Lecture 2: Introduction to ML and Classical Models

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Machine Learning: An Overview

What is Machine Learning?

Machine Learning: An Overview

- Let's review some inspirational quotions ...
  - *"Machine learning is the hot new thing"*

    — John L. Hennessy, President of Stanford (2000–2016)
  - *"A breakthrough in machine learning would be worth ten Microsofts"*

    — Bill Gates, Microsoft Co-Founder
  - *"Computers are able to see, hear and learn. Welcome to the future."*

    — Dave Waters, Professor at University of Oxford
  - *"If software ate the world, models will run it"*

    — Steven A. Cohen and Matthew W. Granade, The Wallstreet Journal, 2018
  - ...

Machine Learning: An Overview

- The main motivation which we develop (computer) programs is to automate various kinds of (often tedious) processes.
- So far, we have learned to program the computers. the analogy that we have been using is something similar to this:

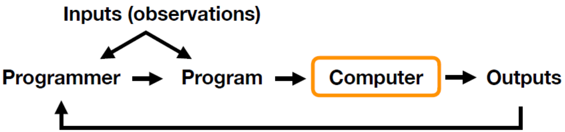

Figure: classical programming paradigm [1].

Machine Learning: An Overview

- The preceding traditional programming paradigm has several disadvantages:
  - what if we don't know waht program should we write for the given data (inputs) ?
  - what if the inputs change dynamically over the time? should we write another program?
- In order to resolve such problems, we should replace the need of developing computer programs "manually"
- In other words, we would like to automate the process of creating programs by informing the computer, the inputs and outputs that it needs:

Figure: ML paradigm [1].

Machine Learning: An Overview

- The preceding model was the main function of Machine Learning paradigm, In fact ML systems use both inputs and outputs to discover the **Rules and Patterns** behind the data
- Now that we are fimiliar with ML paradigm, we would like to define it formally:

  A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

- Here, inputs and outputs would be the experience (E), the main problem(s) that the computer wants to solve, is the class of tasks (T) and finally the performance measure shows how computer succeeded in performing (P)

Categories of Machine Learning

- The three broad categories of ML are summerized in:
  - **Supervised Learning**
  - **Unsupervised Learning**
  - **Reinforcement Learning**

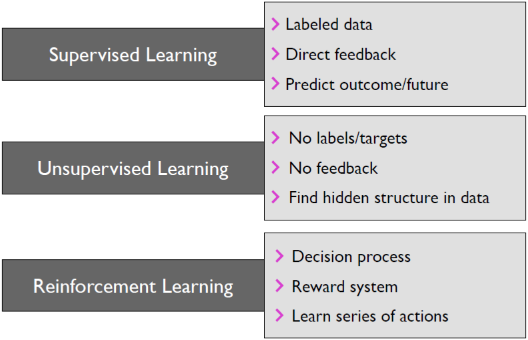

Figure: Categories of ML [2]

Introduction to Supervised Learning

- Supervised learning is the subcategory of machine learning that focuses on learning a **Classification** (Figure left), or **Regression** model (Figure right), that is, learning from labeled training data (i.e., inputs that also contain the desired outputs or targets; basically, "examples" of what we want to predict).

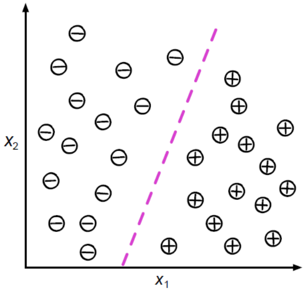

Figure: Illustration of classifi cation problem [2].

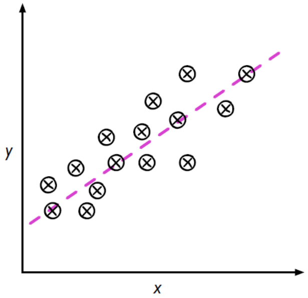

Figure: Illustration of a linear regression model [2].

