MACHINE LEARNING: CONCEPTS AND APPLICATIONS

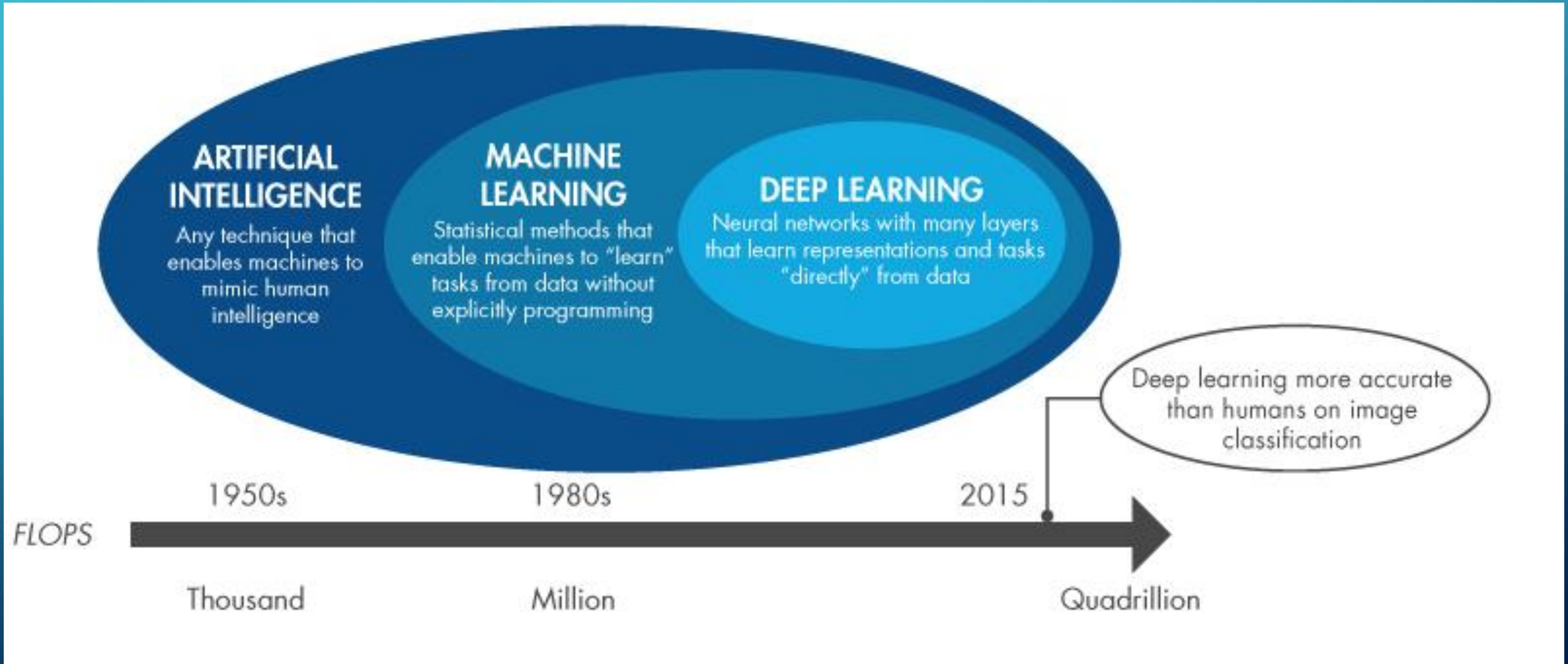
MAYSAM ABBOD

BRUNEL UNIVERSITY LONDON

OUTLINE

- AI and Machine Learning
- History
- How does the Machine Learns
- Supervised Learning
- Unsupervised Learning
- Deep Learning
- Applications
- What the Future Holds!

ARTIFICIAL INTELLIGENCE



WHAT IS MACHINE LEARNING

Machine learning teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

LEARNING FROM DATA

Machine learning algorithms find natural patterns in data that generate insight and help you make better decisions and predictions. They are used every day to make critical decisions in medical diagnosis, stock trading, energy load forecasting, and more. Media sites rely on machine learning to sift through millions of options to give you song or movie recommendations. Retailers use it to gain insight into their customers' purchasing behavior.

REAL WORLD APPLICATIONS

With the rise in big data, machine learning has become particularly important for solving problems in areas like these:

- Computational finance, for credit scoring and algorithmic trading
- Image processing and computer vision, for face recognition, motion detection, and object detection
- Computational biology, for tumour detection, drug discovery, and DNA sequencing
- Energy production, for price and load forecasting
- Automotive, aerospace, and manufacturing, for predictive maintenance
- Natural language processing

HISTORY OF MACHINE LEARNING

- 1950s
  - Samuel's checker player
  - Selfridge's Pandemonium
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- 1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis
  - Scientific discovery with BACON
  - Mathematical discovery with AM

HISTORY OF MACHINE LEARNING

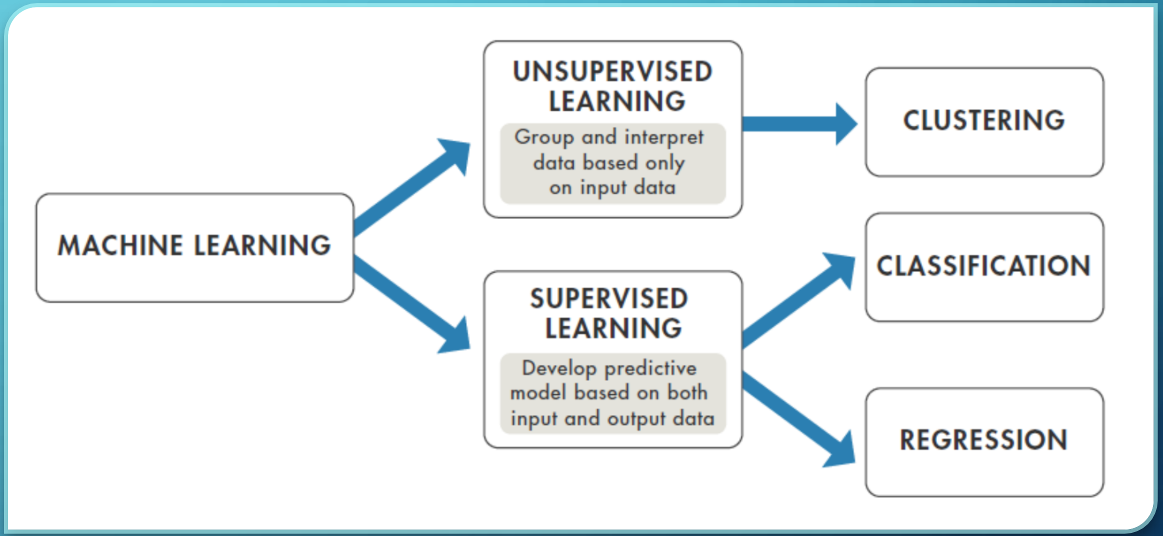
- 1980s:
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem
  - Analogy
  - Cognitive architectures
  - Resurgence of neural networks (connectionism, backpropagation)
  - Valiant's PAC Learning Theory
  - Focus on experimental methodology
- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning

HISTORY OF MACHINE LEARNING

- 2000s
  - Support vector machines & kernel methods
  - Graphical models
  - Statistical relational learning
  - Transfer learning
  - Sequence labelling
  - Collective classification and structured outputs
  - Computer Systems Applications (Compilers, Debugging, Graphics, Security)
  - E-mail management
  - Personalized assistants that learn
  - Learning in robotics and vision
- 2010s
  - Deep learning systems
  - Learning for big data
  - Bayesian methods
  - Multi-task & lifelong learning
  - Applications to vision, speech, social networks, learning to read, etc.

HOW MACHINE LEARNING WORK

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data



SUPERVISED LEARNING

- The aim of supervised machine learning is to build a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.
- Supervised learning uses classification and regression techniques to develop predictive models.
- **Classification techniques** predict discrete responses—for example, whether an email is genuine or spam, or whether a tumour is cancerous or benign. Classification models classify input data into categories. Typical applications include medical imaging, speech recognition, and credit scoring.
- **Regression techniques** predict continuous responses— for example, changes in temperature or fluctuations in power demand. Typical applications include electricity load forecasting and algorithmic trading.

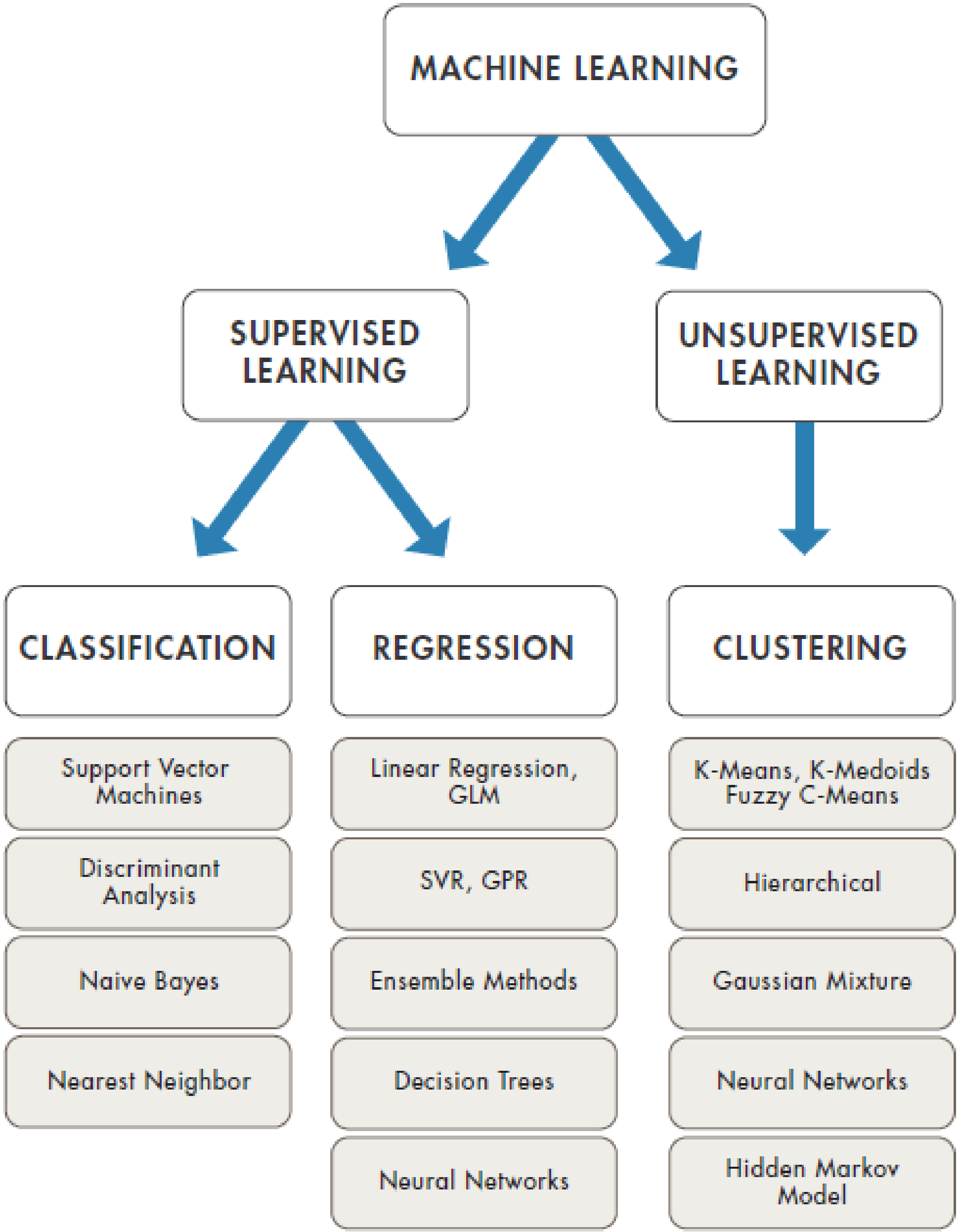
UNSUPERVISED LEARNING

- Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labelled responses.
- **Clustering** is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data.
- Applications for clustering include gene sequence analysis, market research, and object recognition.

WHICH ALGORITHM SHOULD BE USED

- Choosing the right algorithm can seem overwhelming—there are dozens of supervised and unsupervised machine learning algorithms, and each takes a different approach to learning.
- There is no best method or one size fits all. Finding the right algorithm is partly just trial and error—even highly experienced data scientists can't tell whether an algorithm will work without trying it out. But algorithm selection also depends on the size and type of data you're working with, the insights you want to get from the data, and how those insights will be used.

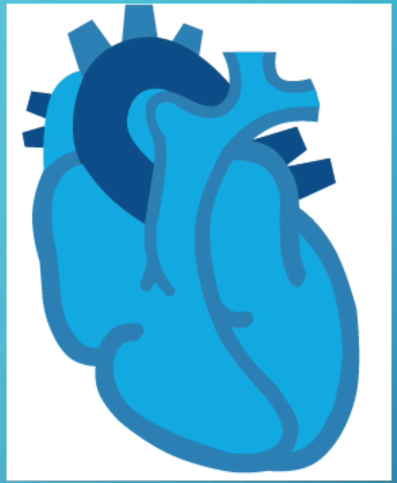
ROAD MAP


MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
CLASSIFICATION
REGRESSION
CLUSTERING
Support Vector Machines
Discriminant Analysis
Naive Bayes
Nearest Neighbor
Linear Regression, GLM
SVR, GPR
Ensemble Methods
Decision Trees
Neural Networks
K-Means, K-Medoids Fuzzy C-Means
Hierarchical
Gaussian Mixture
Neural Networks
Hidden Markov Model

WHEN MACHINE LEARNING SHOULD BE USED

- Consider using machine learning when you have a complex task or problem involving a large amount of data and lots of variables, but no existing formula or equation. For example, machine learning is a good option if you need to handle situations like these:
- Hand-written rules and equations are too complex—as in face recognition and speech recognition.
- The rules of a task are constantly changing—as in fraud detection from transaction records.
- The nature of the data keeps changing, and the program needs to adapt—as in automated trading, energy demand forecasting, and predicting shopping trends.

HEART ATTACK PREDICTION



Using Supervised Learning to Predict Heart Attacks

Suppose clinicians want to predict whether someone will have a heart attack within a year. They have data on previous patients, including age, weight, height, and blood pressure. They know whether the previous patients had heart attacks within a year. So the problem is combining the existing data into a model that can predict whether a new person will have a heart attack within a year.

REAL WORLD PROBLEMS

- **Creating Algorithms that Can Analyse Works of Art**
- Researchers at the Art and Artificial Intelligence Laboratory at Rutgers University wanted to see whether a computer algorithm could classify paintings by style, genre, and artist as easily as a human. They began by identifying visual features for classifying a painting's style. The algorithms they developed classified the styles of paintings in the database with 60% accuracy, outperforming typical non-expert humans.
- The researchers hypothesized that visual features useful for style classification (a supervised learning problem) could also be used to determine artistic influences (an unsupervised problem).
- They used classification algorithms trained on Google images to identify specific objects. They tested the algorithms on more than 1,700 paintings from 66 different artists working over a span of 550 years. The algorithm readily identified connected works, including the influence of Diego Velazquez's "Portrait of Pope Innocent X" on Francis Bacon's "Study After Velazquez's Portrait of Pope Innocent X."

REAL WORLD PROBLEMS

Optimizing HVAC Energy Usage in Large Buildings

The heating, ventilation, and air-conditioning (HVAC) systems in office buildings, hospitals, and other large-scale commercial buildings are often inefficient because they do not take into account changing weather patterns, variable energy costs, or the building's thermal properties.

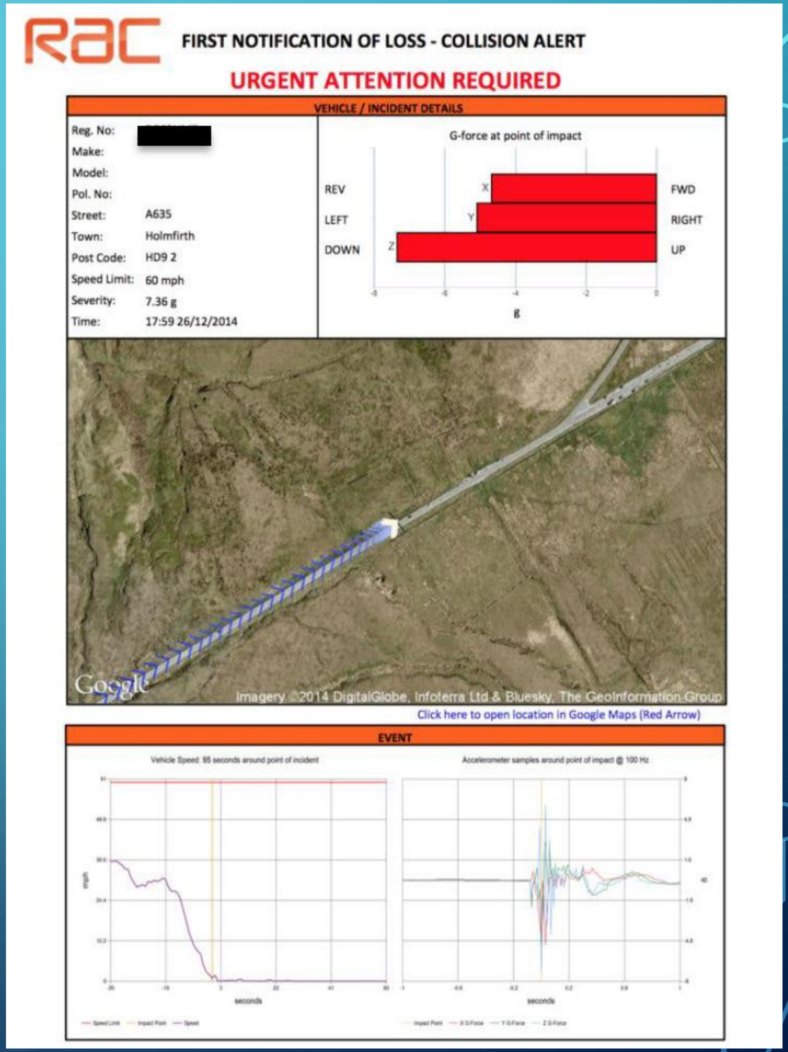
Building IQ's cloud-based software platform addresses this problem. The platform uses advanced algorithms and machine learning methods to continuously process gigabytes of information from power meters, thermometers, and HVAC pressure sensors, as well as weather and energy cost. In particular, machine learning is used to segment data and determine the relative contributions of gas, electric, steam, and solar power to heating and cooling processes. The building IQ platform reduces HVAC energy consumption in large-scale commercial buildings by 10% - 25% during normal operation.

REAL WORLD PROBLEMS

Detecting Low-Speed Car Crashes

With more than 8 million members, the RAC is one of the UK's largest motoring organizations, providing roadside assistance, insurance, and other services to private and business motorists.

To enable rapid response to roadside incidents, reduce crashes, and mitigate insurance costs, the RAC developed an onboard crash sensing system that uses advanced machine learning algorithms to detect low-speed collisions and distinguish these events from more common driving events, such as driving over speed bumps or potholes. Independent tests showed the RAC system to be 92% accurate in detecting test crashes.



OTHER APPLICATIONS

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

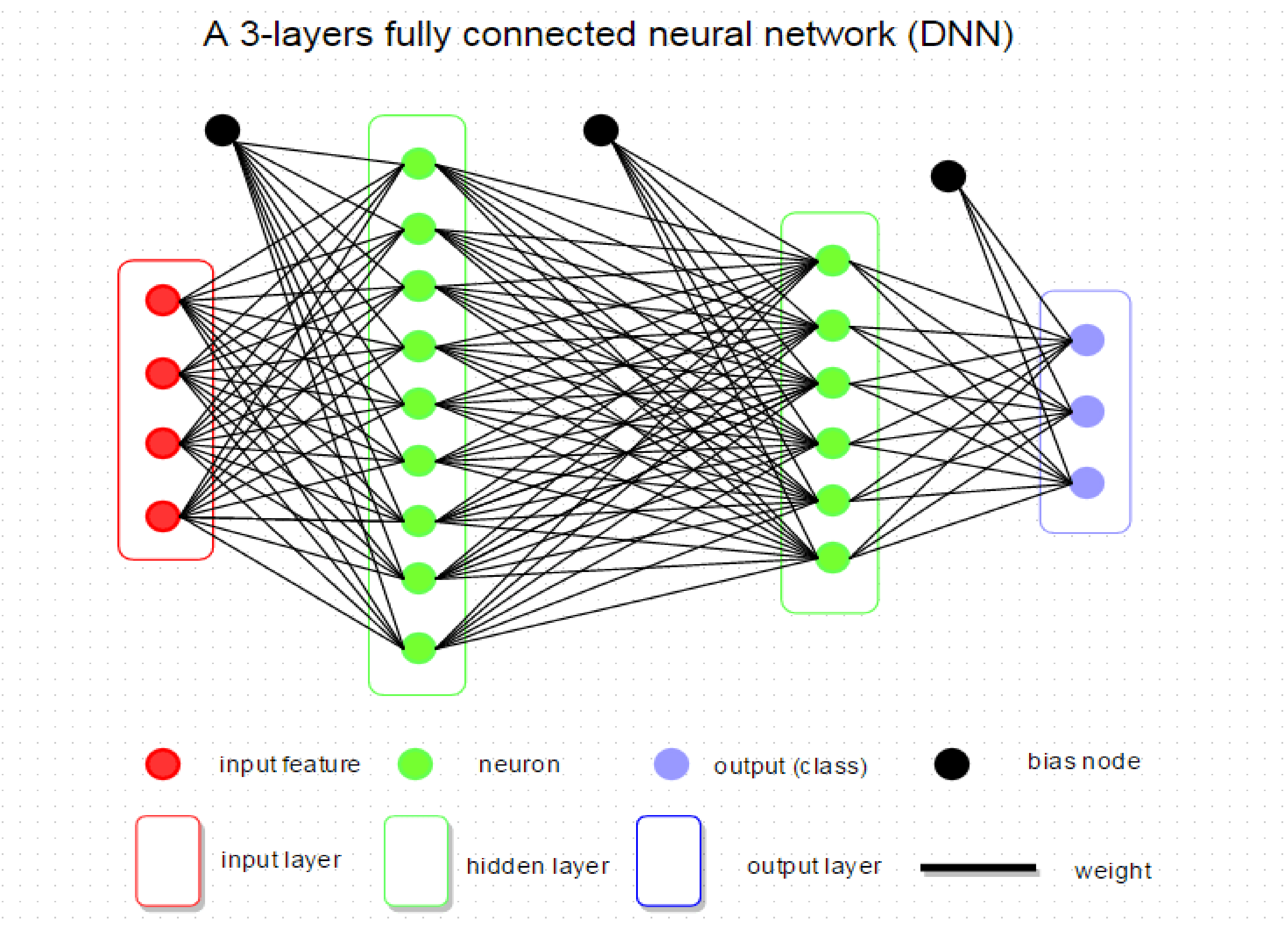
OTHER APPLICATIONS

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- What do you do!

SUPERVISED LEARNING

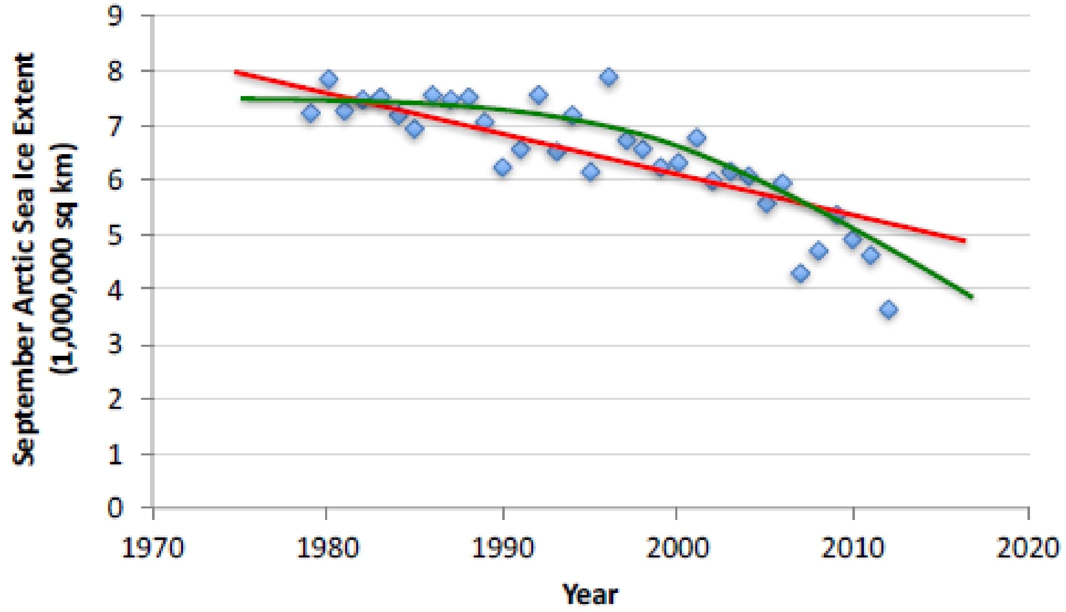
- Neural Networks
- Naïve Bayes Theorem
- Support Vector Machine
- Random Forest
- Decision Trees

TYPICAL NEURAL NETWORK TOPOLOGY



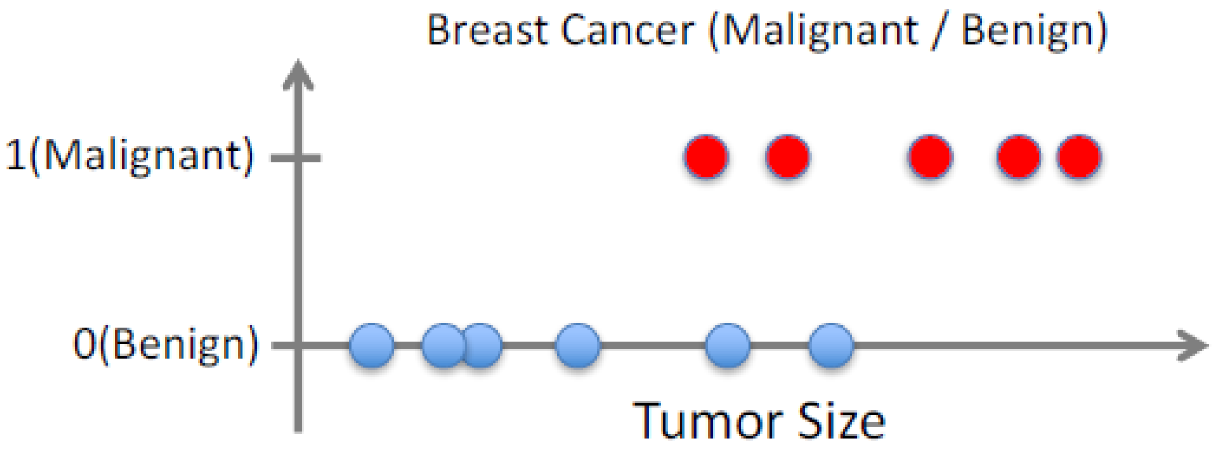
REGRESSION (SUPERVISED)

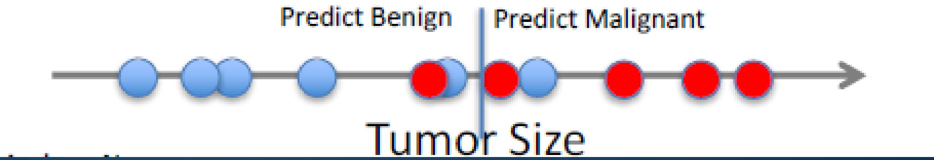
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



CLASSIFICATION (SUPERVISED)

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification

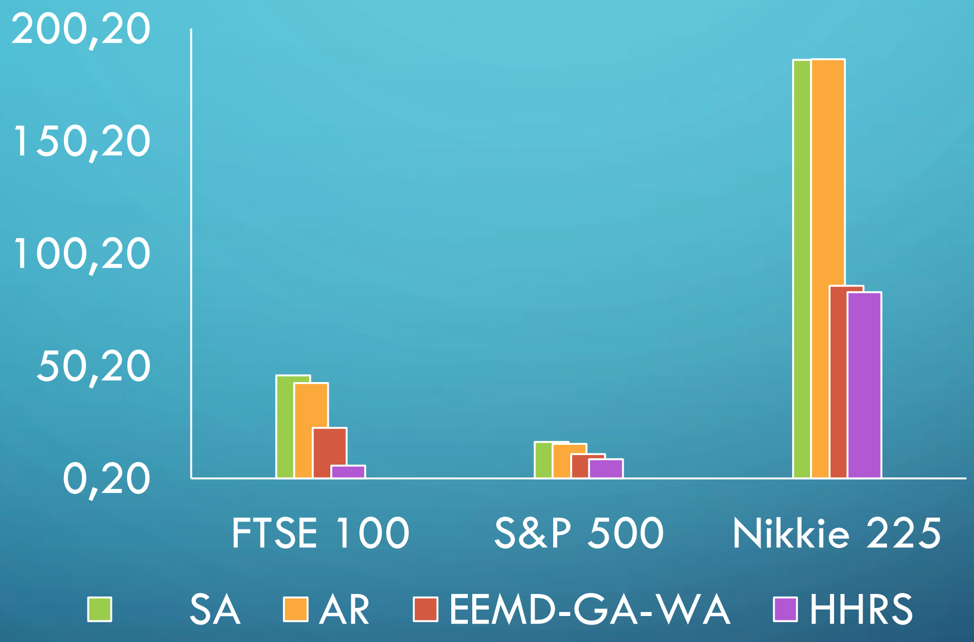




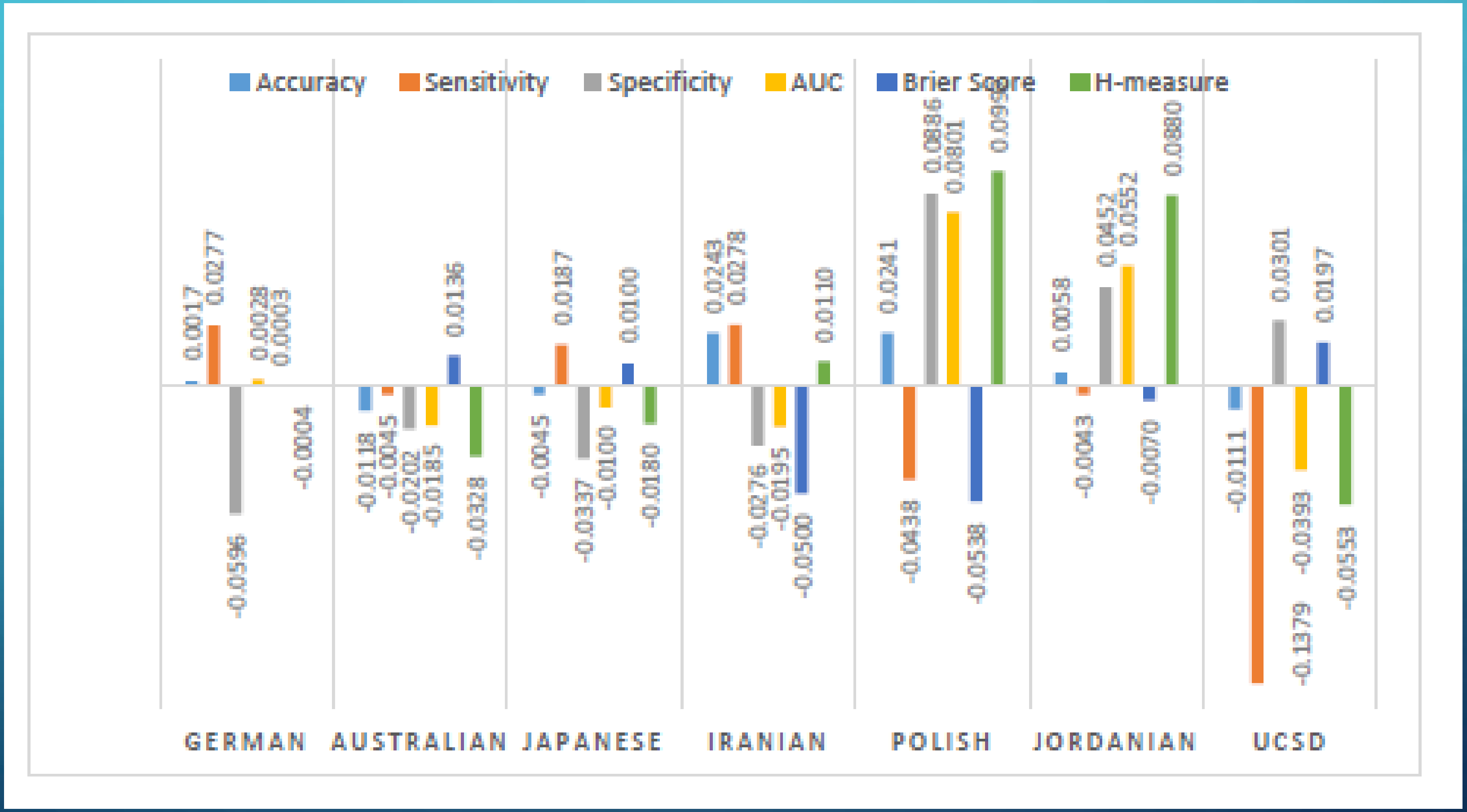
FINANCIAL ENGINEERING

Prediction results

The below figure illustrates the four prediction models SA, AR, EEMD-GA-WA and HHRS. The performance measurement is RMSE



SVM FOR CREDIT SCORING



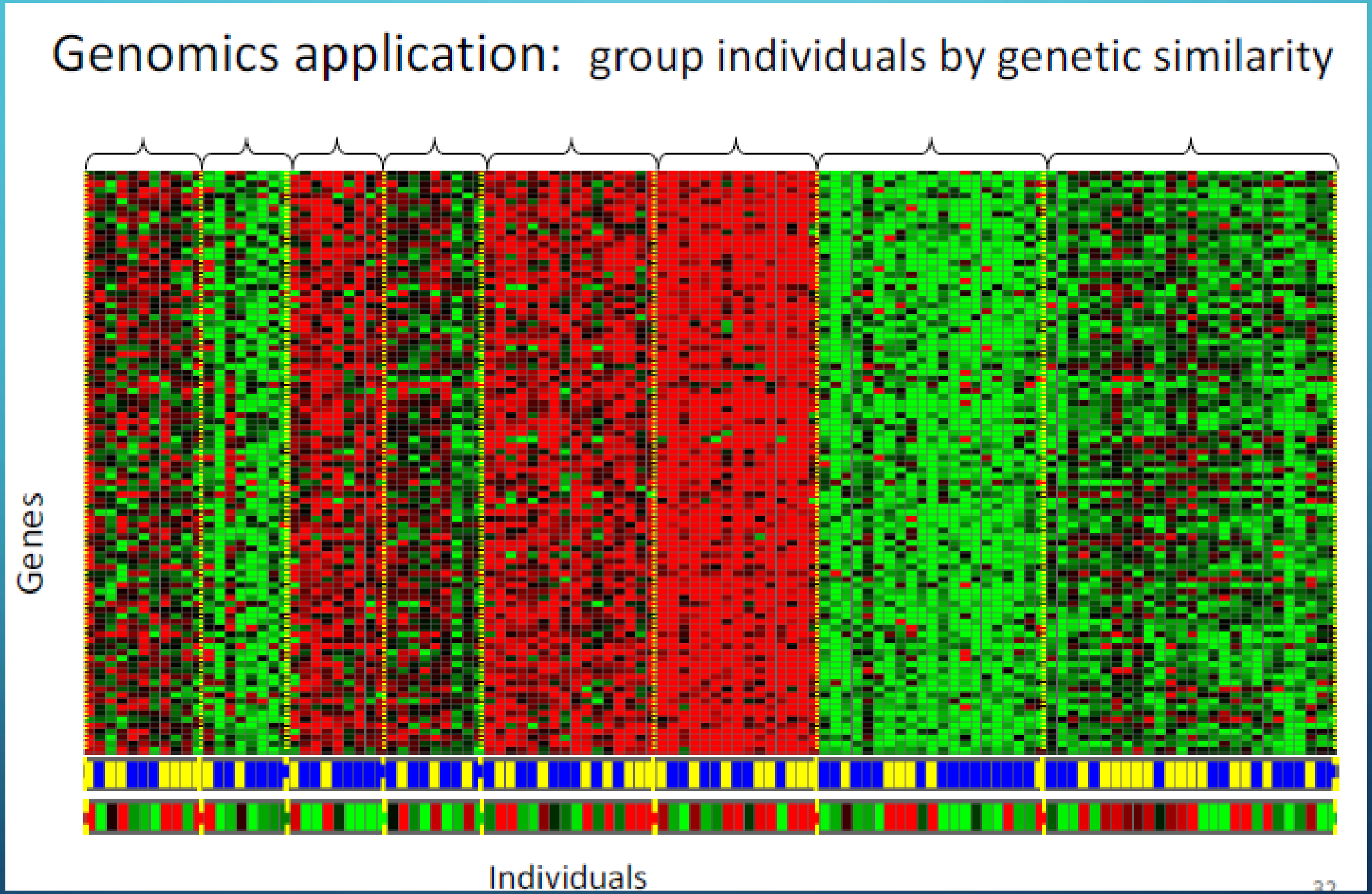
POWER SYSTEMS

- Power System Stability
- Power Systems Control
- Fault Diagnosis
- Security Assessment
- Load Forecasting
- Reactive Power Planning
- State Estimation

UNSUPERVISED LEARNING

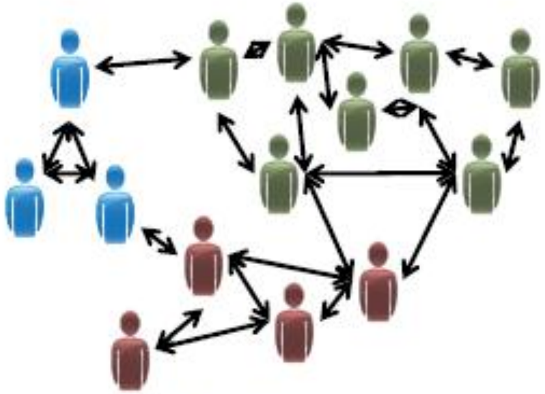
- FUZZY CLUSTERING
- K-MEAN
- HIERARCHICAL CLUSTERING
- SELF-ORGANISING MAP (SOM)

MACHINE LEARNING



MACHINE LEARNING


Organize computing clusters

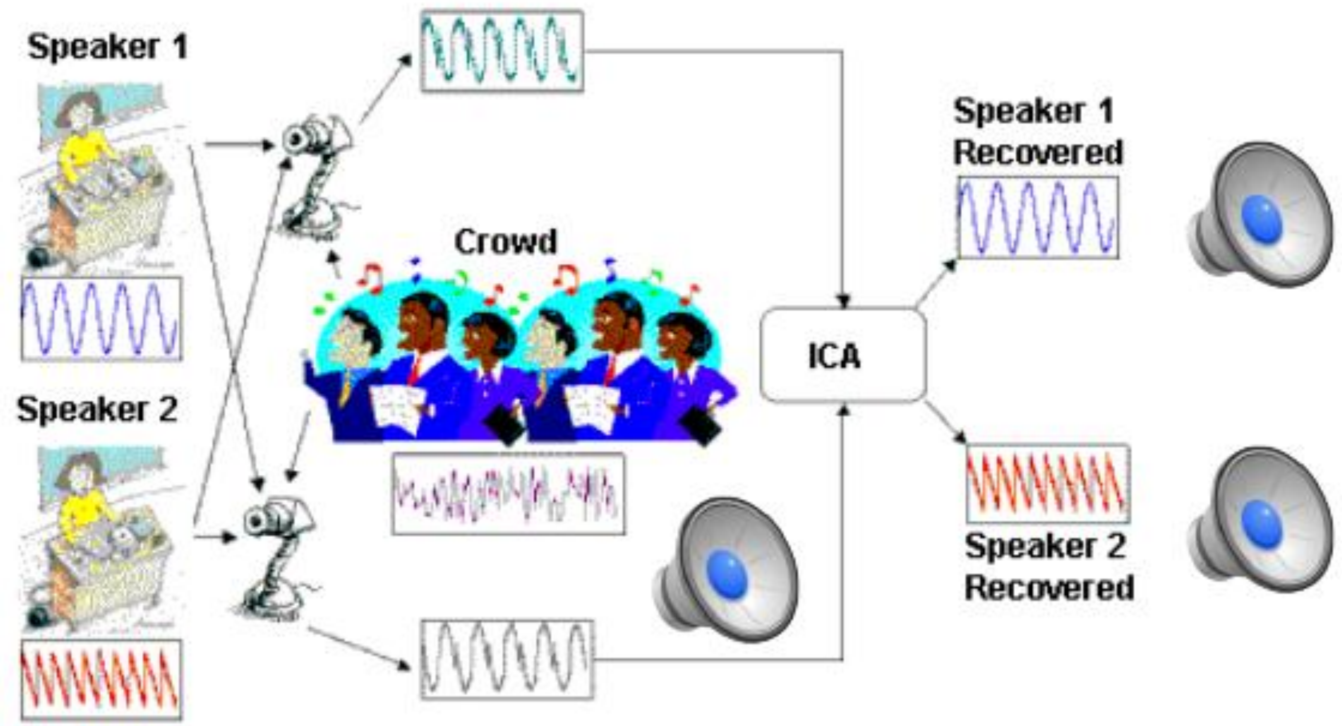

Social network analysis


Market segmentation


Astronomical data analysis

MACHINE LEARNING

- Independent component analysis – separate a combined signal into its original sources



DEEP LEARNING

Deep learning is a type of machine learning in which a model learns to perform classification tasks directly from images, text, or sound. Deep learning is usually implemented using a neural network architecture. The term “deep” refers to the number of layers in the network—the more layers, the deeper the network. Traditional neural networks contain only 2 or 3 layers, while deep networks can have hundreds.

DEEP LEARNING APPLICATIONS

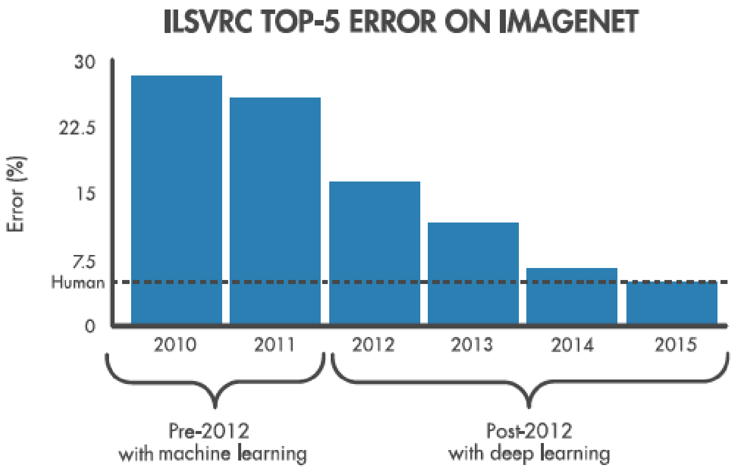
Here are just a few examples of deep learning at work:

- A self-driving vehicle slows down as it approaches a pedestrian crosswalk.
- An ATM rejects a counterfeit bank note.
- A smartphone app gives an instant translation of a foreign street sign.

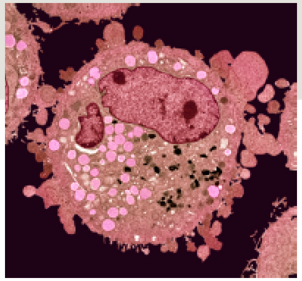
Deep learning is especially well-suited to identification applications such as face recognition, text translation, voice recognition, and advanced driver assistance systems, including, lane classification and traffic sign recognition.

WHAT MAKES DEEP LEARNING STATE OF THE ART

In a word, accuracy. Advanced tools and techniques have dramatically improved deep learning algorithms—to the point where they can outperform humans at classifying images, win against the world's best GO player, or enable a voice-controlled assistant like Amazon Echo® and Google Home to find and download that new song you like.



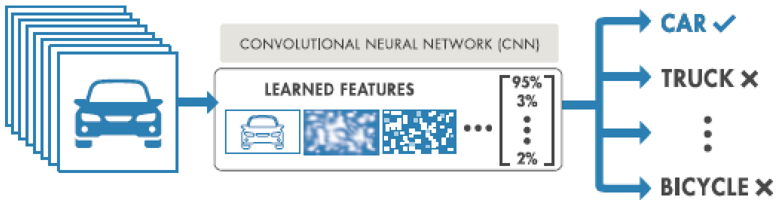
UCLA researchers built an advanced microscope that yields a high-dimensional data set used to train a deep learning network to identify cancer cells in tissue samples.



WHAT MAKES DEEP LEARNING STATE OF THE ART

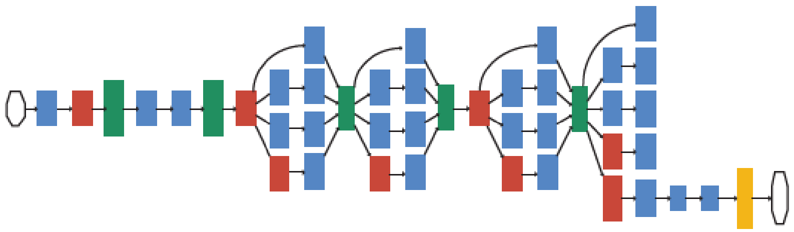
Three technology enablers make this degree of accuracy possible:

Easy access to massive sets of labeled data
Data sets such as ImageNet and PASCAL VoC are freely available, and are useful for training on many different types of objects.



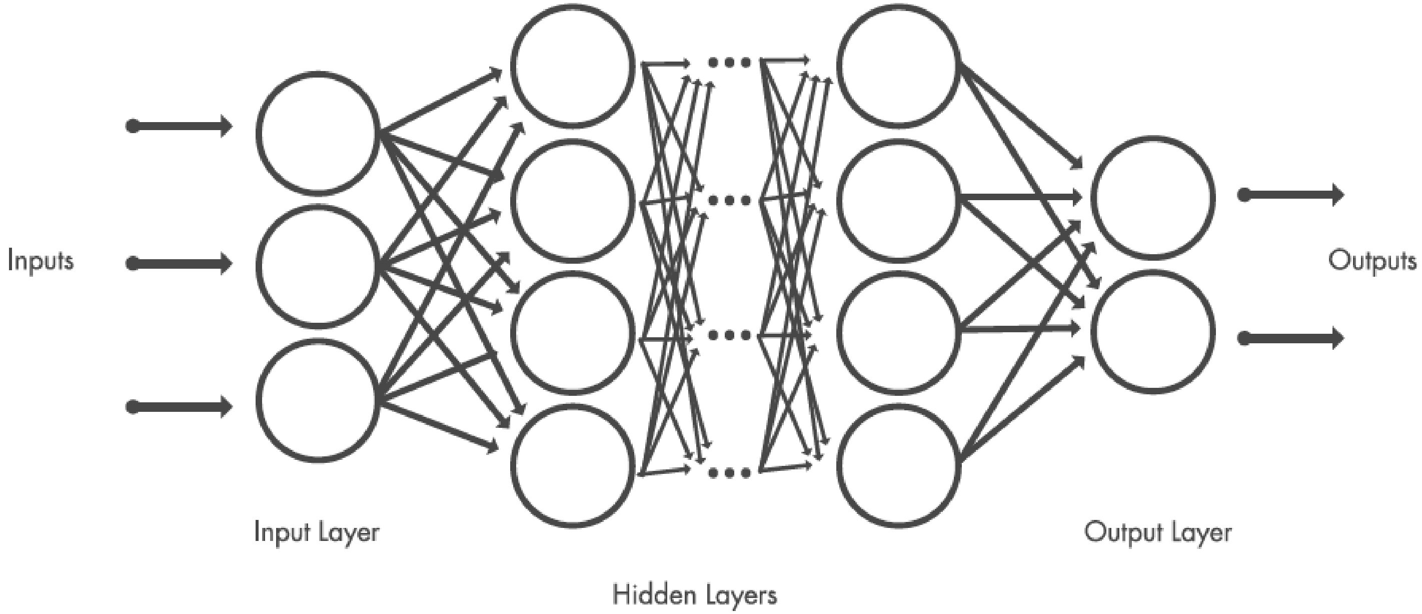
Increased computing power
High-performance GPUs accelerate the training of the massive amounts of data needed for deep learning, reducing training time from weeks to hours.

Pretrained models built by experts
Models such as AlexNet can be retrained to perform new recognition tasks using a technique called *transfer learning*. While AlexNet was trained on 1.3 million high-resolution images to recognize 1000 different objects, accurate transfer learning can be achieved with much smaller datasets.



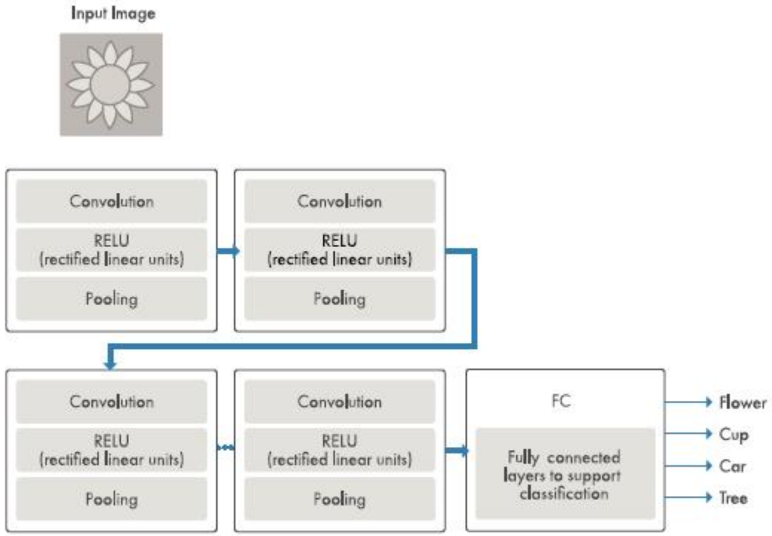
INSIDE DEEP LEARNING NETWORK

A deep neural network combines multiple nonlinear processing layers, using simple elements operating in parallel and inspired by biological nervous systems. It consists of an input layer, several hidden layers, and an output layer. The layers are interconnected via nodes, or neurons, with each hidden layer using the output of the previous layer as its input.



HOW DOES DEEP LEARNING NETWORK LEARN

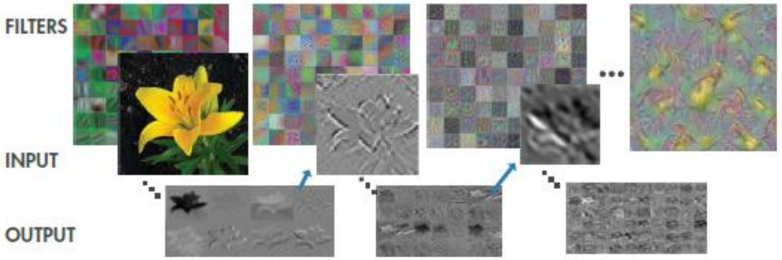
Let's say we have a set of images where each image contains one of four different categories of object, and we want the deep learning network to automatically recognize which object is in each image. We label the images in order to have training data for the network.



Using this training data, the network can then start to understand the object's specific features and associate them with the corresponding category.

Each layer in the network takes in data from the previous layer, transforms it, and passes it on. The network increases the complexity and detail of what it is learning from layer to layer.

Notice that the network learns directly from the data—we have no influence over what features are being learned.



HOW DOES DEEP LEARNING NETWORK LEARN

A convolutional neural network (CNN, or ConvNet) is one of the most popular algorithms for deep learning with images and video.

Like other neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between.

Feature Detection Layers

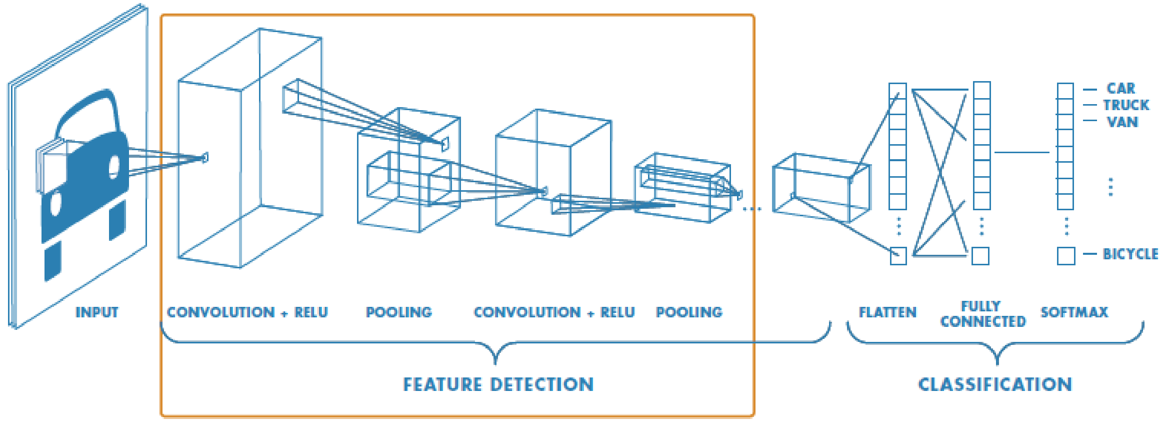
These layers perform one of three types of operations on the data: convolution, pooling, or rectified linear unit (ReLU).

Convolution puts the input images through a set of convolutional filters, each of which activates certain features from the images.

Pooling simplifies the output by performing nonlinear downsampling, reducing the number of parameters that the network needs to learn about.

Rectified linear unit (ReLU) allows for faster and more effective training by mapping negative values to zero and maintaining positive values.

These three operations are repeated over tens or hundreds of layers, with each layer learning to detect different features.



HOW DOES DEEP LEARNING NETWORK LEARN

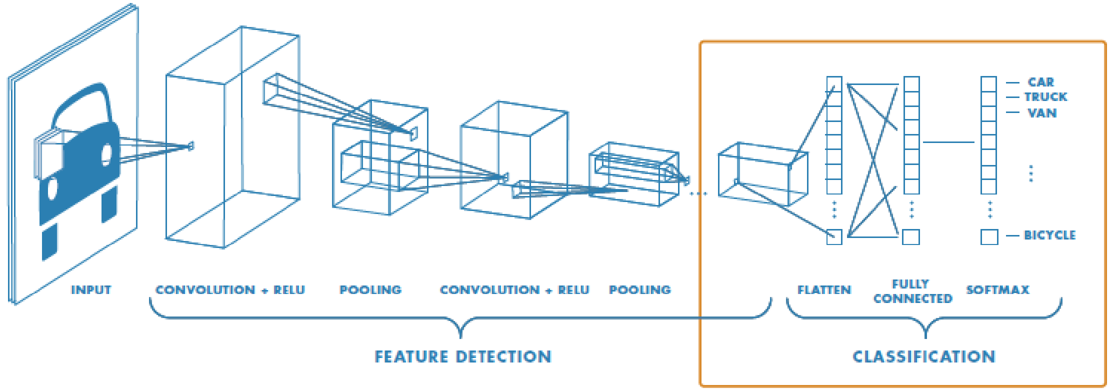
Classification Layers

After feature detection, the architecture of a CNN shifts to classification.

The next-to-last layer is a **fully connected layer** (FC) that outputs a vector of K dimensions where K is the number of classes that the network will be able to predict. This vector contains the probabilities for each class of any image being classified.

The final layer of the CNN architecture uses a **softmax** function to provide the classification output.

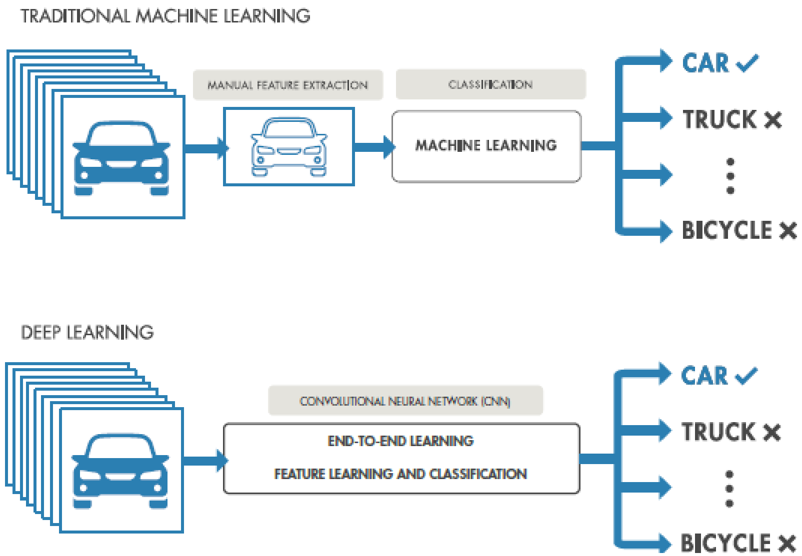
There is no exact formula for selecting layers. The best approach is to try a few and see how well they work—or to use a pretrained network.



HOW DOES DEEP LEARNING NETWORK LEARN

Deep learning is a subtype of machine learning. With machine learning, you manually extract the relevant features of an image. With deep learning, you feed the raw images directly into a deep neural network that learns the features automatically.

Deep learning often requires hundreds of thousands or millions of images for the best results. It's also computationally intensive and requires a high-performance GPU.

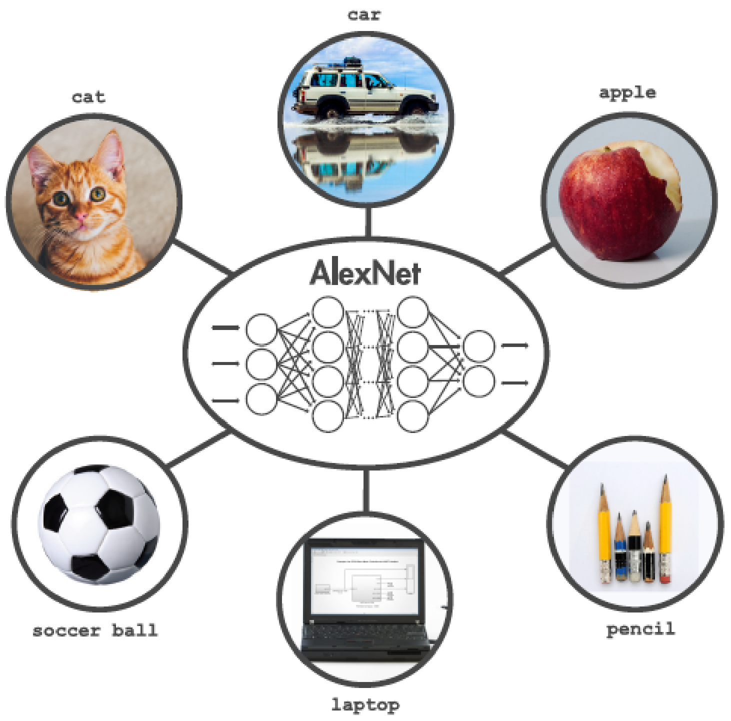


Machine Learning	Deep Learning
+ Good results with small data sets	— Requires very large data sets
+ Quick to train a model	— Computationally intensive
— Need to try different features and classifiers to achieve best results	+ Learns features and classifiers automatically
— Accuracy plateaus	+ Accuracy is unlimited

HOW DOES DEEP LEARNING NETWORK LEARN

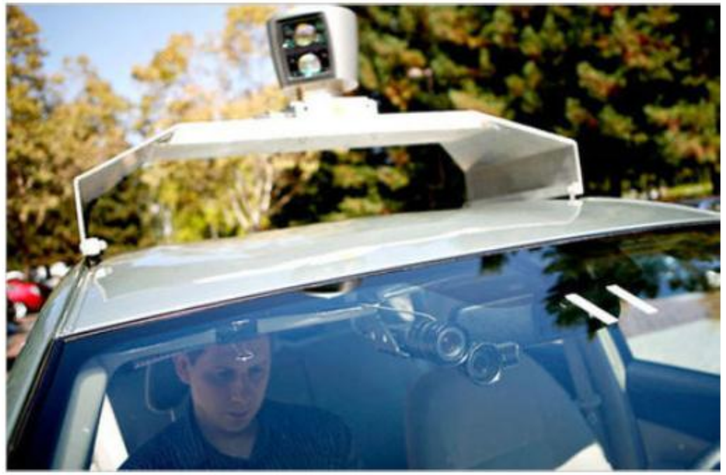
If you're new to deep learning, a quick and easy way to get started is to use an existing network, such as AlexNet, a CNN trained on more than a million images. AlexNet is most commonly used for image classification. It can classify images into 1000 different categories, including keyboards, computer mice, pencils, and other office equipment, as well as various breeds of dogs, cats, horses, and other animals.

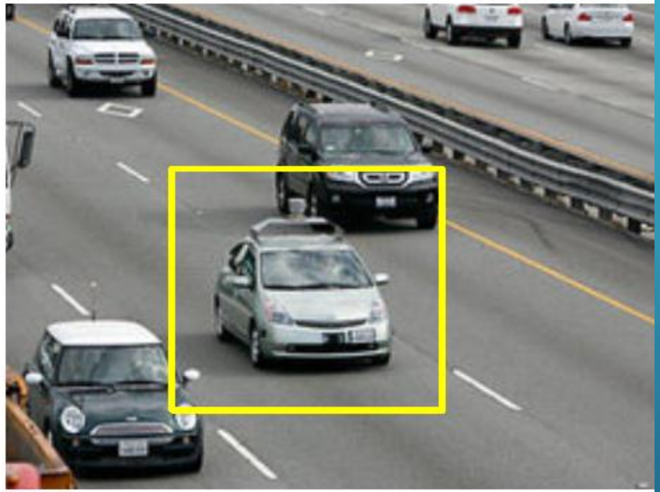
AlexNet was first published in 2012, and has become a well-known model in the research community.



STATE-OF-THE-ART MACHINE LEARNING APPLICATIONS

AUTONOMOUS CARS

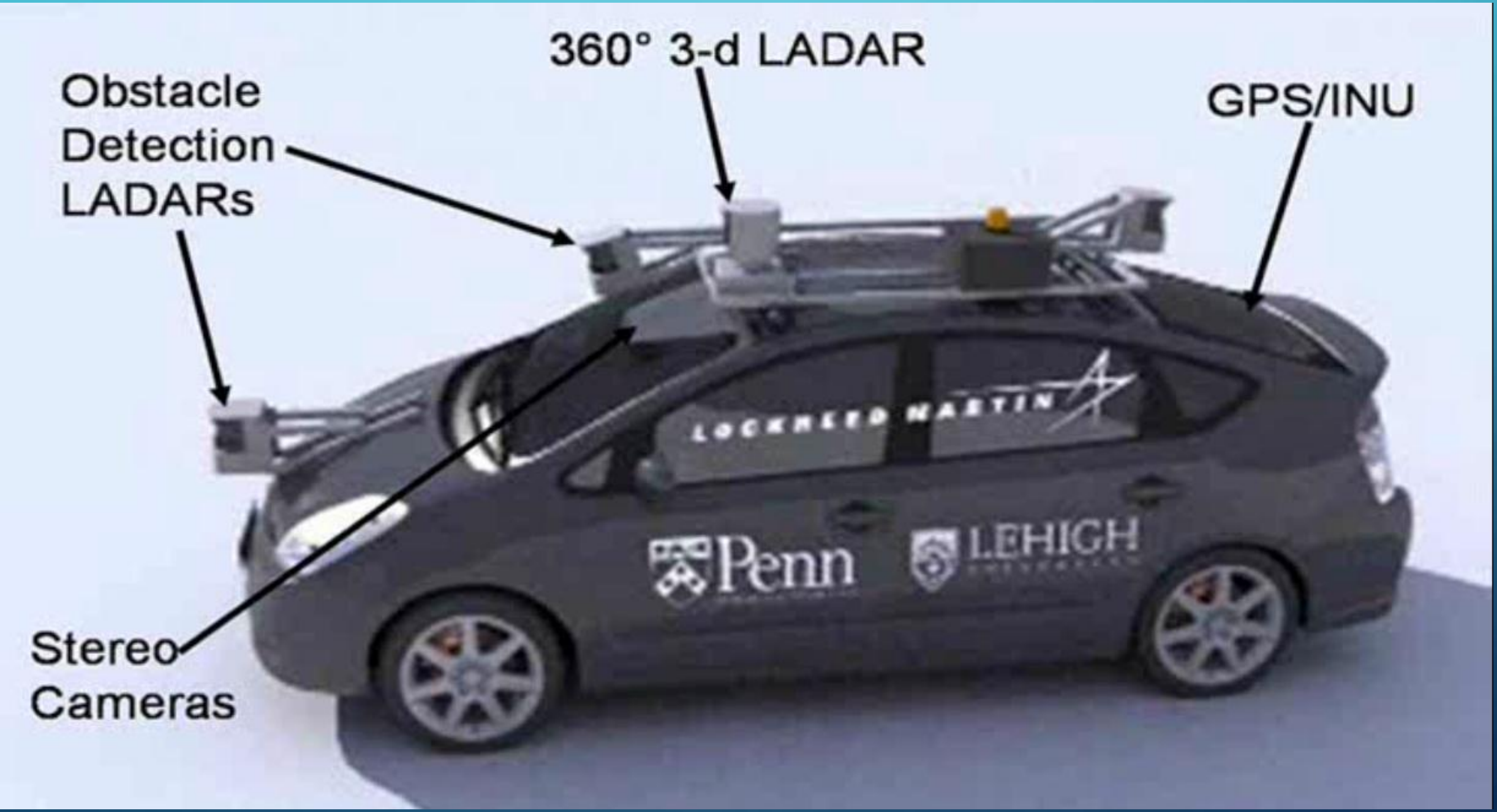




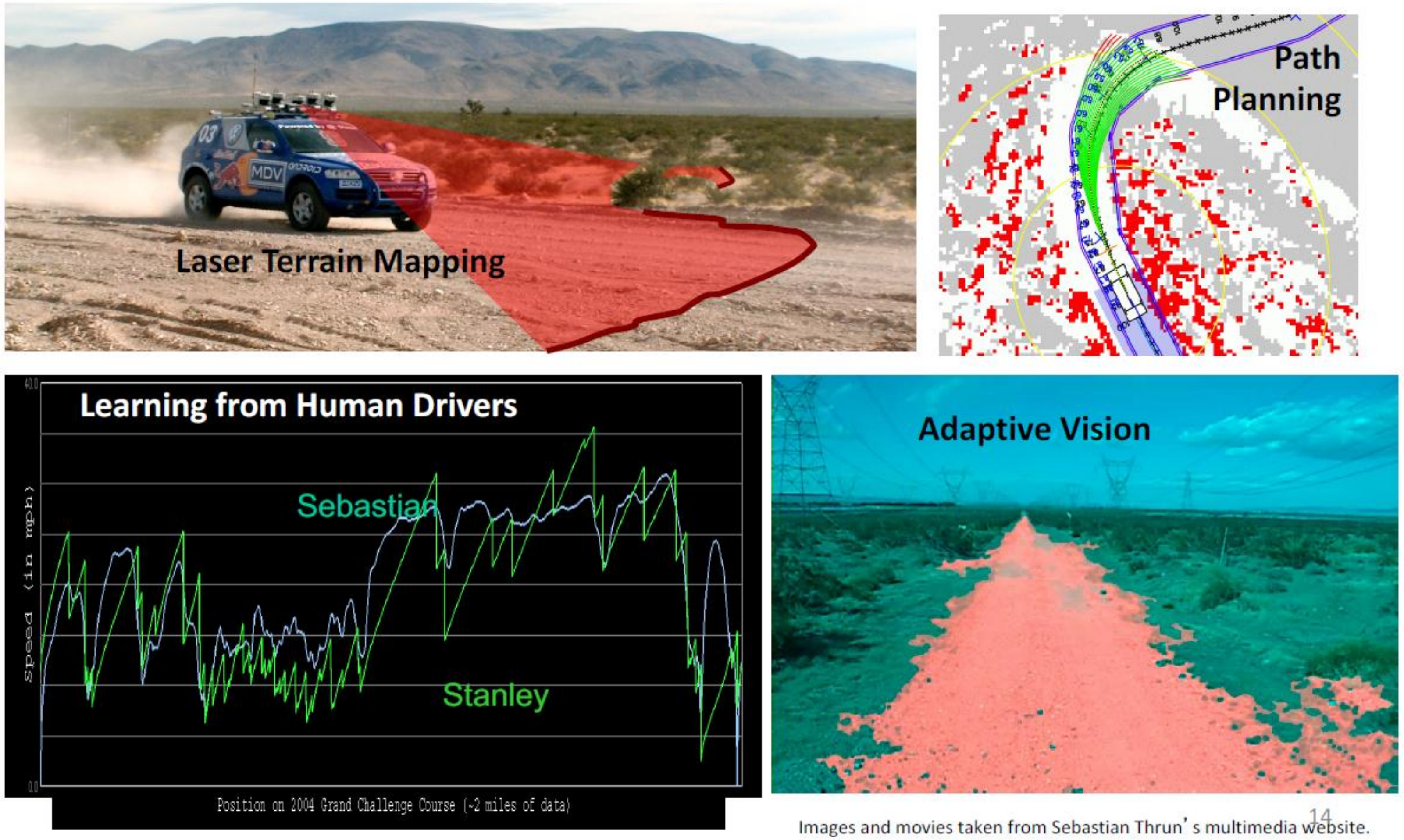
- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Penn's Autonomous Car →
(Ben Franklin Racing Team)

AUTONOMOUS CARS



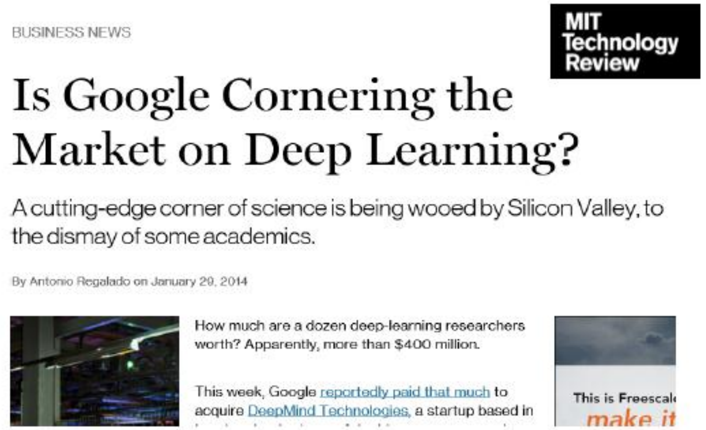
AUTONOMOUS CARS



Images and movies taken from Sebastian Thrun's multimedia website.

SCENE LABELLING USING DEEP LEARNING

DEEP LEARNING


BUSINESS NEWS

MIT Technology Review

Is Google Cornering the Market on Deep Learning?

A cutting-edge corner of science is being wooed by Silicon Valley, to the dismay of some academics.

By Antonio Regalado on January 29, 2014

How much are a dozen deep-learning researchers worth? Apparently, more than $400 million.

This week, Google reportedly paid that much to acquire DeepMind Technologies, a startup based in


BloombergBusinessweek Technology

Acquisitions

The Race to Buy the Human Brains Behind Deep Learning Machines

By Ashlee Vance | January 27, 2014

intelligence projects. "DeepMind is bona fide in terms of its research capabilities and depth," says Peter Lee, who heads Microsoft Research.

According to Lee, Microsoft, Facebook (FB), and Google find themselves in a battle for deep learning talent. Microsoft has gone from four full-time deep learning experts to 70 in the past three years. "We would have more if the talent was there to

