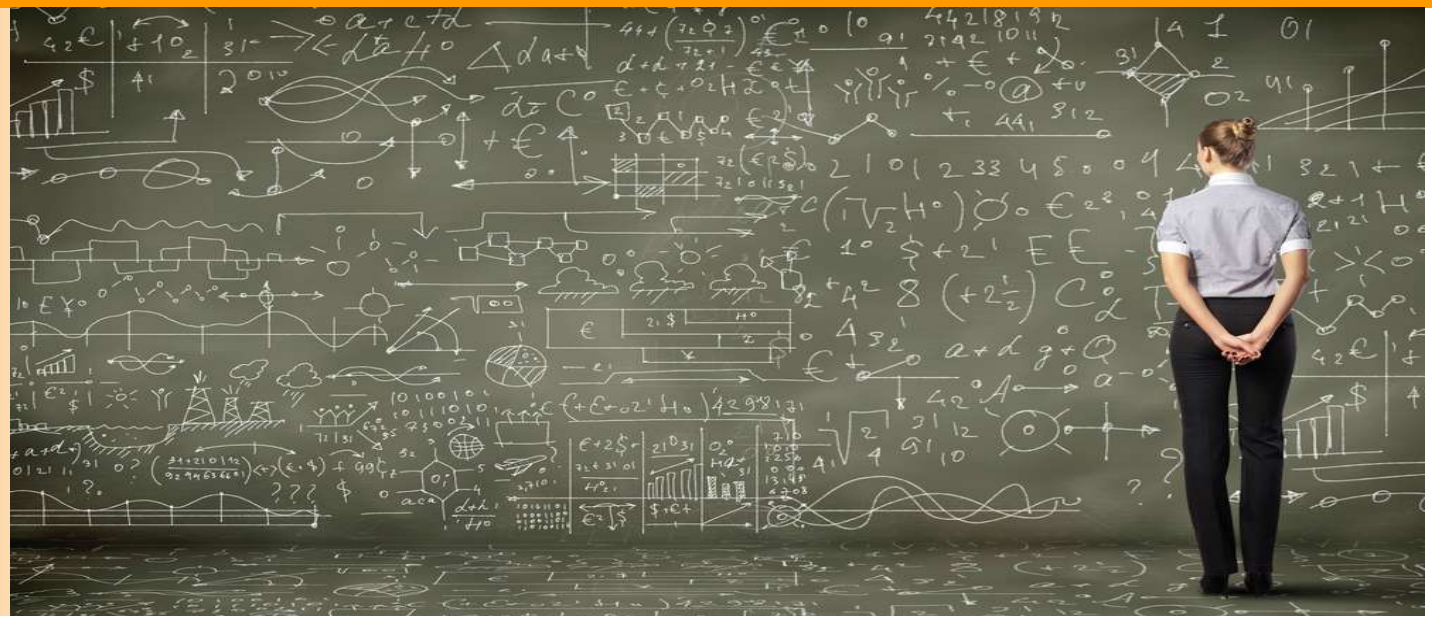
Overview of MAS651: Machine Learning II



About Me

Dr. Lan Wang
Professor

Department of Management Science
Email: lanwang@mbs.miami.edu

Office: KE 404

Office Hours:TBA

- 2019-present Full Professor, Department of Management Science
- 2003-2019 Assistant/Associate/Full Professor, School of Statistics, University of Minnesota.
- Ph.D. in Statistics, Pennsylvania State University, 2003.
- Elected fellow of ASA, IMS
- Co-editor of Annals of Statistics (2022-2024)
- https://sites.google.com/view/lanwang/home

Agenda

- Syllabus

Overview of the course:

- What is machine learning
- What are recommender systems?
- What is deep learning?
- Python

What is machine learning?

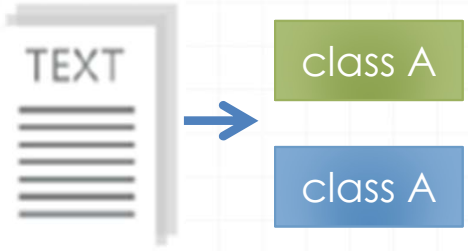
- ML focuses on the use of data and algorithms to imitate the way that humans learn
  - Self-driving car. The essence of machine learning.
  - Online recommendation offers such as those from Amazon and Netflix. Machine learning applications for everyday life.
  - Knowing what customers are saying about your product on social media? Machine learning combined with linguistic rule creation.
  - Fraud detection? One of the more obvious, important uses in our world today.

Types of Learning

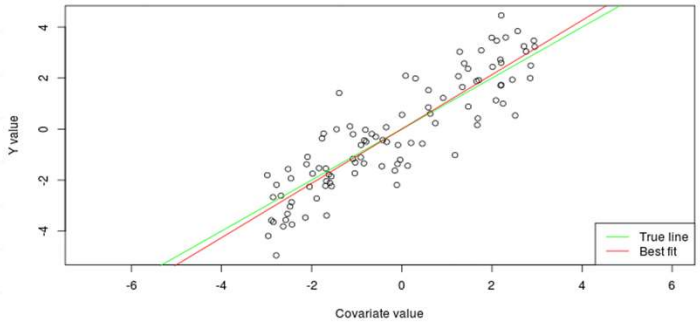
Supervised: Learning with a **labeled training** set
Example: email *classification* with already labeled emails

Unsupervised: Discover **patterns** in **unlabeled** data
Example: *cluster* similar documents based on text

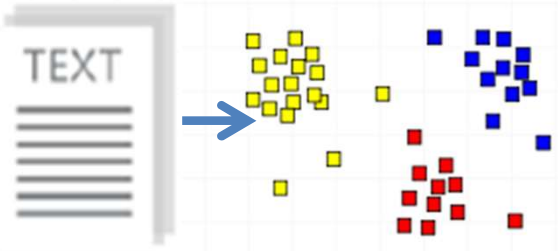
Reinforcement learning: learn to **act** based on **feedback/reward**
Example: learn to play Go, reward: *win or lose*



Classification



Regression



Clustering

Anomaly Detection
Sequence labeling
…

http://mbjoseph.github.io/2013/11/27/measure.html

Unsupervised Learning

- Example in retail: Customer transactions to consumer behavior:

  People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)

- No output of label/predicted value
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.

Learning Associations

- Basket analysis:

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  Example: $P(\text{ chips} \mid \text{beer}) = 0.7$

Market-Basket transactions

TID	*Items*
1	**Bread, Milk**
2	**Bread, Diaper, Beer, Eggs**
3	**Milk, Diaper, Beer, Coke**
4	**Bread, Milk, Diaper, Beer**
5	**Bread, Milk, Diaper, Coke**

Supervised Learning

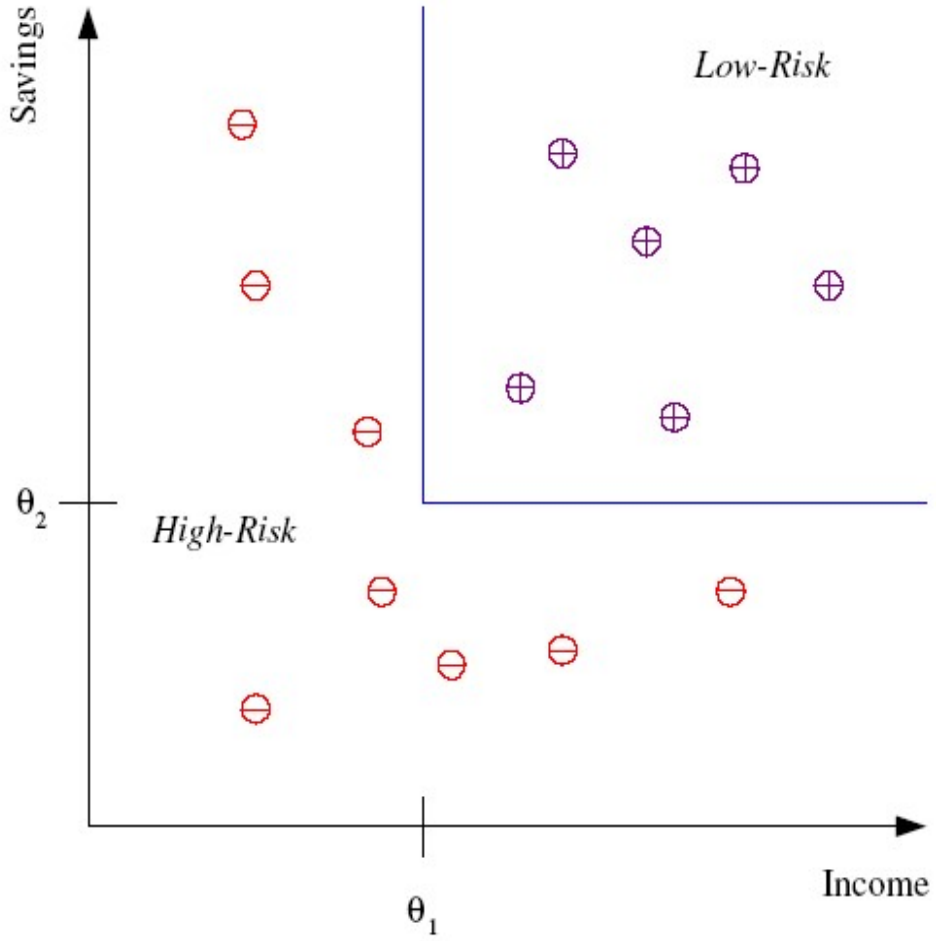
- Example: You can train a model for predicting car prices based on car attributes using historical sales data. That model can then predict the optimal price for new cars that haven't been sold before.

- In supervised learning, the data the algorithm "learns" from comes with the "correct" answers.
- The model created is then used to predict the answer for new, unknown values.

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	target
30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no
43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	no
39	services	married	secondary	no	9374	yes	no	unknown	20	may	273	1	-1	0	unknown	no
43	admin.	married	secondary	no	264	yes	no	cellular	17	apr	113	2	-1	0	unknown	no
36	technician	married	tertiary	no	1109	no	no	cellular	13	aug	328	2	-1	0	unknown	no
20	student	single	secondary	no	502	no	no	cellular	30	apr	261	1	-1	0	unknown	yes
31	blue-collar	married	secondary	no	360	yes	yes	cellular	29	jan	89	1	241	1	failure	no
40	management	married	tertiary	no	194	no	yes	cellular	29	aug	189	2	-1	0	unknown	no
56	technician	married	secondary	no	4073	no	no	cellular	27	aug	239	5	-1	0	unknown	no
37	admin.	single	tertiary	no	2317	yes	no	cellular	20	apr	114	1	152	2	failure	no
25	blue-collar	single	primary	no	-221	yes	no	unknown	23	may	250	1	-1	0	unknown	no
31	services	married	secondary	no	132	no	no	cellular	7	jul	148	1	152	1	other	no
38	management	divorced	unknown	no	0	yes	no	cellular	18	nov	96	2	-1	0	unknown	no

Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

Model

Customer Lifetime Value Modeling

- Customer lifetime value models are among the most important for eCommerce business to employ.
- These models predict the future revenue that an individual customer will bring to your business in a given period.
- With this information, you can focus your marketing efforts to encourage these customers to interact with your brand more often.

Customer churn modeling

- Customer churn modeling can help you identify which of your customers are likely to stop engaging with your business and why.
- The results of a churn model can range from churn risk scores for individual customers to drivers of churn ranked by importance.
- These outputs are essential components of an algorithmic retention strategy because they help optimize discount offers, email campaigns, or other targeted marketing initiatives that keep your high-value customers buying.

Reinforcement Learning

- Topics:
  - Policies: what actions should an agent take in a particular situation
  - Utility estimation: how good is a state (→used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome)
- Applications:
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...

RCH

The New York Times

Godzillium vs. Trumpium: Some Suggestions to Add to the Periodic Table

To Protect Against Zika Virus, Pregnant Women Are Warned About Latin American Trips

THE NEW OLD
F.T.C.'s Lum
Doesn't End
Training De

SCIENCE

Scientists See Promise in Deep-Learning Progr

By JOHN MARKOFF NOV. 23, 2012

BBC Sign in News Sport Weather Sho

NEWS

Home Video World UK Business Tech Science Magazi

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue

Archive > Volume 518 > Issue 7540 > News > Article

NATURE | NEWS

عربي

Game-playing software holds lessons neuroscience

DeepMind computer provides new way to investigate how the brai

Forbes / Tech

Top 20 Stocks for 2016

DEC 29, 2014 @ 11:37 AM 89,471 VIEWS

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World

'Deep learning' technology inspired by human brain

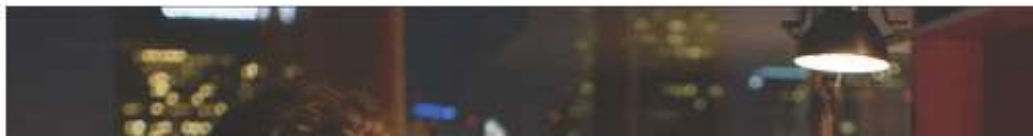
culture business lifestyle fashion environment tech trave

ndroids do dream of electric sheep

up feedback loop in its image recognition neural network - which

Google a step closer to developin machines with human-like intell

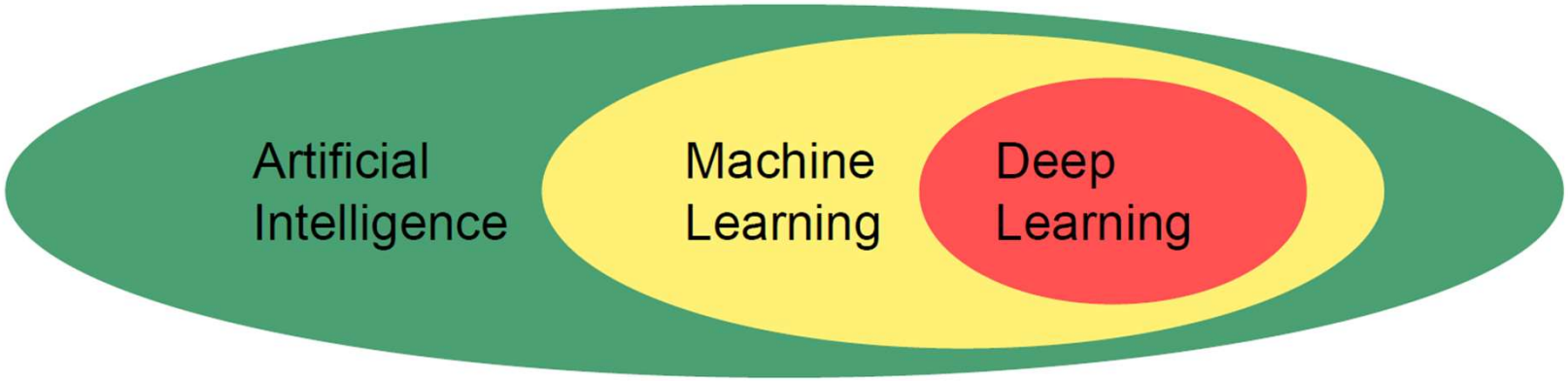
Algorithms developed by Google designed to encode thoughts, co computers with 'common sense' within a decade, says leading AI



What is Deep Learning?

AI: Any technique that enables computers to mimic human behavior

Deep Learning by Y. LeCun et al. Nature 2015

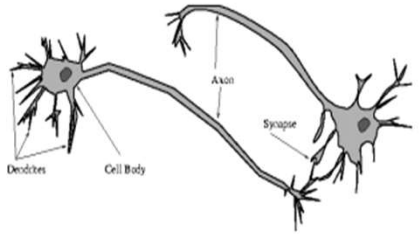


Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.

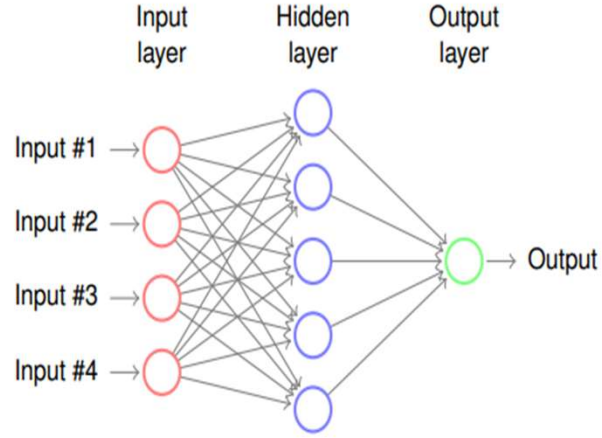
What are neural networks?

Neural networks are computational models inspired by biological neural networks.

- Biological



- Computational



What exactly is deep learning ?

The short answers

1. ‘Deep Learning’ means using a neural network with several layers of nodes between input and output.

2. The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

Why is DL useful?

- Manually designed features are often **over-specified**, **incomplete** and take a **long time to design** and validate
- Learned Features are **easy to adapt**, **fast** to learn
- Deep learning provides a very **flexible**, (almost?) **universal**, learnable framework for representing world, visual and linguistic information.
- Can learn both unsupervised and supervised
- Can utilize large amounts of training data

The traditional model of pattern recognition (since the late 50's)

- Fixed/engineered features (or fixed kernel) + trainable classifier

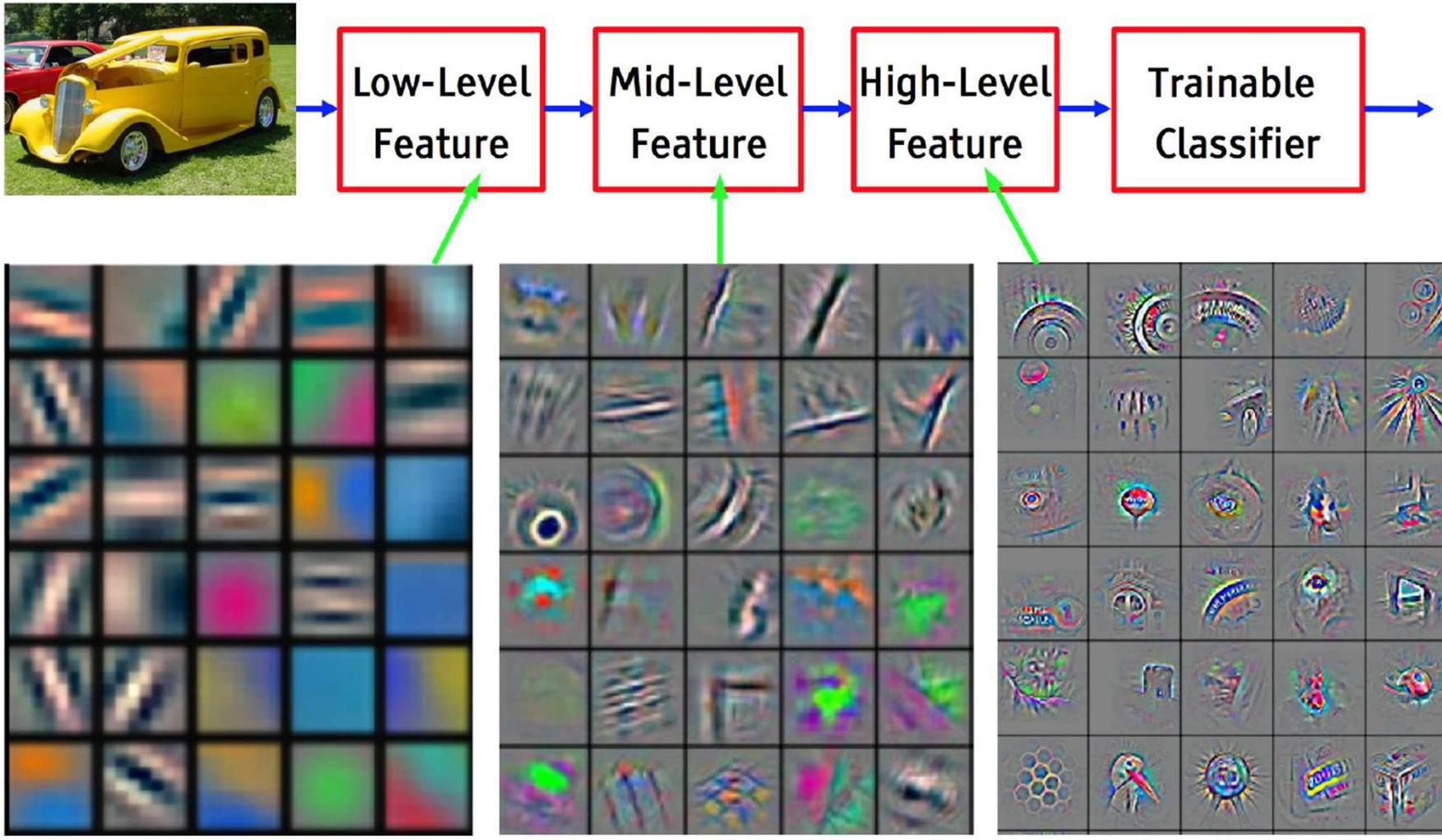
hand-crafted Feature Extractor → "Simple" Trainable Classifier →

End-to-end learning / Feature learning / Deep learning

- Trainable features (or kernel) + trainable classifier

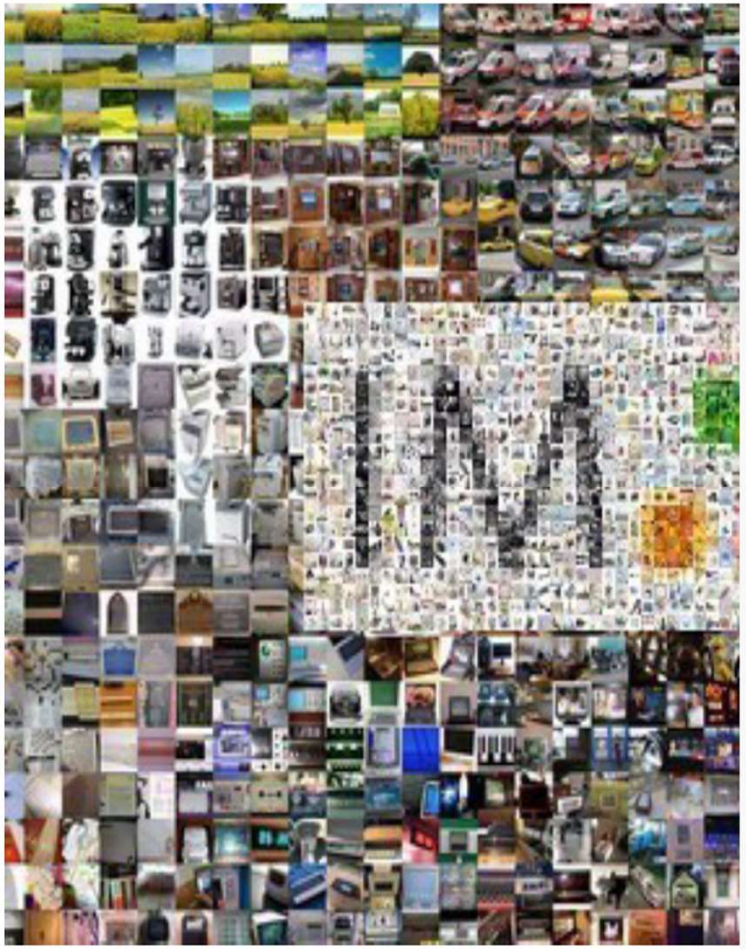
Trainable Feature Extractor → Trainable Classifier →

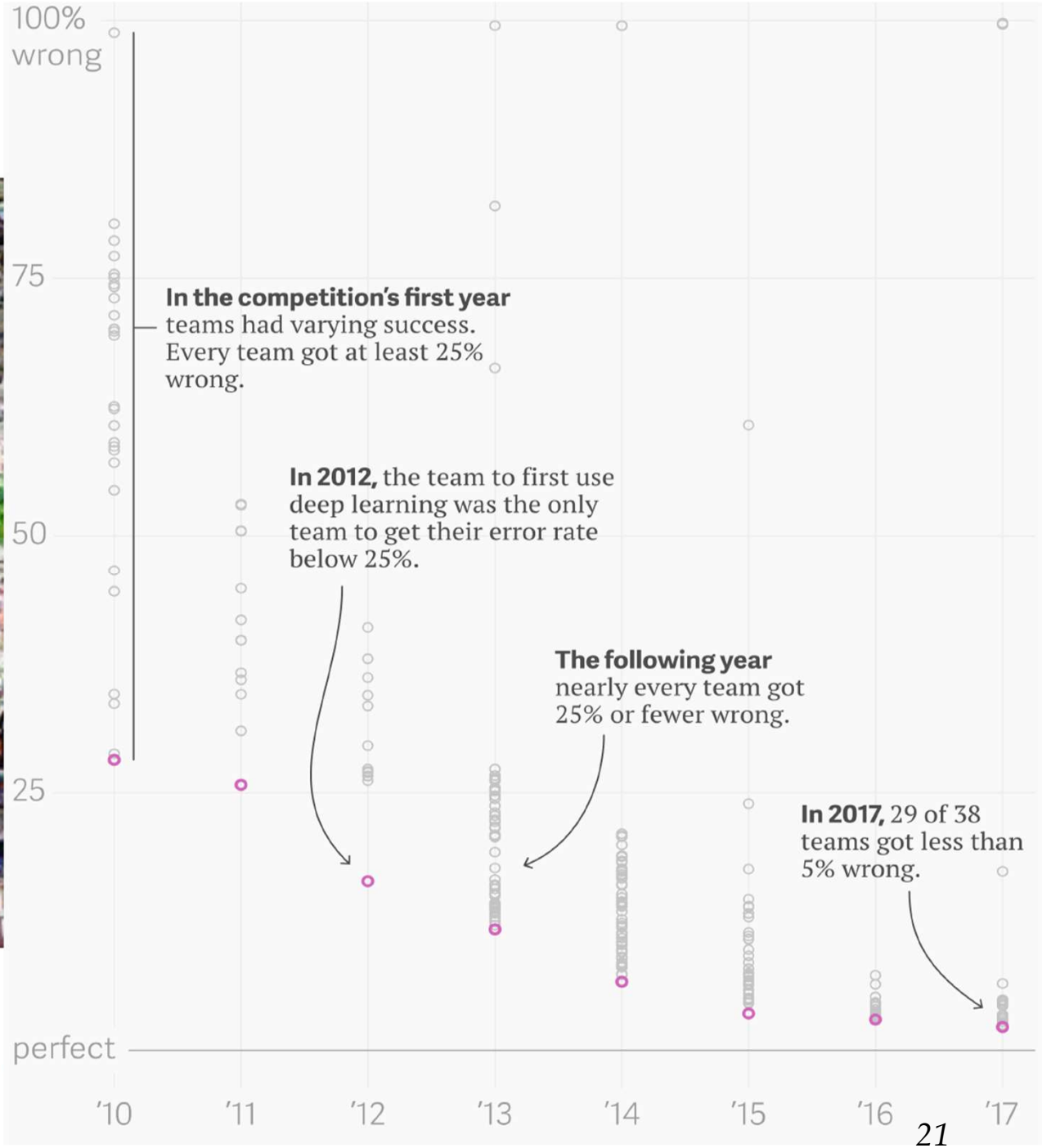
It's deep if it has more than one stage of non-linear feature transformation



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Classify Images





Yanofsky, Quartz
https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

An incomplete timeline of deep learning

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron

1958 New York Times...

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the hu-

ings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be con-

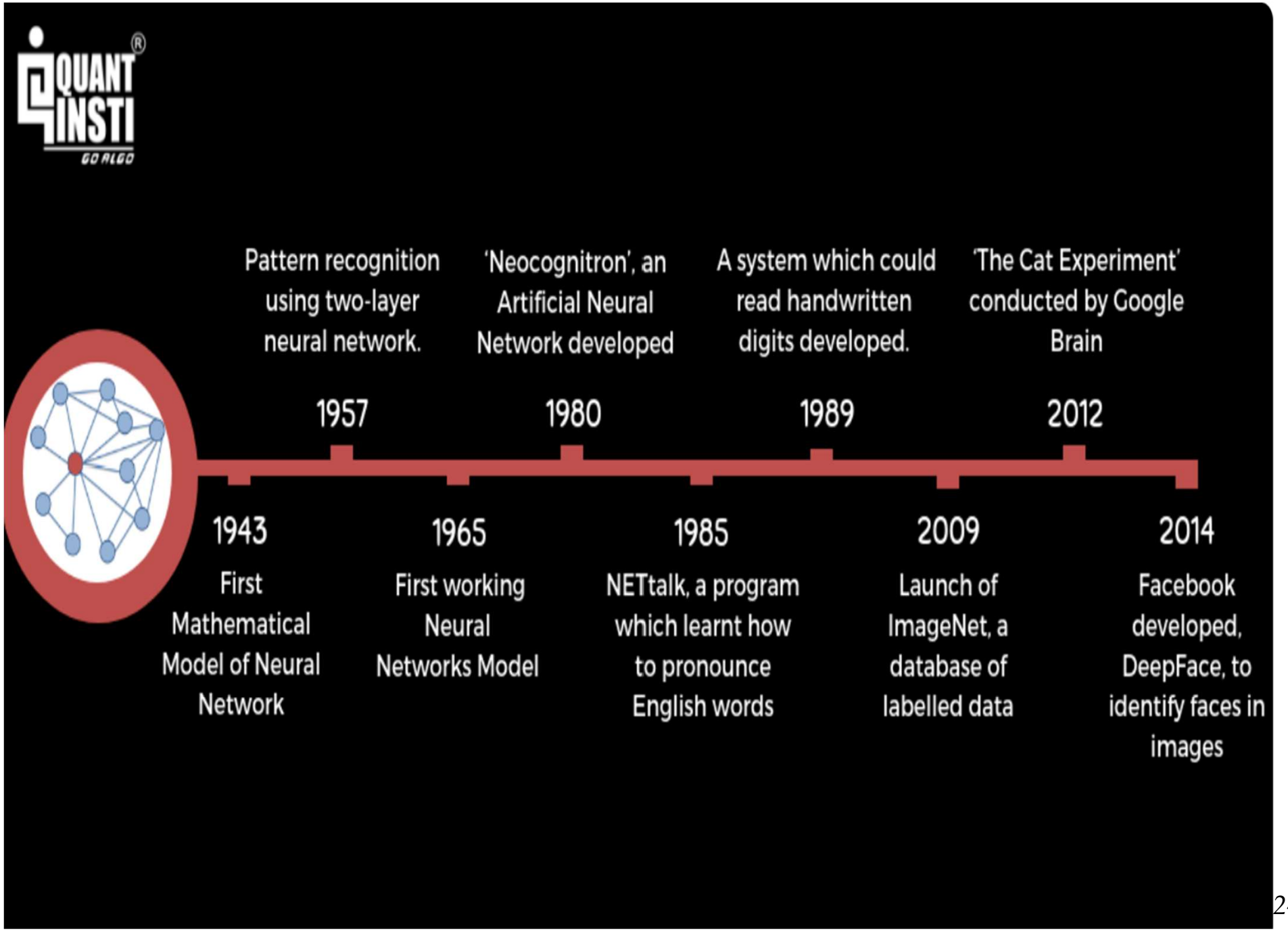
In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.


QUANT INSTI
GO ALGO
1943
First Mathematical Model of Neural Network
1957
Pattern recognition using two-layer neural network.
1965
First working Neural Networks Model
1980
'Neocognitron', an Artificial Neural Network developed
1985
NETtalk, a program which learnt how to pronounce English words
1989
A system which could read handwritten digits developed.
2009
Launch of ImageNet, a database of labelled data
2012
'The Cat Experiment' conducted by Google Brain
2014
Facebook developed, DeepFace, to identify faces in images

- 1957 - Frank Rosenblatt submitted a paper titled ‘The Perceptron: A Perceiving and Recognizing Automaton’, which consisted of an algorithm or a method for pattern recognition using a two-layer neural network.
- 1965 - Alexey Ivakhnenko and V.G. Lapa developed the first working neural network and Alexey Ivakhnenko created an 8-layer deep neural network in 1971 which was demonstrated in the computer identification system, Alpha. This was the actual introduction to deep learning.
- 1980 - Kunihiko Fukushima developed the ‘Neocognitron’, an artificial deep neural network with multiple and convolutional layers to recognize visual patterns.

- 1989 - Yann LeCun, using convolution deep neural network, developed a system which could read handwritten digits.

- 2009 - As deep learning models require a tremendous amount of labelled data to train themselves in supervised learning, Fei-Fei Li launched ImageNet, which is a large database of labelled images.

- 2012 - The results of 'The Cat Experiment' conducted by Google Brain were released. This experiment was based on unsupervised learning in which the deep neural network worked with unlabelled data to recognize patterns and features in the images of cats. However, it could only recognize 15% of images correctly.

- 2014 - Facebook developed, DeepFace, a deep learning system to identify and tag faces of users in the photographs.

Why now?

1. Big Data

- Larger Datasets
- Easier Collection & Storage

IMAGENET

WIKIPEDIA
The Free Encyclopedia

2. Hardware

- Graphics Processing Units (GPUs)
- Massively Parallelizable

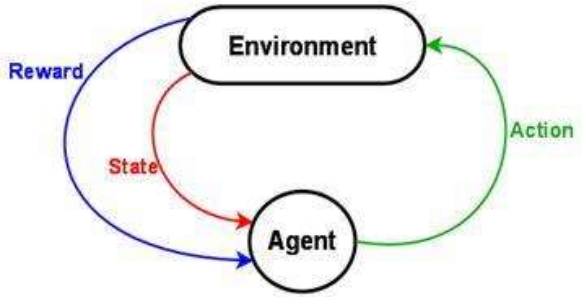
3. Software

- Improved Techniques
- New Models
- Toolboxes

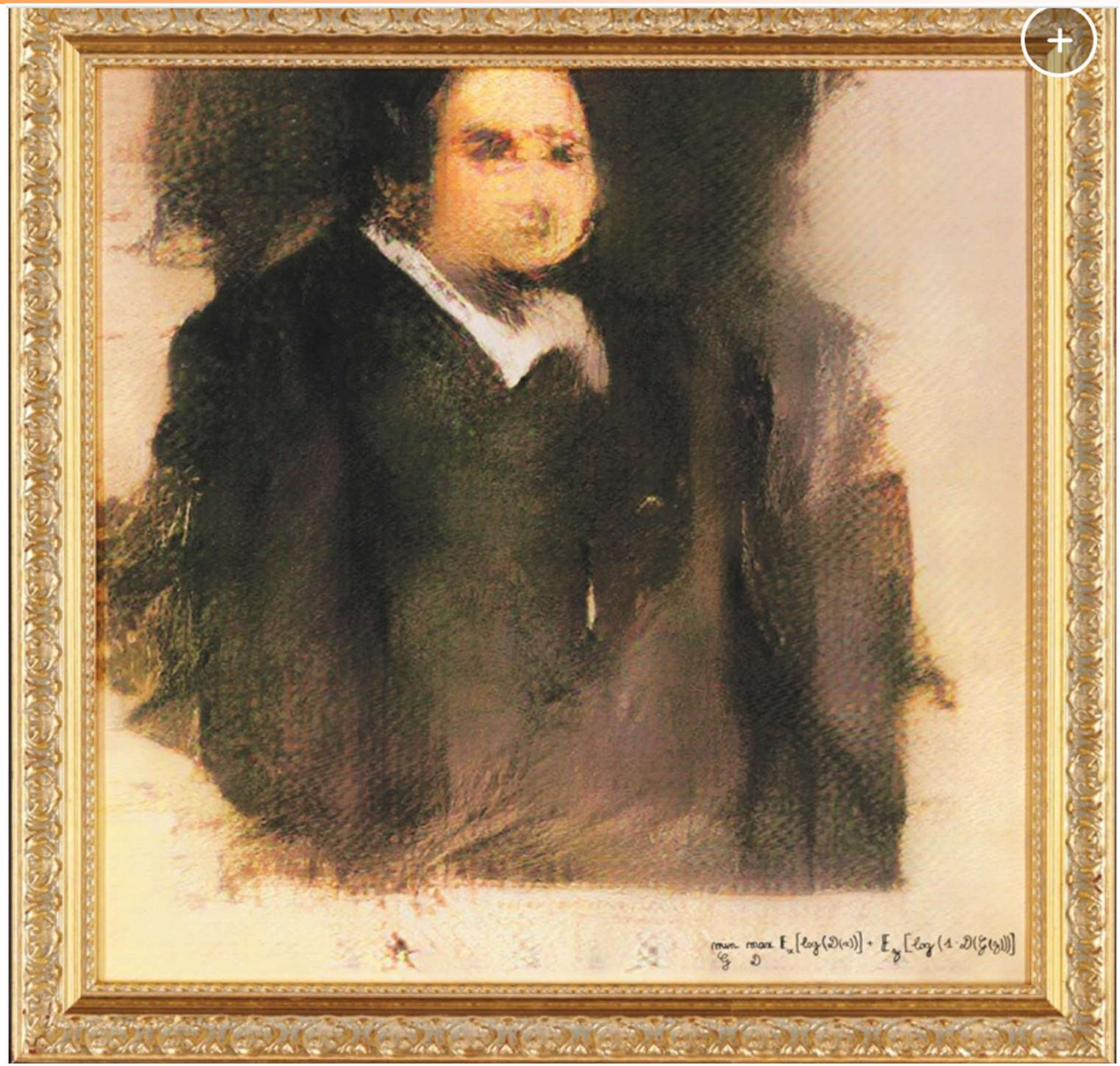
TensorFlow


ALPHAGO
00:01:00
LEE SEDOL
00:01:00

Milestones: Deep Reinforcement Learning



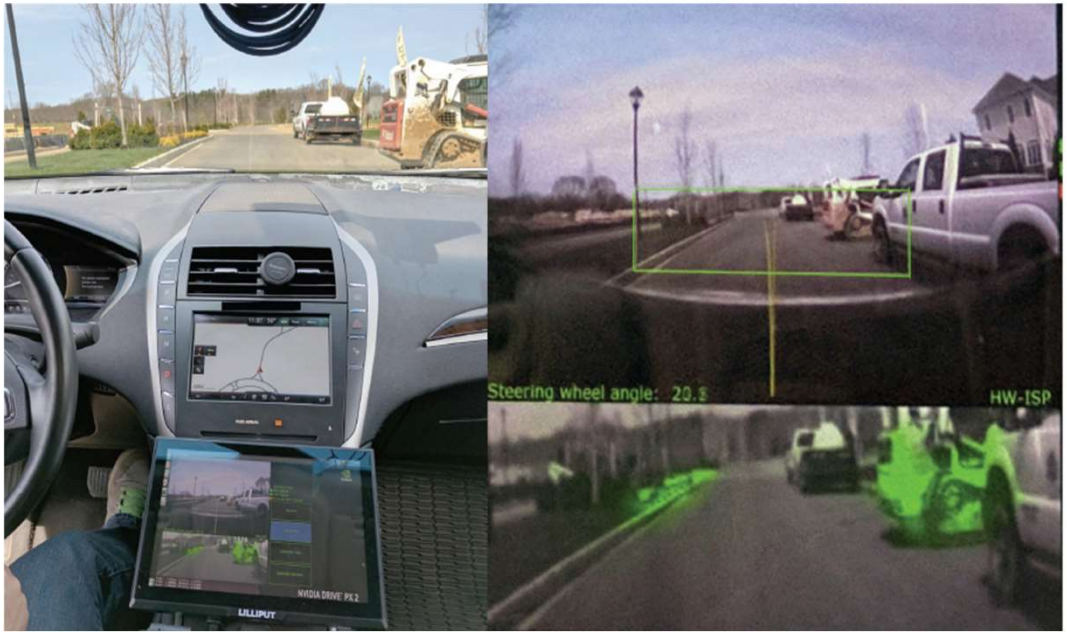
In 2016, Deep Mind's alphaGo defeats former world champion Lee Sedol


min max $E_x[\log(D(x))] + E_y[\log(1 - D(G(y)))]$
G D

- Portrait of Edmond Belamy, 2018, created by GAN (Generative Adversarial Network). Sold for $432,500 on 25 October at Christie's in New York.
- 'The algorithm is composed of two parts,' says Caselles-Dupré. 'On one side is the Generator, on the other the Discriminator. We fed the system with a data set of 15,000 portraits painted between the 14th century to the 20th. The Generator makes a new image based on the set, then the Discriminator tries to spot the difference between a human-made image and one created by the Generator. The aim is to fool the Discriminator into thinking that the new images are real-life portraits. Then we have a result.'

Self-driving cars

- Deep learning crucial for the global success of automotive autonomy



What will we learn in this course?

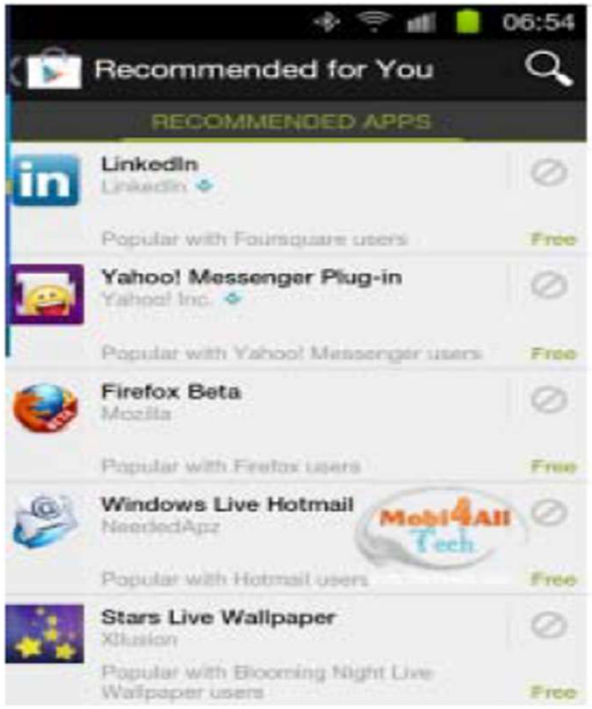
Topic 1: High-dimensional regression and classification

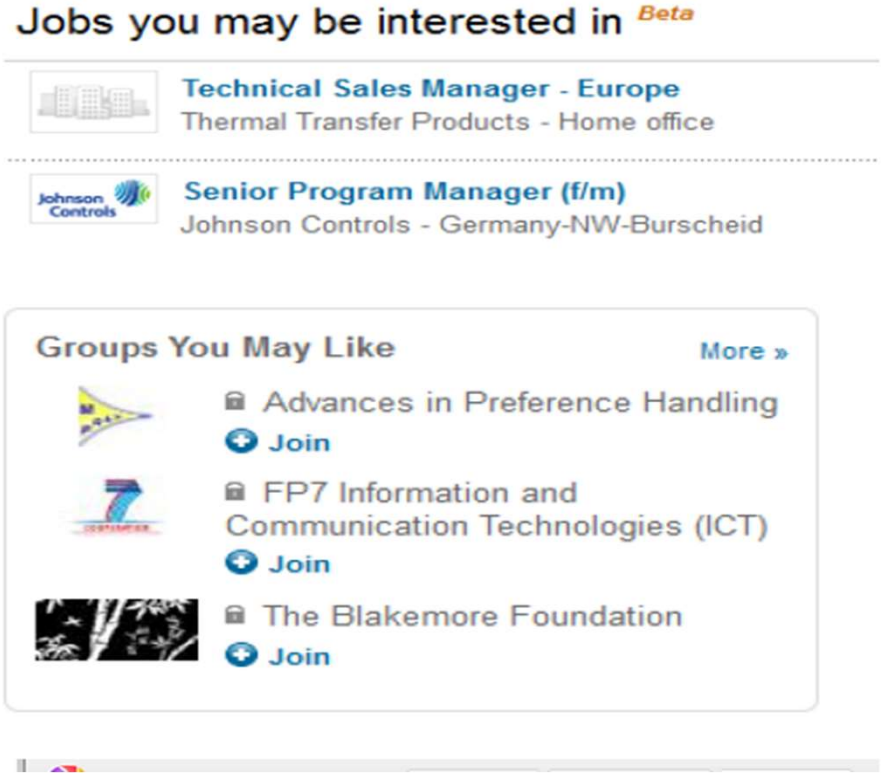
- P>>n number of variables larger than the sample size
- Lasso, penalized logistic regression, random forest, and other methods with Python

Topic 2: Recommender systems

- Basics of recommender systems
- Content-based filtering
- Neighborhood-based collaborative filtering
- Model-based methods
- Implementation in Python

Recommendation system





Why using Recommender Systems?