Supervised Learning

- Given a data set $\mathcal{D} = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \dots, \langle \mathbf{x}_n, y_n \rangle\}$, there exists an unkown function called f which:

$$y = f(\mathbf{x})$$

- The supervised learning final goal is to **Approximate** this unkown function. we call our discovery function a *hypothesis* and we define it:

$$\begin{cases} h : \mathbb{R}^m \to \mathbb{R} \\ h(\mathbf{x}) = y \end{cases}$$

Unsupervised Learning

- In contrast to supervised learning, unsupervised learning is a branch of machine learning that is concerned with unlabeled data. Common tasks in unsupervised learning are **Clustering** analysis and **Dimensionality Reduction**.

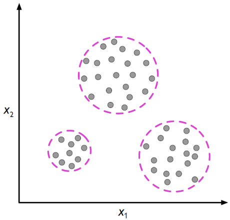

Figure: Illustration of Clustering [3].

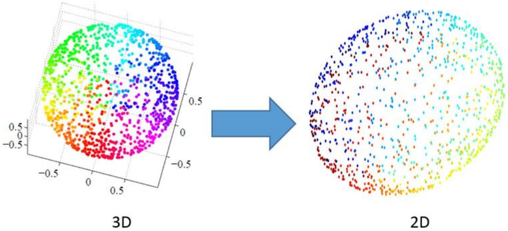

Figure: Illustration of Dimensionality Reduction [4].

Reinforcement Learning

- Reinforcement is the process of learning from rewards while performing a series of actions.

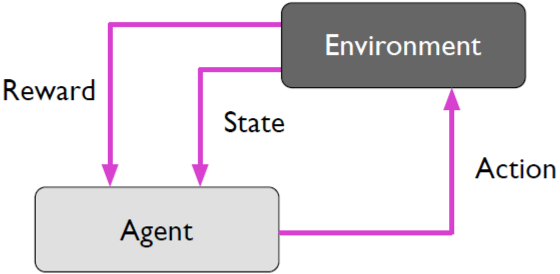

Figure: Illustration of Reinforcement Learning [2].

Classes of Machine Learning Algorithms

- Generalized linear models (e.g., logistic regression)
- Support vector machines (e.g., linear SVM, RBF-kernel SVM)
- Artificial neural networks (e.g., multi-layer perceptrons)
- Tree- or rule-based models (e.g., decision trees)
- Graphical models (e.g., Bayesian networks)
- Ensembles (e.g., Random Forest)
- Instance-based learners (e.g., K-nearest neighbors)

Algorithm Categorization Schemes

- Eager vs Lazy
- Single-Task vs Multi-Task
- Generative vs Discriminant
- Instance-based vs Model-Based
- Parametric vs Non-Parametric
- Batch vs Online

5 Steps To Solve A Machine Learning Problem

- 1. Define the problem to be solved.
- 2. Collect (labeled) data.
- 3. Choose an algorithm class.
- 4. Choose an optimization metric for learning the model.
- 5. Choose a metric for evaluating the model.

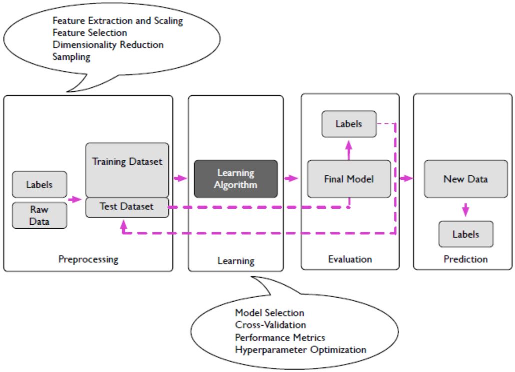

Figure: Learning Process [3].

