Open Issues in Deep Learning

Maria Vakalopoulou
Center for Visual Computing,
CentraleSupélec
Université Paris-Saclay

CentraleSupélec

CENTRE FOR
VISUAL COMPUTING
CENTRE DE VISION NUMÉRIQUE
CENTRALESUPÉLEC | INRIA

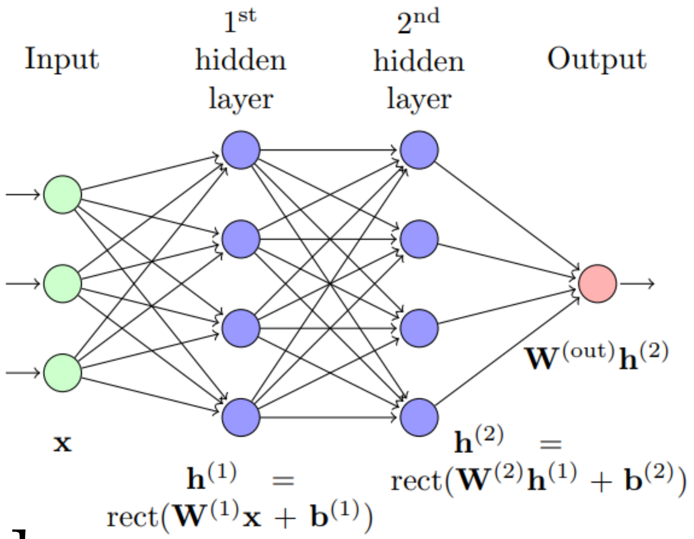
Friday 16 March, 2018

Outline

- Non-convex Optimization
- Adversarial Examples
- Real-life Challenges

Non-Convex Optimization

- Explore the complexity and expression power of deep feedforward MLP networks with rectified linear activations. *[Pascanu et al 2014]*
  - Parameters:
    - k : # hidden layers
    - n_0 : input dim
    - n : # neurons per layer
    - 1:output dim
  - Deep [k hidden layers with n neurons]
  - Shallow [1 hidden layer with kn neurons]
  - Both are piecewise linear functions (due to the their layers)


Input
1st hidden layer
2nd hidden layer
Output
$\mathbf{x}$
$\mathbf{h}^{(1)} = \mathrm{rect}(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})$
$\mathbf{h}^{(2)} = \mathrm{rect}(\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)})$
$\mathbf{W}^{(out)}\mathbf{h}^{(2)}$

