ARTIFICIAL INTELLIGENCE IMPLEMENTATION BOOT CAMP

Our Agenda

- Part 1: AI – Introduction and working definitions
- Part 2: Big Data and its relationship to AI
- Part 3: Implementing machine learning
- Part 4: Machine intelligence and customer experience
- Part 5: Machine intelligence & cyber security
- Part 6: Teaming and internal capability
- Part 7: Python Examples

Part 1: Introduction

Welcome. Did you know…

- AI is more than 40 years old.
- Only 10,000 people in the world can actually “do” serious AI.

Welcome. Did you know…

- Only 23 % of companies have actually deployed AI.
- Only a small percentage (approx.. 4%) are actually performing substantive, state-of-the-art AI operations
- 54% either have no plans to deploy AI applications or have not begun implementing their planned solutions.

– David Kiron, MIT Sloan Management Review

Introductions

- What is your Name and Job Role?
- Your company or team?
- Expectations for the class. Why are you here?
- Name something interesting about yourself.

What to expect from this class

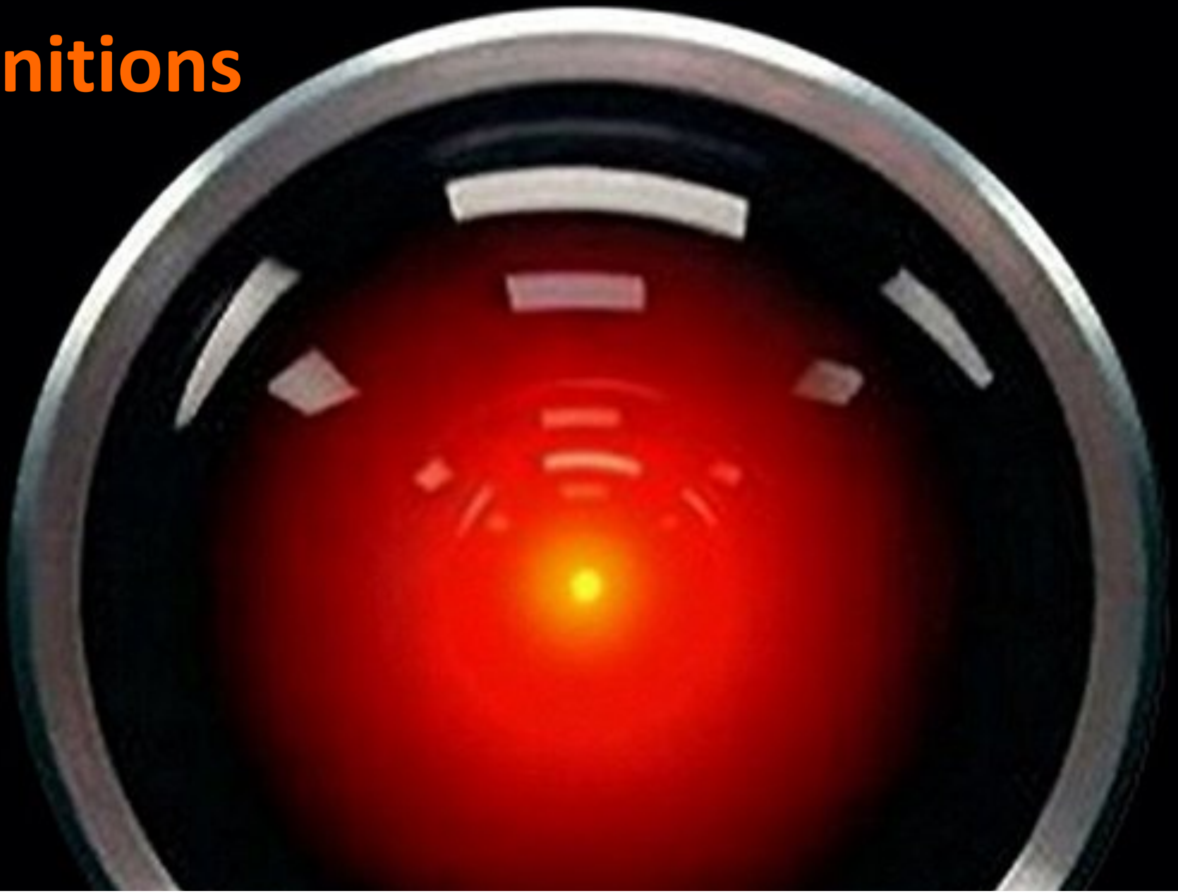
- Flexibility
- Conversations
- Literacy and awareness of what’s possible with AI right now
- General examples of good AI use cases
- Focus on history and current landscape of AI and high-level overview of machine learning
- Focus on how AI and machine learning can help your organization so you can act on what you learn

What not to expect from this class

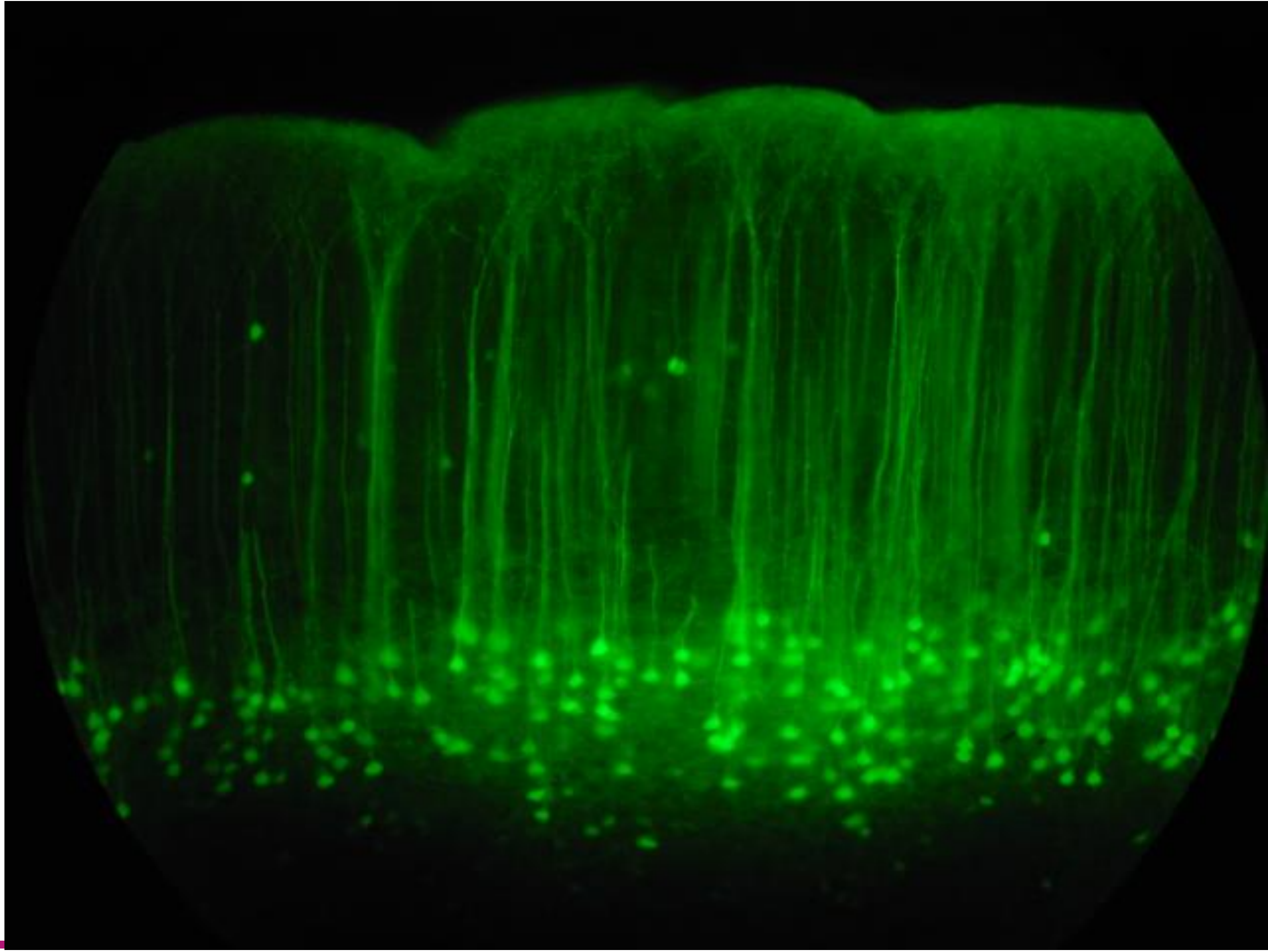
- Prescriptions, formulas, one-size-fits-all solutions
- Perfect and/or mature best practices
- Rigid processes or step-by-step instructions
- Big overnight transformations
- Extended technical discussions or deep focus on any specific technology or tool

AI: Working Definitions

- AI
- Machine Learning
- Deep Learning
- Data Science
- Data Model
- Big Data



What is intelligence?



