

A Review: MACHINE LEARNING ALGORITHM

SUDHIR PRAKASH INGLE

VINIT PRAVIN PHARKADE

Keraleeya Samajam (Regd.) Dombivli's Model College

Abstract:

In this following paper, various known machine learning algorithms have been discussed. These algorithms are used for various purposes such as data mining, predictive analytics, image processing, etc. to name a few. The important advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically or manually.

Keywords:

Machine learning, algorithm

1 Introduction:

Machine learning is used to teach machines how to handle or deal with the data more efficiently. Sometimes after seeing the data, we cannot extract information from the data. In such case, we use machine learning. With the abundance of datasets available, the demand for that the machine learning is always in rise. Many industries from medicine to military apply machine learning to extract the relevant information.

The main purpose of the machine learning is to learn from data. Many studies have been done already on how to make machines learn by themselves. Many mathematicians and programmers apply

several approaches to find out the solution of this particular problem.

2 Types of Learning:

A. Supervised Learning The supervised machine learning algorithms are those which needs external assistance. The input dataset is divided into train data and test dataset. The train dataset has output variable which is needs to be predicted or classified. All algorithms learn some different kind of patterns from the training dataset and apply them to test dataset for the prediction or classification.

1) Decision Tree: Decision trees are those type of trees in which groups attributes by sorting them that based on their values. Decision tree is used for classification purpose mainly. Each tree consists of nodes and various branches. Each nodes represents the attributes in group that is to be classified and each branch is represents a value that the node can take.

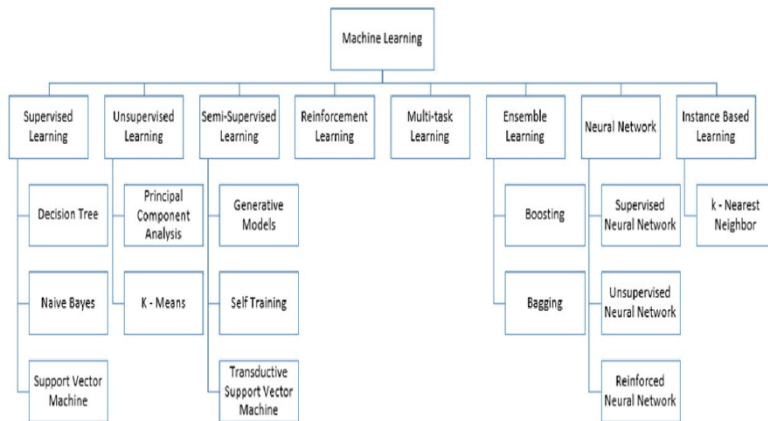


Fig. 1. Types of Learning [2] [3]

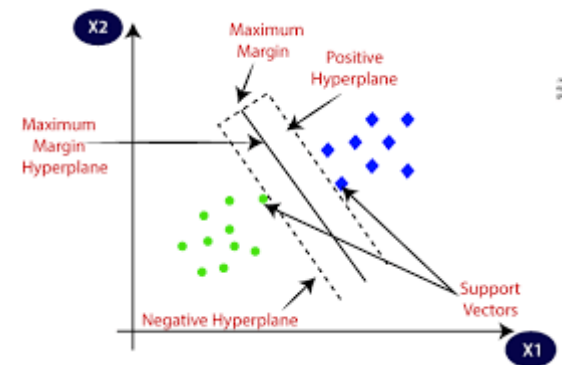
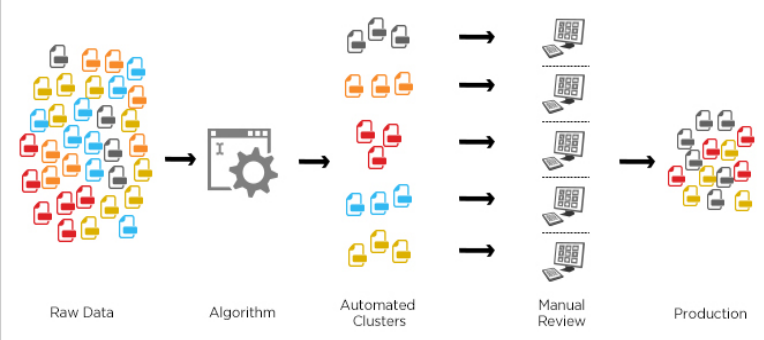


Fig. 2. Working of Support Vector Machine

Unsupervised Learning



B. Unsupervised Learning

The unsupervised learning algorithms are need to learns few features from data. When the new data is introduced, it using the previously learned features to recognize that class of the data. It is mainly used for the clustering and feature reduction. An example of workflow of unsupervised learning is given in fig.3.

Fig. 3. Example of Unsupervised Learning

2) Naive Bayes: Naive Bayes mainly targets text classification industry. It is mainly used for the clustering and classification purposes. The underlying architecture of Naive Bayes is depends on the conditional probability. It creates the trees based on their probability of happening. These trees are also as called as the Bayesian Network.

