Week 1: Introduction

Ekarat Rattagan

August 20, 2025

Course Outline

1. Introduction
2. Linear Regression (Single variable)
3. Linear Regression (Multiple variables), Polynomial
4. Naive Bayes Classification
5. Logistic Regression
6. Regularization
7. Model Evaluation
8. K-NN, Tree
9. Support Vector Machine (SVM)
10. Neural Network I
11. Neural Network II
12. Clustering I (k-means, DBScan)
13. Clustering II (hierarchical clustering)
14. Dimension Reduction (PCA, AutoEncoder)
15. Association Rule learning

Week 1 Outline

1. Introduction to Machine Learning
2. Types of Machine Learning
    1. Supervised Learning
    2. Unsupervised Learning
    3. Reinforcement Learning
3. Machine Learning Process

Machine Learning Definition (1)

Traditional approach

Machine learning approach

Figure: 1 Traditional VS ML approach

Machine Learning Definition (2)

Arthur Samuel (1959)

Machine Learning (ML) is a field of study that gives computers the ability to learn without being explicitly programmed.

Machine Learning Definition (3)

Tom Mitchell (1998) Well-posed learning problem:

> *"A computer program is said to learn from* **Experience E** *with respect to some* **Task T** *and some* **Performance measure P**, *if its performance on* **T**, *as measured by* **P**, *improves with experience* **E**."

Machine Learning Definition - Examples I

- **Task T:** Classifying handwritten digits from images
- **Training experience E:** Dataset of digits given classifications, e.g., MNIST
- **Performance measure P:** Percentage of digits classified correctly

Machine Learning Definition - Examples II

- **Playing checkers**
  - T: Playing checkers
  - P: Percentage of games won against an arbitrary opponent
  - E: Playing practice games against itself
- **Autonomous Driving**
  - T: Driving on four-lane highways using vision sensors
  - P: Average distance traveled before a human-judged error
  - E: A sequence of images and steering commands recorded while observing a human driver
- **Spam Detection**
  - T: Categorize email messages as spam or legitimate
  - P: Percentage of email messages correctly classified
  - E: Database of emails, some with human-given labels

Types of Machine Learning

Supervised Learning

- Given a dataset $D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, where $x_i \in \mathbb{R}^d$ be a vector of attributes, and y_i is a label, if $y_i \in \mathbb{R}$ (A regression problem) or if $y_i \in \{C_1, C_2, ..., C_K\}$ (A classification problem)
- **Objective:** Find parameters θ such that the expected loss $L(y, f(x; \theta))$ is minimized, where f is the model and L is a loss function.

Types of Machine Learning

Unsupervised Learning

- Given a dataset $D = \{x_1, x_2, \ldots, x_N\}$, where $x_i \in \mathbb{R}^d$ be a vector of attributes.
- **Objective:** To understand and extract patterns, structures, or relationships from unlabeled data D.

Unsupervised Learning

Algorithms

- Clustering
- Dimensionality Reduction
- Anomaly Detection

Algorithms

- K-means
- DBScan
- PCA

Reinforcement Learning

- Given a sequence of states, actions, and rewards
- Output an optimal policy (a mapping from states to actions)

