PharmaSUG 2019 - Paper SI-062

Get to the Meat on Machine Learning

Aadesh Shah, GSK

ABSTRACT

You've probably heard of machine learning and artificial intelligence, but are you sure you know what they are? If you're struggling to make sense of them, you're not alone. There's a lot of buzz that makes it hard to tell what's science and what's science fiction. For many of us, machine learning seems futuristic and scary. Recently, though, it's been showing up, as we hear many new presentations about machine learning at different conferences like Phuse, CDISC Interchange as well as the most famous PharmaSUG 😉. YouTube know which videos you would like to watch in your home section, Facebook recommends local event in your area, also recommends friends, LinkedIn recommends you connect with your ex-boss. And while that's all exciting, some of us are still wondering what exactly machine learning is. This paper will walk you through the process basics, work in practice, Machine Learning vs Artificial Intelligence.

INTRODUCTION

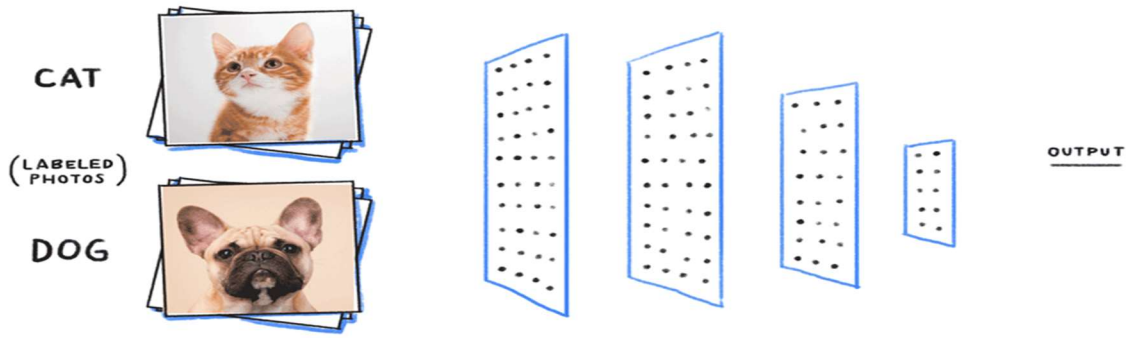
LET'S GET STARTED WITH THE BASICS. WHAT EXACTLY IS MACHINE LEARNING?

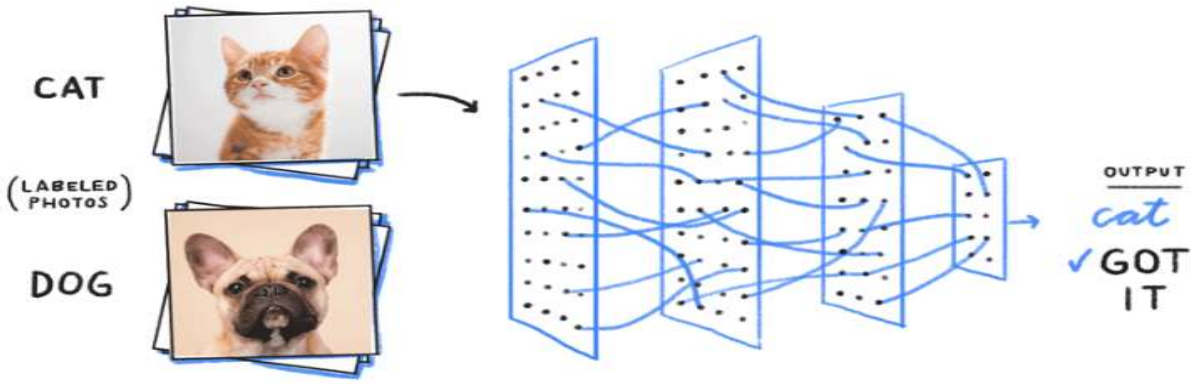
There are many interpretations online on Machine learning, one of them is "how to construct computer programs that automatically improve with experience" and my favorite since its easier to understand is "machine learning takes a bunch of examples, figures out patterns that explain the examples, then uses those patterns to make predictions about new examples."