3) Support Vector Machine: Another and most widely used state of the art machine learning technique is Support Vector Machine which is SVM. It is mainly used for the classification. SVM also works on the principle of margin calculation. It is basically, draw margins between the classes. The margins are drawn in such a fashion that distance between the margin and that of the classes is maximum and hence, minimizing all the classification error.

1)K-Means Clustering: Clustering or the grouping is a type of unsupervised learning technique that when initiates, creates the groups automatically. The items which possesses similar characteristics are put into the same cluster. This algorithm is called as k-means because it creating k distinct clusters. The mean of values in a particular cluster is that the center of that cluster. A clustered data is represented in Fig.3. The algorithm for k-means is given in Fig.4

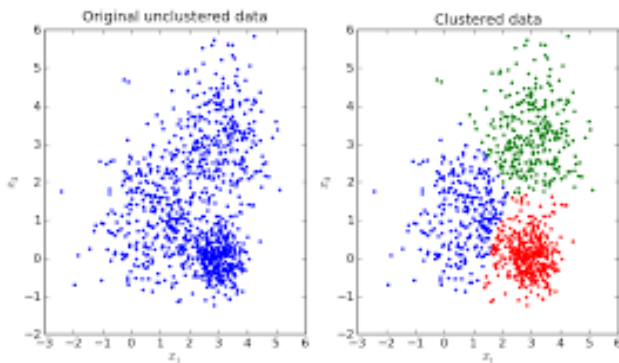
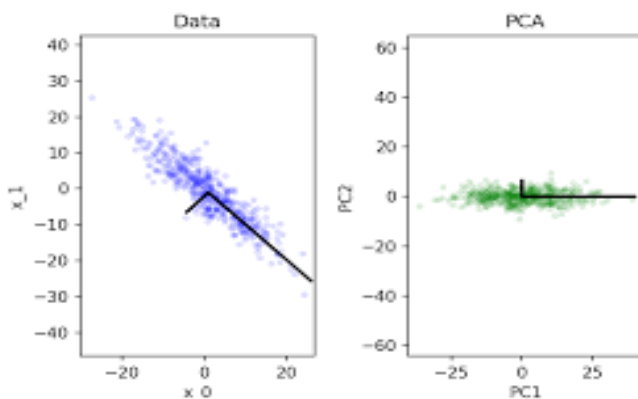


Fig. 4. K-Means Clustering

2) Principal Component Analysis In Principal Component Analysis (PCA), the dimension of data is reduced to make the computations faster and more easier. To understand how PCA actually works, let's take an example of the 2D data. When data is being plot in the graph, it will take up to two axes. PCA is applied on data, the data then will be converted 1D. This is explained in Fig. 5.

Fig. 5. Visualization of data before and after applying PCA



C. Semi - Supervised Learning Semi-supervised learning algorithms is a technique in which it combines the power of both supervised as well as unsupervised learning. It can be fruit-full in those areas of the machine learning and data mining where unlabeled data is already present and getting the labeled data is a slow process.

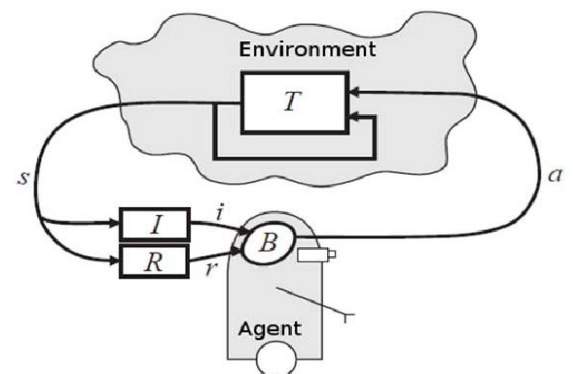
There are many categories of semi-supervised learning, are as follows;

- 1) Generative Models:
- 2) Self-Training
- 3) Transductive SVM

D. Reinforcement Learning

Reinforcement learning is a type of learning in which it makes decisions based on which actions to take such that the outcome is more positive. The learner has no knowledge which actions to take until it is been given a particular situation. The action which is taken by learner may affect situations and that their actions in the future. Reinforcement learning solely depends on main two criteria which is trial and error search and delayed outcome. The general model for reinforcement learning is depicted in Fig. 6.

Fig. 6. The Reinforcement Learning Model



In the given figure, the agent receives an input I , current state is S , state transition R and input function I from the given environment. Based on these particular inputs, the agent generates behavior B and takes an action A which generates proper outcome.

E. Multitask Learning

Multitask learning has a very simple goal of helping other learners to perform better. When the multitask learning algorithms are applied on task, it remembers the procedure that how it solved the problem or how it reaches to particular conclusion. The algorithm then uses these following steps to find out the solution of other similar problem or task. This helping of one algorithm to the another can also be known as inductive transfer mechanism. If learners share their personal experience with each other, the learners can also learn concurrently rather than individually and it can be much faster.