AI: Working Definitions

- **AI**
- Machine Learning
- Deep Learning
- Data Science
- Data Model
- Big Data

Artificial Intelligence - Definition

An intelligent agent that perceives its environment and makes decisions to maximize chances of achieving its goal.

3 Types of AI:

- Artificial Narrow Intelligence
- Artificial General Intelligence (strong AGI)
- Artificial Super Intelligence

Artificial Intelligence - Definition

An intelligent agent that perceives its environment and makes decisions to maximize chances of achieving it’s goal.

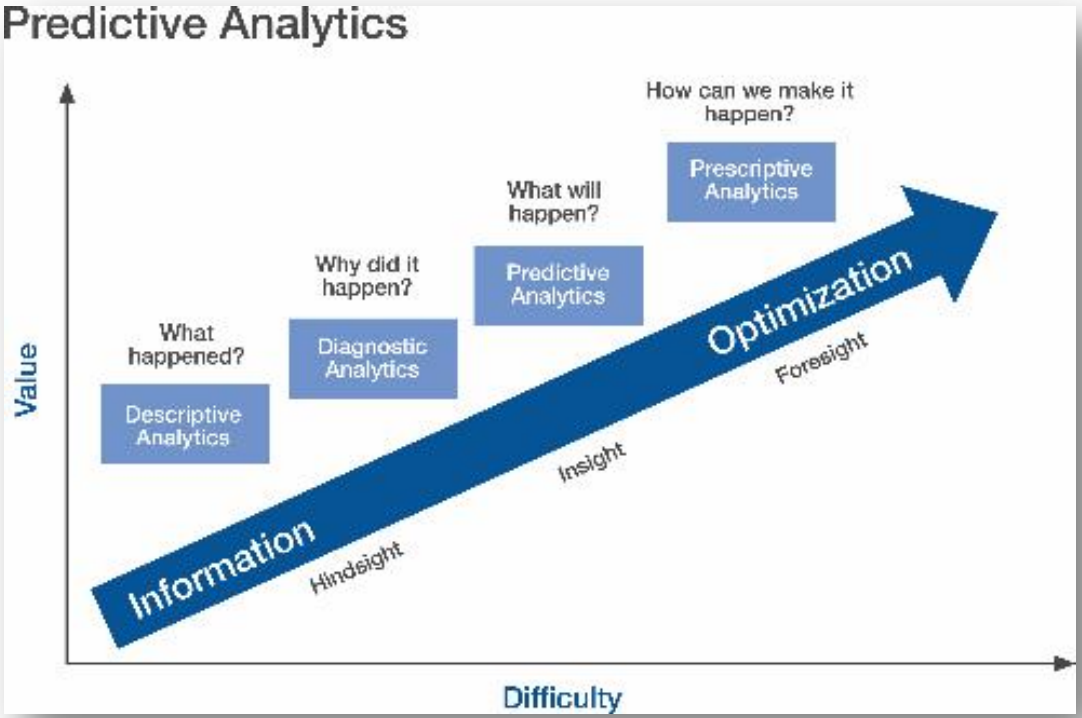
3 Types of AI:

- Artificial Narrow Intelligence
- Artificial General Intelligence(strong AGI)
- Artificial Super Intelligence

Let’s get these out of the way first

AI Stages	**Artificial Narrow Intelligence (ANI)** Execute specific focused tasks, without ability to self-expand functionality	**Artificial General Intelligence (AGI)** Perform broad tasks, reason, and improve capabilities comparable to humans	**Artificial Super Intelligence (ASI)** Demonstrate intelligence beyond human capabilities
Timing	Today	About 2040?	Soon after AGI
Implications	Outperform humans in specific repetitive functions, such as driving, medical diagnosis and financial advice	Compete with humans across all endeavors, such as earning university degrees and convincing humans that it is human	Outperform humans, helping to achieve societal objectives or threatening human race

Levels of Algorithmic Sophistication



AI Today…

Many organizations want to deploy AI but do not have enough, or consistent, or strategically structured enterprise data architecture to do so. Microsoft explained that many orgs simply don’t have the historical data in place in order for AI algorithms to learn enough to be useful. For instance – one company wanted to use AI to predict future failures in their system, but they did not have data on past failures in order to train their AI solution.

Data is the next oil

- Clive Humby, 2006

If you have data, we are in business !!

AI Today…

- Hurdles

- AI solutions are currently incomplete. They are like very young children who need to learn in order to be productive and useful.
- If orgs do not understand how to train their AI solutions, AND the organized stores of data upon which to build: these are big hurdles to adoption.
- In one well-known example, *A BANK* cannot exploit many potential AI applications because the data stores in their organization are so fragmented. So the underlying substrate of data is not in a state that is ready to harness AI.

AI: Working Definitions

- AI
- **Machine Learning**
- Deep Learning
- Data Science
- Big Data

Machine Learning – Definition

Machine learning is a subfield of artificial intelligence. Its goal is to enable computers to learn on their own. A machine's learning algorithm enables it to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models.

Machine Learning - Types

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Example: Supervised Learning for Classification

With supervision, machine models can be effective at learning how to categorize things like:

- Is this email spam or not
- Is this a fraudulent charge
- Is this photo a dog or cat
- Is this fake news or not

What problem are you trying to solve? (think small)

- No such thing as machine learning fairy dust
- Ask small questions first
- Big problems are a mass of small problems, you can make more progress by tackling small ones first
- Chances of triggering change on a small problem are much greater

Machine Learning - Data

- Importance of Data
- Structured vs. Unstructured
- Cleaning Data
- Training, Validation, and Test

Machine Learning – Structured Data

- Pre-defined and machine readable.
- Usually has a relational data model.

- Library catalogues
- Census records
- Weather
- Databases

Lot Frontage	LotArea	House Style	Year Built	Year Remod	Roof Style	Total Bsmt SF	Heating	1stFlr SF	2ndFlr SF	Full Bath	Half Bath	Bedroom	Fire places	Garage Type	Garage Cars	Garage Area	Pool Area	YrSold	SalePrice
65	8450	2Story	2003	2003	Gable	856	GasA	856	854	2	1	3	0	Attchd	2	548	0	2008	208500
80	9600	1Story	1976	1976	Gable	1262	GasA	1262	0	2	0	3	1	Attchd	2	460	0	2007	181500
68	11250	2Story	2001	2002	Gable	920	GasA	920	866	2	1	3	1	Attchd	2	608	0	2008	223500
60	9550	2Story	1915	1970	Gable	756	GasA	961	756	1	0	3	1	Detchd	3	642	0	2006	140000
84	14260	2Story	2000	2000	Gable	1145	GasA	1145	1053	2	1	4	1	Attchd	3	836	0	2008	250000
85	14115	1.5Fin	1993	1995	Gable	796	GasA	796	566	1	1	1	0	Attchd	2	480	0	2009	143000
75	10084	1Story	2004	2005	Gable	1686	GasA	1694	0	2	0	3	1	Attchd	2	636	0	2007	307000
NA	10382	2Story	1973	1973	Gable	1107	GasA	1107	983	2	1	3	2	Attchd	2	484	0	2009	200000
51	6120	1.5Fin	1931	1950	Gable	952	GasA	1022	752	2	0	2	2	Detchd	2	468	0	2008	129900
50	7420	1.5Unf	1939	1950	Gable	991	GasA	1077	0	1	0	2	2	Attchd	1	205	0	2008	118000
70	11200	1Story	1965	1965	Hip	1040	GasA	1040	0	1	0	3	0	Detchd	1	384	0	2008	129500
85	11924	2Story	2005	2006	Hip	1175	GasA	1182	1142	3	0	4	2	BuiltIn	3	736	0	2006	345000

ASPE

Machine Learning – Unstructured Data

- There's no predefined model
- It's often text|audio|video

- Email
- Newspapers
- Health Records
- Books
- PDF Documents

As data and the business problems that can be addressed by it proliferate, our ability to analyze them is falling behind. We don't have enough data scientists, we can't create enough good models, and we can't get them into production. Enter automated machine learning (AutoML), which offers substantial potential for solving the problem. This powerful set of tools can help with a wide variety of ML activities, including preparing data for analysis, performing feature engineering, automatically generating well-fitting models using the best algorithm, and generating code or APIs to help deploy the

We've reached a tipping point that's prompt actively exploring how AI can help them achi transformation goals

With the advent of DevOps and Continuous Delivery, businesses are now looking for real-tin stages of the software delivery cycle.

