Machine Learning Overview

Machine Learning

- It is finally time to dive deep into Machine Learning!
- This Machine Learning Overview section is designed to help get us in the correct frame of mind for the paradigm shift to Machine Learning.
- First, let's quickly review where we are in the Machine Learning Pathway....

ML Pathway

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

Problem to Solve

Real World

Question to Answer

ML Pathway

Problem to Solve

How to fix or change X?

Real World

Question to Answer

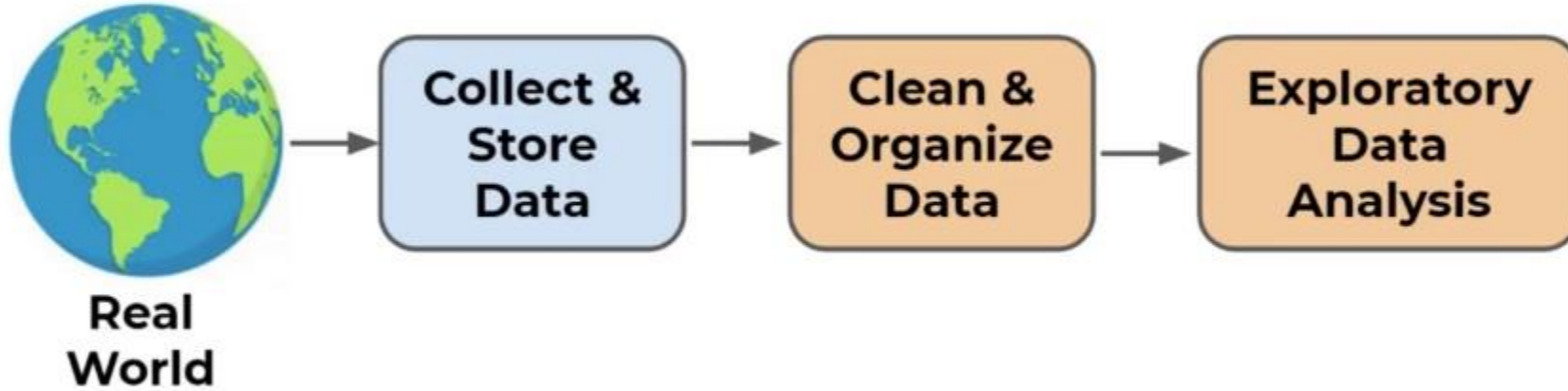How does a change in X affect Y?

ML Pathway

رواد مصر الرقمية

Problem to Solve

How to fix or change X?

Real World

Question to Answer

How does a change in X affect Y?

ML Pathway

ML Pathway

Real World → Collect & Store Data → Clean & Organize Data → Exploratory Data Analysis → Machine Learning Models