Objective Functions

- Maximize the posterior probabilities (e.g., naive Bayes)
- Maximize a fitness function (genetic programming)
- Maximize the total reward/value function (reinforcement learning)
- Maximize information gain/minimize child node impurities (CART decision tree classification)
- Minimize a mean squared error cost (or loss) function (CART, decision tree regression, linear regression, adaptive linear neurons, ...)
- Maximize log-likelihood or minimize cross-entropy loss (or cost) function
- Minimize hinge loss (support vector machine)

Optimization Methods

- Noi

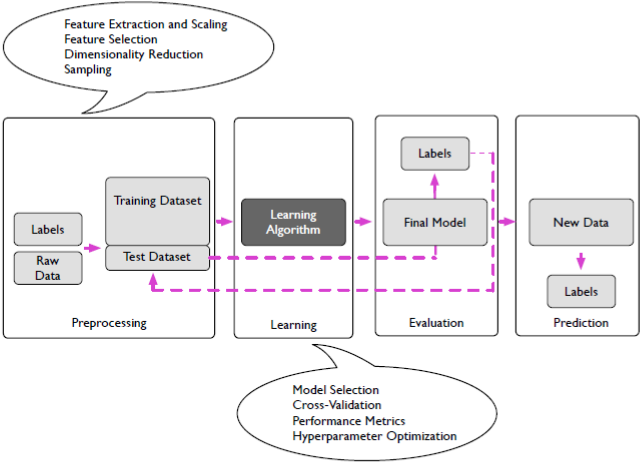

Figure: Learning Process [3].

Glossary

- **Training example**: A row in the table representing the dataset. Synonymous to an observation, training record, training instance, training sample (in some contexts, sample refers to a collection of training examples).
- **Training**: Model fitting, for parametric models similar to parameter estimation.
- **Feature,** x: A column in the table representing the dataset. Synonymous to predictor, variable, input, attribute, independent variable, and covariate.
- **Target**: Synonymous to outcome, output, response variable, dependent variable, (class) label, ground truth.
- **Predicted output,** $\hat{y}$: Use this to distinguish from targets; here, means output from the model.
- **Loss function**: Often used synonymously with cost function; sometimes also called error function. In some contexts the loss for a single data point, whereas the cost function refers to the overall (average or summed) loss over the entire dataset. Sometimes also called empirical risk.

Glossary

- **Hypothesis:**: A hypothesis is a certain function that we believe (or hope) is similar to the true function, the target function that we want to model. In context of *spam* classification, it would be a classification rule we came up with that allows us to separate spam from non-spam emails.
- **Model**: In the machine learning field, the terms *hypothesis* and *model* are often used interchangeably. In other sciences, they can have different meanings: A hypothesis could be the "educated guess" by the scientist, and the model would be the manifestation of this guess to test this hypothesis.
- **Learning algorithm**: Again, our goal is to find or approximate the target function, and the learning algorithm is a set of instructions that tries to model the target function using our training dataset. A learning algorithm comes with a hypothesis space, the set of possible hypothesis it explores to model the unknown target function by formulating the final hypothesis.

Glossary

- **Classifier**: A classifier is a special case of a hypothesis (nowadays, often learned by a machine learning algorithm). A classifier is a hypothesis or discrete-valued function that is used to assign (categorical) class labels to particular data points. In an email classification example, this classifier could be a hypothesis for labeling emails as spam or non-spam. Yet, a hypothesis must not necessarily be synonymous to the term *classifier*. In a different application, our hypothesis could be a function for mapping study time and educational backgrounds of students to their future, continuous-valued, SAT scores - a continuous target variable, suited for regression analysis.
- **Hyperparameters**: Hyperparameters are the *tuning parameters* of a machine learning algorithm - for example, the regularization strength of an L2 penalty in the mean squared error cost function of linear regression, or a value for setting the maximum depth of a decision tree. In contrast, model parameters are the parameters that a learning algorithm fits to the training data - the parameters of the model itself. For example, the weight coeffcients (or slope) of a linear regression line and its bias (or y-axis intercept) term are *model parameters*.

References

- Eager vs Lazy
- Single-Task vs Multi-Task
- Generative vs Discriminant
- Instance-based vs Model-Based
- Parametric vs Non-Parametric
- Batch vs Online

Thank You!

Any Question?