CHAPTER 1: MACHINE LEARNING BASICS

Theory

This chapter explains the basic concepts of machine learning and deep learning such as supervised and unsupervised learning, semi-supervised learning, and reinforcement learning, and linear regression, binary and multi-label classification and so on.

Basic concepts

- What is Machine Learning?
- What is learning?
    - Supervised
    - Unsupervised
    - Semi-supervised
    - Reinforcement
- What is regression?
- What is classification?

Machine Learning

Today machine learning and deep learning is becoming a widespread area of artificial intelligence, including spam filtering, auto driving, face recognition, optical character recognition (OCR), and more.

So what is machine learning?

Machine learning is a field of artificial intelligence that refers to the areas in which algorithms and techniques are developed that allow computers to learn.

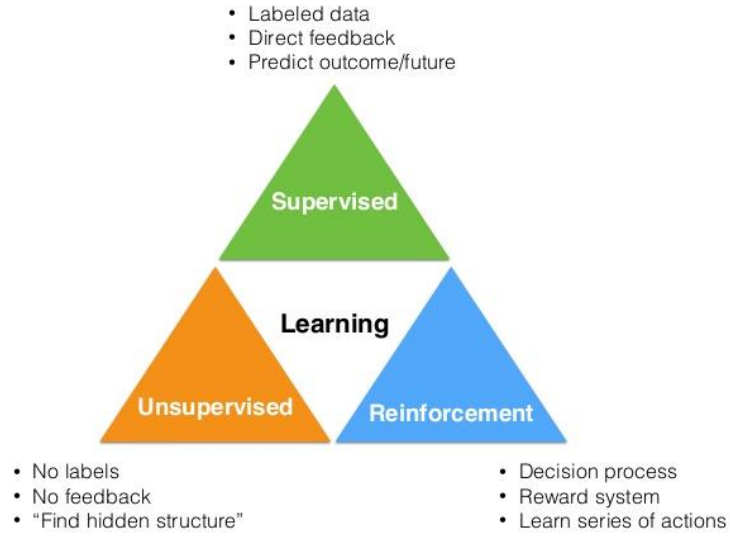
For example, machine learning can be trained to distinguish whether emails received are spam or not.

The key to machine learning lies in representation and generalization. Representation is an evaluation of data, and generalization is processing of data that is not yet known. This is also the field of computational learning

theory. There are various applications of machine learning. Character recognition is the best known example of this.

In the field of data analysis, machine learning is a method used to devise complex models and algorithms that help predictions.

For commercial purposes, this is known as predictive analytics. This analytical model enables researchers, data scientists, engineers, and analysts to "produce reliable and repeatable decisions and results," and "hidden insights" through historical relationships and trends in the data.



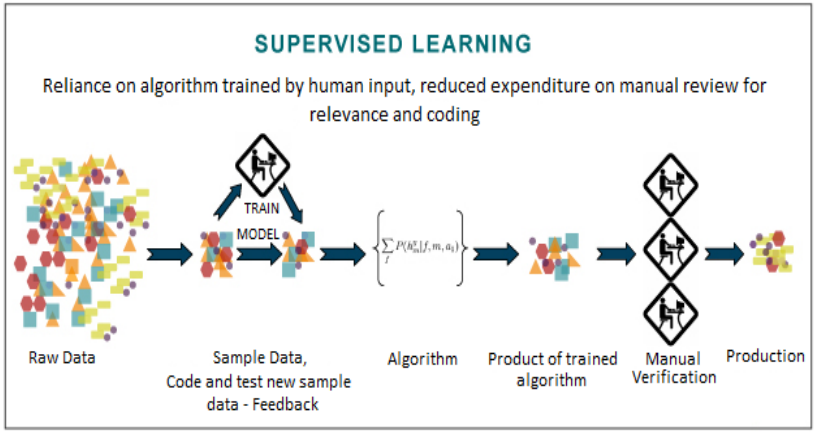
Machine Learning could be divided into supervised, unsupervised and reinforcement learning. Sometimes semi-supervised machine learning is also considered.

