“Machine learning is a core, transformative way by which we're rethinking everything we're doing.“- Google CEO

“A breakthrough in machine learning would be worth ten Microsofts.” (Bill Gates)

“Machine learning is the next Internet.” (Tony Tether, Former Director, DARPA)

“With artificial intelligence we are summoning the demon.” - Elon Musk

“Artificial intelligence could spell the end of the human race." - Stephen Hawking

A Simple Starting Point

Today, humans write computer programs (above left). The customer tells the programmer what functionality is needed. The coder then designs and builds the application. Data is sent to this application which generates output. Tomorrow (above right), a computer will write the computer program for us. Customer will provide the functional requirements and the data to the computer and it will write the application without any human intervention.

The plane you get on this week will be piloted by a computer programmer. Planes are currently programmed using the “as-is” model. A large group of subject matter experts and engineers get together and write code for every possible scenario a plane might encounter. Lightning strikes, splitting and striking both wings. What line of code should we add for that?

“While airlines have long used automation safely to improve efficiency and reduce pilot workload, several recent accidents, including the July 2013 crash of Asiana Airlines Flight 214, have shown that pilots who typically fly with automation can make errors when confronted with an unexpected event or transitioning to manual flying.”
– Inspector General in a letter to the FAA

An Air France plane en route from Brazil to Paris crashed into the Atlantic Ocean after the auto-pilot malfunctioned and crew error caused the plane to stall. All 228 aboard died. In 2014, an AirAsia plane crashed into the Java Sea after the auto-pilot kicked off in bad weather and the pilot’s bad decision put the plane into a stall that led to 162 deaths.

What do these three examples have in common? They forgot to write a line of code for that scenario. Secondly, pilots are becoming shockingly illiterate in the cockpit. When computer code is missing a line and control of the aircraft must be handed off to the pilot, the pilot is completely out of practice. Chelsey Sullenberger made that landing in the Hudson because he was a glider pilot who knew how to land a zero-computer-code, engine off, airplane.

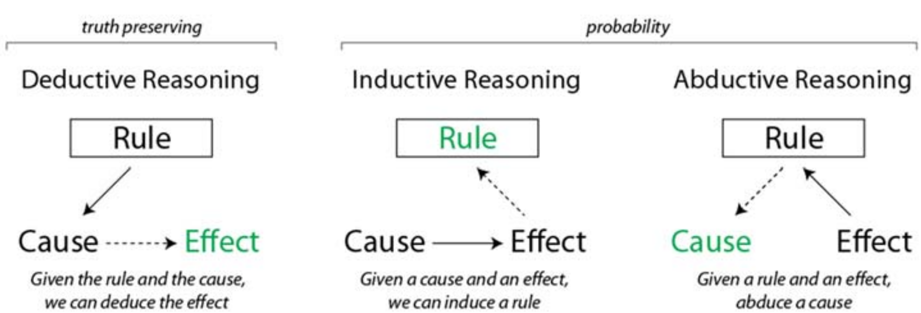
Did you ever imagine a driverless car was possible? In your lifetime? Grandchildren's lifetime? It's here. And the technology is based on machine learning (to-be scenario). When a human gets into an accident, s/he tells the family and 4-8 people may benefit from that learned experience. When a driverless car gets into an accident, it teaches every other car in America within a few seconds. It's not the car that is important. It's the shared experience between thousands or millions of cars. In other words, we have a desired outcome, don't get into an accident. We have data on all previous accidents by all cars. Now, computer, write a program for a safe car. This is called mass learning.

What will happen to the airplane? It will be converted to machine learning. Each plane will talk to every other plane and learn from a shared experience. Now, this is where it gets a little complicated. It's not the plane that matters. It's the environment around the plane that holds the promise of machine learning.

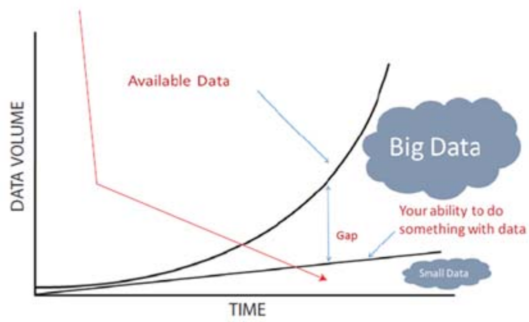
The plane is not learning how to fly better. It's learning how to read the environment (ocean) in any of 1MM+ different scenarios. Sensors will be built into every part of the plane. When pilot tubes are obstructed with ice crystals, the machine will know how to deal with that scenario. If tubes are a new scenario (not learned) and it does cause a stall, the machine will know how to recover from that stall. It will teach every plane on earth, within a few seconds, what to do when pilot tubes freeze. A plane that lands will teach the plane about to land what conditions to expect as it touches down. Same with turbulence. Plane in front educates plane behind it.

"A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E." - Tom Mitchell, Carnegie Mellon

Above is the generally accepted definition of machine learning. Before we leave this section, let's talk about fundamental forms of inference. This will be important when we talk about risk later.



