Machine Learning is Fun

Yuan An

May 21, 2018

The guide *Machine Learning is Fun* [1] is written for anyone who is curious about machine learning but has no idea where to start. The author of this guide is Adam Geitgey.

In the part 1, the author answered the question " What is machine learning " in plain language:

> Machine learning is the idea that there are generic algorithms that can tell you something interesting about a set of data without you having to write any custom code specific to the problem. Instead of writing code, you feed data to the generic algorithm and it builds its builds its won logic based on the data.

Then the author gave a example of the classification algorithm that can be used to recognize handwritten numbers and classify emails into spam and not-spam without changing a line of code (see Fig. 1).

In the following, the author introduced two kinds of Machine Learning Algorithm: Supervised Learning and Unsupervised Learning. In this section, the author used estate agent as example to tell the different of this two algorithm. Su-
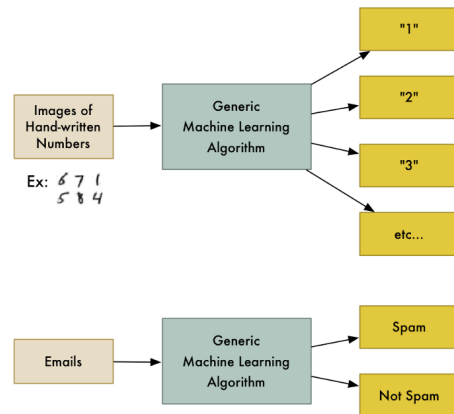

Figure 1: This machine learning algorithm is a black box that can be re-used for lots of different classification problems

pervised algorithm seems like predicting estate sale price on condition that you have known much about knowledge of sold prices of other estates, while unsupervised algorithm without knowing much about sale prices. And now unsupervised learning is becoming increasingly important as the algorithms get better because it can be used without having to label the data with the correct answer.

As a human, we all can approach most any situation and learn how to deal with that situation without any explicit instruc-

tions. And the goal of Strong AI is to have this ability. But current machine learning algorithm aren't that good yet, they only work when focused a very specific, limited problem. So maybe a better definition for learning is figuring out an equation to solve a specific problem based on some example data.

Then the author taught how to 'write' the program, in fact it's in pseudocode way to program. And he compared the program of predicting house sale prices to making soup. As is described above, the learning procedure is to find a best equation to match existing data step by step. But it is impossible for us to try every number in order to match this. So wise mathematicians have figured out lots of clear ways to quickly find good values for those weights without having to try very many.

First, write a simple equation in a easy-to-understand way:

$$Cost = \frac{\sum\limits_{i=1}^{n}(MyGuess(i) - RealAnswer(i))^2}{2n} \tag{1}$$

this is our cost function.

Now rewrite Equ. 1 into a mathematical way:

$$J(\theta) = \frac{1}{2n}\sum_{i=1}^{n}(h_\theta(x^{(i)} - y^{(i)}))^2 \tag{2}$$

where θ (in Equ. 2)is what represents your current weights. $J(\theta)$ means the 'cost for your current weights'.

The Equ. 2 represents how wrong our price estimating function is for the weights we currently have set. We can get the graph of $J(\theta)$ as shown in the following (Fig. 2). In Fig. 2, the lowest point in blue is where
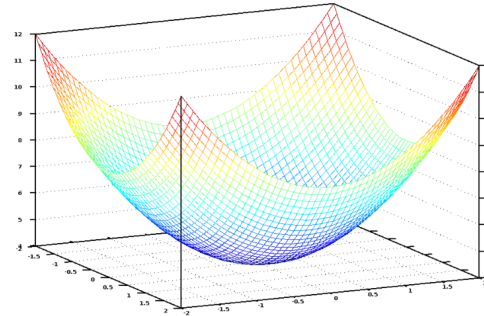

Figure 2: The graph of $J(\theta)$

our cost is the lowest, that is to say the $J(\theta)$ is the least wrong, while the highest points are where we are most wrong. So we can adjust weights to arrive the lowest point.

In the last of this part, the author also told us that machine learning works after satisfying one condition that the problem is actually solvable with the data that you have. And the author also introduced the way to learn machine learning. Except the matched courses of this article, we can attend Andrew Ng's free Machine Learning class on Coursera. We can also play around with tons of machine learning algorithms by downloading and installing SciKit-Learn[1].

References

[1] A. Geitgey, "Machine learning is fun," https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471.

---

[1] a python framework that has "black box" versions of all the standard algorithms.