



Introduction to Machine Learning

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to 'learn' from given data, without being explicitly programmed.

Traditionally, a computer was given instructions on how to perform task based on some given data. With the help of provided data and the specific program the machine will give out certain output.

Whereas, in Machine Learning, the data is given along with an expected output and an algorithm. Based on these provided material the computer will learn the trends in the data and how is the output affected with specific data points using the algorithm. Now that the machine has learned about the data we can provide it some new input and it will directly give us a desired output.

For example, its similar to how we go to school, the teacher provides us with the study material/books (data), and teaches us ways to understand the text in the books (algorithm). We learn from this 'data' using the 'algorithm' the teacher taught us and now on the exam data we are given some new text and we have to make predictions/inferences from the learned material and write out our answers (output).

Learning by Memorisation Approach

Consider a case of building a model that detects spam emails. The machine will memorise the data previously available categorised as 'Spam emails', but now if a new email is received, comparing that previously categorised data and the new email, it might give the result as 'Spam email' if the new email is almost exactly same but it will give the result 'Not spam email' if the email is having different content which for the model is very new compared to what it had learned before

during memorisation. Same analogy in school exams, the teacher taught something, you memorised it but now some new unseen question came on the test and you very obviously not knowing anything other than what you had memorised gave the wrong answer.

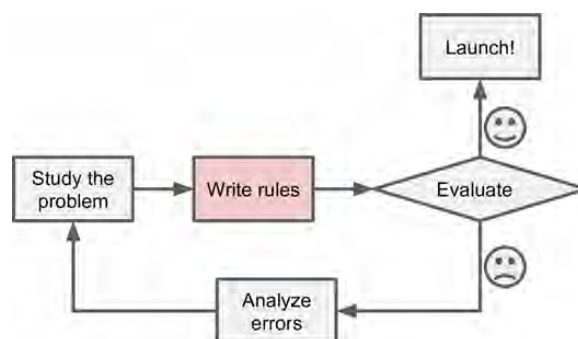
This is a very faulty approach for a computer to learn as the model does not have the ability to label unseen data, thus the approach is not generalisable.

Machine Learning in the words of Tom Mitchell, is a computer program which is said to learn from experience E with respect to some task T and some performance P , if its performance on T , as measured by P , improves with experience E .

Why use Machine Learning ?

Another question can be, if there exists traditional methods to do stuff, why employ ML? Well, consider how you would write up a 'Spam Detection' program using traditional methods:

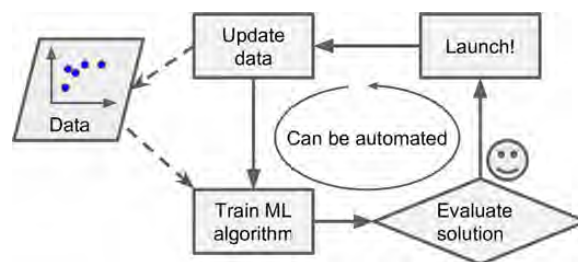
1. Generalise what comes under 'Spam' and 'Non-Spam' category. Find words and patterns that usually are part of the content of a Spam email.
2. Write an algorithm in any programming language that will be able to detect these types of words or patterns.
3. Test the program and subsequently make changes until you get the desired output.



This problem is very subjective in its nature i.e. a spam email can contain N number of patterns and words that can or cannot be classified in 'Spam' category. You will spend your whole life updating the code again and again, and the people who are sending these mails will keep changing the words they use to bypass the spam detection.

Now, if you use Machine Learning approach in this problem:

1. Generate a dataset generalising what comes under 'Spam' and 'Non-Spam' category. Use old spam and not spam emails online to add to this dataset. Or download a pre made dataset by other people from the Internet.
2. Now train the algorithm to generalise what email content comes under spam and what comes under not-spam using the dataset generated.
3. This program being much shorter will easily detect if an email is spam or not. And now if the spammers try to even change the words to bypass it they won't be able to do so because the algorithm will keep updating its dataset on its own as new emails are received, and it will detect any changes on a similar content based emails, hence putting it in 'Spam'.



This is just one example where we can say Machine Learning is great. With recent advancements, ML now even has subfields for specific problems and use cases.