Deductive reasoning is our current programming language. We write a rule that when lightning strikes a plane (cause), please adjust the plane (effect) to minimize any adverse outcomes. This rules-based system is very human centric, because it's black and white. X happens? Do Y. Done. Inductive reasoning causes human beings great consternation, because it's all about probability. X happens. What's the probability the plane will crash, if we do Y. This brings us to big data.

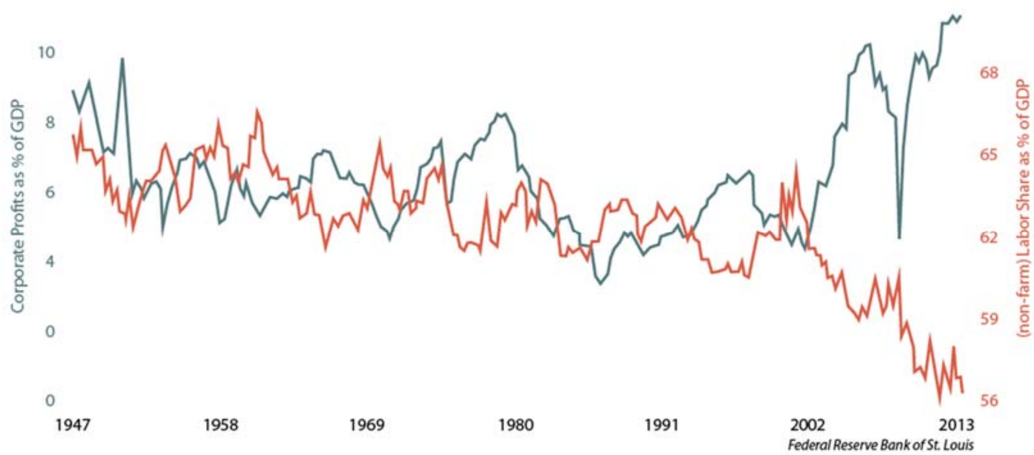


Human beings can only handle so much data. We are at a point in time where data has so far exceeded our cognitive abilities, most data just sits idle and is never used. Amazon owns a machine learning service, yet it's home page recommendations are just more of what you recently browsed. It's not even smart enough to see that I bought a product on Amazon and therefore the home page should not keep recommending it.

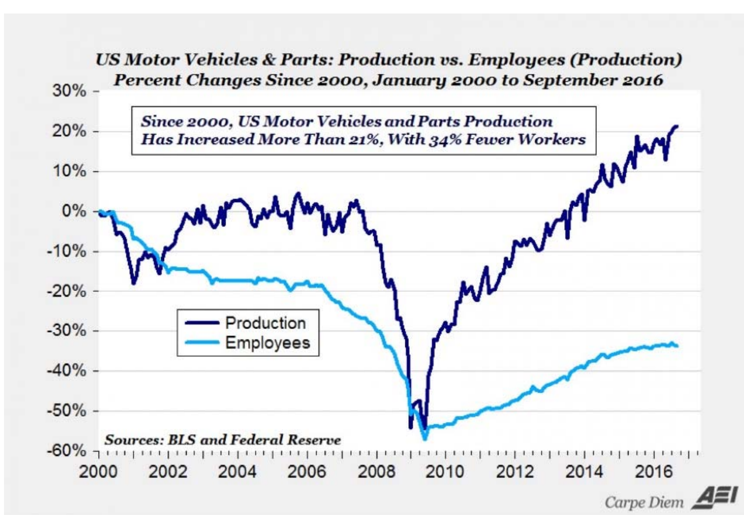
Inductive reasoning and machine learning can use all the data. In fact, machine learning only works with very large datasets, like the millions of scenarios a plane may encounter at any given place on the globe. We've come to a decision time (inflection point) for humans. Humans write the truth, but only over a sliver of data. Machine learning writes a probability equation with a very small chance of failure over all the data. Which should we trust?

The Economics of Machine Learning

The orange line is labor share and the green line is corporate profits.