Let's take movie recommendations, for example. Say a billion people each tell us their ten favorite movies. That's a bunch of examples the computer can use to learn what movies that people like have in common. Then the computer comes up with patterns to explain those examples like maybe, "People who like horror movies don't usually like romances, but people do like movies with the same actors in them." Then if you tell the computer you liked The Shining with Jack Nicholson, it can make a good guess about whether you'd like the romantic comedy Something's Gotta Give with Jack Nicholson, and which other videos to recommend to you on YouTube.

GOT IT. SORT OF. HOW DOES THAT WORK IN PRACTICE, THOUGH?

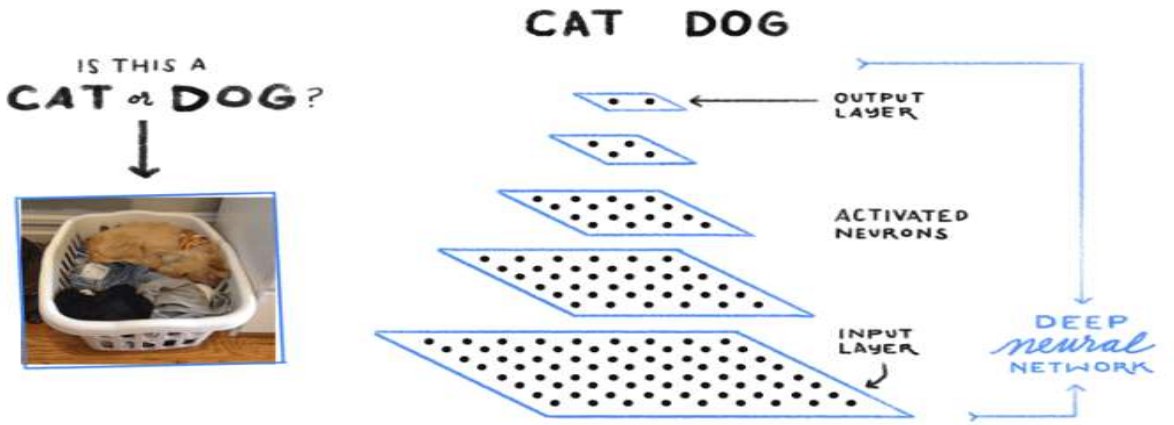
In practice, the patterns that the machine learns can be very complicated and hard to explain in words. Consider Google Photos, which lets you search your photos to find pictures with dogs. How does Google do that? Well, first we get a bunch of examples of photos labeled "dog". We also get a bunch of photos labeled "cat," and photos with about a million other labels, but I won't list them all here 😉

