Basics of Machine Learning

Jimut Bahan Pal

Department
of Computer Science
B.Sc. 2nd year

Machine learning is the method of teaching computers to make and improve predictions or behaviours based on some data. The data could be readings from a robot's sensors as it learns to walk, or the correct output of a program for certain input. Suppose there is a problem to recognize a 3D object from a novel viewpoint in new lighting conditions in a clustered scene. So in this case we don't know the program as we don't know how it is done in our brain, else if we even knew the program it would be horrendously complicated. Another example is to compute the probability that a credit card transaction is fraudulent. In this case there may not be any rule that is simple and reliable; we need to combine very large number of weak rules. Fraud is a moving target; the program needs to be changing. Instead of writing a program by hand for each specific task, we collect a large number of examples that specifies the correct output for a given input. Now, the machine learning algorithm takes the input and produces a program, which might look different from a handwritten program in the case that it may contain millions of numbers. If the program is made right, it works on new problems as perfectly as it works for the old problems we trained on. If the data changes the program can change too by training on the new data. So the reason for machine learning is massive amounts of computations are cheaper than paying someone to write a task specific program.

Some of the examples of task best solved by machine learning is recognizing patterns, for example objects in real scene, this is done by layering of the image and dividing the image into layers and then throwing the layers into a classifier, then recognizing facial identities and facial expressions and even recognizing spoken words. It could be even read to recognize anomalies, like unusual sequences in credit card transactions, unusual patterns of sensor reading in a nuclear power plant. It can even predict future stock prices or currency exchange rates or like which movie will a person like? Nowadays a lot of genetics is done on fruit flies, they are convenient because they breed fast and we already know a lot about them. The MNIST (Modified National Institute of Standards and Technology) database of handwritten digits (Fig. 1) is the machine learning equivalent of fruit flies as they are publicly available and we can learn them in a moderate sized neural net.

Fig. 1 (MNIST database of handwritten digits)

There are even more databases beyond MNIST , for example ImageNet , where there are 1000 different object classes in 1.3 million high-resolution training images from the web, and the best system to classify objects in 2010 competition got 47% error for its first choice and 25% error for its top five choices. Some of the earlier examples of the image classification are shown in Fig. 2(i), (ii), (iii).

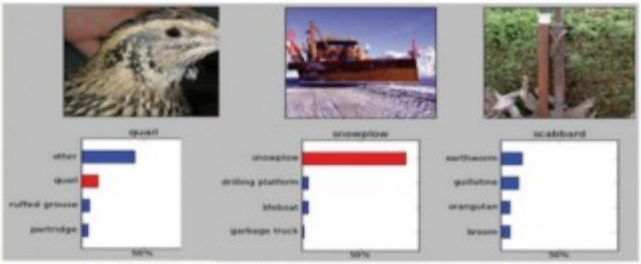

Fig 2(i) Example of earlier version of Net

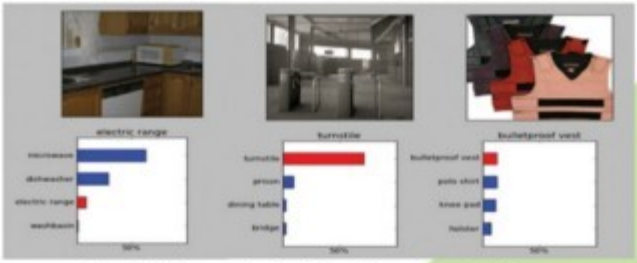

Fig 2(ii) It can deal with wide range of objects

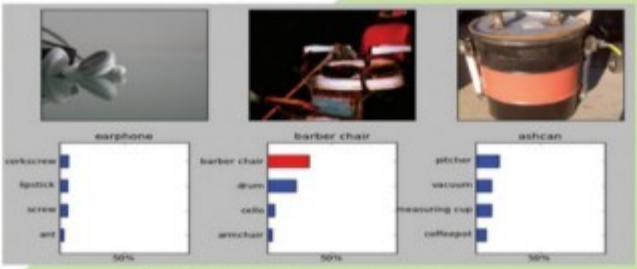

Fig 2(iii) It makes really cool errors

From Fig 2(i) (ii) (iii) we can see that it can correctly predict wide range of objects and it does make some really cool errors like the earphone in the figure looks like ant from above, and some are very awkward errors which can be predicted when we see it in a different way.

Now one of the most widely used applications of speech is Google voice, and one of the widely used applications of image processing is Google lens. So, a speech recognition task has several stages, firstly in the preprocessing stage, it converts the sound wave into a vector of acoustic coefficients. Extract a new vector about every millisecond. In the acoustic model, it uses a few adjacent vectors of acoustic coefficients to place bets on which part of which phoneme is being spoken. Then comes the decoding part, in this it finds the sequence of bets that does the best job of fitting the acoustic data and also fitting a model of the kinds of things people say.

So in the history of the neural networks, deep neural networks pioneered by George Dahl and Abdel Rahman Mohammed are now replacing the previous machine learning method for the acoustic model. After standard post processing using a bi-phone model, a deep net with 8 layers gets 20.7% error rate. The best previous speaker independent result of TIMIT was 24.4% and this required averaging several models. The data from word error rates from MSR, IBM and Google shows that Switchboard(Microsoft Research) requires 309 hours of training data and it has a error of 18.5% in the speech, so English broadcast news of IBM requires 50 hours of training and it gives an error of 17.5% whereas the Google voice search of android 4.1 requires 5,870 hours of training and it's error rate is 12.3% and falling, being the most effective speech recognition system in this planet.

There are different types of learning task:

1. Supervised learning – It learns to predict an output when an input vector is given.

There are two types of supervised learning.
Each training case consists of an input vector x and a target output t.

i. Regression: The target output is a real number or a whole vector of real numbers, for example the price of a stock in a six months time period or the temperature at the noon tomorrow.

ii. Classification: The target output is a class label. The simplest case is a choice between 0 and 1. We can also have multiple alternative labels.

2. Reinforcement learning-Learn to select an action to maximize payoff.

In reinforcement learning, the output is an action or sequence of actions and the only supervisory signal is an occasional scalar reward. The goal in selecting each action is to maximize the expected sum of future rewards. We usually use a discount factor for delayed rewards so that we don't have to look too far into the future. Reinforcement learning is difficult since the rewards are typically delayed and so it is hard to know where we went wrong or right. A scalar reward does not supply much information.

3. Unsupervised learning- Discover a good internal representation of the input.

For about 40 years, unsupervised learning was largely ignored by the machine learning community; some widely used definitions of machine learning actually excluded it. Many researchers thought that clustering was the only form of unsupervised learning. It is hard to say what the aim of unsupervised learning is. One major aim is to create an internal representation of the input that is useful for subsequent supervised or reinforcement learning. We can compute the distance to a surface by using the disparity between two images. But you don't want to learn to compute disparities by stubbing your toe thousands of times. It also forms a compact, low dimensional representation of the input. High dimensional inputs typically live on or near a low dimensional manifold(s). Principal component analysis is a widely used linear method for finding a low-dimensional representation. It provides an economical high-dimensional representation of the input in terms of learned features.