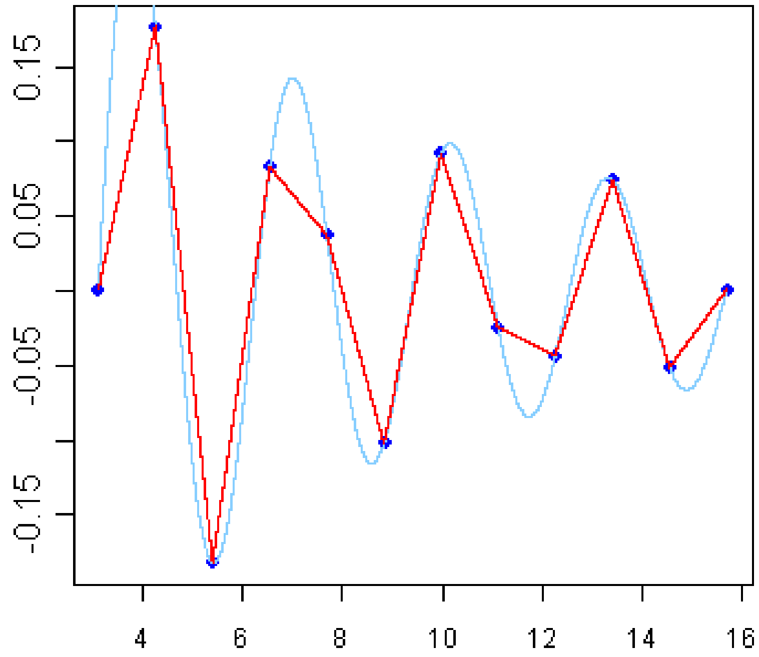
$$s = \sum_{i \in [n]} w_i x_i + b$$

$$\mathrm{rect}(s) = \max\{0, s\}$$

Non-Convex Optimization

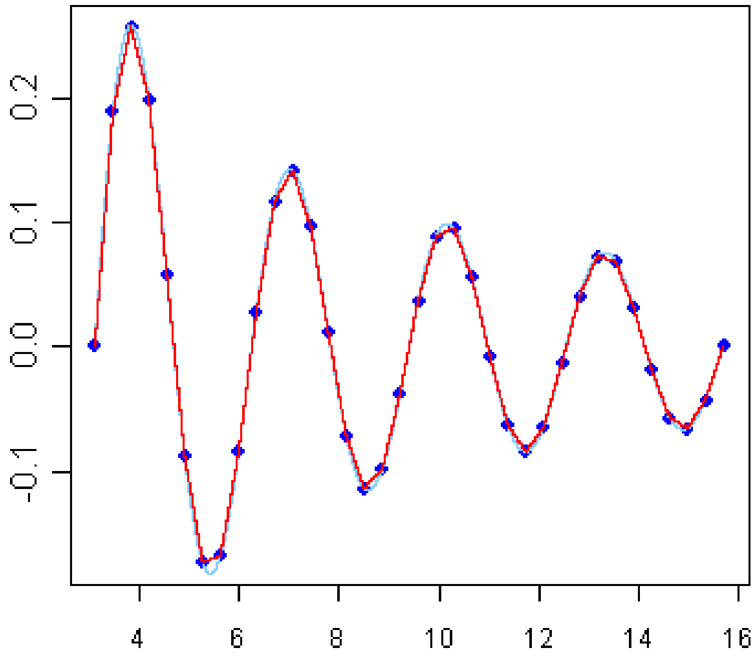
- How well do they approximate arbitrary functions?

12 Points



RMS error = 5.04147E - 2

36 Points



RMS error = 5.64535E - 3

- Corresponds to the number of pieces (linear regions).

Non-Convex Optimization

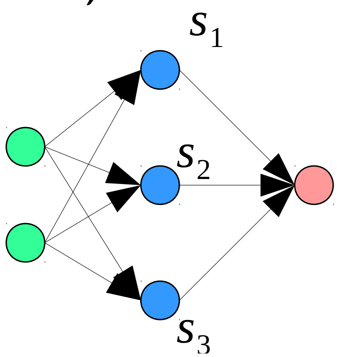
- Give upper bounds on the number of regions with n_0 as constant:

  - Note that the last layer is a linear function → # of regions is not increasing.

- Shallow: $O(k^{n_0} n^{n_0})$ *[Zaslavsky, 1975]*

  Actually, since each rectifying unit acts as a hyperplane dividing the hyperspace (R^{n_0}) the problem is well studied (hyperplane arrangements) and the result is known.