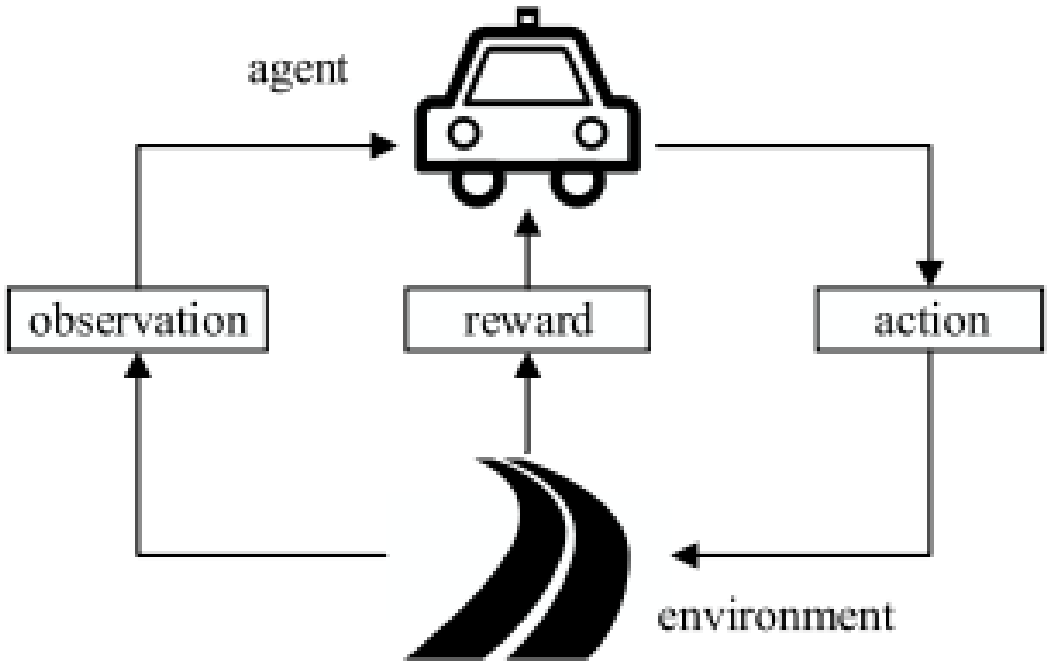

Fig. 1. Interaction between the agent and the environment: at each time step, after the agent observes the environment, it chooses an action according to its policy. After the action is executed, the environment gives a reward signal to the agent and transit to a new state.

Machine Learning Applications

- Email spam detection
- Face detection and recognition
- Sport analytics
- Zip code recognition
- Credit card fraud detection
- Stock prediction
- Smart assistants (e.g., ChatGPT)
- Recommendations
- Self-driving cars

Machine Learning Process

1. What is the desired outcome? (*Business Requirement*)
2. What could the dataset look like? (E)
3. Is this a supervised/unsupervised/reinforcement problem? (T)
4. What algorithms would you use? (*Solutions*)
5. How would you measure success? (P)
6. How to maintain the generated models?
7. What are potential challenges or pitfalls?

Machine Learning Operations

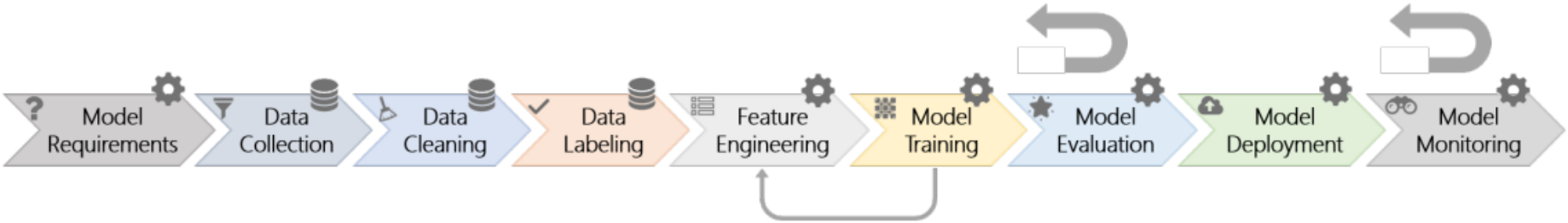

Fig. 1. The nine stages of the machine learning workflow. Some stages are data-oriented (e.g., collection, cleaning, and labeling) and others are model-oriented (e.g., model requirements, feature engineering, training, evaluation, deployment, and monitoring). There are many feedback loops in the workflow. The larger feedback arrows denote that model evaluation and monitoring may loop back to any of the previous stages. The smaller feedback arrow illustrates that model training may loop back to feature engineering (e.g., in representation learning).

Figure: `https://andrewbegel.com/papers/Software_Engineering_for_ML.pdf`

Components of Machine Learning Algorithms

1. Representation
2. Optimization
3. Evaluation

1. Representation

- **Numerical Functions**
  - Linear Regression: $y = \theta_0 + \theta_1 x$ or $\theta^T x$
  - Logistic Regression: $y = \sigma(\theta_0 + \theta_1 x)$ or $\sigma(\theta^T x)$ where σ is an activation function.
- **Symbolic Functions**
  - Decision Tree
  - Rule-based systems (e.g., If $A == B$, then C)
- **Instance-based Functions**
  - k-Nearest Neighbors
- **Probabilistic Graphical Models**
  - Bayesian Networks

2. Optimization

- Gradient Descent
- Stochastic Gradient Descent
- Adam
- AdaGrad
- Newton's Method
- Hessian Free Method
- Conjugate Gradient

3. Evaluation

- Accuracy
- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Precision, Recall, F1-Score, ROC AUC
- Cohen's kappa
- Matthews Correlation Coefficient (MCC)

Reference

- Hastie,T.(2009).The elements of statistical learning: data mining, inference, and prediction.