Supervised Learning

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future data.

Starting from the analysis of a known training dataset, the learning algorithm produces an inferred mathematical and logical function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training.

The machine learning algorithm can also compare its output with the correct data, intended output and find errors in order to modify the model accordingly.

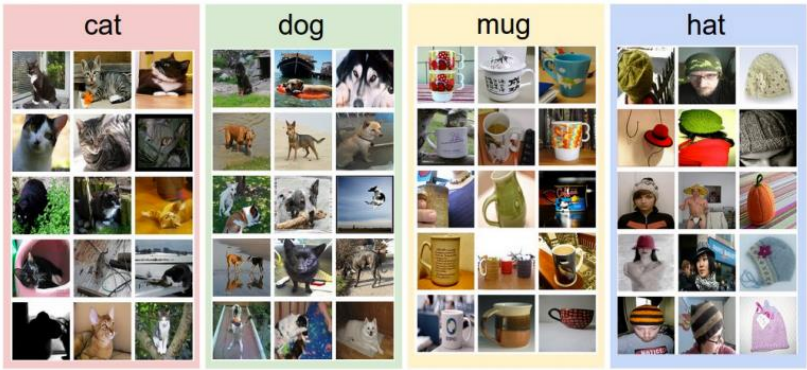


When a person gives data to the computer with a label (y) for each input (x) as a teacher, the computer learns it. It has the advantage of using highly accurate data because it is directly intervened by people.

Instead, there is a labor cost problem because people have to label them directly, and there is a problem that the amount of data that can be obtained is small.

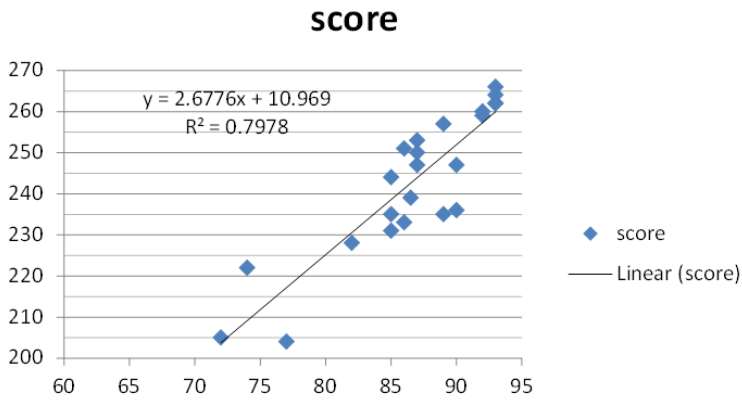
Supervised learning is the most common problem type in Machine Learning.

- Image labelling: learning from tagged images
  The supervised machine learning algorithm could be used for image and data classification. There are lots of image classification models such as Inception V2, Inception V3, Alphanet, Resnet, Mobilenet and so on and its accuracy is over 98% now.



- Email spam filter: learning from labeled (spam or ham) email
  One of simple example of supervised machine learning is spam mail filtering. As you can see, using spam filtering feature, you can distinguish the unwanted spam mails and block them from your inbox.

- Predicting exam score: learning from previous exam score and time spent
  If we have the data set pare of study duration and score, we can predict how many hours will need to meet the high score.



Types of Supervised Learning: Linear Regression

Linear regression is mainly used to solve the problem of supervised learning. Linear regression, logistic regression, and artificial neural networks are also examples.

Linear regression is used to solve the regression problem of supervised learning (new value prediction), and logistic regression and artificial neural networks are used to solve the classification problem of supervised learning.

Linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. If the label y is a real number in machine learning, it is called a regression problem.

Usually when we draw a graph in Excel, it is very common problem. Use the regression function when you want to sprinkle data straight and want to draw a straight line or quadratic curve that best describes it.