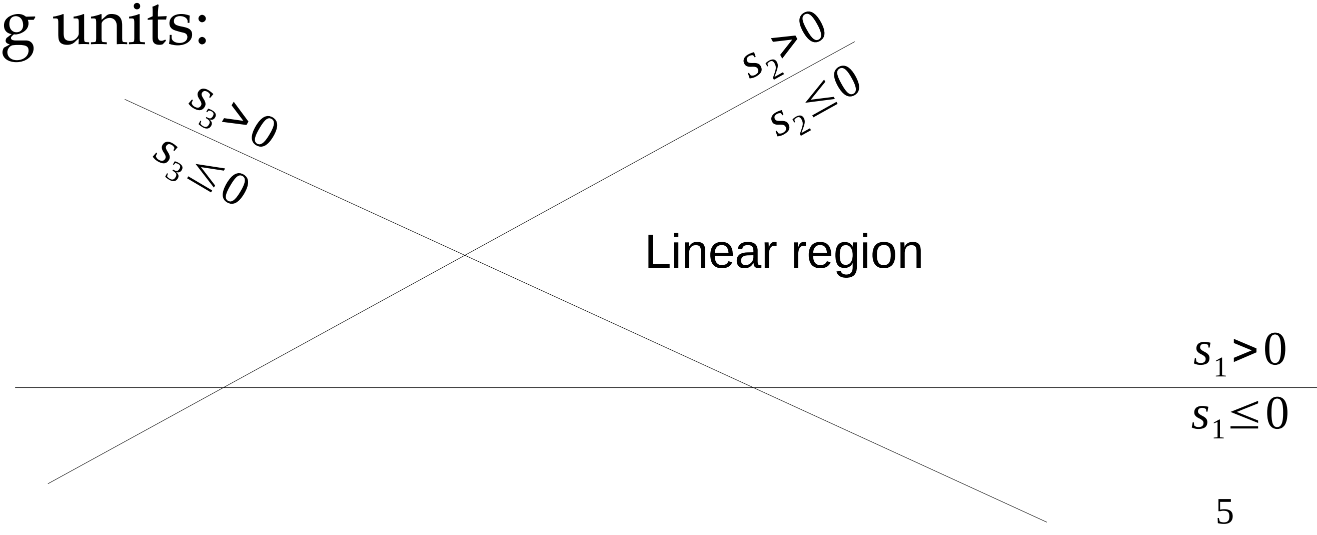
- $n_0=2$, $n=3$, $k=1$ rectifying units:



$$s = \sum_{i \in [n]} w_i x_i + b$$

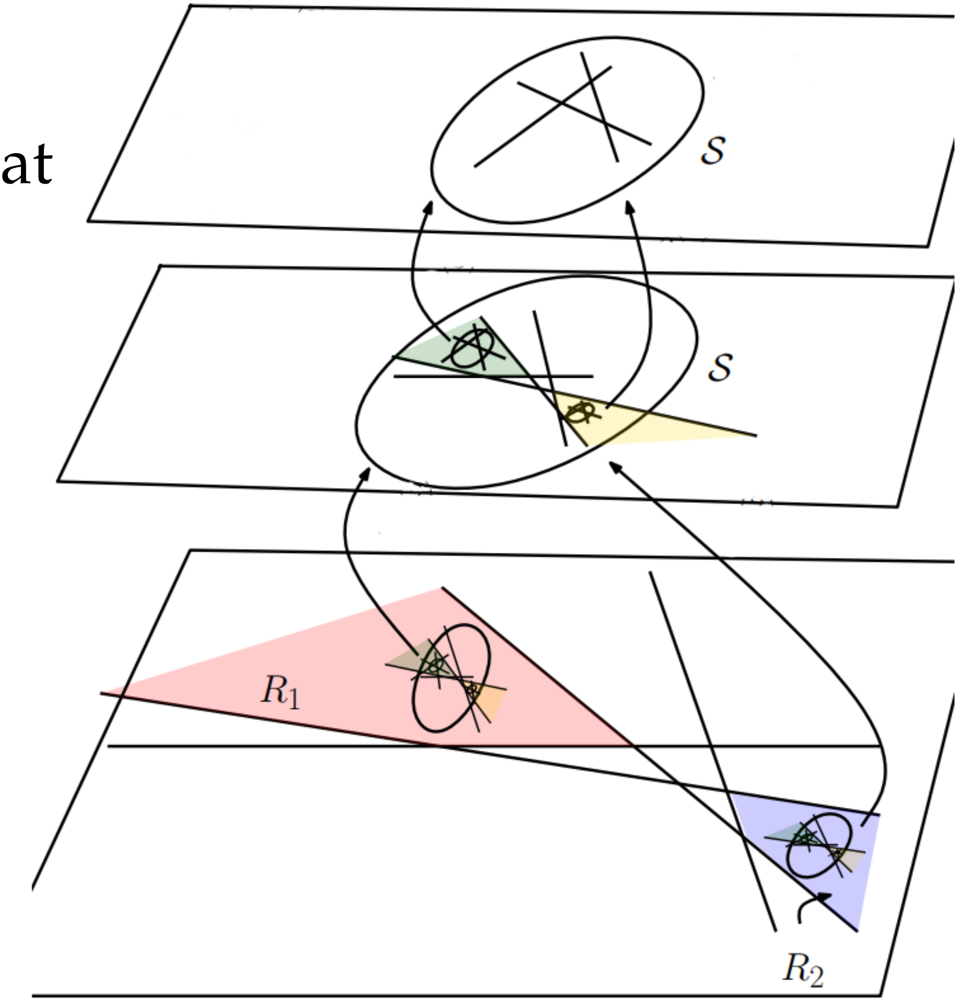
$$\mathrm{rect}(s) = \max\{0, s\}$$

Equation of a plane



Non-Convex Optimization

- Deep: $O([n/n_0]^k n^{n_0}/k)$ → the proof is out the scope of this talk but intuitively the proof shows that each layer can be considered an arrangement and that these arrangements can be embedded into each other.
- This shows that although both networks have the same number of neurons the deep possesses higher expressive power i.e., might approximates functions with possibly more piecewise linear regions.



Non-Convex Optimization

- Try to explain the paradigm of optimizing the highly non-convex neural network objective function through the prism of spin-glass theory. *[Choromanska et al 2015]*
  - For large size networks, most local minimum are equivalent and yield similar performance on a test set.
  - The probability of finding a “bad” local minimum is non-zero for small-size networks and decreases quickly with network size.
  - Trying to find the global minimum on the training set is not useful in practice and may lead to overfitting.

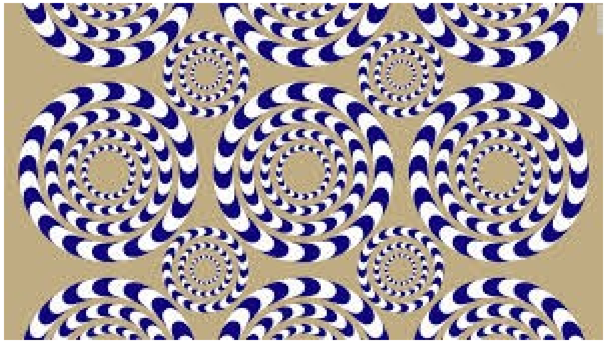
Non-Convex Optimization

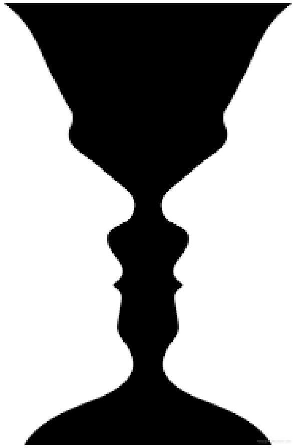
- Try to explain the paradigm of optimizing the highly non-convex neural network objective function through the prism of spin-glass theory. *[Choromanska et al 2015]*
  - For large size networks, most local minimum are equivalent and yield similar performance on a test set.
  - The probability of finding a “bad” local minimum is non-zero for small-size networks and decreases quickly with network size.
  - Trying to find the global minimum on the training set is not useful in practice and may lead to overfitting.

  - **Paradox:** The latent space we are looking for can be expressed using smaller networks. Deep networks after trained can be compressed by removing ~90% parameters to give similar performance. *[Rueda etal 2017, Han etal 2016]*

Adversarial Examples

- What are they?
  - Like optical illusions to humans.

