Intro to ML with Microsoft Azure Machine Learning Studio

Amita Shukla

March 31, 2018

What is Machine Learning?

Machine learning is an umbrella term for a set of techniques that allow you to program computers to do things you don't explicitly program them to do. For example, if you've programmed before, you know the computer does exactly what you tell it to, sometimes frustratingly so.

Machine learning describes the set of situations where you haven't described precisely what output you'd like to receive, and let the computer figure that out for you.
Supervised vs unsupervised:

In supervised machine learning, you train with options and answers. In unsupervised machine learning, you allow the computer to discover structure.

What Kinds of Questions Can We Ask?

Data only answers questions phrased in certain ways.

- **Predict Values** (Regression)
  - Given past prices for an item, how much will it sell for in the future?
  - Which part in my machine will fail first?
- **Find oddities** (Anomaly Detection)
  - Is this an example of credit card fraud?
  - Is this signature forged?
- **Discover structure** (Clustering)
  - What segments can I group my customers by?
  - What do my customers tend to like?
- **Is it A or B?** (Two-Class Classification)
  - Will this customer be a return customer?
  - Is this tweet sarcastic?
- **Which category does it fit in?** (Multi-Class Classification)
  - Sorting handwritten numerals
  - Sorting by color

Today we'll explore an example of Regression (Automobile Price Data) and if we have time, we'll do Binary Classification (Titanic data) and company clustering.

What can the Microsoft Azure ML Studio Do?

ML Studio has a drag-and-drop interface much like MIT's Scratch that will automatically run your chosen algorithms on your data. It's a very black-box approach and so a good intro.

ML Studio does not clean your data (at least not automatically). ML Studio does not filter your data in any way.

ML Studio will give you results whether or not your algorithms make any sense! Sometimes the form of your results will give you a hint that you need to change your algorithm–but not always!

Linear Regressions

You've probably seen linear regression before as 'best-fit line'. Given some data, which you plot in two dimensions, draw the best fit line. There are many other types of regressions: you may have also seen quadratic regressions and polynomial regressions, which are exactly the same idea!

Don't be intimidated by longer names for it like 'Ordinary Least Squares Linear Regression'!

More formally, in the Ordinary Least Squares Linear Regression, you try to minimize both the x and y distance of each point from the line.

In our Automobile Price Data example (example from Microsoft), we'll do just this.

Multi-Class Classification

When you hear people talk about 'neural networks', it is usually in this context. (Neural networks is a subset of machine learning). This usually works like this:

- You give the computer a dataset with known examples (training set)
- You test your model's performance and adjust its parameters accordingly
- You test new unknown data with your network

The YouTube channel 3Blue1Brown has the best explanation of this that I've ever seen.

This usually works in two ways: batch learning and online learning. Batch learning means one cycle of the above; online learning means that the system learns incrementally instead of being trained just once.

Further Resources

Coursera class by Andrew Ng
3Blue1Brown YouTube channel
Microsoft Azure AI examples https://gallery.azure.ai/experiments