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 are using, is something similar to this:

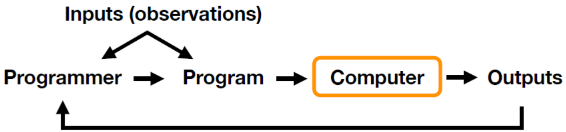


Figure: Classical Programming Paradigm source.

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 source.

Categories of Machine Learning

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

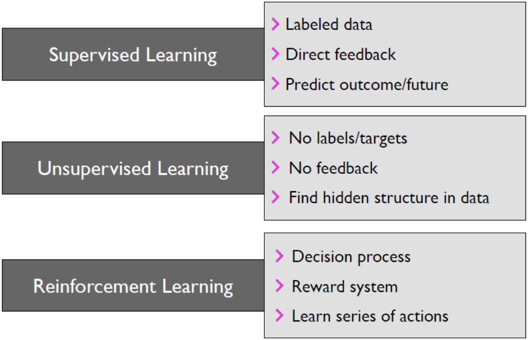


Figure: Categories of ML [1].

Introduction to Supervised Learning

- Supervised learning is the subcategory of machine learning that focuses on learning from labeled training data, our main goal in supervised learning is summerized in one of these categories:
  - **Classification**: predicting the discrete values such as male/female, etc.
  - **Regression**: predicting the continuous values such as price, age, etc.

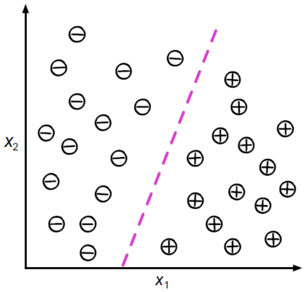


Figure: Illustration of Classifi cation Problem [1].

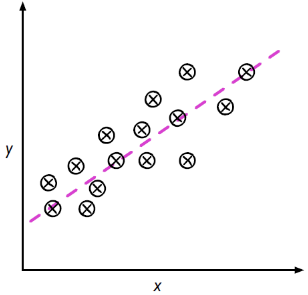


Figure: Illustration of Regression Model [1].

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**.

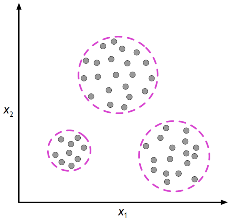


Figure: Illustration of Clustering [1].

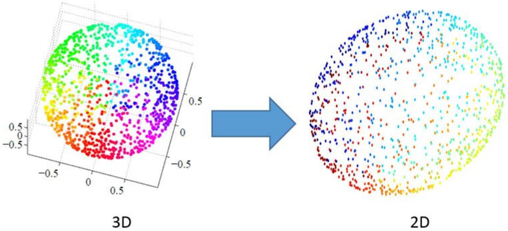


Figure: Illustration of Dimensionality Reduction source.

Reinforcement Learning

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

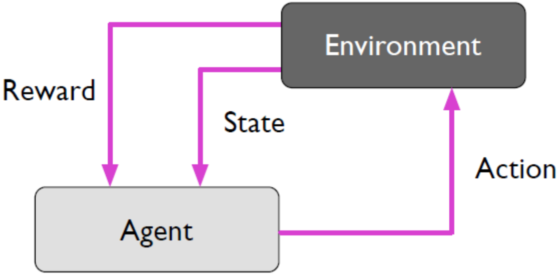
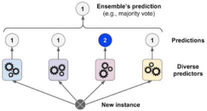


Figure: Illustration of Reinforcement Learning [1].

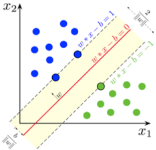
Classes of Machine Learning Algorithms

- 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)

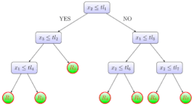
Classes of Machine Learning Algorithms



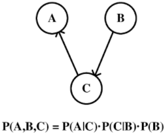
(a) Ensemble Learning source.



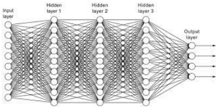
(b) Support Vector Machine source.



(c) Decision Tree source.

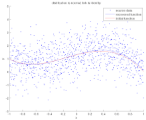


(d) Graphical Models source.



(e) Neural Networks source.

(f) K-Nearest Neighbors source.



(g) Generalized Linear Models source.

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.

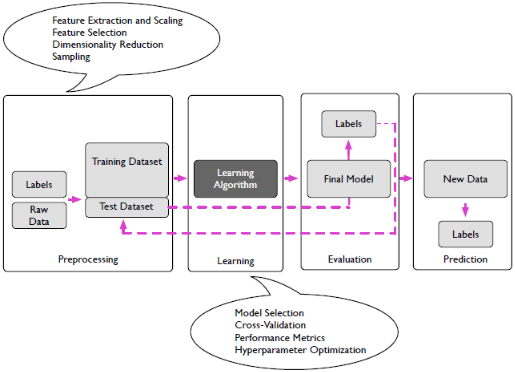


Figure: Learning Process [1].

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

- Combinatorial search, greedy search (e.g., decision trees over, not within nodes);
- Unconstrained convex optimization (e.g., logistic regression);
- Constrained convex optimization (e.g., SVM);
- Nonconvex optimization, here: using backpropagation, chain rule, reverse autodi. (e.g., neural networks).
- Constrained nonconvex optimization (semi-adversarial networks, not covered in this course)

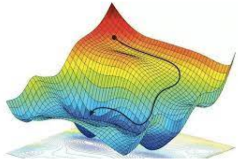


Figure: Gradient Descent source.

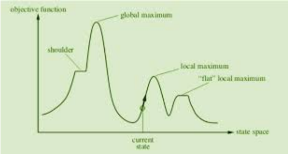


Figure: Hill Climbing source.

Evaluation

- There are several different evaluation metric to assess the performance of a model, some of them are:
  - Accuracy (1-Error)
  - ROC AUC
  - Precision
  - Recall
  - (Cross) Entropy
  - Likelihood
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - L-norms
  - ...

Glossary

- **Training example**: A row in the table representing the dataset.
- **Training**: Model fitting, for parametric models similar to parameter estimation.
- **Feature,** x: A column in the table representing the dataset.
- **Predicted output,** $\hat{y}$: Use this to distinguish from targets; here, means output from the model.
- **Loss function**: Often used synonymously with cost function.
- **Hypothesis:**: A hypothesis is a certain function that we believe (or hope) is similar to the true function.
- **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.
- **Hyperparameters**: Hyperparameters are the *tuning parameters* of a machine learning algorithm.
- **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.

References

- **[1]**. *Raschka, Sebastian, and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning With Python, Scikit-learn, and TensorFlow 2, 3rd Edition. 3rd ed., Packt Publishing, 2019.*

Thank You!

Any Question?