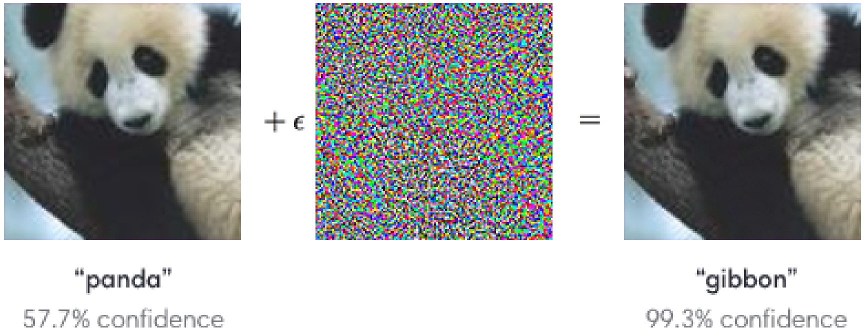


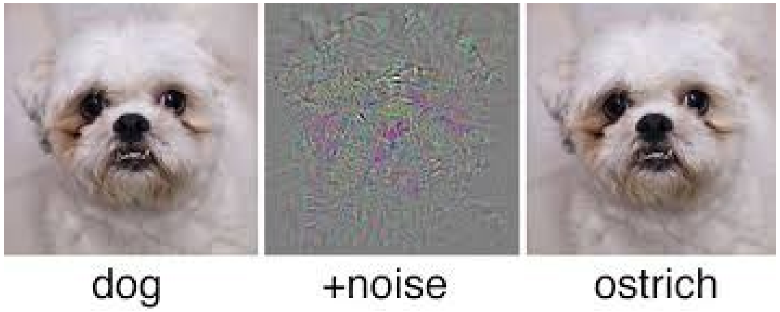


- Why they are important?
  - Create security issues for the deployment of deep leaning in the physical world: autonomous cars, automatic speech recognition, reinforcement learning algorithms, etc

Adversarial Examples

- What are they?
  - Like optical illusions to humans.





- Given a trained model f, an original input sample x, generating an adversarial example x' can be described as a constrained optimization problem.

$$\begin{aligned} \min_{x'} \quad & \|x' - x\| \\ s.t. \quad & f(x') = l', \\ & f(x) = l, \\ & l \neq l', \\ & x' \in [0,1], \end{aligned}$$

$$\eta = x' - x$$

$$\begin{aligned} \min_{x'} \quad & J(f(x'), l') \\ s.t. \quad & \|\eta\| \leq \epsilon, \\ & f(x) = l, \\ & l \neq l', \end{aligned}$$