Supervised Learning: *Predict an Outcome*
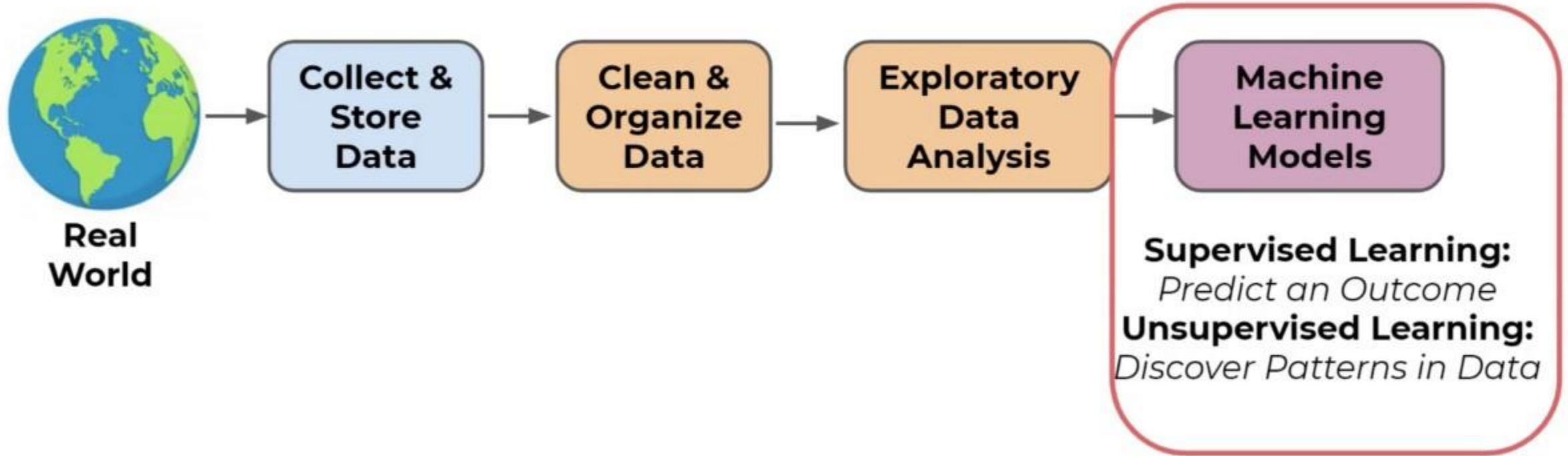**Unsupervised Learning:** *Discover Patterns in Data*
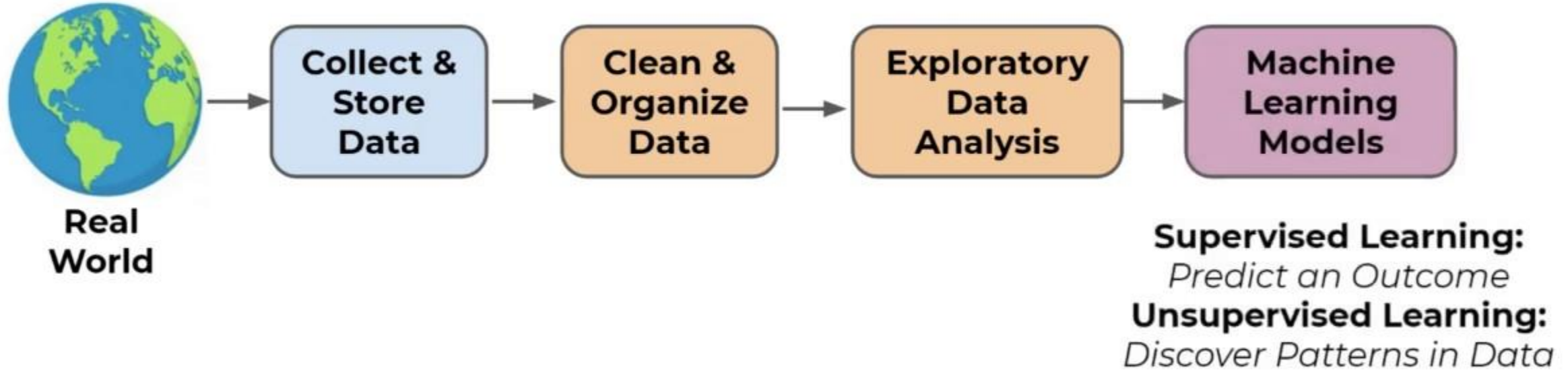
ML Pathway

Machine Learning

- Our main goals in ML Overview section:
  - Problems solved by Machine Learning
  - Types of Machine Learning
    - Supervised Learning
    - Unsupervised Learning
  - ML Process for Supervised Learning
  - Discussion on Companion Book

Machine Learning

- Our main goals in ML Overview section:
  - No coding in this section!
  - Purely a discussion on critically important ideas applied to ML problems.

Machine Learning

- Many other relevant topics will be discussed later in the course as we "discover" them, including:
  - Bias-Variance Trade-off
  - Cross-validation
  - Feature Engineering
  - Scikit-learn
  - Performance Metrics and much more!