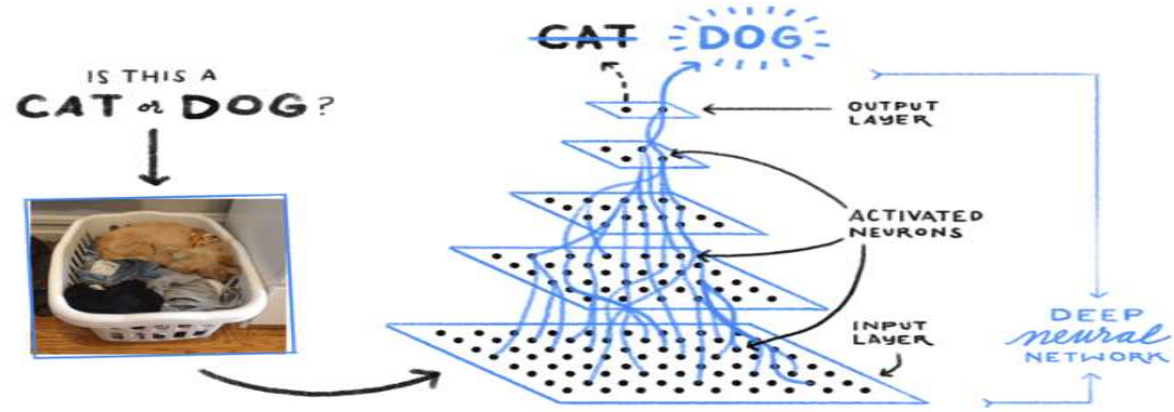




Then the computer looks for patterns of pixels and patterns of colors that help it guess if it's a cat or dog or something else. First, it just makes a random guess at what good patterns might be to identify dogs. Then it looks at an example dog image and sees if its current patterns get it right. If it's mistakenly calling a cat a dog, then it makes some tiny adjustments to the patterns it's using. Then it looks at a cat image, and again tweaks its patterns to try to get that one right. And it repeats this about a billion times: look at an example, and if it's not getting it right, tweak the patterns it's using to do a better job on that one example.

In the end, the patterns form a machine-learned model, such as a deep neural network, that can (mostly) correctly identify dogs and cats and fire fighters and many, many other things.





WHAT ARE NEURAL NETWORKS?

Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret sensory data through a kind of labeling raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text or time series, must be translated.

Neural networks help us cluster and classify. You can think of them as a clustering and classification layer on top of the data you store and manage. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labeled dataset to train on.

IS MACHINE LEARNING THE SAME THING AS ARTIFICIAL INTELLIGENCE?

Well, actually, these words can mean different things to different people, but essentially, artificial intelligence (AI) is a loose term for computer programs that try to solve the kind of problems that humans find easy, like telling a story about what's happening in a picture. One of the cool things that humans also do easily is learn from examples. And that's what machine learning programs try to do, too: teach computers to learn from examples.

The cool thing is when we figure out how to make these computer programs, we can sometimes scale them up to handle a lot of data really fast, and then we can solve really hard problems like routing everyone through traffic simultaneously, optimizing energy usage nationwide, and of course my favorite standardizing CDISC data.

SO, WHY IS EVERYONE MAKING SUCH A BIG DEAL ABOUT MACHINE LEARNING NOW?

Machine learning is not brand new and has its roots in 18th century statistics. But you're right it has really heated up lately for three reasons.

First off, we need a huge number of examples to teach computers how to make good predictions, even about stuff you or I would find easy (like finding a dog in a photo). With all the activity on the internet, we've now got a rich source of examples computers can learn from. For example, there are now millions of dog photos labeled as "dog" on websites around the world, in every language.

But it's not enough to have a lot of examples. You can't just show a bunch of photos of dogs to a webcam and expect it to learn anything — the computer needs a learning program. And lately many companies have made some exciting breakthroughs in how complicated and powerful those machine learning programs can be.

However, many programs are still not perfect, and computers are still pretty dumb, so we have to see a lot of examples a bunch of times to tweak a lot of digital knobs to get it right. That all takes a huge amount of computing power, and fancy parallelized processing. But new software and hardware advancements have made that possible, too.

CONCLUSION WHAT DOES MACHINE LEARNING LOOK LIKE IN TEN YEARS?

One thing the whole field is working on is how to learn faster from fewer examples. One approach to that is giving machines more common sense, which in the field is called "regularization."

What does common sense look like to a machine? Well one thing it means is that, in general, if an example only changes a little, the machine shouldn't totally change its mind. For example, a photo of a dog with a cowboy hat is still a dog.

We enforce this kind of common sense in the learning program by making the machine learning insensitive to small, unimportant changes, like a cowboy hat. While that's easy to say, if you do it wrong, you make the machine not sensitive enough to important changes! So, this a balancing act.

REFERENCES

Google Labs Machine Learning and Deep Learning.

https://skymind.ai/wiki/neural-network

https://ai.google/education/

https://www.google.com/about/stories/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

https://www.linkedin.com/in/aadesh-shah-a4b19811a

Aadesh Shah
Clinical Programming Manager
Biostatistics (US)
GSK Research & Development

GSK
1250 S. Collegeville Road, Collegeville, Pennsylvania, 19426-0989, United States
Email aadesh.x.shah@gsk.com
Tel +16109175509

gsk.com | Twitter | YouTube | Facebook | Flickr