Value for the customer

- Find things that are interesting
- Narrow down the set of choices
- Help me explore the space of options
- Discover new things
- Entertainment
- …

Value for the provider

- Additional and probably unique personalized service for the customer
- Increase trust and customer loyalty
- Increase sales, click trough rates, conversion etc.
- Opportunities for promotion, persuasion
- Obtain more knowledge about customers

Movie Recommender System: MovieLens Dataset

- The MovieLens dataset is hosted by the GroupLens website. Several versions are available. We will use the MovieLens 100K dataset [Herlocker et al., 1999].

- This dataset is comprised of 100,000 ratings, ranging from 1 to 5 stars, from 943 users on 1682 movies. It has been cleaned up so that each user has rated at least 20 movies.

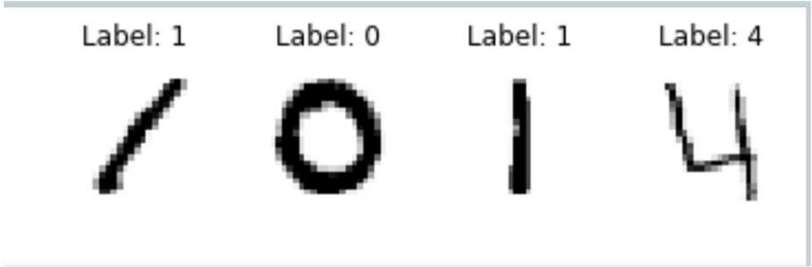
Topic 3: deep learning

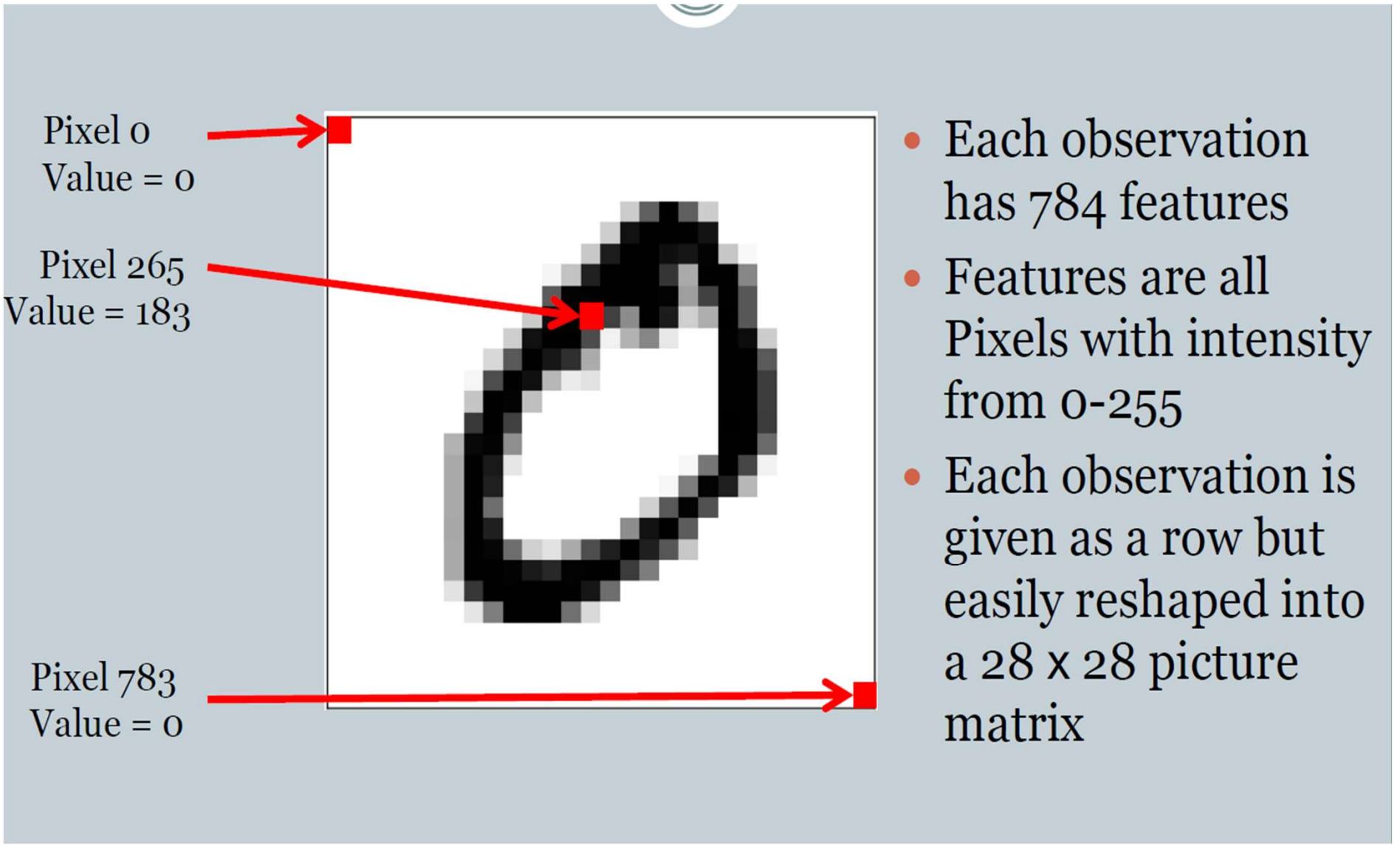
- Neural networks
- Deep learning basics: architecture, optimization, tuning
- Convolutional Neural Network (CNN)
- Implementation in Python

Handwriting recognition: MNIST Data

- We have 70,000 observations
  - 42,000 observations of training data
  - 28,000 observations of testing data

- Each observation represents an image of a handwritten digit 0-9:



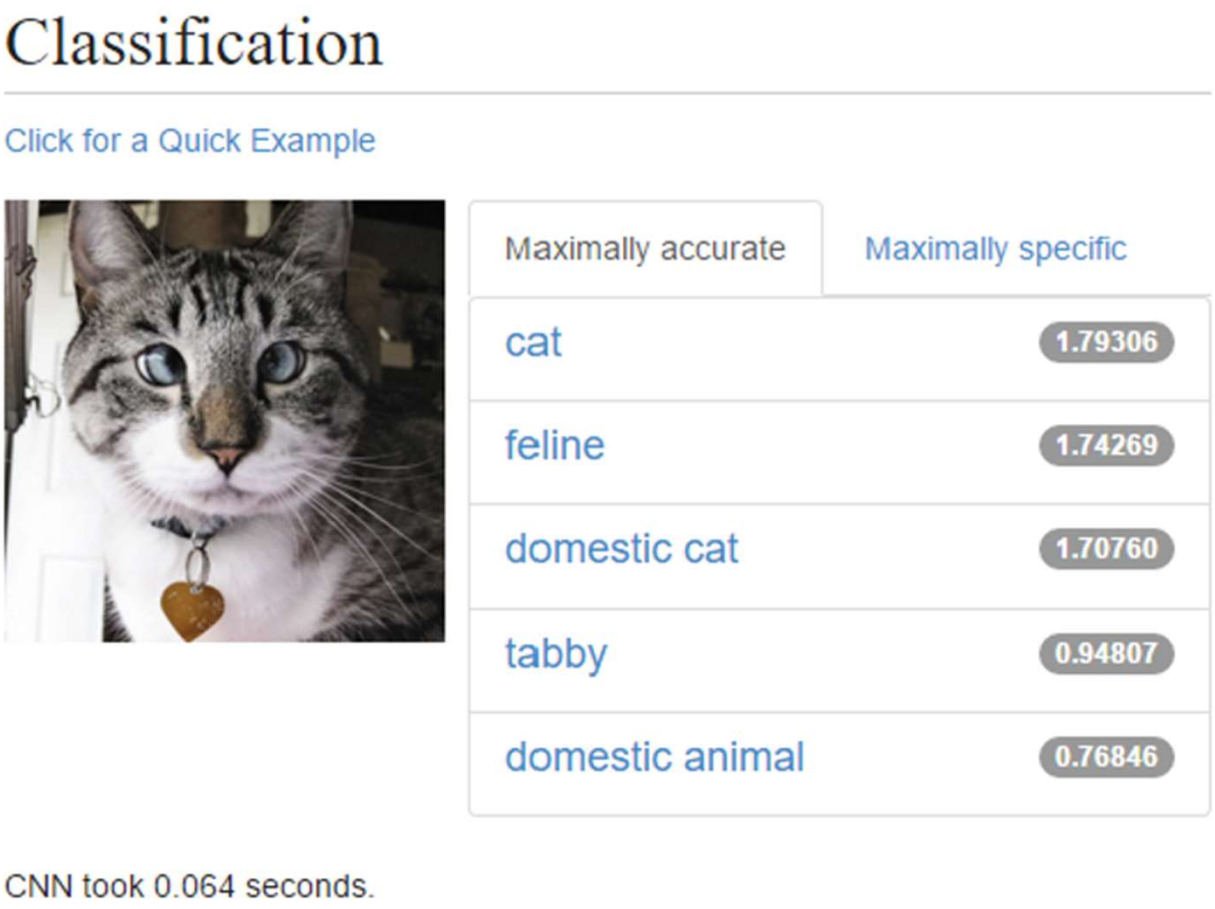


- Each observation has 784 features
- Features are all Pixels with intensity from 0-255
- Each observation is given as a row but easily reshaped into a 28 x 28 picture matrix

CNN's: what are they for?

- When you have data that doesn't neatly align into columns
  - Images that you want to find features within
  - Machine translation
  - Sentence classification
  - Sentiment analysis
- They can find features that aren't in a specific spot
  - Like a stop sign in a picture
  - Or words within a sentence

CNN for image classification



Try out a live demo at
http://demo.caffe.berkeleyvision.org/

Learn Political Affliations

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Congressional Voting Records Data Set

Download: Data Folder, Data Set Description

Abstract: 1984 United Stated Congressional Voting Records; Classify as Republican or Democrat

Data Set Characteristics:	Multivariate	**Number of Instances:**	435	**Area:**	Social
Attribute Characteristics:	Categorical	**Number of Attributes:**	16	**Date Donated**	1987-04-27
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	202231

The data set records the votes cast by the US House of Representatives during a particular session of Congress in 1984.

	party	handicapped-infants	water-project-cost-sharing	adoption-of-the-budget-resolution	physician-fee-freeze	el-salvador-aid	religious-groups-in-schools	anti-satellite-test-ban	aid-to-nicaraguan-contras	mx-missle	immigration
0	republican	n	y	n	y	y	y	n	n	n	y
1	republican	n	y	n	y	y	y	n	n	n	n
2	democrat	NaN	y	y	NaN	y	y	n	n	n	n
3	democrat	n	y	y	n	NaN	y	n	n	n	n
4	democrat	y	y	y	n	y	y	n	n	n	n

Sentiment analysis: Movie Reviews Data

- The dataset is the Large Movie Review Dataset often referred to as the IMDB dataset.

- The Large Movie Review Dataset (often referred to as the IMDB dataset) contains 25,000 highly polar moving reviews (good or bad) for training and the same amount again for testing. The problem is to determine whether a given moving review has a positive or negative sentiment.

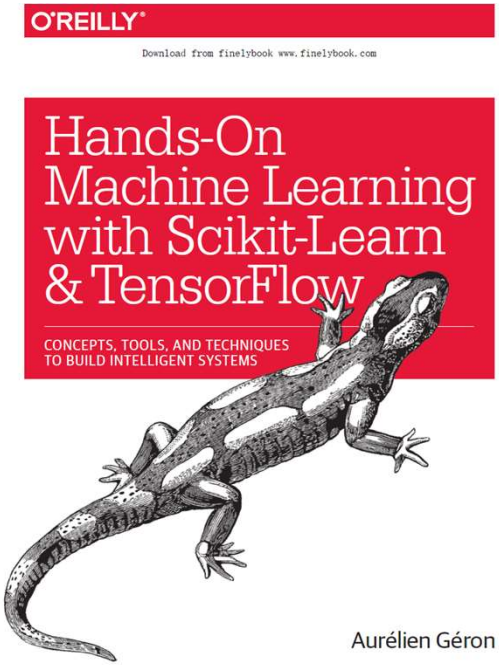
- The data was collected by Stanford researchers and was used in a 2011 paper where a split of 50/50 of the data was used for training and test. An accuracy of 88.89% was achieved.

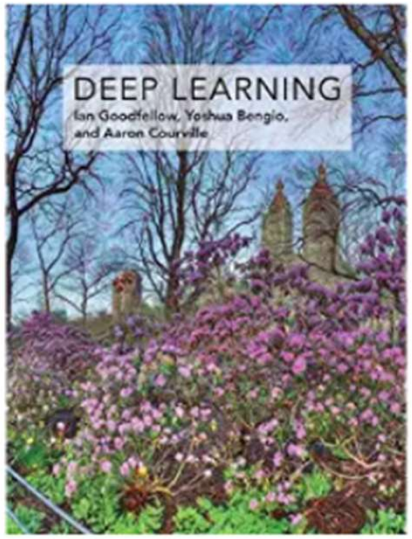
- The data was also used as the basis for a Kaggle competition titled “Bag of Words Meets Bags of Popcorn” in late 2014 to early 2015. Accuracy was achieved above 97% with winners achieving 99%.

Other topics (if we have time)

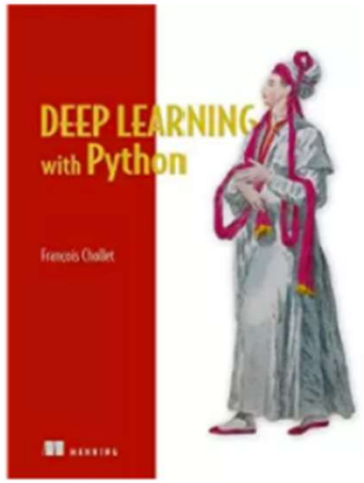
- Reinforcement learning
- Causal inference

References Books





Deep Learning (Adaptive Computation and Machine Learning series) Nov 18, 2016



Deep Learning with Python
Oct 31, 2017