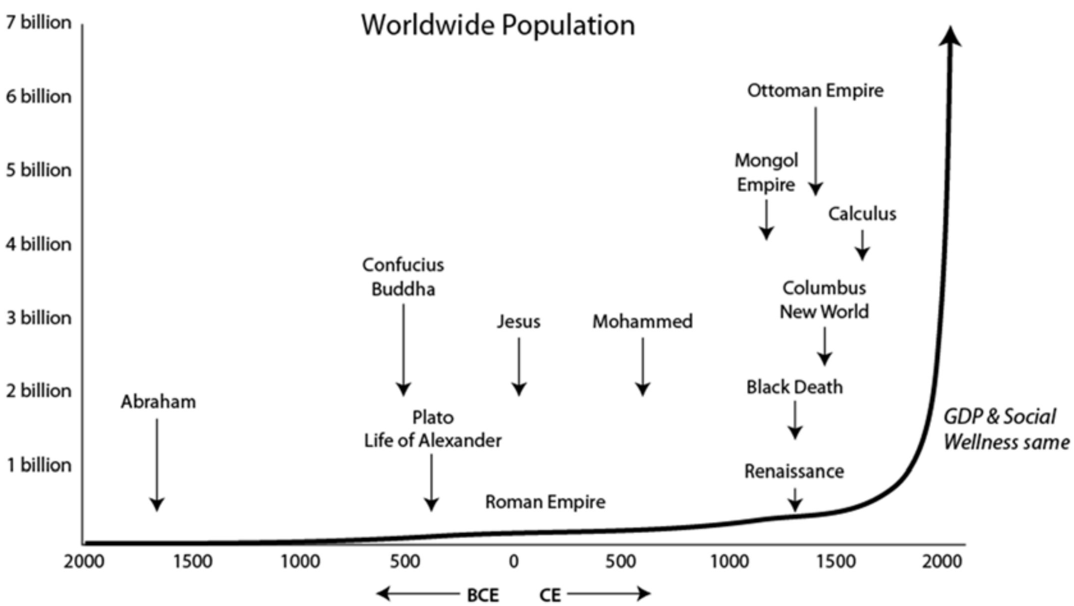


Since World War II, profits showed a give and take relationship with employee needs. Around 2000, we started to see a "great divergence" between the lines. The 2008 financial crisis had an impact on profits, but the recovery was swift. Many believe these lines will never cross again. Employees no longer drive profit. In 1990, the three largest companies in Detroit had a market capitalization of $36 billion while collectively employing ~1.2 million workers. In 2014, the three leading companies of Silicon Valley employed ~137,000 workers with a combined market capitalization of $1.09 trillion. Operating income per employee at Google is ~6x greater than IBM and ~12x greater than General Motors.



The above chart should surprise no one. Nor should the below one. Unemployment stands at 4.9%, but does not take into account our labor force has been shrinking since the late 1990s. Why? Technology. And this is just the beginning.

Another way to approach this subject it to look at changes to worldwide population, GDP and Social Wellness.



There is no event in human history that comes close to the impact of the industrial revolution. None of these earlier events impacted human "potential" or productivity. The industrial revolution was about muscles and power. Instead of using human muscles to pick that field or transport ourselves from town to town, we started using machines. This event gave us the above hockey stick in population, GDP and social wellness.

The next revolution will be a cognitive revolution. What the industrial revolution did for our muscles, Artificial Intelligence (AI) and machine learning will do for our minds. Many believe this is the endgame. If humans are not needed for their biceps or their intelligence, they become irrelevant? This is the view of Elon Musk, "Robots will take your jobs, government will have to pay your wage." Companies always had to worry about competitors and governments. Today, the greatest threat to entire industries is a takeover by technology.

The Death of the Auditor

This section should be titled "The Death of Rate per Hour," but let's use accounting as an example of the coming cognitive revolution. According to the Oxford Martin School at the University of Oxford, Accountants and Auditors have a 94% chance of becoming computerized. 47% of total U.S. employment is at risk. This should not be discounted considering the driverless car… nothing is impossible anymore with today's technology.

What if a person buying a car, the dealership selling that car, the distributor delivering that car, the manufacturer building that car and the supplier providing parts for that car all use a single finance system, a single ledger. There's nothing to audit because the intermediaries are gone. For the first time in human history, people will be able to transact peer-to-peer with full trust. Conventional banking will change dramatically. There's nothing to "settle." Assets and liabilities will be stored digitally and replicated across the world. This is the promise of a new technology called Blockchain, the internet of value, the trust protocol. There will be complete trust (a digital signature) between a debit and credit on the (shared) ledger. The debit may be recorded by one counterparty and the credit another counterparty. If Ford Motors says it sold X cars last month, one only need look at the blockchain for confirmation. Financial reports can be generated with a single click at any time throughout the year. The "trust" verification provided by auditors is no longer needed. Instead, accounting firms will transition to financial technology firms.

Tax returns will also be fully automated. The government is on the blockchain. All taxes are automatically computed and withdrawn in real-time. Machines calculate everything for everyone using a single program that cannot be hacked or exploited. Why? Blockchain is a distributed network. Hack a block in Hawaii? The other 999,999 replicated chains immediately see it and kick out the hacker. One would have to hack 1MM+ computers at the same exact time. Who owns the blockchain? A government? Wall Street? No. We all own it just like we all own the internet.

Here is an example. Why does Airbnb exist? To clear transactions. It's an intermediary that goes away in a blockchain world. Anyone wanting to rent their home shows ownership on the chain. A customer enters into an agreement and transaction on that chain. Airbnb is irrelevant. Uber? Same thing. You ask the blockchain for a ride to the airport and anyone on the chain can provide that service. Zero intermediary needed. Transaction histories? All in the blockchain. And the financial services industry knows it's coming…

"Silicon Valley is coming. There are hundreds of startups with a lot of brains and money working on various alternatives to traditional banking." - Jamie Dimon

Financial Tech (FinTech) firms in China will soon have more customers than the Chinese banking system. The text messaging service, WeChat, created WeBank. Baidu is creating an online bank. Baidu is the equivalent of Google getting into banking. ATM machines will disappear, as will credit/debit cards. A cashless society means bad news for only one group – criminals.

