MAN VS. MACHINE

APPLICATIONS OF MACHINE LEARNING IN FINANCE

WALT POHL

Machine Learning: The Hype

- Machine Learning is the latest hyped technology, joining a long list of heavily-sold innovations:

    - Dot coms
    - Internet 2.0
    - Cloud computing
    - Big Data

- So is everyone right to be so excited?
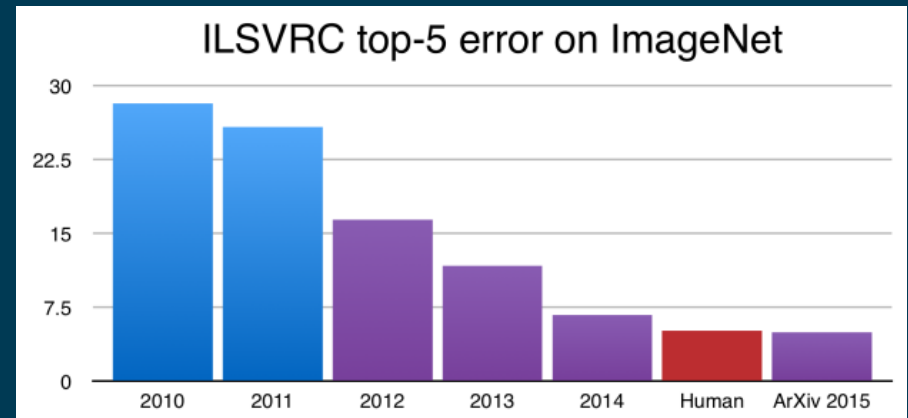
ImageNet

ImageNet
(http://image-net.org/)
is a database of
14,000,000 images
tagged with their
contents.

The ImageNet Challenge

The annual ILSVRC contest challenges researchers to identify the contents of the photo via computer.

In 2015 the winner achieved better-than-human performance.

ILSVRC top-5 error on ImageNet

Games

In 1997 IBM's Deep Blue beat world champion Kasparov at chess.

Interest turned to a Japanese board game known as "Go".

Progress in Go

- As of 2014, *no* computer program had ever beat a Go professional. People thought the Go equivalent of Deep Blue was ten years away.
- In 2015, Google premiered a program, AlphaGo, that beat the European Go champion 5-0.
- By 2017, the latest version of AlphaGo had beaten every top player in the world.

Progress in… Chess?

- Deep Blue was carefully designed by humans to contain much human chess knowledge.
- AlphaGo works differently. It *learns* by playing itself over and over.
- Google tried the same self-training technique on chess. After 24 hours of playing against itself, the same algorithm beat the top chess computer program.
- So humans have nothing left to teach the computer.

What Is Machine Learning?

- *Machine learning* is simply the art of using a computer to identify patterns in data.
- It is like statistics, except that you are flexible on how you try to fit the data.
- The danger in this flexibility is that you will *overfit*: find patterns that aren't real.
- To save yourself you use Occam's Razor: prefer simple explanations over complicated ones.

Deep Learning

- The spectacular successes have been using *deep learning*.
- A *neural network* is loosely modeled on neurons in the brain.
- A deep neural network uses many intermediate layers between input and output.
- The intermediate layers somehow capture higher-level information. (We don't know exactly how.)
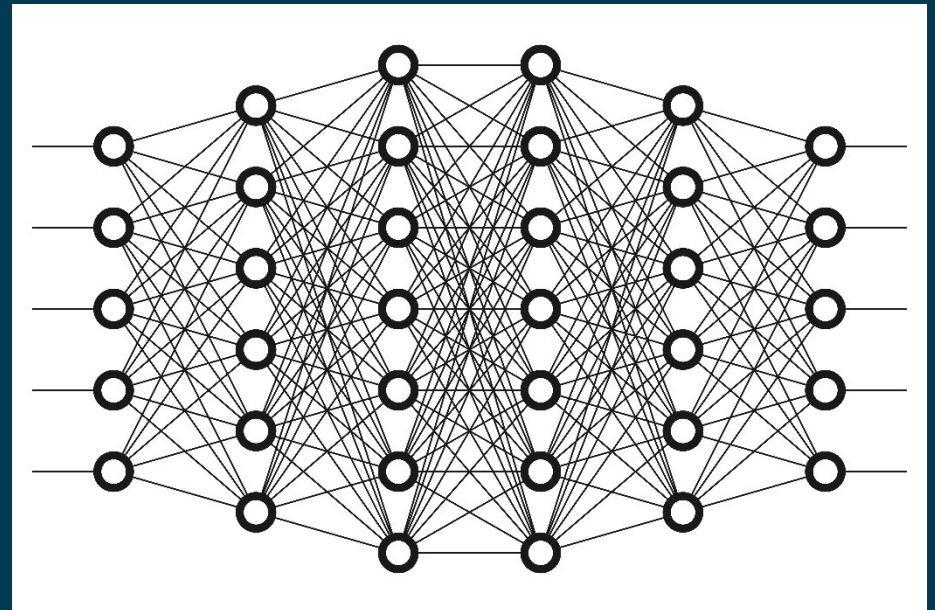
Future Application: Tesla Autopilot