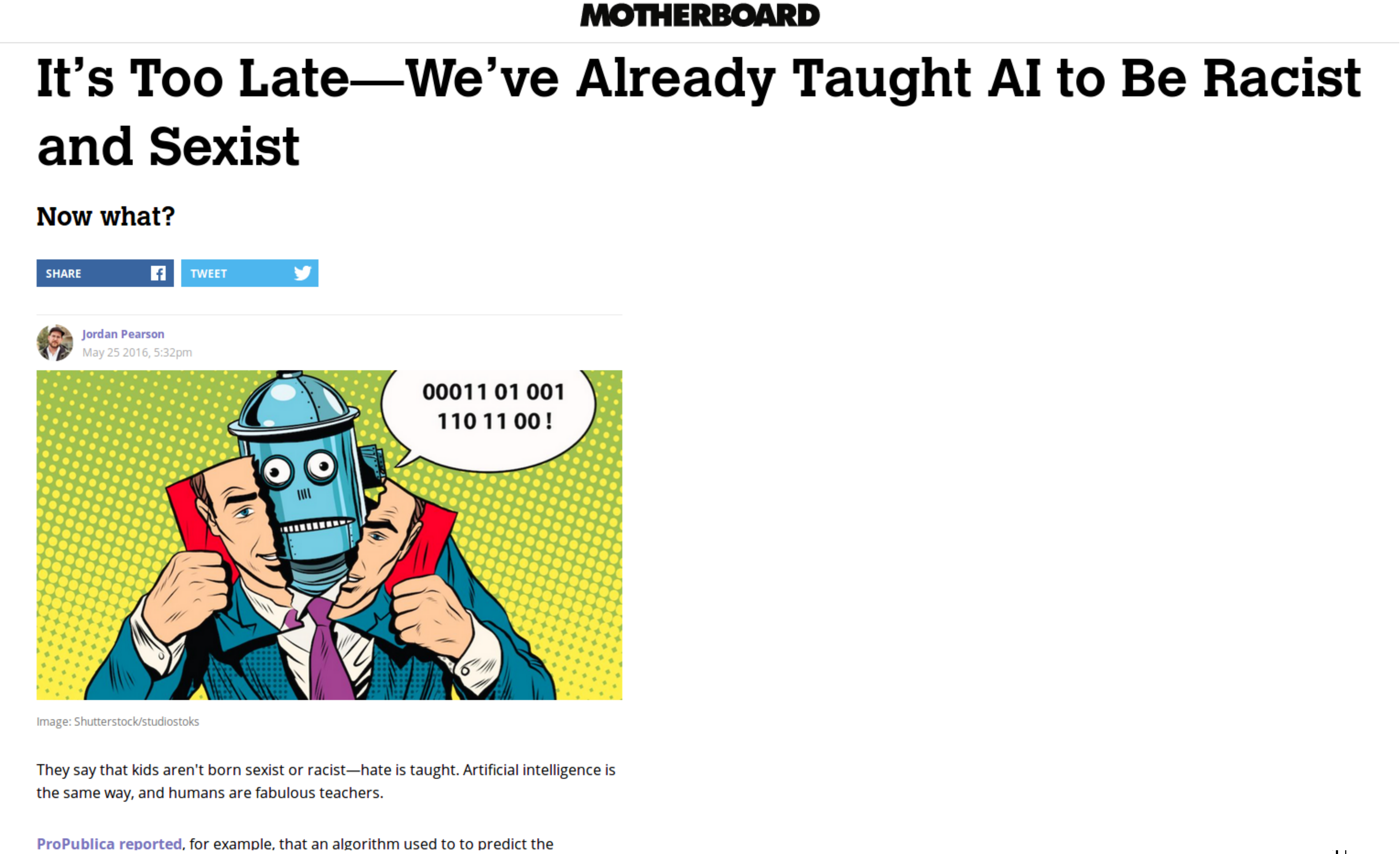
Adversarial Examples

- How to create adversarial examples?
  - Variety of methods. *[Yuan et al 2017],[Tramer et al 2018]*

- How do they attack to the networks?
  - Most adversarial examples construct techniques using the gradient to attack knowing the model. (white-box attacks)
  - When the model is not known. (black-box attacks)

- How to fight against them?
  - Open research topic, no generic way to fight against them.
  - Adversarial Training.
  - Defensive distillation.

Real-life Challenges



MOTHERBOARD

It’s Too Late—We’ve Already Taught AI to Be Racist and Sexist

Now what?

SHARE TWEET

Jordan Pearson
May 25 2016, 5:32pm

Image: Shutterstock/studiostoks

They say that kids aren't born sexist or racist—hate is taught. Artificial intelligence is the same way, and humans are fabulous teachers.

ProPublica reported, for example, that an algorithm used to to predict the

Real-life Challenges

MOTHERBOARD

It’s Too Late—We’ve Already Taught AI to Be Racist and Sexist

Now what?