"Everyone talks about how Robo Advisors can't connect with clients. I actually believe those kinds of tools are like an ATM machine. We are all going to have it."
– Larry Fink, Blackrock, August 2015

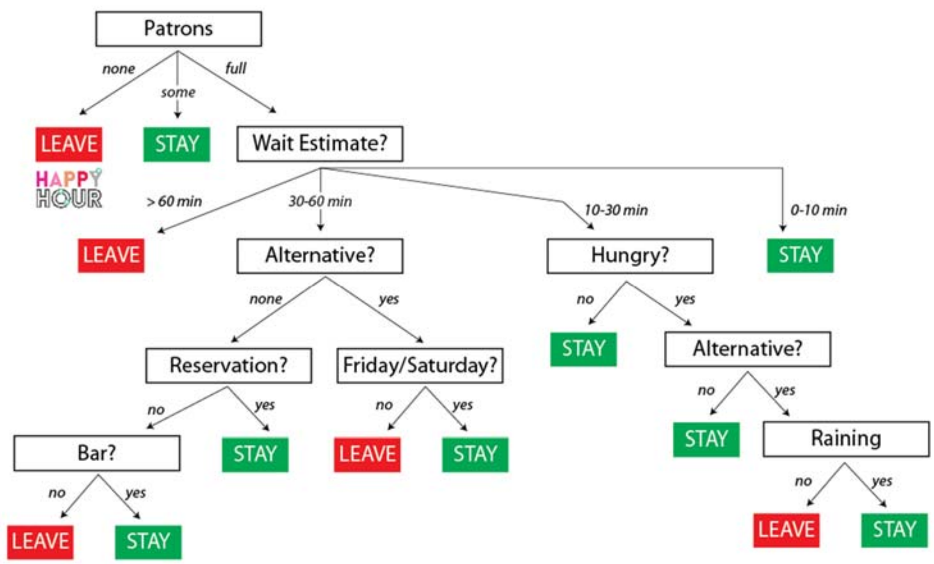
The Robo-Advisor
While today's leaders may be able to kick the "blockchain" can down the road, they won't avoid the robotization of human tasks. The certified financial planner is being replaced with a computer. Fidelity Investments has rolled out a national pilot for its robo service. Blackrock bought a robo firm. Schwab, Vanguard and others are already doling out robo-advice. Facebook is investing heavily in computerized chat bots. Buying a home? A computer will assist you. Need legal advice. A computer. Call centers? A computer. Therapy? A computer.

Robo advice is highly profitable. A single technology platform can service thousands of customers. The advice is custom tailored to each client. Millennials prefer a robo-advisor and will let the application make investment decisions without even checking in with the client. It's a brave new world.

Machine Learning – Part Two

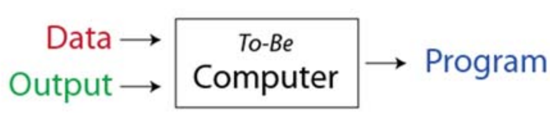
We covered the definitions and took a break to look at ML's coming impact on the economy. Let's dive a little deeper into the mechanics of the coming cognitive revolution.

We are all familiar with the two simple concepts of predict and classify. On the left, we see a linear regression where we take any x value and try to predict the y value. On the right, we are trying to split the data and put the next input into the right class. Already, we can see complexities as we move to big data, lots of x values, lots of inputs. Decision trees are an example of a classification algorithm.



Let's take a restaurant. We want to predict whether the next person coming in the door will stay and eat (class 1) or choose to leave (class 0). But why did we pick whether the restaurant is crowded as the top node? Why didn't we start with the type of food (French, Thai, Burgers, etc.) not even mentioned in the above tree? Why didn't we make the weather or the hunger of the person the first decision in the tree? What if we could run all variations of a decision tree and combine them (random forest) into a stronger predictor.

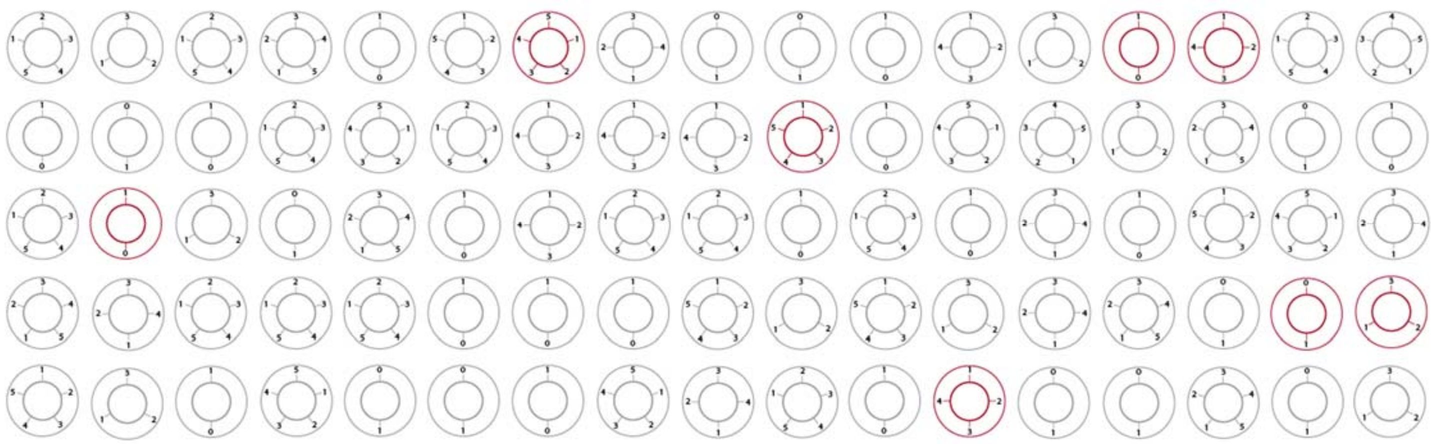
To do this with conventional programming is possible, but what if we have 200 features to choose from with that top node? And how do we place a value on the importance of any one tree? This gets very complicated very fast.