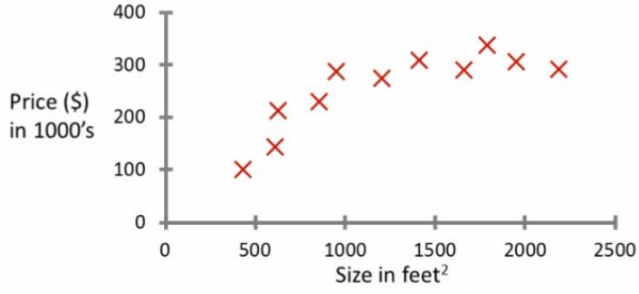
If you think carefully, the data consists of a pair of input (x) and real label (y), and matching y to a new arbitrary input (x) is a straight line or curve. Among the regression analysis methods of statistics, linear regression is a representative example.

Examples:

- Predicting final exam score based on time spent

x (hours)	y (score)
10	90
9	80
3	50
2	30

- Housing price prediction



Types of Supervised Learning: Binary classification

In machine learning, binary or binomial classification is the task of classifying the elements of a given set into two classes (predicting which class each one belongs to) on the basis of a classification rule

Binary classification is dichotomization applied to practical purposes, and in many practical binary classification problems, the two groups are not symmetric – rather than overall accuracy, the relative proportion of different types of errors is of interest.

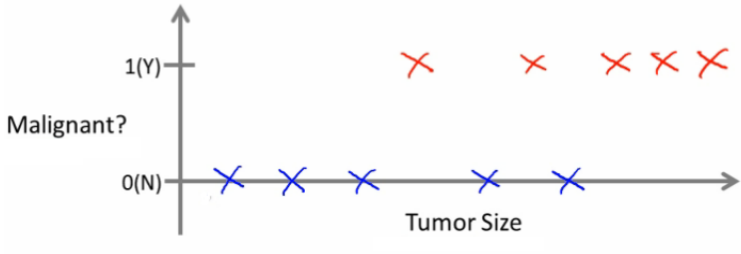
For example, in medical testing, a false positive (detecting a disease when it is not present) is considered differently from a false negative (not detecting a disease when it is present)

Examples:

- Pass/non-pass based on time spent

x (hours)	y (pass/fail)
10	P
9	P
3	F
2	F

- Decide breast cancer as malignant or benign based on tumor size



Types of Supervised Learning: Multi-label classification

Multi-label classification which is strongly related problem of multi-output classification are variants of the classification problem where multiple labels may be assigned to each instance. Multi-label classification is a generalization of multiclass classification.

Formally, multi-label classification is the problem of finding a model that maps inputs x to binary vectors y (assigning a value of 0 or 1 for each element (label) in y).

Examples:

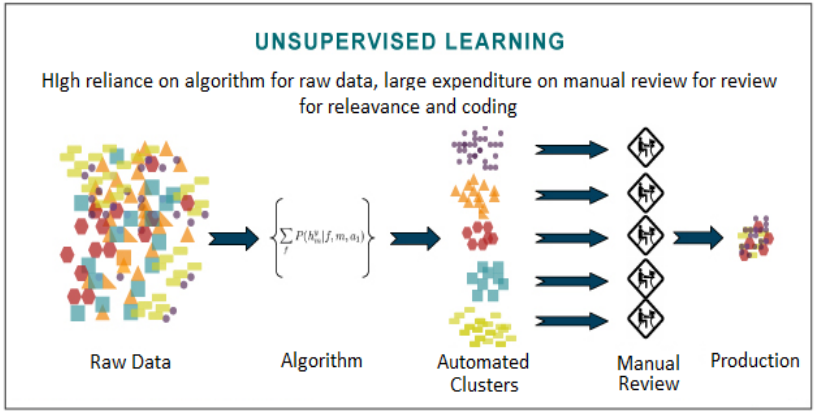
- Letter grade (A, B, C, E and F) based on time spent

x (hours)	y (grade)
10	A
9	B
3	D
2	F

Unsupervised Learning

In contrast with supervised machine learning, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data.

