On Machine Learning
And an Exploration of its Significance

Machine learning has long held a characteristic of being relatively easy to identify, but hard to define. In part, this is due to the fact that machine learning ("ML") overlaps with many other popular fields such as data-mining, statistics, algorithms, and artificial intelligence, among others. How do we identify when ML is taking place? When can we say a machine is learning?

Credit: www.v3.co.uk

In 1997, Tom Mitchell of Carnegie Mellon University provided the definition, abridged; "A computer program is said to learn from experience with respect to some task and some performance measure, if the programs performance on the task improves with the experience gained" [(1)]

So we see that ML involves training a program to make its own decisions, rather than telling it explicitly what to do. We can take the classic example of email-filtering to illustrate this.

How could one approach creating a spam filter to classify emails? At the highest level, we need a program that can take in an email, and classify the email as either spam or ham (good email) based on some number of factors. One of these factors could be the counts of how many times, if at all, specific word patterns appear, like the receiver's name, or 'Viagra'. We could also look at send times (an email sent at 4 a.m. is less likely to come from a real person), word count, average word length,

Credit: www.gwmac.com

etc. Collecting this data and giving it to the program is called 'training', since our program is 'learning' how to make decisions from the data.

Credit: www.mariner-usa.com

Once we have collected enough data, we have a good idea about the sort of traits spam emails have compared to ham emails. When we get a new email, we check out its traits, and compare that to our training data to decide which category the new email fits.

It's interesting to note that we did no direct calculations to tell us whether the email was spam or ham, but instead relied on the analysis of data to help make the decision. Machine Learning can often help provide predictive models for situations that are infeasible to directly calculate. A remarkable example of this is the story of a particular Target store in Minneapolis. [(2)]

One day, a father walked into the store angry that his teen daughter had begun to receive coupons from Target for nursery furniture, maternity clothing, and other items for soon-to-be mothers. In his eyes, Target was trying to encourage teen pregnancy. The manager apologized and even called back a few days later to follow-up and make sure all was well. At this point, the father surprisingly offered his own apology to the store manager; he had since found out that his daughter was indeed pregnant.

What had happened was that Target's purchase-tracking algorithm, developed by Andrew Pole, noticed that the teen was buying unscented lotions and particular vitamin supplements – items that had been found to correlate with purchaser pregnancy based on training data. Targets algorithm noticed these purchases and sent the teen the appropriate coupons. A machine learning algorithm was able to identify a girl was pregnant before her own father.

(1) https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer