Learning From Data:

Evolution and Revolution

Yaser S. Abu-Mostafa

California Institute of Technology
&
Machine Learning Consultants LLC
Outline

- Buzzwords/Terminology
- Historical Perspective
- The Essence
- The Revolution
- Challenges
Outline

- Buzzwords/Terminology
- Historical Perspective
- The Essence
- The Revolution
- Challenges
What are all these buzzwords?

Artificial Intelligence

**Machine Learning**

Big Data

Deep Learning

Neural Networks

Data Science

Pattern Recognition

Data Mining
Here is the dictionary!
Here is the dictionary!

Deep Learning = Neural Networks
Here is the dictionary!

Deep Learning $=$ Neural Networks

Neural Networks $\subseteq$ Machine Learning
Here is the dictionary!

Deep Learning = Neural Networks

Neural Networks ⊆ Machine Learning

Machine Learning ≈ Artificial Intelligence
Here is the dictionary!

Deep Learning $=\,$ Neural Networks

Neural Networks $\subset\,$ Machine Learning

Machine Learning $\approx\,$ Artificial Intelligence

All other buzzwords $\approx\,$ Machine Learning
Machine Learning $\rightarrow$ Artificial Intelligence

Different levels that capture our notion of intelligence:

- Performing complex tasks
- Learning new skills
- Innovation
- Taking Over / Rebellion

Revolution

Evolution
Machine Learning $\implies$ Artificial Intelligence

Different levels that capture our notion of intelligence:

- Performing complex tasks
  - 
  - 
  - 
  -
Machine Learning $\implies$ Artificial Intelligence

Different levels that capture our notion of intelligence:

- Performing complex tasks
- Learning new skills
Machine Learning $\implies$ Artificial Intelligence

Different levels that capture our notion of intelligence:

- Performing complex tasks
- Learning new skills
- Innovation
Different levels that capture our notion of intelligence:

- Performing complex tasks
- Learning new skills
- Innovation
- Taking Over / Rebellion
Machine Learning $\implies$ Artificial Intelligence

Different levels that capture our notion of intelligence:

- Performing complex tasks
- Learning new skills
- Innovation
- Taking Over / Rebellion

Evolution

Revolution

Hype?
Maybe it’s not hype! 😊

**Stephen Hawking:** “the development of full artificial intelligence could spell **the end of the human race**.”

**Elon Musk:** “I think we should be very careful about artificial intelligence. If I were to guess at what **our biggest existential threat** is, it’s probably that.”
Learning is the key buzzword

*Jeopardy*’s Watson (IBM) is a one-task machine. Big task, but one task.
Outline

- Buzzwords/Terminology
- Historical Perspective
- The Essence
- The Revolution
- Challenges
From Evolution to Revolution

Over the past 7 years, ML models and applications have moved

- From -

- **Low Hanging Fruit:**
  Almost any application can use ML for immediate benefit from the data.

- To -

- **Very Ambitious Goals:**
  Advanced ML creates super-human performance in intelligent tasks.

It was a long, bumpy road that led to this achievement.
Timeline of Ups and Downs

**ML**: Machine Learning  **AI**: Artificial Intelligence

- **1950's**: ML/AI is the great future.
Timeline of Ups and Downs

**ML**: Machine Learning    **AI**: Artificial Intelligence

- **1950′s**: ML/AI is the great future.
- **1970′s**: AI became a bad word.
Timeline of Ups and Downs

**ML**: Machine Learning  **AI**: Artificial Intelligence

- **1950’s**: ML/AI is the great future.
- **1970’s**: AI became a bad word.
- **1980’s**: Neural Networks are the great future.
Timeline of Ups and Downs

**ML**: Machine Learning  **AI**: Artificial Intelligence

- **1950’s**: ML/AI is the great future.
- **1970’s**: AI became a bad word.
- **1980’s**: Neural Networks are the great future.
- **1990’s**: Neural Networks are not that good.
Timeline of Ups and Downs

ML: Machine Learning  AI: Artificial Intelligence

- **1950’s**: ML/AI is the great future.
- **1970’s**: AI became a bad word.
- **1980’s**: Neural Networks are the great future.
- **1990’s**: Neural Networks are not that good.
- **2010’s**: Neural networks are great after all.
Timeline of Ups and Downs

**ML**: Machine Learning  **AI**: Artificial Intelligence

- **1950’s**: ML/AI is the great future.
- **1970’s**: AI became a bad word.
- **1980’s**: Neural Networks are the great future.
- **1990’s**: Neural Networks are not that good.
- **2010’s**: Neural networks are great after all.
- **Now**: ML/AI is the great future.
- Then and Now -

**CNN**

1989 by Bell Labs

**AlexNet**

2012 by Alex Krizhevsky
- Then and Now -

**TD Gammon**
1992 by G. Tesauro (IBM)

**AlphaZero**
2017 by DeepMind
1st Wave of Success: Financial Applications

- Market forecasting

- Financial model calibration

- Consumer and corporate credit assessment.
Biggest Early Success of Neural Networks

Detection of credit-card fraud - huge commercial success:

#### NEURAL NETWORKS: HECHT-NIELSON NEUROCOMPUTERS WINS FUNDS

**CBR STAFF WRITER**

**13TH AUGUST 1987**

The radically new computing technology of neural networking, which mimics in simple form the operation of the central nervous system of vertebrates, has taken a significant step forward with four venture capital firms coming forward to finance one of the first start-up companies in the field. The company is Hecht-Nielsen Neurocomputer Corp, San Diego, which [...]
2nd Wave of Success: E-commerce

- Recommender systems (Amazon, fashion, ...)