- Tesla Autopilot records video from your car when you drive.
- Given enough data, the hope is that this will lead to self-driving cars.
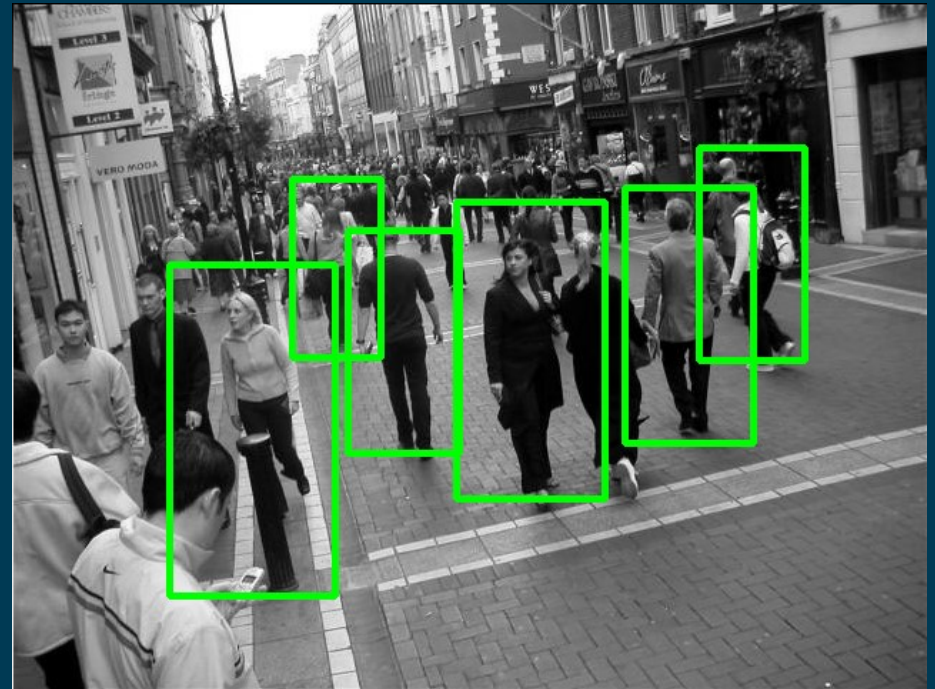
Why Now?

- Machine learning is not new. Many techniques are 30 to 60 years old.

- So what changed?
  - Computers are more powerful
  - We have much more data.
  - Some improvements in algorithms

- More importantly is that people now believe that you can solve problems this way, so they are solving problems this way.

Machine Learning's Big Successes

- Given *enough* data, the algorithms will find a pattern that is there.

- Examples:
  - Image recognition
  - Voice recognition
  - Translation
  - Playing two-player games such as Go.

- All of these examples we know a pattern is there because humans put it there.

- What is less clear: what if we don't already know if there's a pattern there?

Most Promising Finance Applications

But what about banking and finance?

Some are obvious:

- Credit scoring
- Automated trading
- Sentiment analysis – Using Big Data to predict how the market will react to news before it reacts.

Some are less obvious.

Credit Scoring

- Credit scoring is using an algorithm to determine the probability of default of a borrower.
- Even the oldest versions of credit scoring use simple machine learning techniques.
- Modern machine learning just allows you to take more information into account.

Automated Trading

- There are many quant funds, and they keep their strategies close to the vest. There are a few exceptions:

- Two Sigma

  - Explicitly uses machine learning.
  - Already has $50 billion under management.

- Numerai

  - Numerai posts encypted finance data on the web.
  - Data scientists compete to analyze the returns.
  - Numerai trades on the winning strategies.

- Quantopian

  - Another attempt at crowd-sourcing investment strategies.

Sentiment Analysis

- High frequency trading means algorithms need to trade on news or investor sentiment faster than any human can react.

- Thomson Reuters sells (machine-learned) news and sentiment signals to customers.

- New frontier: extract investor sentiment from Twitter's 6000 tweets a second.

Less Obvious: Product Cross-Selling

Retailers know from data what products go together?

Can we use it for financial product cross-selling?

Market Basket Example

? Where should detergents be placed in the Store to maximize their sales?

? Are window cleaning products purchased when detergents and orange juice are bought together?

? Is soda typically purchased with bananas? Does the brand of soda make a difference?

? How are the demographics of the neighborhood affecting what customers are buying?

Image source: deepclimate.org

Less Obvious: Financial Crime

Can we detect financial crimes as they happen?

Credit Suisse, CIA-Funded Palantir to Target Rogue Bankers

By **Vogeli Voegeli**

March 22, 2016, 2:55 PM GMT+1
Updated on March 22, 2016, 5:15 PM GMT+1

→ Bank says it started working with Palantir after Adoboli case

→ Signac venture aims to detect unauthorized trading, misconduct

Future Opportunities

Introducing machine learning in your business isn't even that hard:

- Machine learning does not *require* big data. Businesses already do not even take advantage of "small data".
- Traditional statistics provides effective tools for small data.
- Machine learning techniques are largely fancier versions of traditional techniques.

Making the Most of Your Opportunities

- The businesses that have succeeded with machine learning – the Googles, the Facebooks – succeeded on the strength of their *internal* expertise.

- Data is messy. It requires domain expertise before algorithms can be brought to bear.

- You can't just have some consultants parachute in and slather some machine learning on top of your business.

Machine Learning and Academia

- Machine learning techniques comes out of a computer science tradition.

- They are not designed with business goals in mind. Much research is needed.

- Example: algorithms that pick stock don't recognize that stock is part of a portfolio, and that portfolio risk is what matters.

- Example: most algorithms are not designed with the time dimension in mind.

Risks

- More search is needed into the risks.

- One key risk: The algorithms only care about patterns in the data. They don't understand **causality**.
  - If the world changes on you, the patterns you rely on may disappear.
  - If the training sample contains prejudice, the algorithms will learn that prejudice (as Microsoft learning the hard way with their Twitter chat bot).

- There's also the obvious risk…

Don't Build Skynet