Our output is 1 or 0. We call this a binary classification problem. Will the customer stay and eat or will the customer leave? The features (data) are hunger level, day of week, time of day, desire for a cocktail, type of food, cost of meal, wait time, spacing of tables, etc. for past customers. The program is the decision tree or a group of decision trees that predicts whether the next customer will be a 1 or 0, based on their feature set. Not only can there be many trees, there will be many nodes in each tree.

Predicting the Outcome

Assume you have customer (feature) data for 10,000 customers. We would have 10,000 rows of data and 201 columns. 200 columns would be input features and 1 column would be our "label" or outcome. Row #1 had this combination of features and stayed for dinner. Row #2 had this set and left without eating. And some features like appetite may be hungry or not hungry (binary) or subtler, like a degree of hunger. Cuisine would be a multi-option feature.



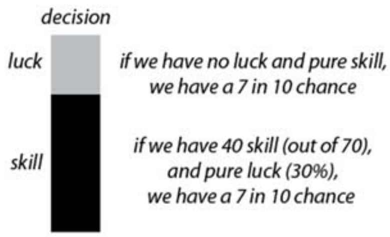
Think of the features as dials. One dial may have two choices (1 or 0, hungry or not). Another dial may have 3. Another 5. And there may be 200 dials. Machine learning will use 10,000 rows to create the optimal combination of knob values to predict an outcome. And the red knobs above may be eliminated altogether as not helpful to the outcome.

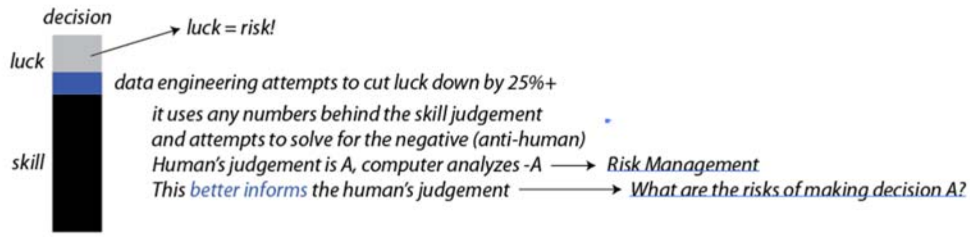
And we can test each machine model. We might give 5,000 rows to train a random forest model. We use the other 5,000 for testing, withhold the outcome label (stayed or left restaurant) and let the model predict each outcome. If the model does a lousy job and only predicts correctly 50% of the time (coin toss), the model is discarded and machine learning goes back and tunes the dials. And it will do this over and over, until it gives us the model with the highest rate of correctly predicting stay or go. We now use that model, not only to predict the next customer's behavior, but to make physical changes to our restaurant that will increase the odds of staying to eat. We might expand the bar. We might have a happy hour. We might open earlier. We might have a lighter-side menu.

Conventional software programming will be obsolete. Hundreds of features. Millions of rows of data. Multiple options for each of the features. Multiple top nodes, multiple secondary+ nodes. And this is just a single machine model.

Nassim Taleb and Clay Christensen

In this section, we will talk about risk and disruptive innovation. Risk management is the lynchpin of machine learning. Basically, humans get it wrong, a lot. Human brains are very intuitive, but can't absorb and process millions of data points to make an informed decision. This leaves a degree of luck in every human decision.





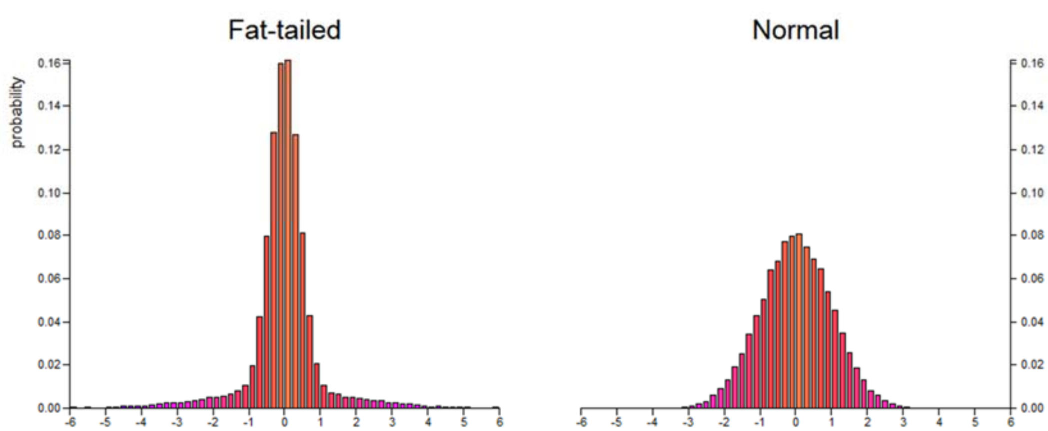
IARPA sponsors a geopolitical forecasting tournament, taking the best minds in America and asking them to forecast the probability of future world events. The Good Judgement Project estimates that about 70% of these teams' output is skill and the other 30% is based on luck (randomness). You may disagree, but none of us believes forecasting is 100% skill. Some even think 70% is too high a number for good judgement.

