edX **Microsoft:** DAT210x Programming with Python for Data Science

Help

Video
⬚ Bookmark this page

Machine Learning

Start of transcript. Skip to the end.

▶

0:00 / 1:56   ▶ 1.25x   🔊   ⤢   CC   "

Machine learning is not black magic, nor is it rocket science.

Rather, machine learning is the name given to a class of

generalizable algorithms that are data driven.

Now when we say data driven, we mean that rules and

solutions are derived by examining data.

Now this process works, because patterns exist everywhere.

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Machine Learning In General

The traditional way programmers get computers to do something is by providing a very detailed list of instructions specific to the task they wish to accomplish. This is known as hard coding. A different approach would be to get the computer to accomplish the same task without

specifically being programmed to, and instead by training it against data samples. Machine Learning is the name given to generalizable algorithms that enable a computer to carry out a task by examining data rather than hard programming.

With machine learning, results are attributed more towards your *data* than to the *algorithm*, since it's the data that guides the computer with what to do. In other words, the algorithm is generic; but the data is specific to the problem being solved. You might feed data into a machine learning algorithm about food your wife likes and food she doesn't like, and the task the algorithm accomplishes is learning how to differentiate between the two. On the other hand, she might feed the same algorithm data about clothes you like / dislike. Without altering a single line of code, the same algorithm can solve the new task based solely on the data!

Arthur Samuel coined the term "machine learning" in the 1950's, noting that a computer can exhibit a behavior just as a pet would after learning through intense training. This *seeming* ability for computers to 'learn' and 'discover', as well as their known ability to effortlessly crunch through millions of records, is precisely why machine learning and other aspects of data science are being applied to so many fields.

Where we would fail due to our inherent inability to comprehend even medium amounts of data (or to not get very, very bored while doing so), computers excel spectacularly. Further, with advances in storage capacity and computing speeds, machine learning performance increases by the day.

Types of Machine Learning

There are three main types of machine learning, however in this course you'll be focusing on two: **unsupervised learning** and **supervised learning**. These two types enable a computer to do a generic task by examining reams of data. To illustrate the difference between the two, let's do a case study:

Let's say your name is Angie, and your husband's name is Craig. You're both organized people and enjoy maintaining lists. In fact, you have a list of every company you've ever done business with, and some details about them, such as how reliable they were. Craig also has a list. His list is of every transaction he's ever made, along with some details about those transactions, such as the items transacted and their costs.

Unsupervised Learning

You and Craig decide to get into the computer resale business. Being data driven people, you wish to do so in an educated way, so you use your lists. Going through yours, you notice the companies that sell computers seem to all be located around some place called 'Silicon Valley'. You also notice the companies with the best reviews are those that pay attention to sleek case designs. Oh, except this one company called 'Fast-Computers', which has awesome reviews, but has computers that look like they were designed by an 8-year old.

Unsupervised learning is similar to this. Given a lot of data, the computer hasn't the slightest idea what any of it means. Yet, it's still able to figure out if there are any meaningful groupings and patterns within the data, along with instances where that data seems out of place!

Supervised Learning

Craig approaches the computer resale challenge differently. All he's really interested in is:

1. How much should you price a computer at, given its stats?

2. What price should you buy a computer for, to maximize your profits?

Looking through his data list, he discovers a correlation between a computer's processor speed, its storage space, and its cost. In fact, Craig is able to calculate a precise equation that models this! He's also able to create a set of "IF *this* AND *that* THEN *transact*" rules to decide when it's best to buy or sell a computer.

Unlike unsupervised learning where the computer had no idea what any of the data meant, with supervised learning, the computer is in charge of taking your data and then fitting rules and equations to it. Once it learns these generalized rules through a process called *modeling*, the rules can be applied to data the computer has never seen before.

Machine Learning And You

Part of being a great programmer is the ability to take complex ideas and abstract them down into solvable constituents. Machine learning operates similarly. Starting with your expertise in an area, look for interesting problems to tackle. Break those problems down into smaller pieces, so that they're either entirely solvable with machine learning, or at least partly so.

Not every problem is machine solvable, nor should every problem even be approached with machine learning! If your issue is directly solvable through some simple means, such as a few yes / no decisions, or if it does not require examining loads of data, there is probably a more fitting solution for you than machine learning. However if your issue is prohibitive on large scales, or requires fine tuning multiple knobs of potentially correlated parameters, then machine learning to the rescue. In the next chapter, we'll check out the types of problems machine learning is strong at solving.