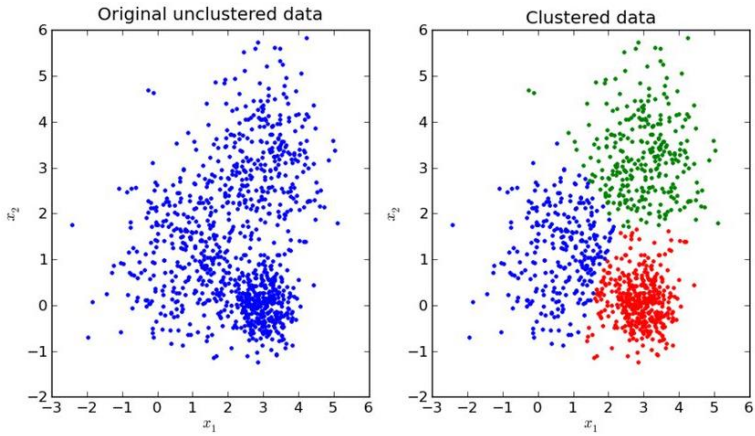
The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.



Since the examples given to the learner are unlabeled, there is no evaluation of the accuracy of the structure that is output by the relevant algorithm—which is one way of distinguishing unsupervised learning from supervised learning and reinforcement learning.

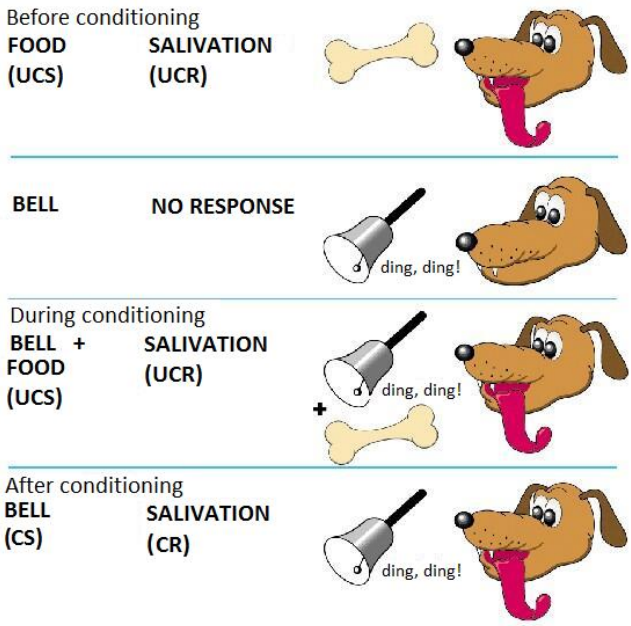
Approaches to unsupervised learning include:

- Clustering
    - K-means
    - Mixture models
    - Hierarchical clustering



- Anomaly detection

- Neural Networks
    - Hebbian Learning
    - Generative Adversarial Networks



Semi-supervised Learning

Semi-supervised machine learning fall somewhere in between supervised and unsupervised learning, since they use labeled and unlabeled data for training together – typically a small amount of labeled data and a large amount of unlabeled data.

The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

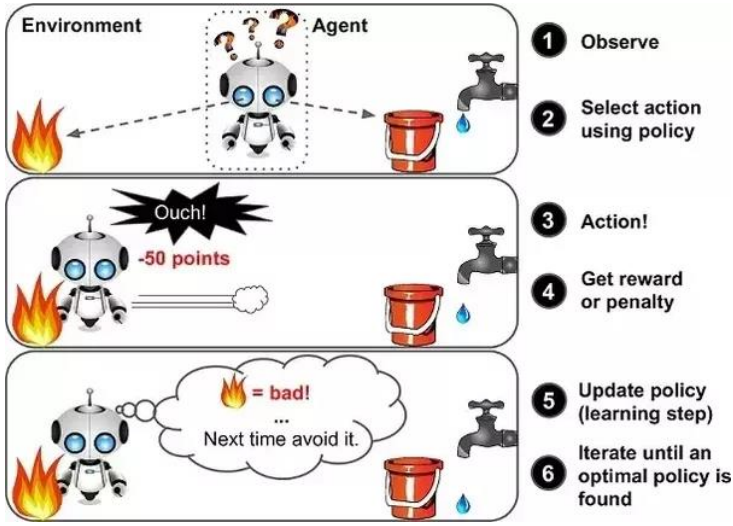
Reinforcement Learning

Reinforcement learning is to learn what action is best to take in the current state.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or

rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning.

