International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 6 | ISSN : 2456-3307

A Review on Hypothesis Representation in Machine Learning

Helga V Lobo

Department of Computer Science Engineering, Centre for P.G Studies, VTU, Belagavi, Karnataka, India

ABSTRACT

Search through a large hypothesis space is the basis for Machine Learning. The process of learning is said to be complete, when the search arrives at a hypothesis, whichprovides best approximation to the target function. Such hypothesis not only fits the available training examples, but also extends beyond it. This review presents different approaches available for representing a hypothesis and the search of hypothesis space to arrive at the best hypothesis.
Keywords: Machine Learning, Hypothesis Representation.

I. INTRODUCTION

The field of machine learning tries to add the learning ability toa computer. It is an attempt to mimic the learning ability of human brain. A machine is said to learn from experience, with respect to some class of task, if its performancein doing the said task improves with experience [1]. The process of learningis completed when the machine is successful in arriving at a hypothesis, which best approximates the target function.The 'target function', is a function which gives right classification for each input ,with respect to some class of task. A hypothesis is like the experience gained, which helps the machine to correctly classify unseen instances or the input beyond the training set, which belongs to the same class, for which it was trained. A hypothesis can be represented in different ways. For each possible hypothesis representation, there can be number of possible hypothesis(h), which forms the hypothesis space(H).This paper provides an overview of different ways in which hypothesis can be represented and the method of search used for each of these representations, to arrive at the best hypothesis from the possible hypothesis.

II. Hypothesis Representations and Hypothesis Space Search

A. Functions

In this simplest approach, the machine is said to complete the learning process by deriving a function through the training examples given to it. Here the hypothesis is a *"function"*. This function not onlygives right classification for the training data set, but also beyond it. This function may be asimple linear, quadratic or complex. Following is an example of simple linear function, where x_i represent an attribute, and $w_i$represents weight given to the attribute x_i .

$$\mathbf{h(b)=w_0+w_1*x_1+w_2*x_2+w_3*x_3+w_4*x_4+.......+w_n*x_n}$$

b-is the input given to the hypothesis or the function.

h(b)- is the output value or the classification.

h(b) may also be replaced by $\hat{V}$(b), where $\hat{V}$(b) is the best approximation of the actual target function. The hypothesis *h,* can actually be described by the learned weights $w_o.....w_n$.In this approach the set of possible weight vectors forms the hypothesis space H[1].

$$H = \{\overrightarrow{w} | \overrightarrow{w} \in R^{n+1}\}$$

R-is a vector of real numbers of length (n+1)

Here the search for the best hypothesis corresponds to, the search for the appropriate weights for a said function. The weights chosen must minimize the error between the value of the actual target function and the hypothesis output. Several algorithms exist to fit appropriate weightsto the function. Least Mean Squares or LMS training rule [1] is one of the algorithm that tunes the weights for the given function, to find the best possible weight vector or the best hypothesis.

An example use of this hypothesis representation is choosing the right move in checkers playing program. Each attribute represents a particular scenario in the game and the value of weights represent the importance given to such scenario.

B. Constraint vector

In constraint vector approach, a hypothesis is represented by vector of constraints on the attributes.The constraint vector looks as follows,

<c1, c2, c3...cn>

Here $c_i$is aconstraint on attribute *i*. Constraint can take up the values **∅ ,?,** or any single value that the attribute must take.The most general hypothesis is< ?,?,?,?,?> and the most specific hypothesis takes the form<**∅,∅,∅,∅>.**The most general hypothesis will classify any given input as positive, and the specific hypothesis will classifyany given inputas negative. A hypothesis spaceis formed by various combination of conjunction of constraints.This kind of hypothesis representation is used in concept learning[2].

In this approach of hypothesis representation, algorithms like List-Then–Eliminate, Candidate Elimination algorithm are used to arrive at a best hypothesis. This kind of hypothesisrepresentation is useful in inferring boolean valued functions [1]. An example may be, classifying a day to be suitable for sport or not based on the attributes like, air temperature, humidity, wind etc.

C. Trees

The tree representation of hypothesis is used in decision tree learning method [3]. The process of learning is said to be complete when a tree is derived successfully from the training data given.The derived tree approximates a discrete valued target function [1]. The tree classifies the unseen instance by sorting the instance from root to some leaf node. Each node of the tree corresponds to test on some attribute and the values on braches correspond to, values that can be taken by the attribute.In this form of hypothesis representation, the hypothesis space is formed by all possible trees,that can be constructed using the set of attributes. The process of learning is said to be complete when the best tree is derived from the training data, which perfectly classifies the unseen instances.Figure 1. shows a learned tree.

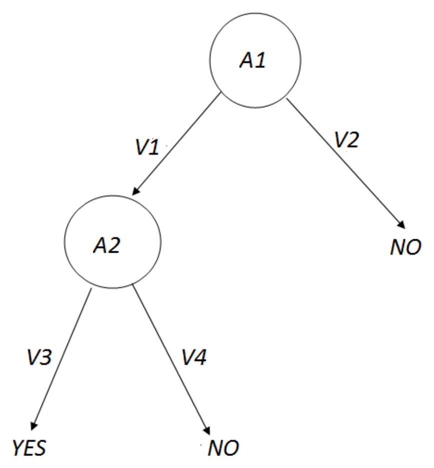

Figure 1. An example of learned tree.