Although Artificial Intelligence (AI) is not really new as a concept, applying AI techniques to reality just the past couple years. Down the line, AI is bound to become part of our day-to-d prior to that, let us take a look at how AI can help us achieve our quality objectives.

Day after day, QA Engineers face a plethora of difficulties and waste a lot of time to find a new additions, the existing code which has already gone through the testing process may sh

AI: Working Definitions

- AI
- Machine Learning
- **Deep Learning**
- Data Science
- Big Data

Deep Learning – Definition

A subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multi-layered artificial neural networks to vast amounts of data. The levels in these learned statistical models correspond to different levels of concepts, where higher level concepts are defined from lower level ones, and the same lower level concepts can help to defined many higher-level concepts.

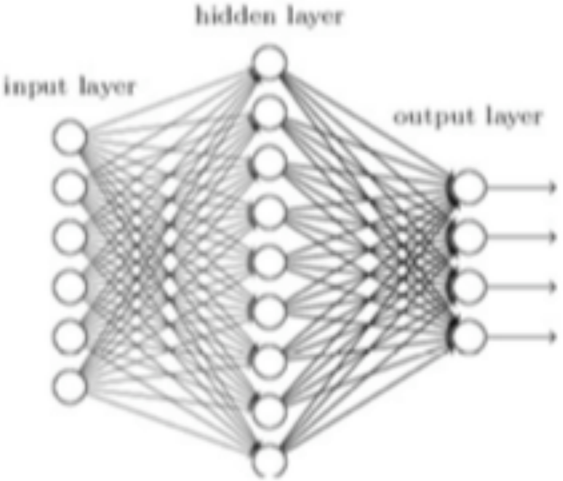
Neural Networks – How they learn

How do they learn?

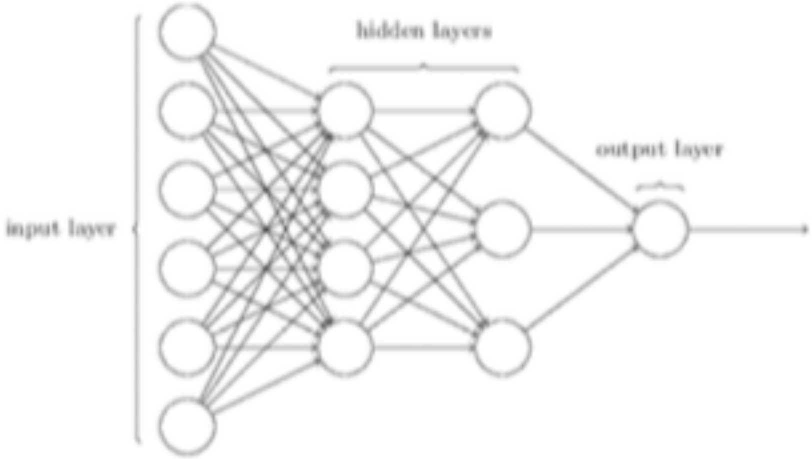
- Setup Network (defined schedule) up so that it takes in inputs and produces output
- If the output does not match desired output network creates an error signal
- Network will pass the error signal backwards through the network. Learning occurs by adjusting the weights to match the inputs and outputs

Neural Network – Deep Learning

- Not Deep



Deep Learning



Neural Networks – What are they good at?

- Automatic speech translation
- Recognize traffic signs > people
- Google maps read all building #s
- Baidu: image understanding
- Sentiment analysis
- Read Chinese ~ native levels
- Write sentences as image captions
- Medicine: cancer research & insights
- Medical decisions & diagnostics in 15 minutes

Neural Networks

- **Association between human brains and neural networks is inspired, but weak**
- **Neural nets seem particularly good at human-like tasks such as:**
  - Recognizing images
  - Recognizing speech
  - Recent work over past few years suggest they are good at predicting

AI – Challenges of "Black Box" vs. "White Box"

- Sophisticated AI applications can operate in ways that are opaque to humans, even designers of the technology. For example, deep learning that has been used for self-driving cars is not understood.
- This presents problems for adoption, trust, buy-in, compliance, troubleshooting, and "Explainability"

AI: Working Definitions

- AI
- Machine Learning
- Deep Learning
- **Data Science**
- Big Data

Data Science - Definition

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is the same concept as data mining and big data: "use the most powerful hardware, the most powerful programming systems, and the most efficient algorithms to solve problems".

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Data Science - Definition

In 2012, when Harvard Business Review called it "The Sexiest Job of the 21st Century", the term "data science" became a buzzword. It became conflated with:

- **Business analytics**
- **Business intelligence**
- **Predictive modeling**
- **Statistics** *(Nate Silver referred to data science as a "sexed up term for statistics.")*

While many university programs now offer a data science degree, there exists no consensus on a definition or suitable curriculum contents.

Many data-science and big-data projects fail to deliver useful results, often as a result of poor management and utilization of resources.

Who is a real data scientist?

Data scientists are highly qualified and can be hard to hire and retain. They are the key professional role behind AI and advanced analytics.

AI: Working Definitions

- AI
- Machine Learning
- Deep Learning
- Data Science
- **Big Data – We will spend some time on it in the next section of class**

Part 2: Big Data

What is Big Data?

The 3 V's

- **Volume**
  Large Quantities (think gigabytes, terabytes of information daily)
- **Velocity**
  Speed of generation and/or need to be processed quickly
- **Variety**
  Data might be structured, unstructured, varying sources
- **And some additional considerations…**

Other Considerations

- **Variability**

  Is your data consistent? Does it have holes in it?
- **Veracity**

  How accurate are your data sets?
- **Complexity**

  The more complex your data, the “bigger” it is. Do you have multiple sources that much be linked, connected, or correlated in order to grasp it’s meaning?

New Data Sources

- Logs
- Sensors / RFID
- Location / Geospatial
- Video / Recognition
- Social / Natural Language