There is great risk (50-50) in luck. Tomorrow, you will make a decision that impacts your organization. You will either be right or wrong. Humans inside corporations are incented not to make decisions for this very reason. Worse, they rely on group think to spread the blame if luck is not on their side. This isn't the worse part. Luck mitigation leads to escalating commitments to losing propositions. Entrepreneurs, instead, are incented to make decisions because they are starting from such a small base, the start-up. They have no choice. Feast or famine. The clock is ticking.

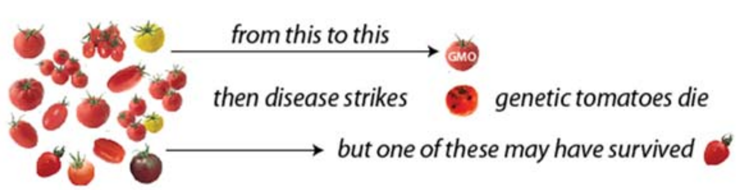
Nassim Taleb

Machine learning and artificial intelligence reduce the risk of luck (of humans getting it wrong). And this is where we bring in Nassim Taleb. The 2008 Financial Crisis was a black swan event. None of the models had a line of "code" that said humans would be incented to sell people homes they could not afford, sending defaults up because owners could not handle a normal, cyclical income shock. This leads to what's called the "risk of ruin." A bank can make money 40 years in a row and lose it all in a single year. 40 years of empirical evidence that making money is a given, just before going off the cliff. This is what we call fat tail events.

The normal distribution produces lots of variance around the average, but few events that are 3+ deviations from the mean. A fat tail distribution has the same standard deviation, but is thicker at the ends (risk of ruin, sigma events). Outcomes for both are great most of the time.



Nassim makes this point beautifully when talking about genetically modified foods. Let's use the Flavr Savr Tomato as an example. We need more tomatoes to feed our population. We need these tomatoes to ripen faster, stay red longer and be insect proof. In the past, we have relied on natural selection (normal distribution) and selective breeding. Today, we splice the DNA of a fish into a tomato's DNA. Are we producing more crops? Yes. Did the new tomato kill anyone? No. But let's not put it in the win column so fast.



There are more than 7,500 tomato varieties. If a super bug comes, it might wipe out some of the varieties, but not all the varieties. The Favr Savr may become so popular (under 1 deviation conditions) it could squeeze out this genetic diversity and leave us all at risk, relying on fewer or one variety of GMO tomato. Machine learning will model these risks.

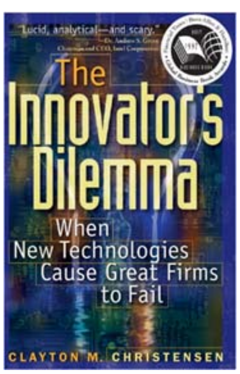
The only saving grace in business today is your competition is just as flawed as you are when it comes to human judgement and decision making. But the first companies to seize on the opportunity of using machine learning to manage risk and improve decision making will put the rest out of business very quickly. This is a revolution, not an incremental change. CEOs need to understand this right now. Inductive reasoning will be the new norm.

In 2007, McKinsey & Company had a meeting. They saw the ceiling on general, rate per hour, consulting. This realization gave us McKinsey Solutions, a technology service provider within McKinsey. This is a radical change for consulting. For 100+ years, the model may have changed, but it was always based on deploying human capital to clients. McKinsey, in a sense, is cannibalizing their current revenue stream, which is selling their judgement.

Clay Christensen

Enter Clay Christensen and disruptive innovation. You must be prepared to destroy your revenue stream if you want to stay in business. IBM knows this better than anyone. In our lifetimes, we have seen a shift from mainframes to personal computers (sold to Lenovo) to professional services and now IBM is betting the 105 year old company on a machine learning technology called Watson. Microsoft recently reorganized and committed 5,000 employees to its Artificial Intelligence division.

All are making huge investments in machine learning. Some may think this is more hype than reality. After all, we've been talking about artificial intelligence for decades (HAL 9000). But this time it's different. In the past, we have tried to replicate a fully adult brain. Today, we are creating toddler brains and letting them (machine) learn their way to adulthood. Deep Blue beat Garry Kasparov in chess back in 1996. Today, a machine can learn to beat a grand master in 72 hours. This year, Google DeepMind won what is considered the hardest game on earth, AlphaGo. How did it do it? Google gave it an outcome (label). 1 is winning. 0 is losing. Ok, go play yourself millions of times and try not to lose.