Machine Learning

- Machine Learning Sections
  - Section for Type of Algorithm
    - Intuition and Mathematical Theory
    - Example code-along of application of Algorithm
    - Expansion of Algorithm
    - Project Exercise
    - Project Exercise Solution

Machine Learning

- Machine Learning Sections
  - Exception for Linear Regression
    - Intuition and Mathematical Theory
    - Simple Linear Regression
    - Scikit-learn and Linear Regression
    - Regularization
  - "Discovering" additional ML topics

Machine Learning

- Machine Learning Sections
  - "Discovering" additional ML topics
    - Performance Metrics
    - Feature Engineering
    - Cross-validation
  - Revisit Linear Regression to combine discovered ML ideas for Project Exercise.

Why Machine Learning?

Machine Learning

- Machine learning in general is the study of statistical computer algorithms that improve automatically through data.

- This means unlike typical computer algorithms that rely on human input for what approach to take, ML algorithms infer best approach from the data itself.

Machine Learning

- Machine learning is a subset of Artificial Intelligence.
- ML algorithms are not explicitly programmed on which decisions to make.
- Instead the algorithm is designed to infer from the data the most optimal choices to make.

Machine Learning

- What kinds of problems can ML solve?
  - Credit Scoring
  - Insurance Risk
  - Price Forecasting
  - Spam Filtering
  - Customer Segmentation
  - Much more!

Machine Learning

- Structure of ML Problem framing:
    - Given **features** from a data set **obtain** a desired **label**.
    - ML algorithms are often called "estimators" since they are estimating the desired **label** or output.

Machine Learning

- How can ML be so robust in solving all sorts of problems?
- Machine learning algorithms rely on data and a set of statistical methods to learn what features are important in data.

Machine Learning

- Simple Example:
  - Predict the price a house should sell at given its current features (Area,Bedrooms,Bathrooms,etc...)

Machine Learning

- House Price Prediction
  - Typical Algorithm
    - Human user defines an algorithm to manually set values of importance for each feature.

- **House Price Prediction**
  - ML Algorithm
    - Algorithm automatically determines importance of each feature from existing data

Machine Learning

- Why machine learning?
  - Many complex problems are only solvable with machine learning techniques.
  - Problems such as spam email or handwriting identification require ML for an effective solution.

Machine Learning

- Why not just use machine learning for everything?
    - Major caveat to effective ML is good data.
    - Majority of development time is spent cleaning and organizing data, **not** implementing ML algorithms.

Machine Learning

- Do we develop our own ML algorithms?
    - Rare to have a need to manually develop and implement a new ML algorithm, since these techniques are well documented and developed.

Machine Learning

- There are two main types of Machine Learning we will cover in upcoming sections:
  - Supervised Learning
  - Unsupervised Learning

Machine Learning

- **Supervised Learning**
  - Using **historical** and **labeled** data, the machine learning model predicts a value.
- **Unsupervised Learning**
  - Applied to **unlabeled** data, the machine learning model discovers possible patterns in the data.

Machine Learning

- Supervised Learning
    - Requires **historical labeled** data:
        - Historical
            - Known results and data from the past.
        - Labeled
            - The desired output is known.

Machine Learning

- Supervised Learning
  - Two main label types
    - Categorical Value to Predict
      - Classification Task
    - Continuous Value to Predict
      - Regression Task

Machine Learning

- Supervised Learning
    - Classification Tasks
        - Predict an assigned category
            - Cancerous vs. Benign Tumor
            - Fulfillment vs. Credit Default
            - Assigning Image Category
                - Handwriting Recognition

Machine Learning

- Supervised Learning
  - Regression Tasks
    - Predict a continuous value
      - Future prices
      - Electricity loads
      - Test scores

Machine Learning

- **Unsupervised Learning**
  - Group and interpret data without a label.
  - Example:
    - Clustering customers into separate groups based off their behaviour features.

Machine Learning

- Unsupervised Learning
    - Major downside is because there was no historical "correct" label, it is much harder to evaluate performance of an unsupervised learning algorithm.

Machine Learning

- Machine Learning Sections
  - We first focus on supervised learning to build an understanding of machine learning capabilities.
  - Then shift focus to unsupervised learning for clustering and dimensionality reduction.

Machine Learning

- Finally, before we dive into coding and linear regression in the next section, let's have a deep dive into the entire Supervised Machine Learning process to set ourselves up for success!