Power Sensors

Driving Sensors

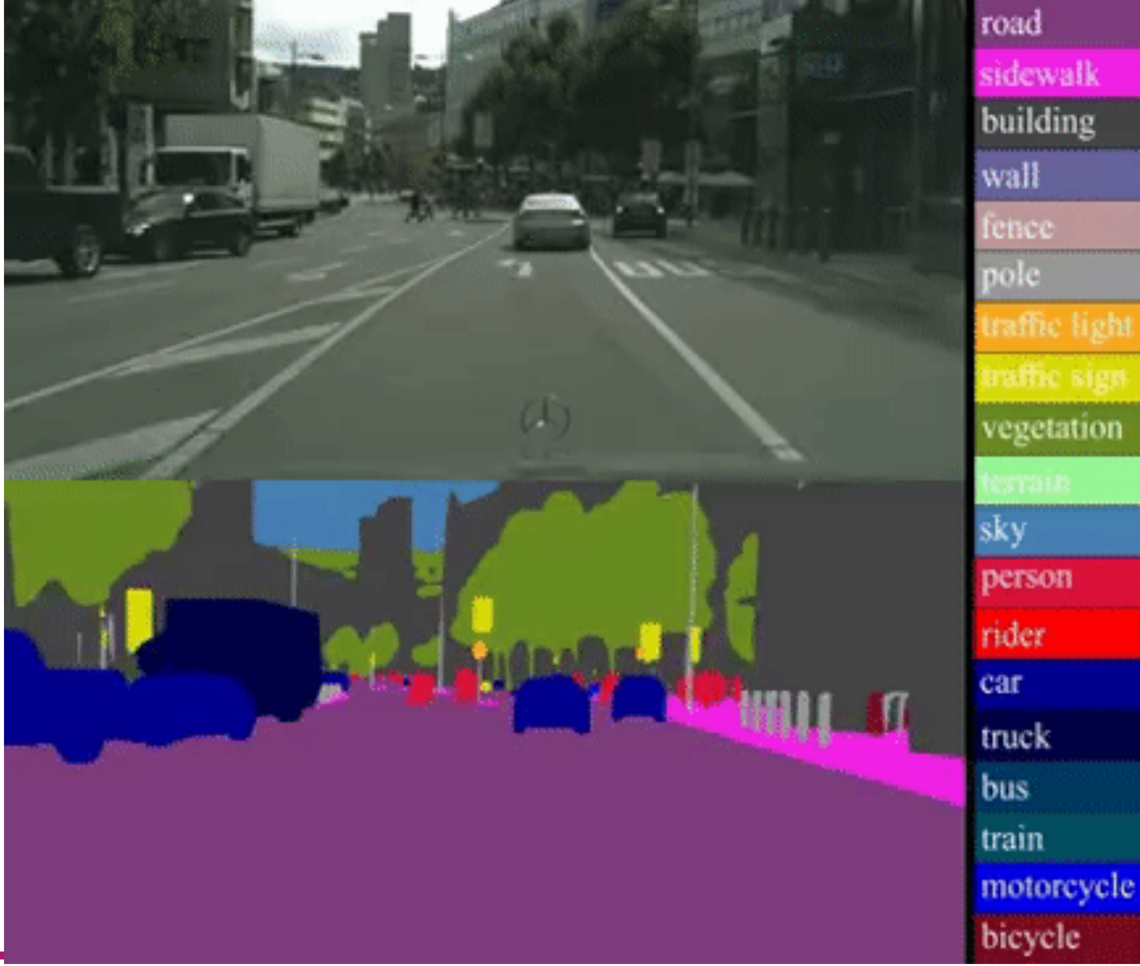
Helmet Sensors

IoT – Internet of Things

E.g. Sensors

Machine-generated
Collection
Utilization

Machine vision



The big data prerequisite

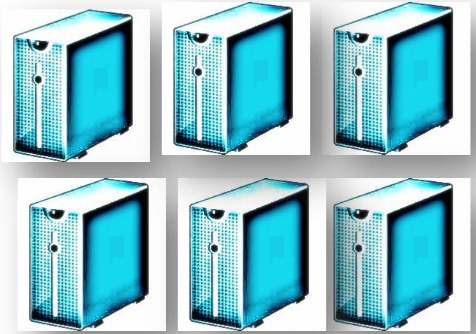
15 years ago…

- Big, exotic supercomputers (Cray J90 pictured)
- 2.5 million dollars
- Big, important projects
- Wait lists…PhDs

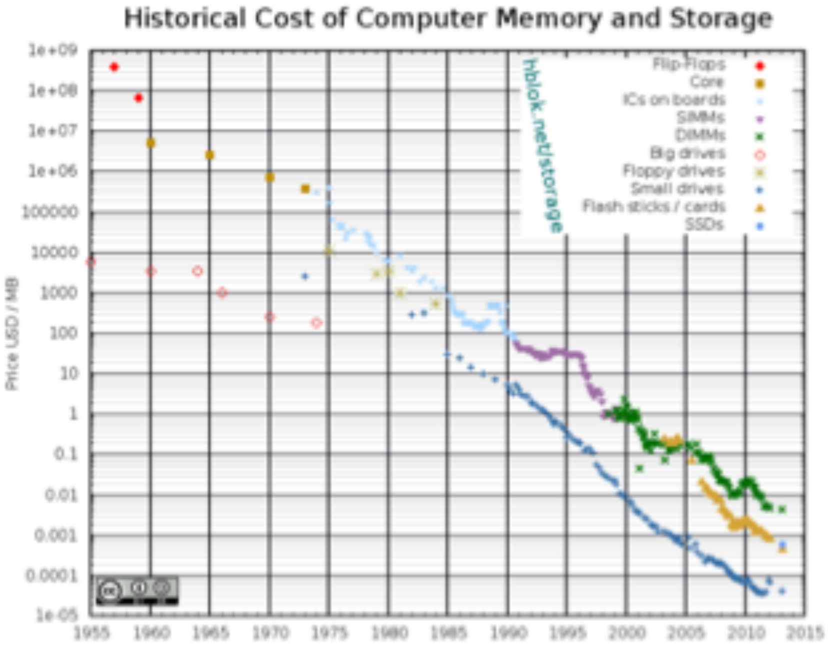
Virtualization

2006

Today



Compute capability, costs, and performance



Applications of Big Data

What can we do with it?

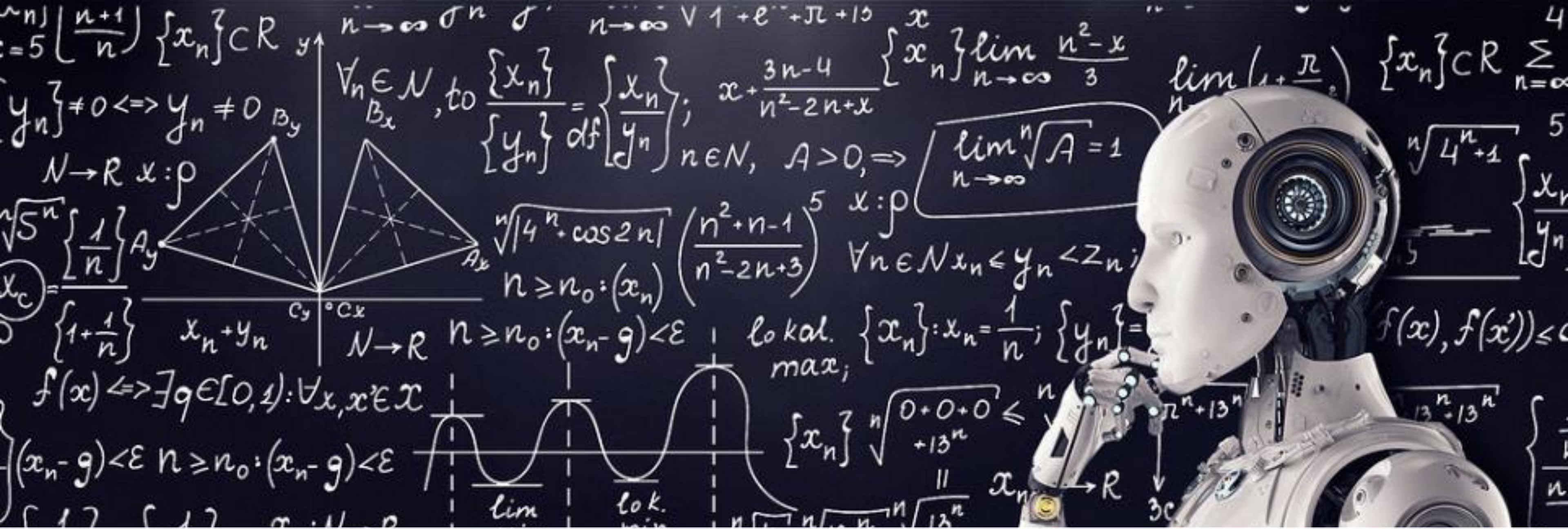
- Data warehousing
- Business intelligence
- Analytics
- Predictive statistics
- Data science

In Summary

- There are many, many sources of data
- The speed of data creation is increasing
- We need to process it
- We need to make sense of it
- We can use it to establish advantage

Big data and AI: the main takeaway

- Effective AI is usually powered by machine learning
- Requires existing datasets to learn
- While there are some interesting things AI can do with local machines and small datasets, in general the most exciting applications rely on very large datasets



Part 3: Implementing Machine Learning

Pillars of a Successful AI Team

- Algorithms
- Modeling
- Business Case & Business Integration
- Domain Expertise
- Data
- Automation
- Scalability

Cross Industry Standard Process for Data Mining (CRISP-DM)

Tools

Machine Learning: Two main types

- Supervised
- Unsupervised

Supervised Learning – Components

- Create a function *f* that can map input X to an output Y with minimal loss.
- Features(input or X) – independent inputs of some specific variable
- Labels(output or Y) – dependent variables you are predicting from a feature or set of features.