Clay tells us companies are incented not to chase their lowest profitable work. The Big Three auto makers gave the bottom to Japan. Then Japan moved upscale (Lexus+) and took a bigger chunk of General Motors. Korea (Hyundai) started going after the bottom rungs of Toyota. And now China is going after Korea. It never ends, but we know what it does to the leaders. It erodes their market share and profitability. Like McKinsey, the leaders hit a ceiling.

Andy Grove met Clay and asked about disruptive innovation. The outcome of that conversation was the Celeron processor, a very low end CPU designed to keep competitors from coming in on the bottom rung. It worked great, but recently, Intel gave NVIDIA an opening in video cards. These "gaming" cards are so powerful now, they are being used for machine learning hardware.

This is why Fortune 500 companies disappear (i.e. Kodak). Before they know it, their business has been cannibalized. They gave away their least profitable work, which created an entry point to come after the entire business.

What do Nassim and Clay teach us about machine learning? ML can model all scenarios (whether a tomato taking over crops or pilot tubes getting blocked). ML gets ahead of humans and will keep us safe and on the right course. Clay tells us many companies are about to fall. They will do business the way they have always done it and ignore the coming cognitive revolution. It will take a special leader to both understand ML and embrace it.

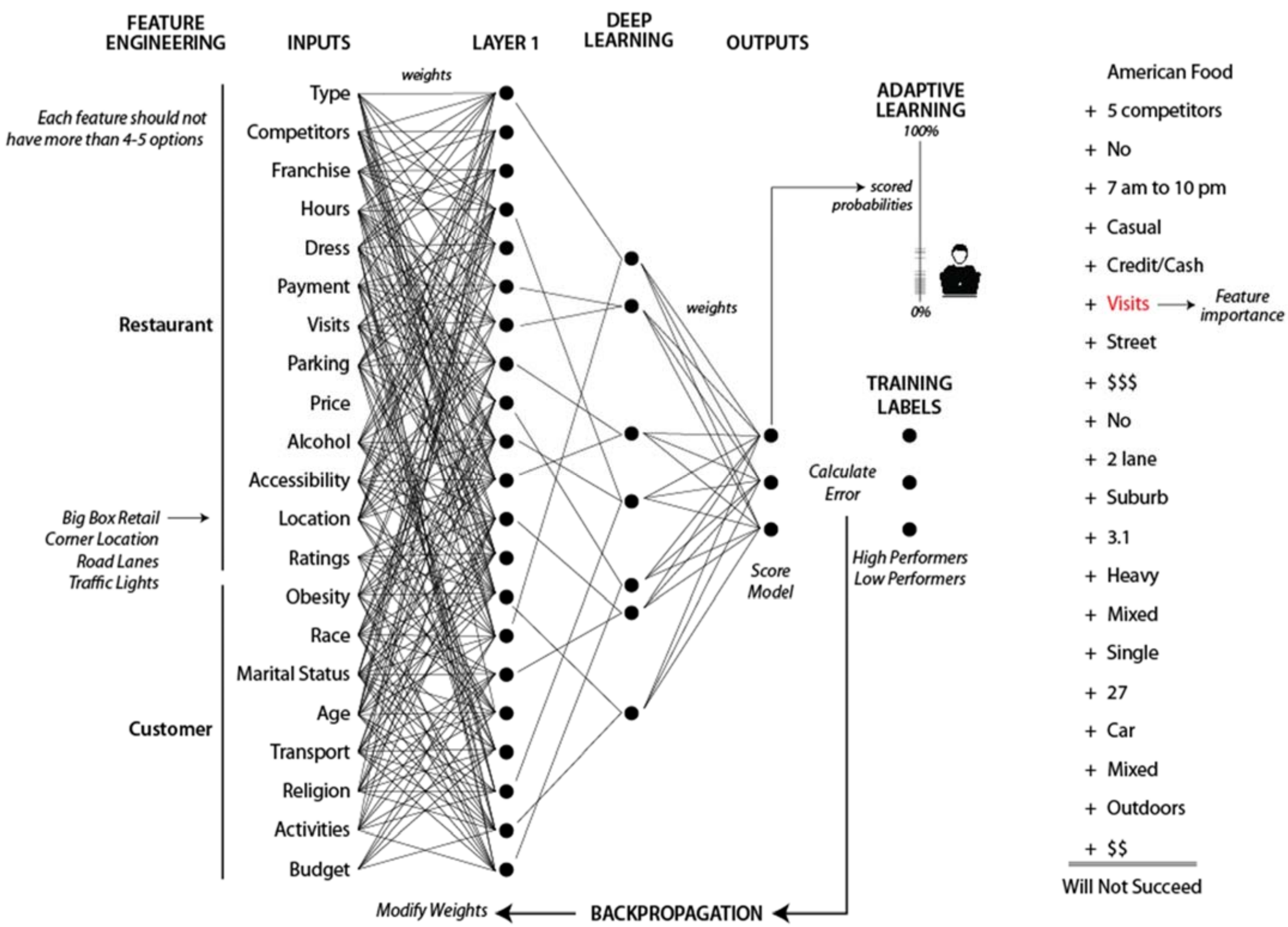
Machine Learning – Part Three

	inductive training set Supervised Learning	*descriptive no training set* Unsupervised Learning
Distinct Values *Discrete*	*classification or categorization*	*clustering*
Any Value *Continuous*	*regression*	*dimensionality reduction*