- Profiling
Famous ML e-commerce problem

2006 - 2009

US$1,000,000 Prize for the first 10% improvement
3rd Wave of Success: Perception Tasks

The last 7 years witnessed a huge ML surge in computer vision and other perception tasks.

speech recognition  object detection  machine translation

The revival of neural networks, with more layers this time.
The comeback of Neural Networks

The dominant ML models over the past 40 years:

- **1980’s**: Neural Networks

- **1990’s**: Support Vector Machines

- **2000’s**: Boosting Algorithms

Outline

- Buzzwords/Terminology
- Historical Perspective
  - The Essence
- The Revolution
- Challenges
1. What Machine Learning Does

The technical core of all these fields:

- *Machine Learning*
- *Artificial Intelligence*
- *Data Mining*
- *Pattern Recognition*

“Automated detection of a pattern based on the data”
Example: Credit Approval

Given the data of an applicant:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>23 years</td>
</tr>
<tr>
<td>gender</td>
<td>male</td>
</tr>
<tr>
<td>annual salary</td>
<td>$30,000</td>
</tr>
<tr>
<td>years in residence</td>
<td>1 year</td>
</tr>
<tr>
<td>years in job</td>
<td>1 year</td>
</tr>
<tr>
<td>current debt</td>
<td>$15,000</td>
</tr>
</tbody>
</table>

should we extend credit?
Human Solution versus Machine Learning

Historical records of good and bad customers used to decide the boundary between credit approval and denial.

The learning algorithm constructs the boundary based on the data.
2. When should ML be used?

ML is the technology of choice when:

- A pattern exists.
- We cannot pin it down mathematically.
- We have a representative data set.
2. When should ML be used?

ML is the technology of choice when:

- A pattern exists.
- We cannot pin it down mathematically. \(\times\) We can be just lazy 😊
- We have a representative data set.
3. The building block

The Deep Neural Network - most successful ML model to date:
3. The building block

The Deep Neural Network - most successful ML model to date:

(a) **Expressive:** Higher and higher level representations capture complex patterns.
3. The building block

The Deep Neural Network - most successful ML model to date:

(b) Composable: Chain rule (backpropagation) allows for various compositions.
3. The building block

The Deep Neural Network - most successful ML model to date:

(c) Flexible: Architecture easily modified to incorporate different functionalities.
3. The building block

The Deep Neural Network - most successful ML model to date:

(d) Specialization: CNN (vision), LSTM (NLP), GAN (generative), Embedding.
3. The building block

The Deep Neural Network - most successful ML model to date:

(d) Specialization: CNN (vision), LSTM (NLP), GAN (generative), Embedding.
3. The building block

The Deep Neural Network - most successful ML model to date:

(e) **Soft Objective**: Reinforcement learning and related paradigms.
3. The building block

The **Deep Neural Network** - most successful ML model to date:

(f) Generalizes well: Not susceptible to overparameterization!!!
3. The building block

The Deep Neural Network - most successful ML model to date:

(f) Generalizes well: Not susceptible to overparameterization!!!
Outline

- Buzzwords/Terminology
- Historical Perspective
- The Essence
- The Revolution
- Challenges
1st Factor: Learned Features

Automated Feature Extraction:

- Getting our ‘wisdom’ out of the way!
- We still sneak in by designing the architecture.
2nd Factor: Jump in computation speed

Commercially available specialized hardware

Gain in ML speed can be more than 2 orders of magnitude
3rd Factor: Elaborate data resources

Using multiple data sources:

For example, using movie preferences, Facebook posts, Amazon purchases, etc.

to profile a person.
4th Factor: Crowdsourcing

Public and Free Software/Hardware Resources:

- TensorFlow
- GitHub
- Google Colab
- AWS

No barrier to entry; a great opportunity for researchers in the developing world.
Main Achievement: Superior “Intelligence”

**From:** Replicating human skills  **To:** Beating human intelligence

ML system can discover novel patterns and strategies beyond human capacity.
Upcoming Breakthrough

Self-driving cars will have a huge economic and social impact.
Outline

- Buzzwords/Terminology
- Historical Perspective
- The Essence
- The Revolution
- Challenges
1. Technical Challenge

How does the neural network avoid overfitting?

There are some partial answers (optimization method, special minima). The dilemma:

For some problems, the network overfits. In some cases, it overfits terribly.
2. Practical Challenges

Two interrelated challenges:

- **Bias**: Inadvertently allowing gender and other biases.

- **Interpretability**: Digging into the ‘black box’ of a neural network

  Understanding what the network is doing & Avoiding irrelevant traps

Latest Google effort: **TCAV** (Testing with Concept Activation Vectors)
3. Human Challenges

1. **Security Risks:**
   - Hacking on steroids - Super Intelligence

2. **Social Risks:**
   - Replacing human workers - Human interaction
Conclusions
Conclusions

- ML/AI will replace “routine intelligence” in the next 20 years.
Conclusions

- ML/AI will replace “routine intelligence” in the next 20 years.

- Profound impact on the economy and on security.
Conclusions

- ML/AI will replace “routine intelligence” in the next 20 years.

- Profound impact on the economy and on security.

- The fast change will have serious social ramifications.
Conclusions

• ML/AI will replace “routine intelligence” in the next 20 years.

• Profound impact on the economy and on security.

• The fast change will have serious social ramifications.

• If you ignore AI, you will become medieval (technologically).
Conclusions

- ML/AI will replace “routine intelligence” in the **next 20 years**.
- Profound impact on the economy and on security.
- The fast change will have serious social ramifications.
- If you ignore AI, you will become **medieval** (technologically).
- AI is **not** Big Brother!
Further Reading

- **Online Lectures:** [http://work.caltech.edu/telecourse](http://work.caltech.edu/telecourse)

- **Book:** *Learning From Data*