Supervised Learning – Types

- **Continuous Variable Prediction** – predicting a continuous value from a feature or set of features. Predict income based on years of education. E.g. Regression
- **Discrete Variable Prediction** – assign a discrete value from a feature or set of features. Predict whether income is higher or lower than 50k based on years of education. E.g. classification

Unsupervised Learning

- Initially no labels, only features
- AI finds the underlying structure of data
- AI groups data into new groups

Unsupervised Learning - Types

- Clustering
- Dimensionality Reduction

Common Algorithms

- K-means Cluster
- Linear Regression
- Decision Trees
- Random Forest
- Naïve Bayes
- Neural Networks

Unsupervised Learning - Clustering

K-Means Clustering

- Common type of clustering
- Create data points such that points in different clusters are dissimilar while points within a cluster are similar.
- Cluster data into K groups
- Larger K creates smaller groups with more granularity
- Lower K creates larger groups with no granularity

Unsupervised Learning - Clustering

K-means Steps

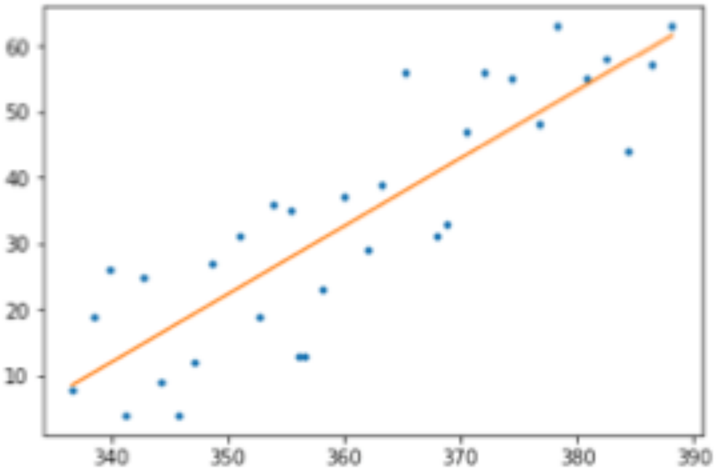
- Step 1: K centroids are randomly selected
- Step 2: Assign data point to closest centroid's cluster (most commonly Euclidean distance)
- Step 3: Distance between the points and the closest centroid is minimized with each iterations of the algorithm by moving the centroid.
- Step 4: Move the centroids to the center of clusters
- Step 5: Repeat steps 3 and 4 until centroid stops moving a lot. This is called converging.

Demos:

- https://www.youtube.com/watch?v=BVFG7fd1H30

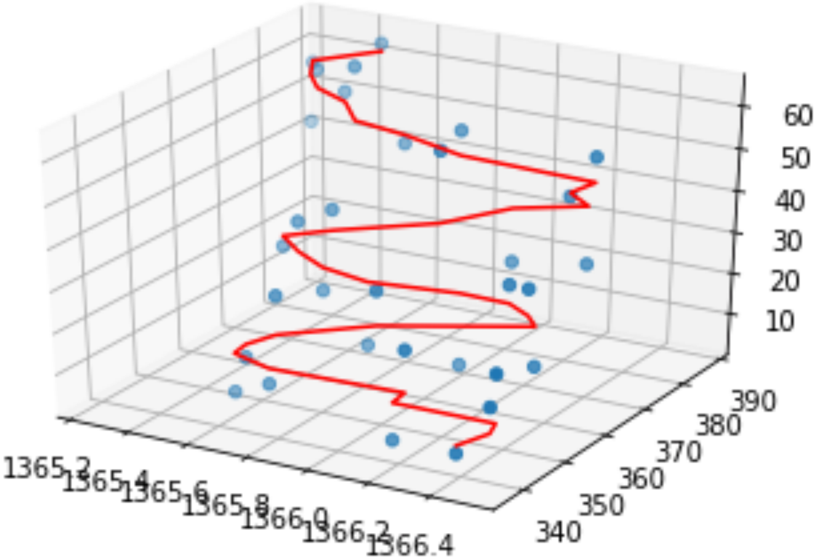
Linear Regression

	Year	CO2	Temp
0	1979	336.67	8
1	1980	338.57	19
2	1981	339.92	26
3	1982	341.30	4
4	1983	342.71	25

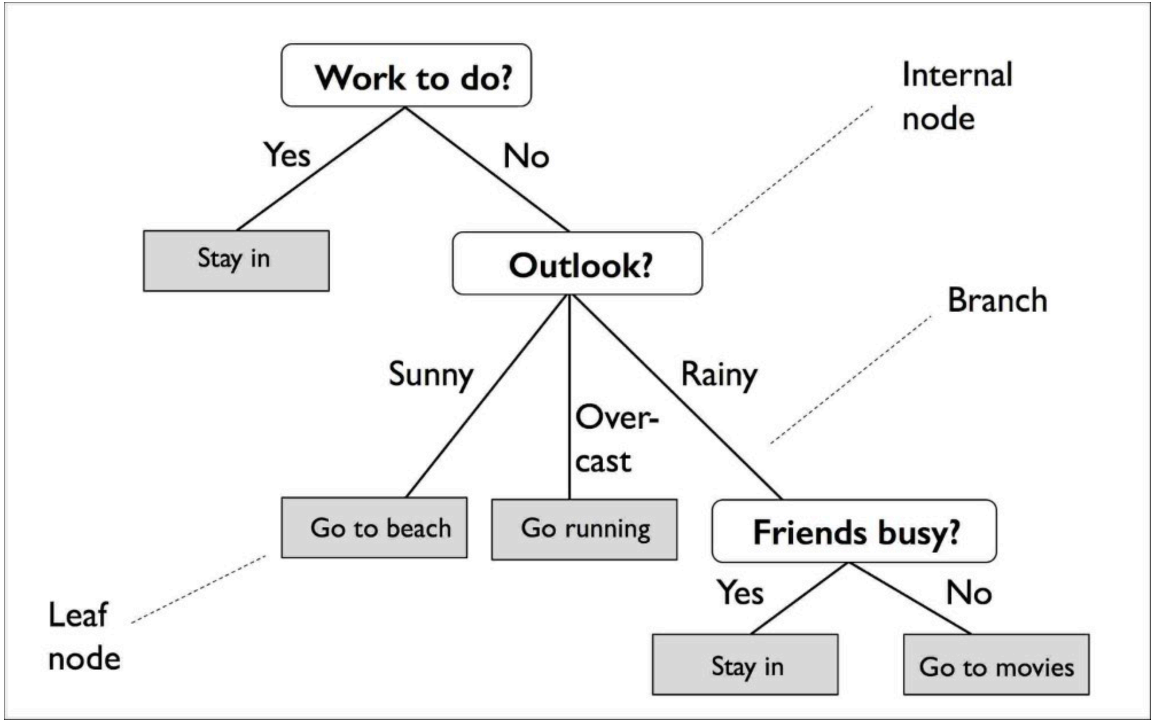


Multiple Linear Regression

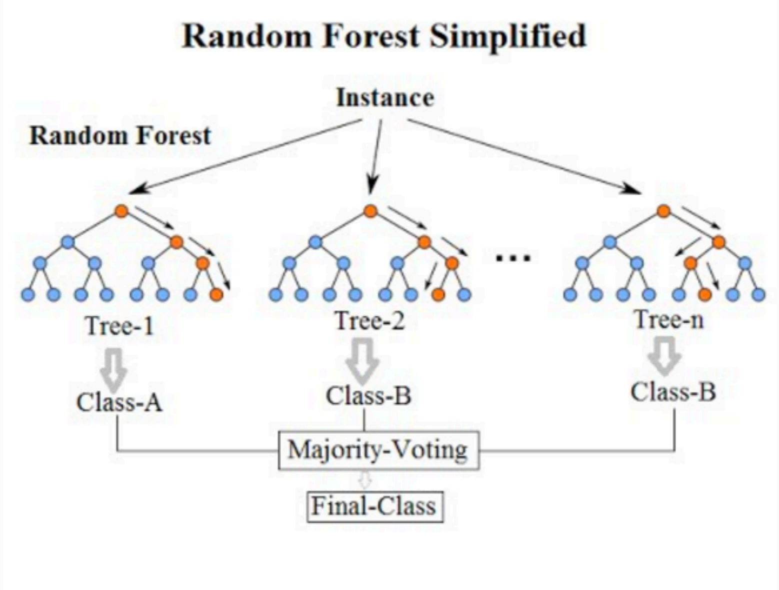
	Temp	CO2	Solar
0	8	336.67	1366.43
1	19	338.57	1366.51
2	26	339.92	1366.51
3	4	341.30	1366.16
4	25	342.71	1366.18