SHARE TWEET

Jordan Pearson
May 25 2016, 5:32pm

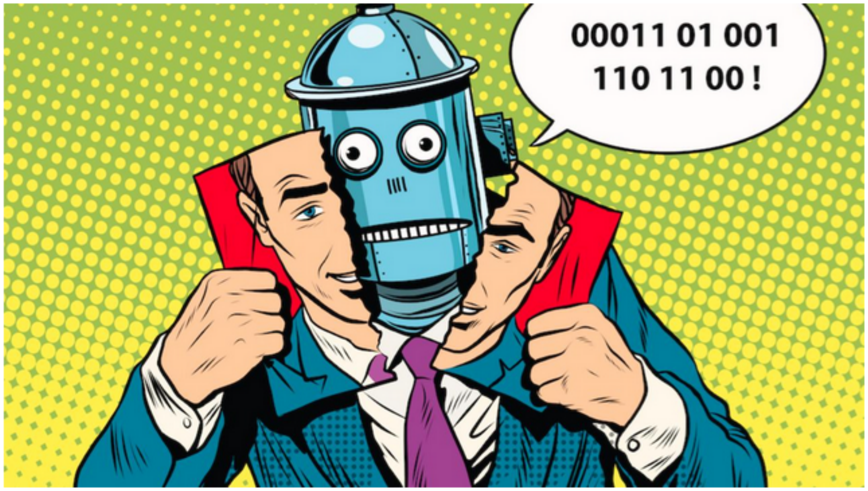


Image: Shutterstock/studiostoks

They say that kids aren't born sexist or racist—hate is taught. Artificial intelligence is the same way, and humans are fabulous teachers.