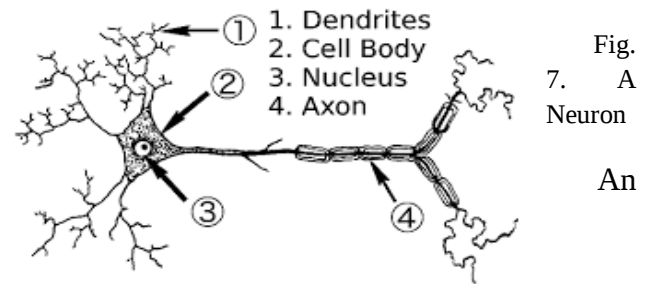
F. Ensemble Learning

When various individual learners are combined to form single learner then that one particular type of learning is called as ensemble learning. The individual learner may be neural network, decision tree, Naive Bayes, etc. Ensemble learning is very hot topic since 1990s. It has been observed that, a collection of all learners is almost always better at doing a very particular job rather than individual learners. Two popular Ensemble learning techniques are as follows;

- 1) Boosting
- 2) Bagging

G. Neural Network Learning

The Neural Network (artificial neural network or ANN) is derived from biological concept of neurons. Neuron is a cell like structure in a brain. To understand more neural network, one must understand how a neuron actually works. A neuron has mainly four parts. They are Dendrites, Nucleus, Soma and Axon.



artificial neural network behaves the same way. It works on three layers. The input layer takes input (much like dendrites). The hidden layer processes the input like soma and axon. Finally, the output layer sends all the calculated output like dendrite terminals. There are basically 3 types of artificial neural network: Supervised neural network, Unsupervised neural network, Reinforced neural network.

H. Instance-Based Learning

In Instance-Based learning, the learner learns very particular type of pattern. It tries to apply same pattern to newly fed data. Hence the name Instance-Based. It is a type of lazy learner which waits for its test data to arrive and then it acts on it together with training data. The complexity of learning algorithm increases with the size of the given data.

1) K-Nearest Neighbor:

In k-nearest neighbor or KNN, the training data which is well-labeled is fed into learner. When the test data is introduced to the learner, it compares both of the data. k most correlated data is taken from the training set. The majority of k is taken that which serves as the new class for test data. The pseudo code for KNN is given in Fig. 8

Let $W = \{x_1, x_2, \dots, x_n\}$ be a set of n labeled samples. The algorithm is as follows:

```

BEGIN
  Input  $y$ , of unknown classification.
  Set  $K, 1 \leq K \leq n$ .
  Initialize  $i = 1$ .
  DO UNTIL ( $K$ -nearest neighbors found)
    Compute distance from  $y$  to  $x_i$ .
    IF ( $i \leq K$ ) THEN
      Include  $x_i$  in the set of  $K$ -nearest neighbors
    ELSE IF ( $x_i$  is closer to  $y$  than any previous nearest neighbor) THEN
      Delete farthest in the set of  $K$ -nearest neighbors
      Include  $x_i$  in the set of  $K$ -nearest neighbors.
    END IF
    Increment  $i$ .
  END DO UNTIL
  Determine the majority class represented in the set of  $K$ -nearest neighbors.
  IF (a tie exists) THEN
    Compute sum of distances of neighbors in each class which tied.
    IF (no tie occurs) THEN
      Classify  $y$  in the class of minimum sum
    ELSE
      Classify  $y$  in the class of last minimum found.
    END IF
  ELSE
    Classify  $y$  in the majority class.
  END IF
END

```

III. CONCLUSION

This paper surveys various of machine learning algorithms. Today each and every person is already using machine learning knowingly or unknowingly. From getting a recommended product in to the online shopping to updating photos on social networking sites. This paper gives an perfect introduction to most of the machine learning algorithms.

IV. REFERENCES

- [1] "Handwriting recognition", https://en.wikipedia.org/wiki/Handwriting_recognition
- [2] "What can a digit recognizer be used for?", <https://www.quora.com/What-can-a-digit-recognizer-beused-for>

- [3] "Handwritten Digit Recognition Using Deep Learning", Anuj Dutt and Aashi Dutt.

- [4] "Basic Overview of Convolutional Neural Network" <https://medium.com/dataseries/basic-overview-ofconvolutional-neural-network-cnn-4fcc7dbb4f17>

- [5] <https://www.researchgate.net/publication/343010636>

Handwritten Digit Recognition using Machine and Deep Learning Algorithms Article in International Journal of Computer Applications · July 2020

- [6] <https://arxiv.org/pdf/2106.12614>

- [7] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0

- [8] M. Welling, "A First Encounter with Machine Learning"

- [9] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-96174-2

- [10] S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249- [11] L. Rokach, O. Maimon, "Top – Down Induction of Decision Trees Classifiers – A Survey", IEEE Transactions on Systems,

- [12] D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation"

- [13] https://webdocs.cs.ualberta.ca/~greiner/C651/Homework2_Fall2008.html

[14] Ayon Dey / (IJCSIT) International Journal of
Computer Science and Information Technologies,
Vol. 7 (3) , 2016, 1174-1179