Decision trees



Random Forest



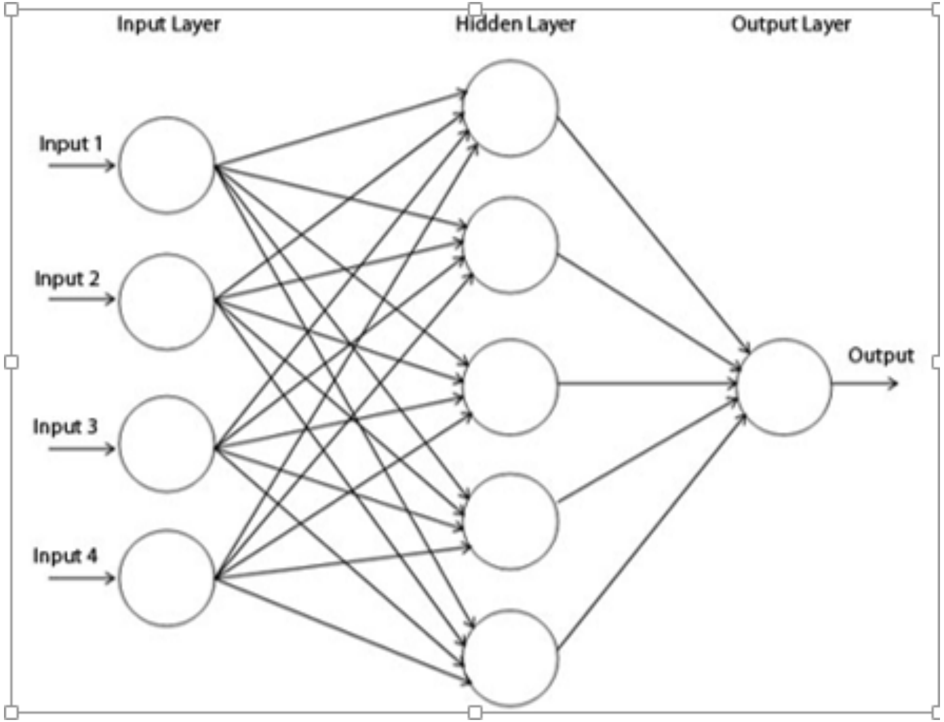
Naïve Bayes

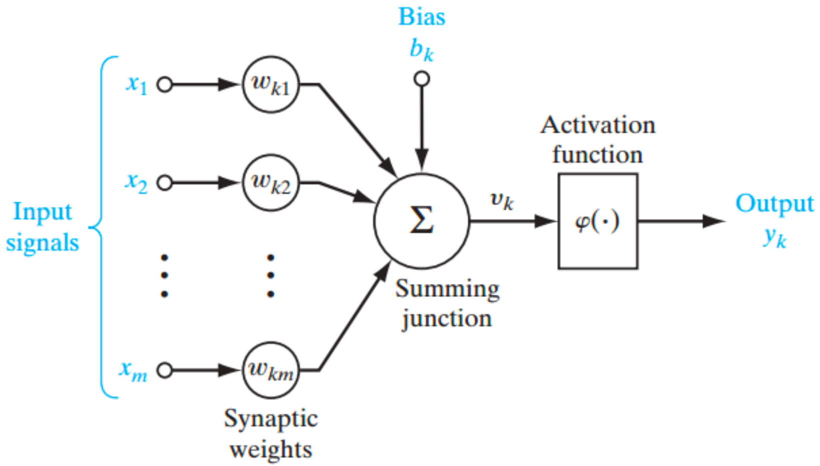
- Built on Bayes Theorem
- Assumes observations are independent (naïve)
- Scales well to very large datasets
- Predictions are either very close to 0 or 1, thus good in a relative manner

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

$$p(C_k|x) \propto p(C_k)\prod_{i=1}^{n} p(x_i|C_k)$$

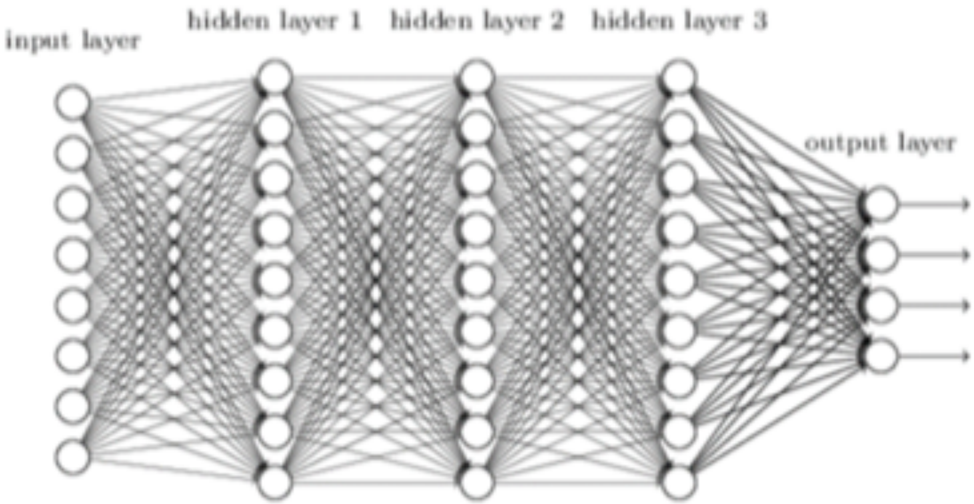
Neural Network





Deep Neural Network – Deep Learning

Deep NN has
More than one
hidden layer



Common Methodologies

- Split Train and Test
- Cross validation

Train, Validate, Test

- Split data into 3 groups randomly
  - True random can be hard
  - 80/10/10 is a good rule-of-thumb
- Validate set makes sure you don't overfit and sets hyperparameters
- Test set is a final sanity check on your hyperparameters

Practical Problems

- Model overfitting/underfitting
- Handle Missing Values
- Handle Nominal Values
- Handle imbalanced data

Model Overfitting

- Extreme overfitting = memorizing every data in the training set
- High accuracy on training data
- Low accuracy on new data!

Model Overfitting - Overcome

One of the reasons for overfitting is highly non-linear decision boundary, that means high degree polynomials are contributing to the decision boundary the most.

- **Regularization**
  - Penalty for higher order polynomials

  E.g. $y = f1(x) = x^1 + x^2$ is a non linear function

  $y = f2(x) = x^1 + x^2 + x^3 + x^4 + x^5$ is highly non-linear

  If f2(x) leads to overfitting, penalize higher order of x to decrease non-linearity

Model Overfitting - Overcome

- Values or settings that impact how the model trains, and the potential for overfitting
- Need to train model many times with different hyperparameters to find ideal setup

Handle Missing Values

Missing values are common in real data

- Remove records with missing features
- Use descriptive statistics to impute values
  - Use mean/median/mode
  - Use value that are more common
- More advanced – Predict and impute

Handle Nominal Values

Models understands only numeric values (except few like decision trees)

- Encode nominal values
  - Create vectors of 0 and 1 of size = number of distinct values
  - One hot encoding
- Advanced – use word embedding models

Imbalanced Data

Cancer prediction classifier [cancer=1, no cancer=0]

Prior probability of having cancer = 0.1% means 1 in 1000 suffer from cancer.

If you train a model that predicts 0 for everything. What is your model accuracy ?

For non cancer class, accuracy = 99.9%

For cancer class, accuracy = 0%

Handle Imbalanced Data

- Equal sample of all classes
- Oversample and provide lesser weights during training (can be used mostly in sequential models like Gradient Boosted Trees)
- Generate synthetic data

Accuracy Metrics

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root mean square error (RMSE)
- Precision, Recall, F-Score
- AUC/ROC curve

Mean Absolute Error

- Measure of difference between two continuous variables.

MAE = sum(y-x) over n

Mean Squared Error

- Measures the average of the squares of the errors that is, the average squared difference between the estimated value and the actual value.

MSE = square(sum(y-x)) over n

- Root mean squared error (RMSE): Take the square root of MSE

Precision, Recall, F-Score

True positives (TP):- actual label=1, predicted label=1

False positives (FP):- actual label=0, predicted label=1

True negatives (TN):- actual label=0, predicted label=0

False negatives (FN):- actual label=1, predicted label=0

- **Precision**:- TP/(TP+FP) = TP/(actual results)
- **Recall**:- TP/(TP+FN) = TP/(predicted results)

Precision, Recall, F-Score

Accuracy: (true positive + true negative)/total

THIS IS RISKY !!