ProPublica reported, for example, that an algorithm used to to predict the

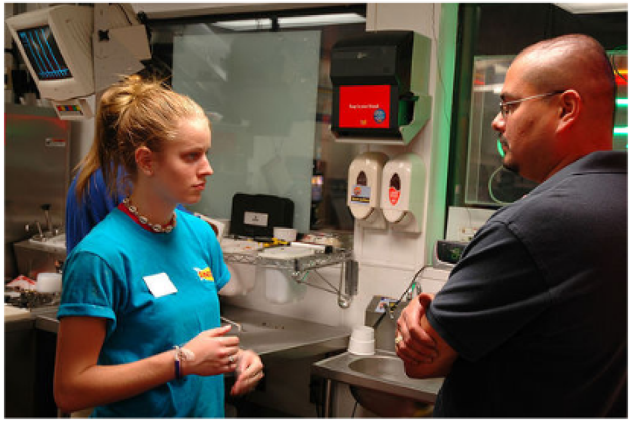
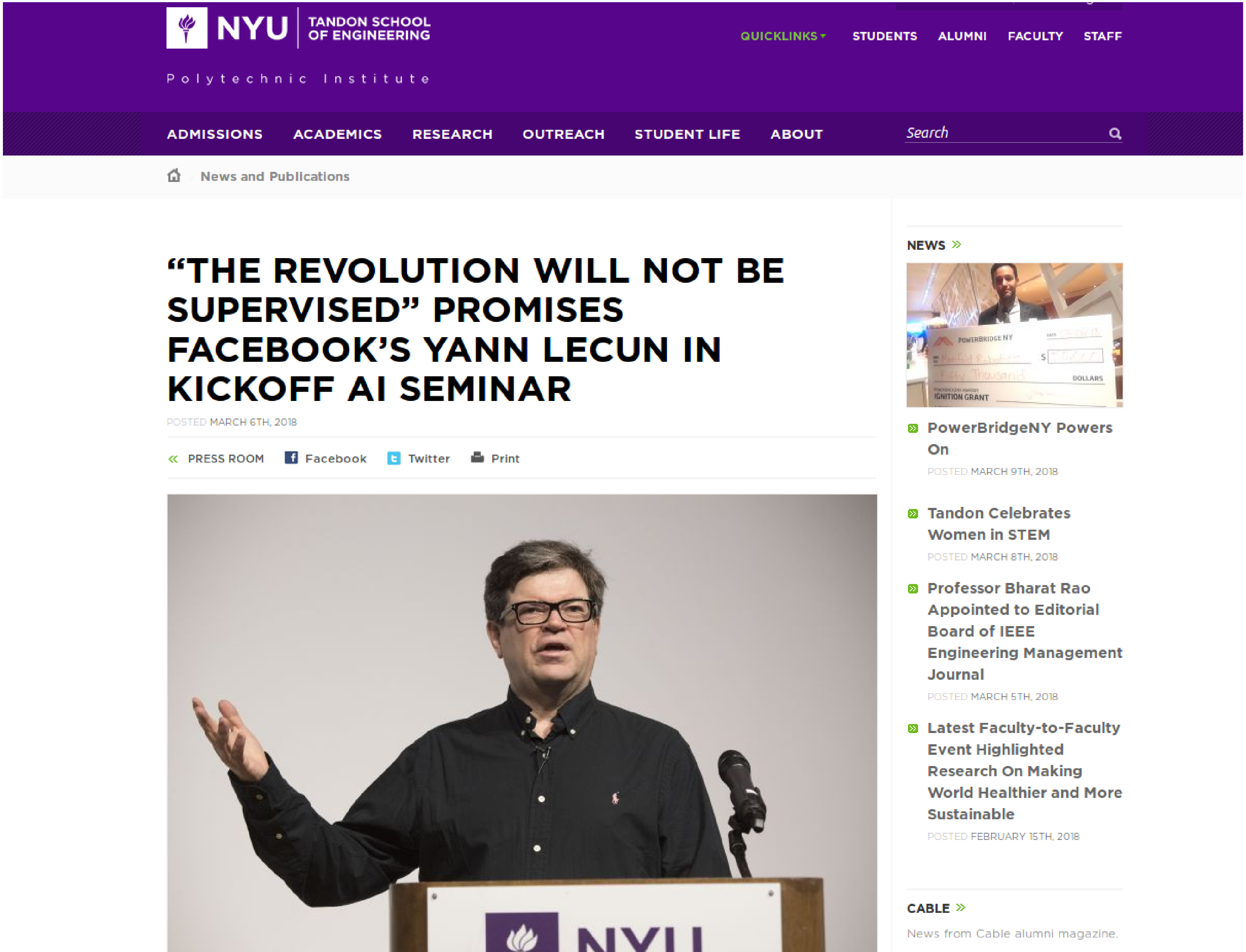


Image 8063007 from Flickr30K dataset

Proposed captions from network trained on Flickr30k *[Miltenburg et al 2016]*

- A blond girl and a bald man with his arms crossed are standing inside looking at each other.
- A **worker** is being *scolded* by her **boss** in stern lecture.
- A **manager** talks to an **employee** about job performance.
- A hot, blond girl getting criticized by **her boss**.
- Sonic employees talking about work.

Real-life Challenges



"THE REVOLUTION WILL NOT BE SUPERVISED" PROMISES FACEBOOK'S YANN LECUN IN KICKOFF AI SEMINAR

POSTED MARCH 6TH, 2018

- PowerBridgeNY Powers On

  POSTED MARCH 9TH, 2018

- Tandon Celebrates Women in STEM

  POSTED MARCH 8TH, 2018

- Professor Bharat Rao Appointed to Editorial Board of IEEE Engineering Management Journal

  POSTED MARCH 5TH, 2018

- Latest Faculty-to-Faculty Event Highlighted Research On Making World Healthier and More Sustainable

  POSTED FEBRUARY 15TH, 2018

CABLE »

News from Cable alumni magazine.

Open Issues in Deep Learning

THANK YOU

CentraleSupélec

CENTRE FOR
VISUAL COMPUTING
CENTRE DE VISION NUMÉRIQUE
CENTRALESUPÉLEC | INRIA

Friday 16 March, 2018