**Introduction to Machine Learning**

12/02/2020

This document has been developed by F. Atiyya Shaw and accompanies the R training script provided. Please direct questions/comments to [atiyya@gatech.edu](mailto:atiyya@gatech.edu).

**Table of Contents**

[1. Overview 1](#_Toc57466957)

[2. Supervised learning 2](#_Toc57466958)

[3. References 2](#_Toc57466959)

## Overview

* **Supervised learning**
  + In supervised machine learning, we have labeled data that can be used to train the machine learning algorithms (a.k.a. models or functions).
    - What does labeled data mean?
      * Each entry in the dataset has both characteristics (i.e., independent variables, features, attributes, x variables) and outcomes (i.e., dependent variables, outcomes, labels, y variables).
      * Examples:
        + We have tail length, weight, fur length, etc., and then the animal associated with each entry (e.g., dog or cat).
        + Iris dataset: species is the outcome variables or labels, and then we have attributes like petal width, sepal length, etc.
  + When the output is continuous, we use regression algorithms, and when the output is categorial, we use classification algorithms.
    - Most machine learning algorithms have both classification and regression types.
* **Unsupervised learning**
  + In unsupervised learning, we have a lack of labeled data. As such, unsupervised machine learning algorithms aim only to find patterns in the data and learn which characteristics are most helpful in delineating which datapoints are more or less similar to each other.
* **Semi-supervised learning**
  + In semi-supervised machine learning, the dataset has both labelled and unlabeled data. Often this can combine the best attributes of supervised and unsupervised learning to yield superior performance than doing one of these options independently.
* **Reinforcement learning** (sometimes classified as a type of supervised learning)
  + Reinforcement learning uses rewards or punishments for success or failure at a task. It is connected to control theory, Markov decision processes, and game theory.

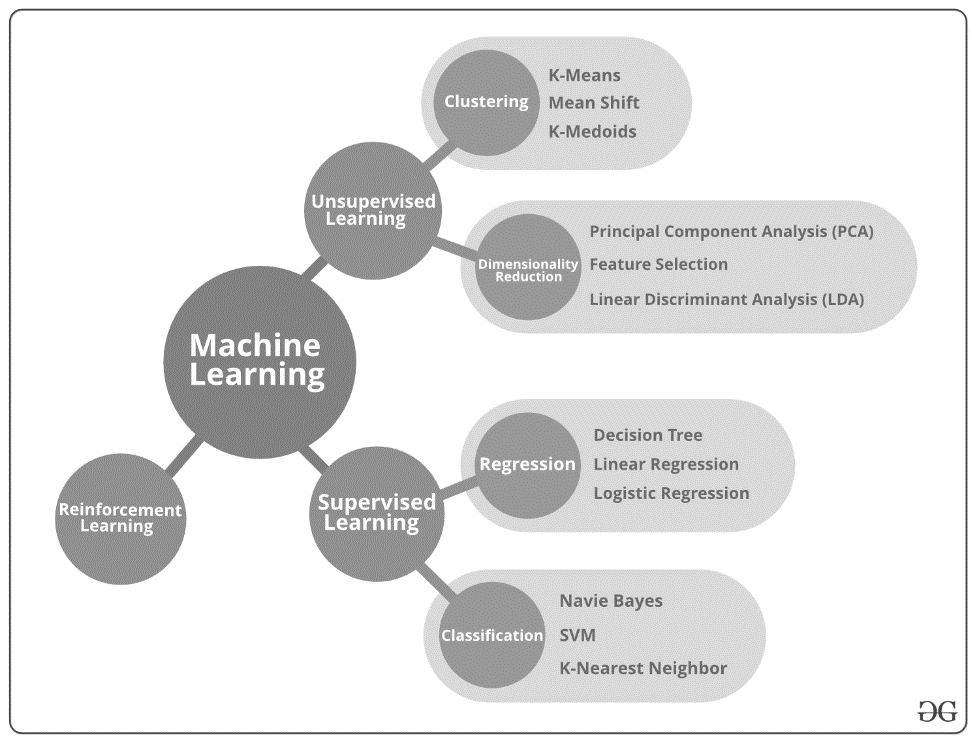


Figure 1. Types of Machine Learning   
Source: Top 10 Algorithms every Machine Learning Engineer should know - GeeksforGeeks

## Supervised learning

* We consider there to be three components of the supervised machine learning process that can be adjusted (see Figure 2). These are:
  + Features/attributes/independent variables
  + Algorithms/functions/models
  + Labels/outcome/dependent variables
* Supervised learning has two main steps/phases:
  + Training step
    - Trains the model/algorithm
  + Testing/validation step
    - Predict new outputs using the trained algorithm
  + We iterate between the training and testing phases, using the performance to adjust the three components mentioned previously.
* Here is the general sequence of steps that we recommend following as you begin your supervised machine learning journey:
  + #01. Define problem that we are trying to solve
  + # 02. Discuss, select, and prepare coding environment
  + # 03. Import/load + prepare data for use in coding environment
  + # 04. Explore + visualize + pre-process data
  + # 05. Run and evaluate algorithms
  + # 06. Tune/improve algorithms

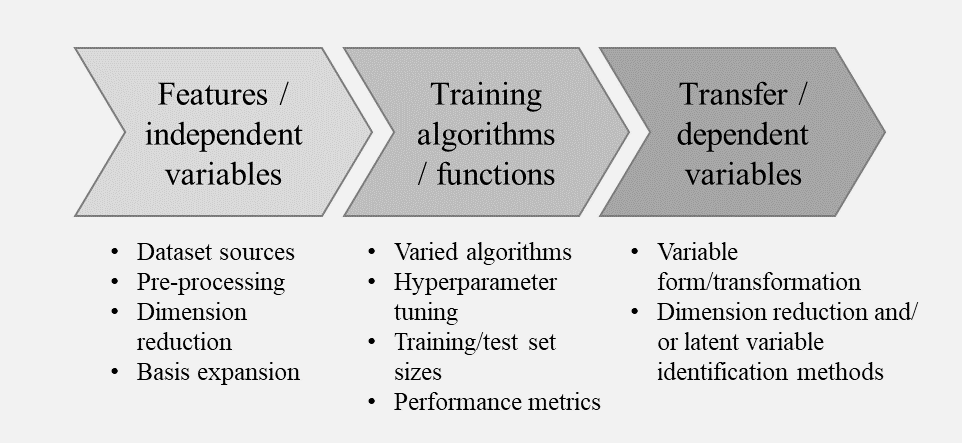


Figure 2. Overview of components and sample parameters in machine learning process

## For more information

* Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.
* National Academy of Sciences. (2018). *The Frontiers of Machine Learning: 2017 Raymond and Beverly Sackler U.S.-U.K. Scientific Forum*. Washington, DC: The National Academies Press.
* https://machinelearningmastery.com/machine-learning-in-r-step-by-step/
* https://www.datacamp.com/community/tutorials/machine-learning-in-r
* https://lgatto.github.io/IntroMachineLearningWithR/an-introduction-to-machine-learning-with-r.html
* [Chapter 5 Supervised Learning | An Introduction to Machine Learning with R (lgatto.github.io)](https://lgatto.github.io/IntroMachineLearningWithR/supervised-learning.html#introduction-1)
* <https://developers.google.com/machine-learning/crash-course/>
* <https://playground.tensorflow.org>