Consider cancer example, accuracy will be 99.9% but the model performs 0% in detecting cancer

F-Score: 2*(PR)/(P+R)

ROC, AUC

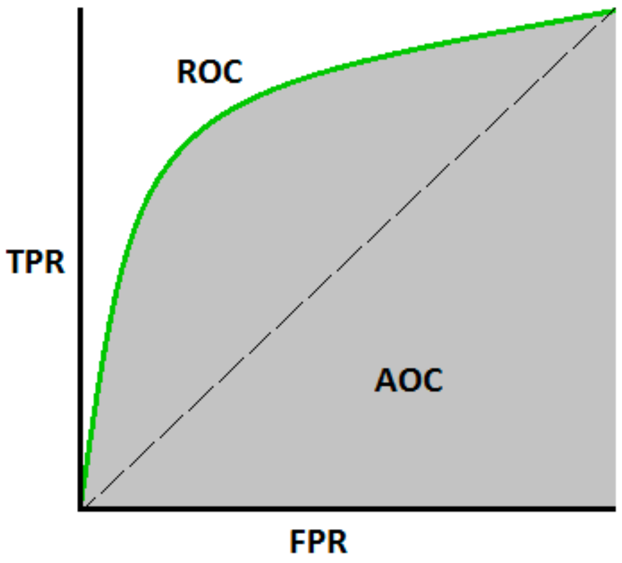
True positive rate (TPR or recall):

TP/(TP+FN)

False positive rate (FPR): FP/(TN+FP)

AUC: area under the curve

ROC: Receiver operating characteristics



Encoding Domain Expertise

- **Feature Engineering** - simple algorithms to combine two or more datapoints
- Ways to build collaboration between business units
- Often related to KPIs
- Can also uncover hidden business assumptions

Machine Learning Model Management

- Every day you get new data and new ideas
- Models should be constantly retrained
- Need to record which model made which prediction
- Might have multiple models in production (A/B testing)

Where is AI valuable?

The #1 barrier to AI is often not the tech itself – it's how to figure out how to use AI for the applications it is good at performing.

So what do many high-profile uses of AI have in common?

Unfortunately, they suck. Not from a technological perspective, as technical capabilities are growing and exciting. Google's self-driving car initiative has been underway for more than 10 years now.

But consider end users: is there any demonstrable value? Do we go any faster compared to a human-driven Uber?

Where (and why) does AI suck?

AI offers "subhuman automation"

- Usually can't handle long tail situations
- **Context** – can't fill in gaps in information as a human, because humans can empathize
- **Expertise** – can't reason or explain as a human could
- **Trust** – can't gain confidence

What does AI do well?

There is however good progress with transformative potential. Treat AI as a tool for "superhuman insight." When applied correctly AI has the ability to outperform any individual or team:

- **Data** – Can handle what's beyond the scope of a person to digest
- **Analysis** – ability to compute in ways humans cannot
- **Complexity** – Can detect patterns too complex for humans to envision
- **Speed** – it's a no brainer…this is where AI dominates over humans

Example: Self-driving cars? Consider instead driver assistance and navigation

Example: Medical recommendations? Consider instead imaging: AI is better than human doctors at identifying benign vs. malignant skin tumors

A few successful examples

Rethink AI for “superhuman insight.” Consider some successful examples:

- Starsky robotics for semi-autonomous trucks
- IBM Watson for oncology – (physician recommendation systems)
- Better results come from leveraging AI and humans effectively
- Encourage adoption

Part 4: Machine Intelligence as Part of Customer Expérience

AI Transforming Customer Experience

- Empowering Self Service
  - Chat bots
  - Question & Answer bots
- Personalization
- Better service by analyzing pain points
- 24*7 support
- Customized Offers
- Customer Segmentation & personalized services

AI Customer Experience Use Cases

- Smart email content curation
- Improved customer experience - Agent assist
- Intelligent chatbots
- Increased productivity
- AI assisted customer insights
- No human loop – AI can work without break

AI and customer experience

Gartner predicts that in the future the vast majority of customer interactions with a company with happen without human customer service rep. They also posit that a well-designed chatbot can handle 80% of customer inquiries. Consider just a few tasks that are fully serviceable by an autonomous agent:

- **Book flights**
- **Buy tickets**
- **Order food**
- **Schedule rides**
- **Purchase products**
- **Send order updates**
- **Book appointments**

First hand customer data

- Topic Modeling – define intent
- Understanding Pain points
- Did the agent solve the problem ?
- Most optimal way that solved the problem
- Use chat, call transcripts, reviews, forums

Personalization

- Customer profiler
- Personalized services and offers
  - Personalized content
- Intelligent and personalized call routing

CLV and Churn

- Customer Lifetime value prediction
  - Based on heuristics
- Churn analysis
  - Targeted customers
  - Reason behind Churn will provide more insights

Part 5: Machine Intelligence & Cyber Security

How can ML help with security

- 61% of enterprises say they cannot detect breach attempts without the use of AI technologies
- 48% say their budgets for AI in Cybersecurity will increase by an average of 29% in Fiscal Year 2020
- Breach attempts are proliferating with CISCO reporting that in 2018, they blocked seven trillion threats on behalf of their customers

Most common use-cases

- Automated Network Analysis
- Email Scanning
- Fraud detection
- Malware Detection
- Intrusion Detection
- Scoring risk in a network
- User/machine behavioral analysis
- Anomaly Detection

Automated Network Scanning

- Huge volume of data
- Most Malwares and cyber-attackers operate over the network
- Monitoring communication systems
- Detect attempted installations of malware
- ML use a combination of - Keyword matching, statistics monitoring, anomaly detection, check http header values, embedded DNS requests – which can be easily overlooked by humans or rule based engines

Email Scanning

- Spam email classification
- Phish email classification
- Use of NLP to check word choice and grammar
- Simulation clicks on all links in the email
- Examine resulting pages for signs of phishing
- Use computer vision simulate the look and feel of phish emails
- Go over attachments

User behavior Modeling

- Some AI based cyber defenses model the behavior of users on the system
- Detect and remediate account takeover attacks
- Checks stolen identity
- Initiate account lockout

Antivirus

- Traditional anti-virus were signature based
- Need manual deployment of new signatures
- Signature list grows exponentially
- AI – detect unusual behavior by programs rather than matching signatures
  - Most malwares are designed to do things different from the standard operation

Fraud Detection

- Incorporating timeseries data often breaks independence assumptions
  - Fewer assumptions = more expensive and slower projects
- Secondary benefit of better understanding customer habits.

Modeling previous breaches

"It has been said critically that there is a tendency in many armies to spend the peace time studying how to fight the last war."

- Lt. Col. J. L. Schley, 1929

Part 6: Teaming and Internal Capabilities

Teaming and Internal Capabilities

PART 1 – “Human Architecture”

Data analysis roles

Champions

- Executive decision makers
- Rely on analytics to understand business process & decisions
- Support and advocate AI driven business capabilities

Professionals

- Knowledgeable in advanced analytics
- Create models and hold advanced degrees

Semiprofessionals

- Apply the models and algorithms
- Understand the business through its data and use.

Amateurs

- Basic understanding in order to perform job tasks.
- Consumers of analyses

Data analysis roles

Data engineers	Wrangle data, ETL, Manage data lake
Business analysts	Liaison between the tech and the business stakeholders
Data scientists	Have advanced degrees in an area of business use and know statistics
Accountants and financial analysts	Focus on internal financials & models for market development
Data visualization specialists	Strong design skills (graphic designer). Design dashboards and reports

Data analysis roles

- Data engineers
- Business analysts
- Data scientists
- Accountants and financial analysts
- Data visualization specialists
- **Domain experts**

Not a role specific to data, but is usually a necessary collaborator in order to inform and define business requirements so engineers and data scientists can build and implement a useful model.

Teaming and Internal Capabilities

PART 2 –
The Technology Ecosystem

Technology ecosystems are evolving from the relational databases to big data and AI systems. Where are we today?