A1, A2 correspond to the test attribute. V1, V2 and V3, V4 are the values that can be taken by the attributes A1 and A2 respectively. The values of the leaf noderepresent the classification made by learned tree.

Search for the best hypothesis or the tree is done by algorithms like ID3 and C4.5[1] .The search involves the choice of right attribute at each level.

D. ANN

The hypothesis representation using artificial neural networks(ANN)[6], is used to learn real-valued, discrete valued, and vector valued target functions[1]. Learned hypothesis is a ANN. ANN is inspired by biological learning system, ie human brain. ANN is a network, whichdenselyinterconnectsseveral simple units, just like the human brain, which contains denselyinterconnected set of neurons. Each unit of ANN takes number of real valued inputs and produces a single real valued output, which may be input to many other units [1].Following figure shows a simple neural network. It has an input layer, hidden layer and an output layer. Each layer is formed by several simple units. These units may be implemented as simple perceptron unit or sigmoid unit [1].

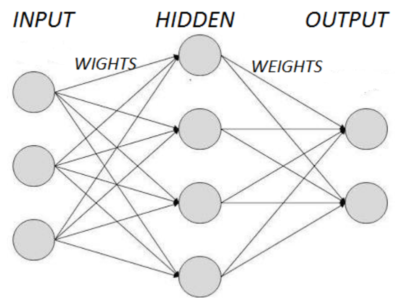

Figure 2. Simple neural network

The learning is complete when a ANN is successfully derived from the training data. The units are interconnected in layers. Learning involves search for the right weights at each layer. This may be accomplished using algorithms like perceptron training rule and delta rule.ANNs can be cyclic or acyclic, directed or undirected graph. Hypothesis representation using ANN is best suited when training data is noisy and complex, such as input of camera and microphone.

E. Bit Strings

Hypothesis representation using bit string is used in, learning using Genetic Algorithms[4].This form of hypothesis representation can be easily manipulated by genetic operators. The bit strings are capable of representing very complex hypothesis. They are capable of easily representing sets of if-then rules.This is possible by right choice of substring for each rule precondition and postcondition. As an example, the hypothesis consisting of two rules

IFa1=T Λ a2=F THEN c=T; IF a2=T THEN c=F, can be represented by the string,

a1	a2	c	a1	a2	c
10	01	1	11	10	0

Hypothesis space search is performed using Genetic algorithms [5]. Initially a population of hypothesis is maintained. The operators like crossover and mutation are used to act on the fittest hypothesis, which are probabilistically selected, to form the new generation of hypothesis. The process is repeated until sufficiently fit hypothesis are discovered. This kind of hypothesis representation, along with genetic algorithms eliminate the problem of local minima[1].

F. Computer Programs

Computer programs themselves can be used to represent hypothesis [1][4]. These programs can be represented by trees corresponding to parse tree of the program. Figure shows an example tree for the function cos(x)+$\sqrt{x^2+y}$. The node in the tree corresponds to function call and the descendents of the node represent arguments to the function.

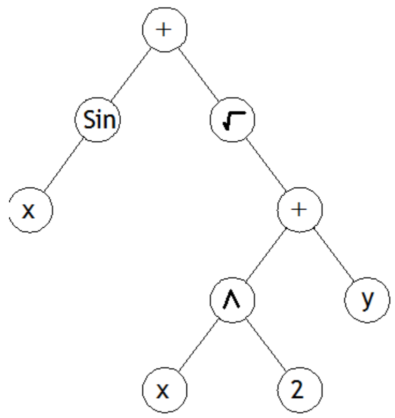

Figure 3. A program tree

The best hypothesis is derived by iterative selection, crossover and mutation of individual hypothesis in the population. The search terminates when the fitness of the hypothesis reaches a predetermined threshold. The fitness is measured by executing the program on a set of training data.

III. CONCLUSION

This paper has presented, existing modelsfor representing hypothesis in machine learning. Selection of a particular hypothesis representation depends on, the kind of machine learning task. A suitable hypothesis representation may be chosen for the given task, based on the training time available and speed of evaluation required by the hypothesis.

IV. REFERENCES

[1]. Tom M Mitchell, "Machine Learning",ISBN-13:978-0070428072.

[2]. Mitchell T M(1979).Version Spaces: An approach to concept learning, (Ph.DDissertation).Electrical Engineering Dept., Standford, USA

[3]. Fayyad, U.M.(1991). On the induction of decision trees for multiple concept learning, (Ph.D Dissertation).EECSDept, University of Michigan.

[4]. DeJong,K.A.(1975).An analysis of behavior of a class of genetic adaptive systems((Ph.D Dissertation). University of Michigan.

[5]. Genetic algorithms in search, optimization , and machine learning. MA: Addison-Wesley.

[6]. Waibel,A.,Hanazawa,T.,Hinton.,Shikano,K.,&Lang,K .(1989).Phoneme recognition using time delay neural networks. IEEE Transactions on Acoustics,Speech and Signal Processing,328-339

[7]. Sebag, M(1994).Using Constraints to build version spaces .Proceedings of the 1994 European Conference on Machine learning. Springer-Verlag.