We have only begun to scratch the surface of machine learning with our decision tree example in part two above. There are two main types of machine learning, supervised and unsupervised. Supervised is where we know the outcome labels (1 = stay and eat, 0 = leave the restaurant). Unsupervised learning has no outcome label. Let's say Netflix wants to recommend movies to its users. It can take the 200, maybe 1000, features about each movie and have machine learning automatically build genres. This is called clustering. Netflix does not need to "set" or even know the genres in advance. An unsupervised model will organize all the movies for us. When a user logs in and watches a certain type of Western, the model will recommend another movie from one of the 200+ Western movie genres (clusters) that match the movie previously watched.

While a movie may have 1000 features, not all of these features are equally important and there will be redundancy between some features. Dimensionality reduction eliminates these unwanted features. Reducing features increases the potential for any one feature to play a (more) meaningful role in the model. And it dramatically reduces the times to calculate the model, which with today's limited hardware, plays an important role.

We cannot conclude this document without looking at the most promising of all the machine models – deep learning. And to understand deep learning, we must first talk about neural networks.



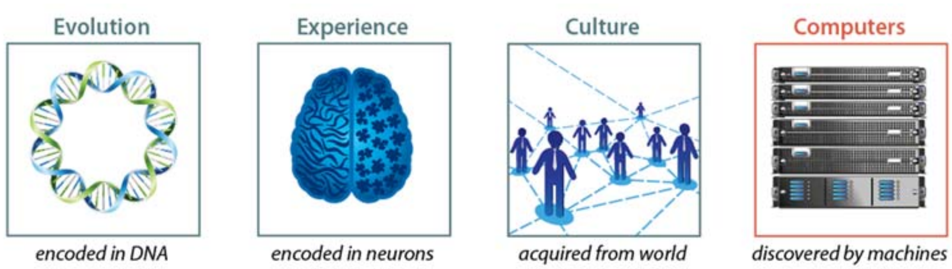
60% of restaurants fail within the first year, 80% within 5 years. Machine learning will change these outcomes dramatically. If you want to open a restaurant, first gather a lot of features about the community and citizens. We can also identify successful restaurants (1 = successful, 0 = closed). We put these features through a model that will focus on the relationships between features. For example, the type of restaurant and the number of competitors for each type is a good (feature) relationship to study. Serving alcohol and the restaurant's proximity to a large workplace or a church or a school is a good relationship to study.

Within these relationships between features, we assign weights to each feature. We then send the result of that to the next layer. That layer takes the outputs of layer one and repeats the process. Ultimately, we get down to our outcomes (restaurant succeeded or failed). If our network does not predict correctly (error), we go back (back propagation) and tweak the weights. Eventually, with the right weights between features, we get the desired outcome. We predict the success of a restaurant with 90% accuracy using our 200 different features. Deep learning is where the model adds its own layer(s) and generates additional features, based on earlier inputs. Adaptive learning is a hybrid between machines and humans, where the humans "validate" outcomes which are fed back into the network as training data.

This is why it's called machine learning. In the old paradigm, all the code is written in advance (as-is) and never changes. Data goes in, outputs come out. But with machine learning, each new row of inputs tunes the model (computer program). Models can create their own features over time. And models can keep track of their recommendations. If the model predicts success and the ultimate outcome is failure, the model will retrain based on that new output.

Conclusion

Data is the new electricity. We started with the internet of information (world wide web). Next, we focused on the internet of communication (encryption). We will soon focus on the internet of value (blockchain). And the internet of things (sensors) is well under way. But there is one internet that stands above all the rest, the internet of data (machine learning and AI). All internets generate and consume data.



The internet of data will surpass the industrial revolution as the single most important event in human history. Computers will learn more than evolution, experience and culture combined. And it will learn much faster.

The internet of data will cure cancer and other diseases. It will dramatically reduce the cost of healthcare. The internet of data will reduce risk, eliminate regulation and improve decision making. The internet of data and value will bring people out of poverty and remove safe havens for criminals. The internet of data and things will monitor and repair our infrastructure. The internet of data will track a blood diamond to your fiancé's finger. It will help us avoid, prepare and overcome natural disasters. The internet of data will eliminate the need for many companies. Skyscrapers will be relics. Earth will transition to a peer-to-peer, sharing economy. Intermediaries will not be needed.

Is Elon Musk right? After the industrial revolution took our muscles and the cognitive revolution takes our brains, what we will be left with is creativity, emotional intelligence and social well-being. The internet of data cannot create. It can only manage what it has seen. Without the burden of managing the earth, all human energy can be spent making the earth a better place to live.