- Relational databases
- NoSQL databases
- Big data tools
- Machine learning
- Visualization and reporting tools

- **Structured data management**
- **Often the combination of:**
  - ETL/discovery/transformation/data prep tools
  - Analytical tools for statistics and modeling
  - Operational processing, governance, data security and recovery tools
- **Available in open source and license packages**

Technology ecosystems: NoSQL, Graphs

- Relational databases
- **NoSQL databases**
- Big data tools
- Machine learning
- Visualization and reporting tools

- **Structured and unstructured data**
- **Heavy on the unstructured**
- **Compromise consistency (from relational) for easier scaling and speed**
- **Graph databases can be queried in many different ways.**

Technology ecosystems: Big data

- Relational databases
- NoSQL databases
- **Big data tools**
- Machine learning
- Visualization and reporting tools

- **Parallelize the processing to trillions of rows and many variable types of data and data types. Lots of use cases that NoSQL and SQL databases were not designed to.**
- **Often merge and have relational techniques that are brought into the mix to be able to get the value of both tools working together.**

Technology ecosystems: statistical tools

Relational databases

NoSQL databases

Big data tools

Machine learning

Visualization and reporting tools

- **Analysis tools for modelling and calculation**
- **Sometimes libraries for general use languages (NumPy, SciPy for Python)**
- **Sometimes programming languages all their own (R, SAS, SPSS, Spark)**

Technology ecosystems – statistical tools

- Relational databases
- NoSQL databases
- Big data tools
- Machine learning
- Visualization and reporting tools

python

NumPy

SciPy.org Sponsored By ENTHOUGHT

Technology ecosystems: statistical tools

- Relational databases
- NoSQL databases
- Big data tools
- Machine learning
- Visualization and reporting tools

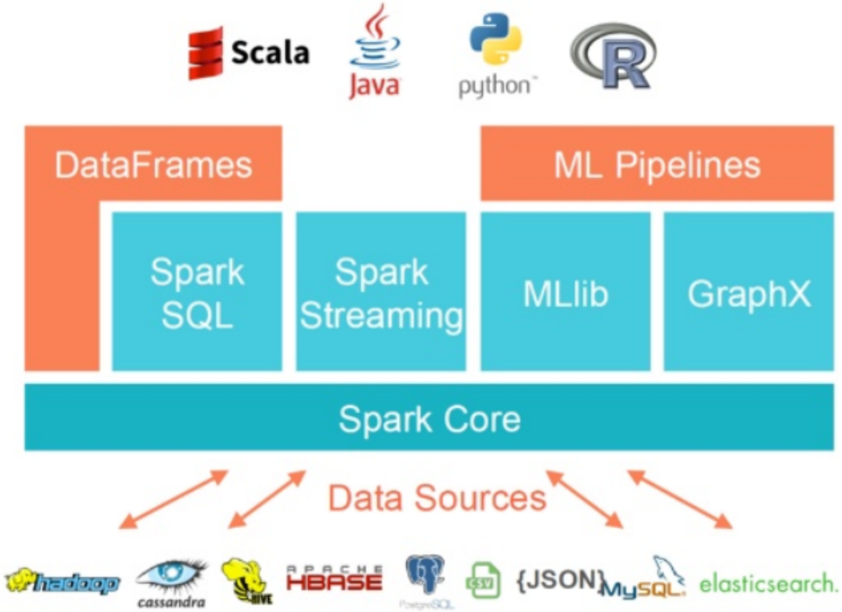
Technology ecosystems: machine learning

- Relational databases
- NoSQL databases
- Big data tools
- **Machine learning**
- Visualization and reporting tools

- **Blending statistics with computer science**
- **From pattern recognition to artificial intelligence**
- **Make data-driven predictions and decisions**

Technology ecosystems: machine learning

- Relational databases
- NoSQL databases
- Big data tools
- **Machine learning**
- Visualization and reporting tools



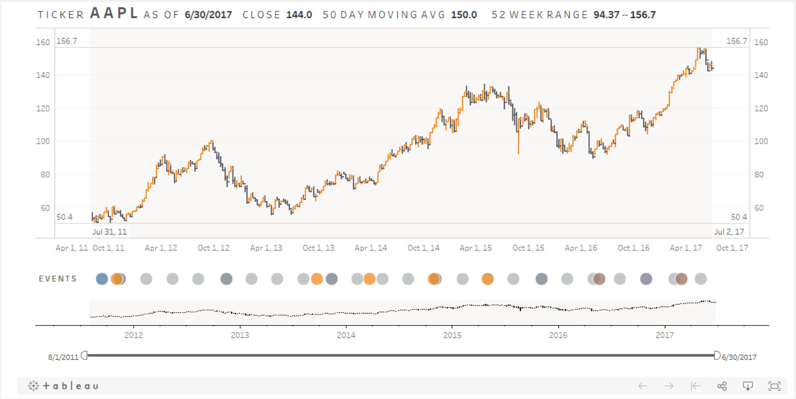
Technology ecosystems: visualization and reporting

- Relational databases
- NoSQL databases
- Big data tools
- Machine learning
- **Visualization and reporting tools**

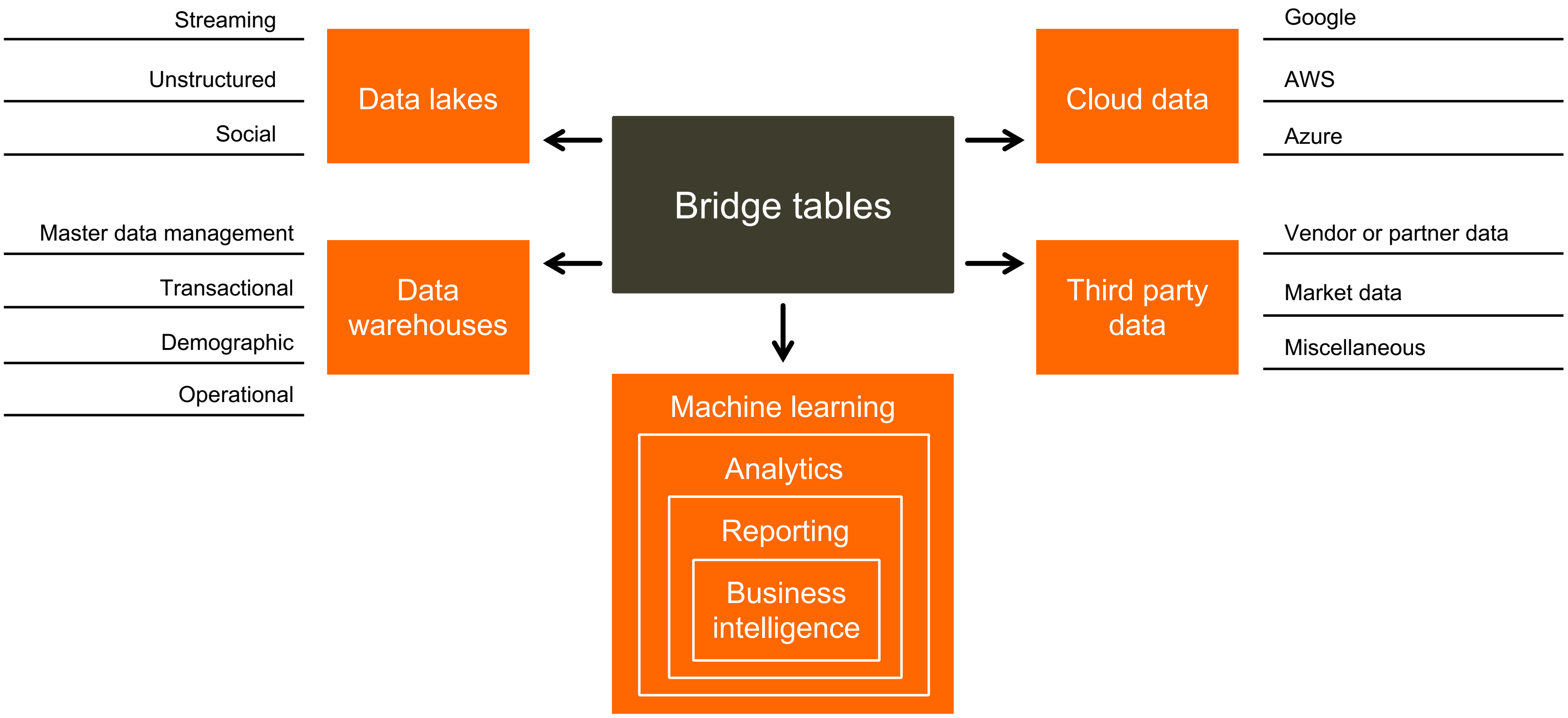
- **Allow easy and/or advanced visualization of data sets and results.**
- **Quickly extract and present useful information.**
- **Examples:**
  - Excel
  - Tableau
  - D3
  - PowerBI

Technology ecosystems: visualization and reporting

- Relational databases
- NoSQL databases
- Big data tools
- Machine learning
- **Visualization and reporting tools**



Technology ecosystems: Putting it all together



Part 7: Python Examples

https://github.com/shekhar2010us/ai-ainomy

Examples for

- Linear Regression
- Classification
- Clustering
- Recommender System

Pre-requisites

- Python3
- Preferably Anaconda
- Jupyter Notebook
- Pip
- Libraries – Matplotlib, numpy, scipy, sklearn, seaborn, pandas

Run Examples

After pre-requisites

- Run command: **Jupyter notebook** (to open Jupyter)
- All data in project/data/*
- All code in project/notebook/*.ipynb
- Run *.ipynb one by one

Thank you for your time!

Please take a moment to give us your feedback on the class using the participant evaluation link.

We hope you will turn to ASPE for any future learning needs on how to put emergent technologies to practical use!