This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

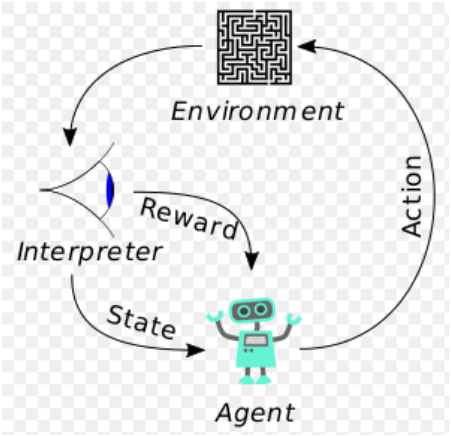


The problem, due to its generality, is studied in many other disciplines, such as control theory, game theory, operations research, information theory, simulation-based optimization, swarm intelligence, multi-agent systems, statistics and genetic algorithms and so on.

Reinforcement learning differs from standard supervised learning in that correct input / output pairs are never provided nor sub-optimal actions explicitly modified.

Instead the focus is on real-time performance, which involves finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge). The exploration vs exploitation trade-off in reinforcement learning has been most thoroughly studied through the multi-armed bandit problem and in finite Markov decision process (MDP).

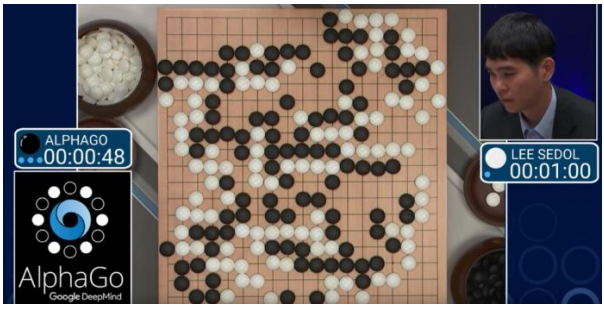
The typical framing of a Reinforcement Learning (RL) scenario is as follows.



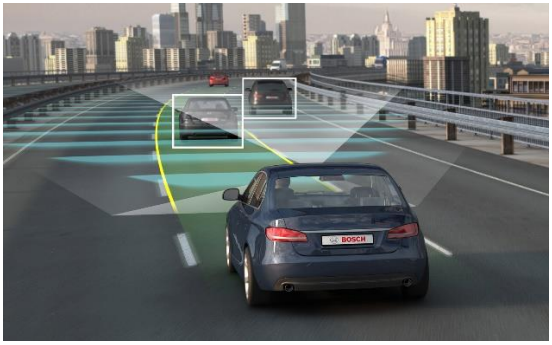
An agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

Reinforcement learning could be used in various area including:

- Game tactics such as chess and go



- Autonomous vehicle



- Decision making

AIM

Our goals are to understand and utilize the following contents through this learning.

- Basic understanding of machine learning
    - Supervised learning
    - Unsupervised learning
    - Semi-supervised learning
    - Reinforcement learning
- Solve your problems using machine learning tools
    - Tensorflow
    - Python

LAB EXERCISE 1

“There are no activities required for this lab”

SUMMARY

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.

Supervised learning is the learning with labeled examples – training set.

Unsupervised machine learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data.

Reinforcement learning(RL) is an area of machine learning inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

REFERENCES

- https://en.wikipedia.org/wiki/machine_learning
- http://www.holehouse.org/mlclass
- https://en.wikipedia.org/wiki/regression

INDEX