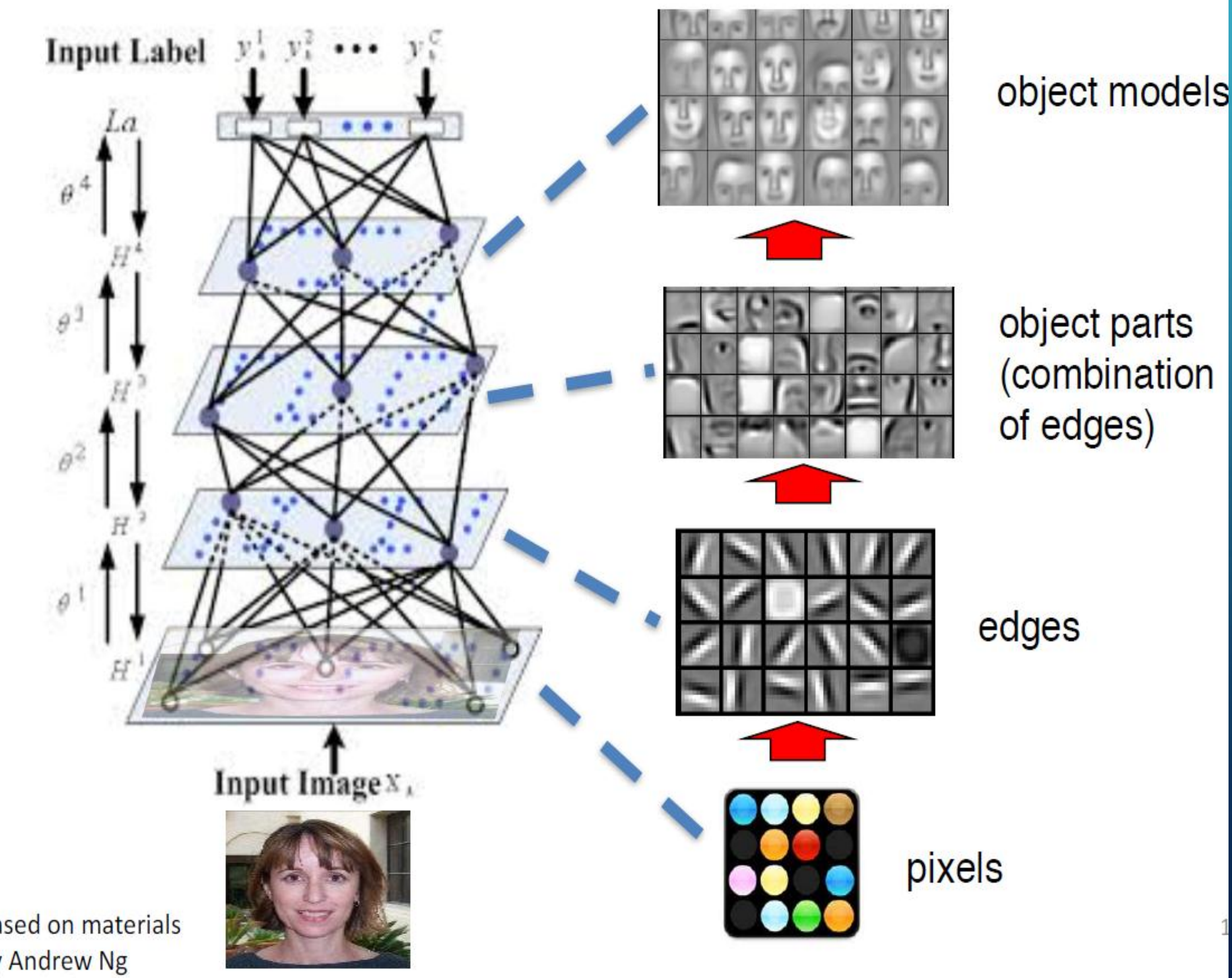

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY

INNOVATION INSIGHTS community content featured

Deep Learning's Role in the Age of Robots

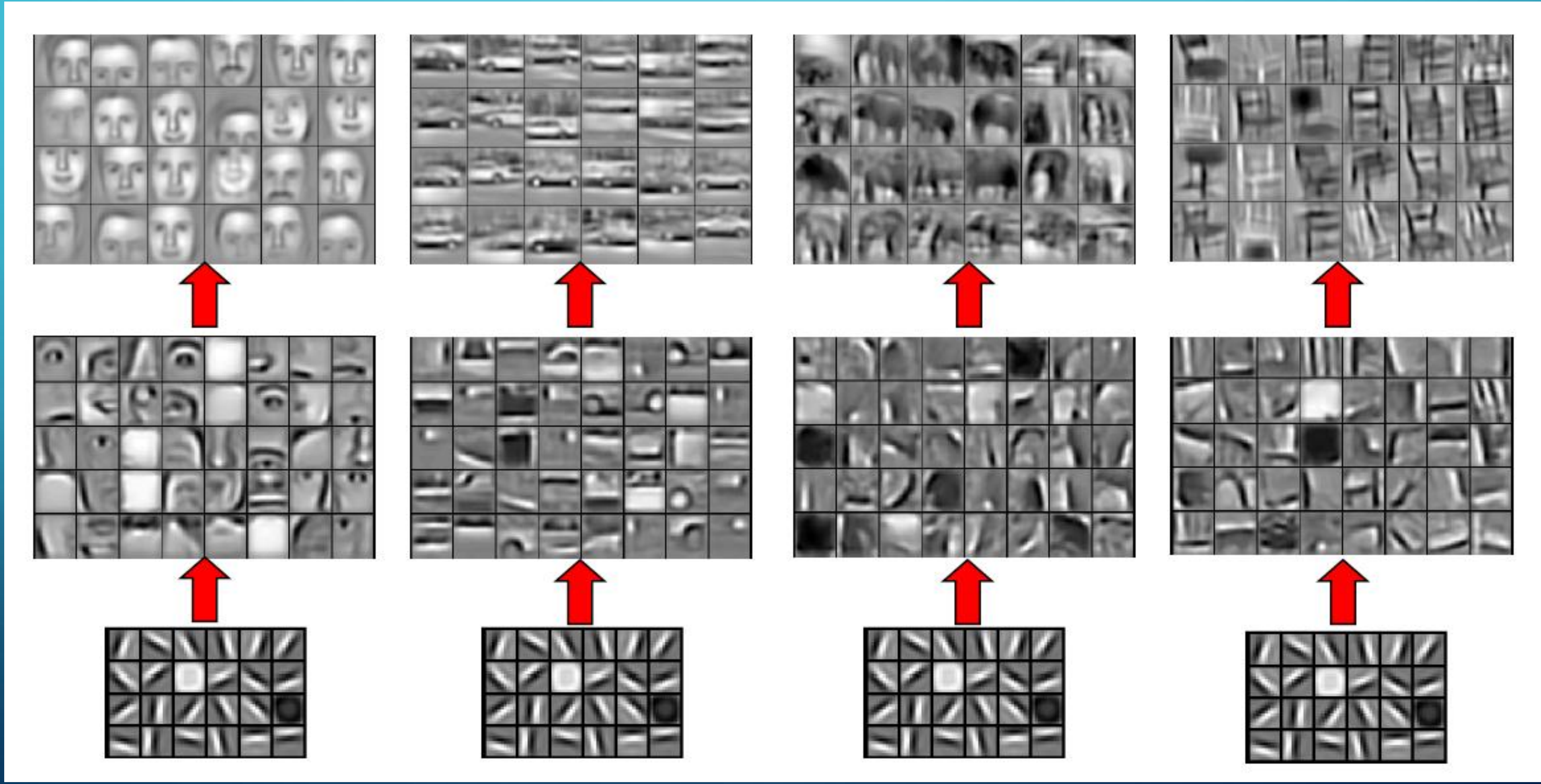
BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM

DEEP BELIEF FOR FACE RECOGNITION



Based on materials by Andrew Ng

LEARNING OF OBJECTS

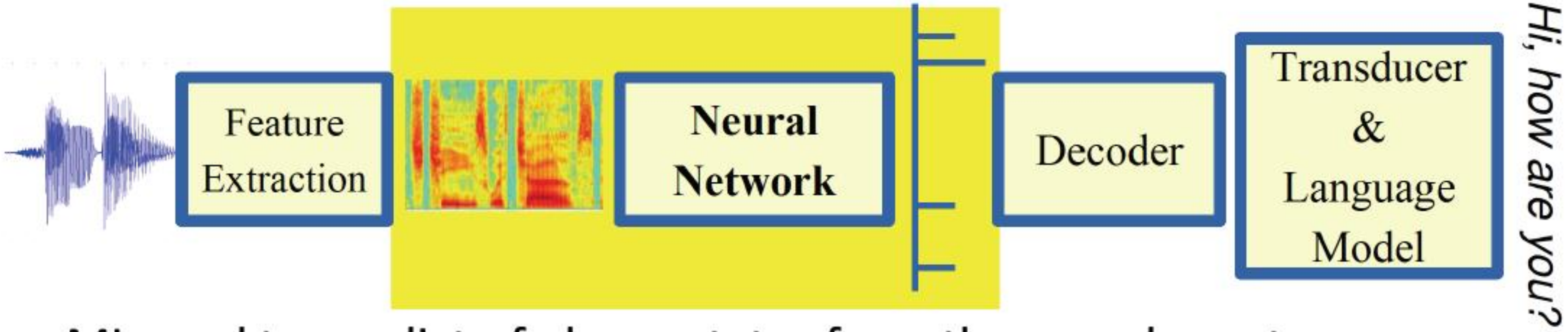


INFERENCE USING DEEP LEARNING

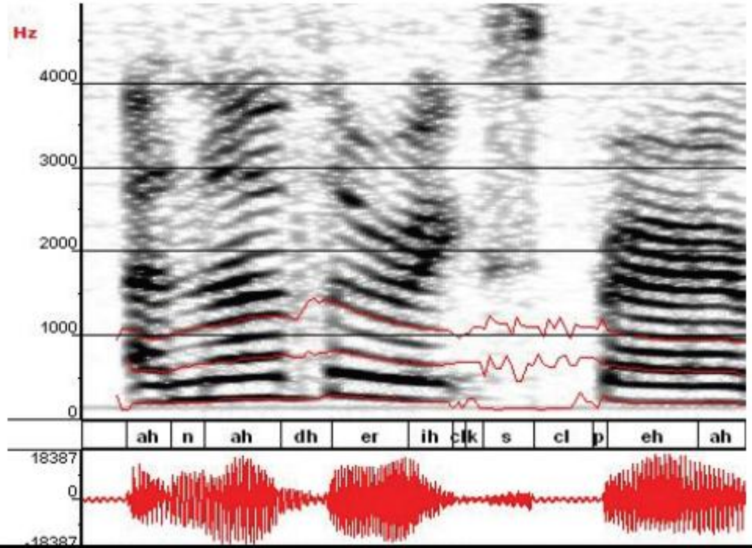
Generating posterior samples from faces by "filling in" experiments (cf. Lee and Mumford, 2003). Combine bottom-up and top-down inference.

SPEECH RECOGNITION

A Typical Speech Recognition System



ML used to predict of phone states from the sound spectrogram



Deep learning has state-of-the-art results

# Hidden Layers	1	2	4	8	10	12
Word Error Rate %	16.0	12.8	11.4	10.9	11.0	11.1

Baseline GMM performance = 15.4%

[Zeiler et al. "On rectified linear units for speech recognition" ICASSP 2013]

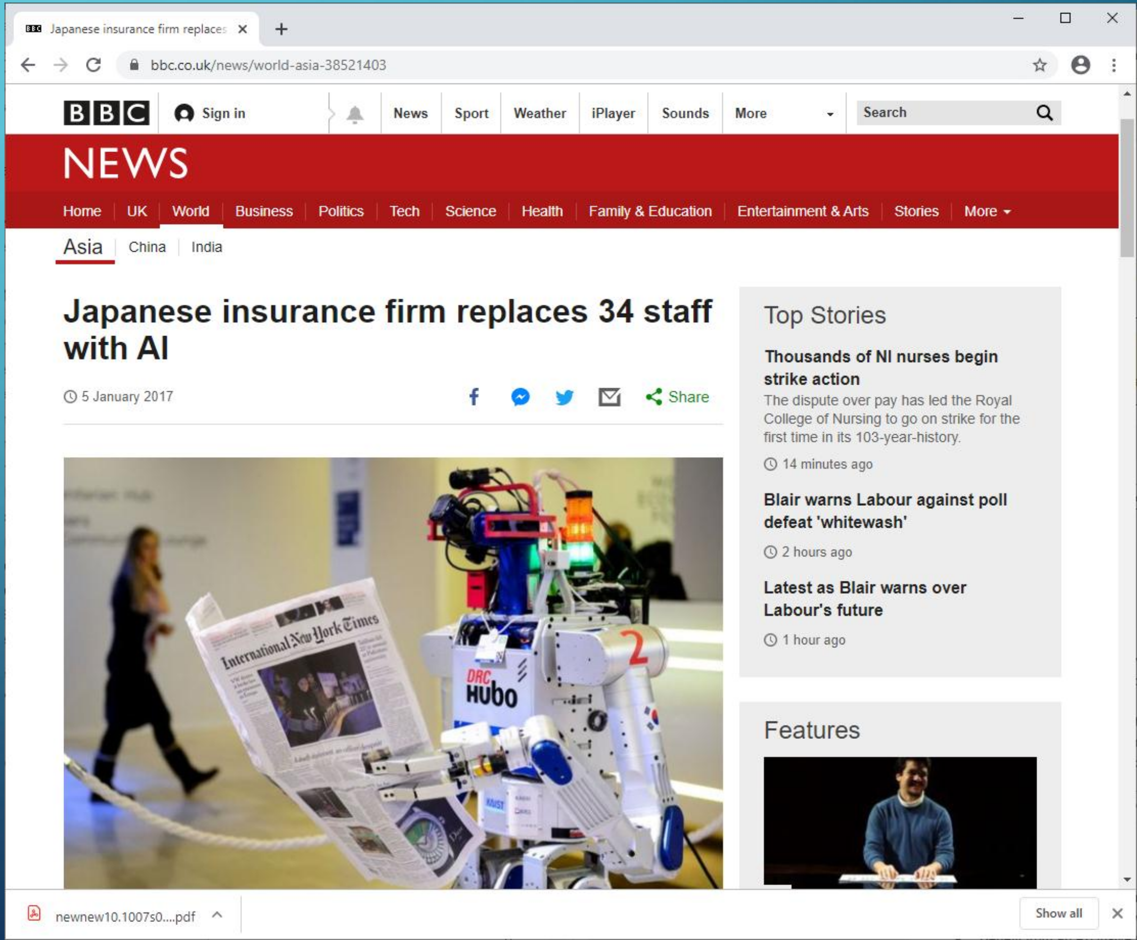
NEW TECHNOLOGIES

- Phones
- Cameras
- Amazon Alexa
- Google Net
- Microsoft Cortana
- Apple Siri

AI IS TAKING OVER HUMAN

- Telemarketing
- Bookkeeping Clerks
- Compensation and Benefits Managers
- Receptionists
- Couriers
- Proofreaders
- Computer Support Specialists
- Market Research Analysts
- Advertising Salespeople
- Retail Salespeople

- Human Resource Managers
- Sales Managers
- Marketing Managers
- Public Relations Managers
- Chief Executives
- Event Planners
- Writers
- Software Developers
- Editors
- Graphic Designers



THE FUTURE

TECHNOLOGY REALISES FICTION

Star Trek and Captain Kirk Inspired Mobile Phone Inventor, Martin Cooper 1973

Star Trek...LOS ANGELES - DECEMBER 29: William Shatner as Captain James T. Kirk. Original Air Date: 29 December 1967. (Photo by CBS via Getty Images)

MACHINES THAT CAN LEARN