binary Classification**
    - If the probability is greater than a certain threshold (commonly 0.5), the input is classified as class 1; otherwise, it is classified as class 0.

Machine learning

Predict whether a student will pass (1) or fail (0) an exam

Data:

Hours Studied (x)	Pass/Fail (y)
1	0
2	0
3	0
4	1
5	1
6	1
7	1

1. Log-oods:
    1. $\log(1/1-p)=\beta 0+\beta 1$ Hours Studied
2. Sigmoid Function:
    1. $p=\sigma(\beta 0+\beta 1 \cdot \text{Hours Studied}) = 1 / (1+ e^{-(\beta 0+\beta 1 \cdot \text{Hours Studied})})$
3. Binary Classification:
    1. If $p \geq 0.5$ predict Pass (1). If $p<0.5$ predict Fail (0).
4. Let's asume model has been trained and the parameters are :
    1. $\beta 0 = -6$ y $\beta 1 = 1$

Machine learning

Predict whether a student will pass (1) or fail (0) an exam

Data:

Hours Studied (x)	Pass/Fail (y)
1	0
2	0
3	0
4	1
5	1
6	1
7	1

1.Prediction Example

1.For a student who studied 3 Hours

2.Log-odds:

1. $z=\log(1/1-p)=\beta 0+\beta 1\ 3 = -6 + 1*3 = -3$

3. Sigmoid Function:

1.$p=\sigma(-3) = 1 / (1+ e^{-(-3)}) = 1/(1+e^3) = 1/21.08 = 0.04$

4. Binary Classification:

1. If $0.04<0.5$ the model predict Fail (0).

Book:

Machine learning with R

Questions?

Practice!!!