Science Teaching & Machine Learning

Dr. Scott Hawley, Belmont Univ.

10am Wed March 15, JAAC 4110

Science Teaching & Machine Learning

Dr. Scott Hawley, Belmont Univ.

National Science Teachers Association, Mar 15 2017

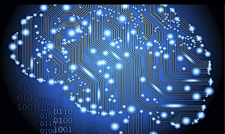
Foreword

- Machine Learning aka “Deep Learning” is transforming many segments of society
- Relies on long-established principles of *scientific data analysis* and *teaching*
  - You’re actually already using these!
- Let’s explore the interplay between these fields…

Terminology

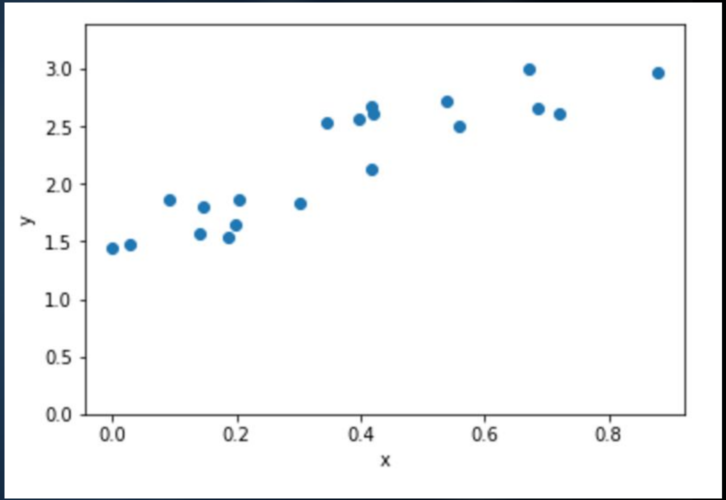
- **Artificial Intelligence (A.I.)**
- **"3 waves of AI" (John Launchbury, DARPA)**

  1. **Handcrafted Knowledge:** Human expertise "programmed in". e.g., TurboTax. Don't learn.
  2. ★ **Statistical Learning:** stat. inference driven by optimization. **"Machine Learning" for this talk.** No context or 'reasoning'.
  3. **Contextual Adaptation:** Contextual models, Perception, & Reason. (Not there yet.)



Paradigm: Data Modeling

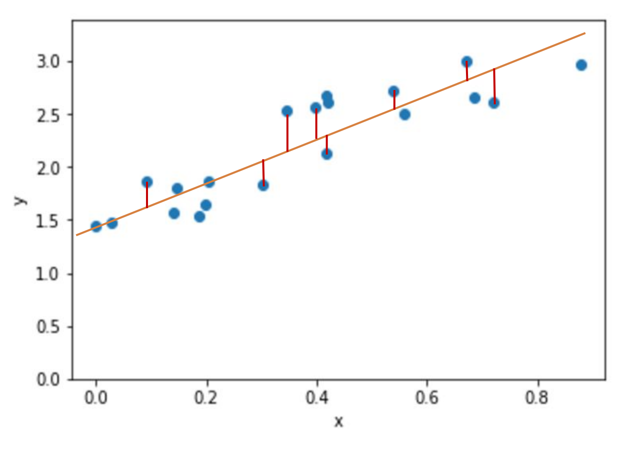
- You're given a bunch of data points (x_i, y_i)
- You want to "fit" a line:
  - Find values of parameters *m* and *b* s.t. $y=mx+b$ passes "thru" the data



- This is something we frequently do "in lab", but how does it work?

Linear Least Squares

- **Choose an "error" function: Squared Error**
  - **For each (x_i,y_i), square the difference between "prediction" $\tilde{y}_i = mx_i + b$ and "true value" y_i, and add squared-differences up**

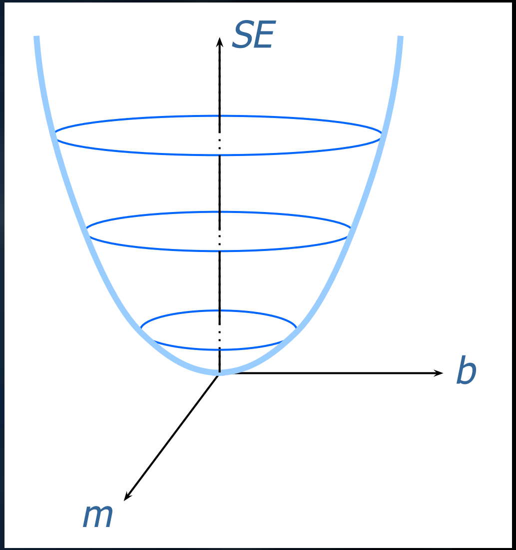


$$SE = \sum_i (\tilde{y}_i - y_i)^2 = \sum_i (mx_i + b - y_i)^2$$

- **Strategy: Try to *minimize* error**

Minimization: Gradient Descent

- *SE* defines a *paraboloid* in *(m,b)*
- Choose initial guess for *(m,b)*
- Find out what the ‘slope’ is there
- Choose new *(m,b)* by going “in the downhill direction”:



$$m_{new} = m_{old} - \alpha \frac{\partial(SE)}{\partial m}$$

$$b_{new} = b_{old} - \alpha \frac{\partial(SE)}{\partial b}$$

…where α is called the “learning rate,” something you choose.

Gradient Descent p.2

- **Just to be explicit…**

$$SE = \sum_i (\tilde{y}_i - y_i)^2 = \sum_i (mx_i + b - y_i)^2$$

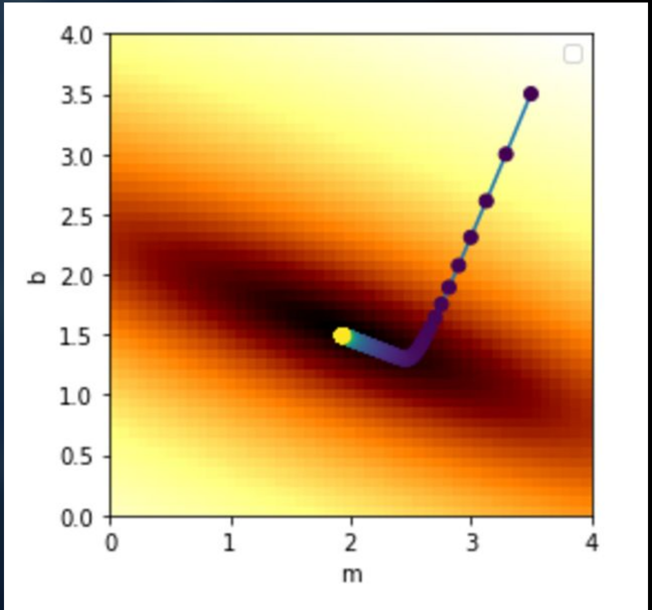
$$m_{new} = m_{old} - \alpha \frac{\partial(SE)}{\partial \mathrm{m}}$$

$$= m_{old} - 2\alpha \textstyle\sum_i (mx_i + b - y_i)\, x_i$$

$$b_{new} = b_{old} - \alpha \frac{\partial(SE)}{\partial b}$$

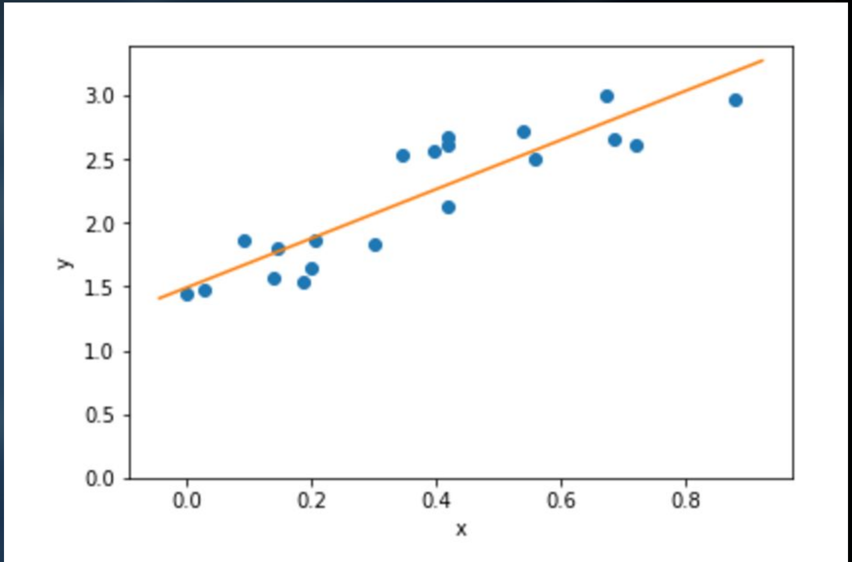
$$= b_{old} - 2\alpha \textstyle\sum_i (mx_i + b - y_i)\,(1)$$

Algorithm's progress:



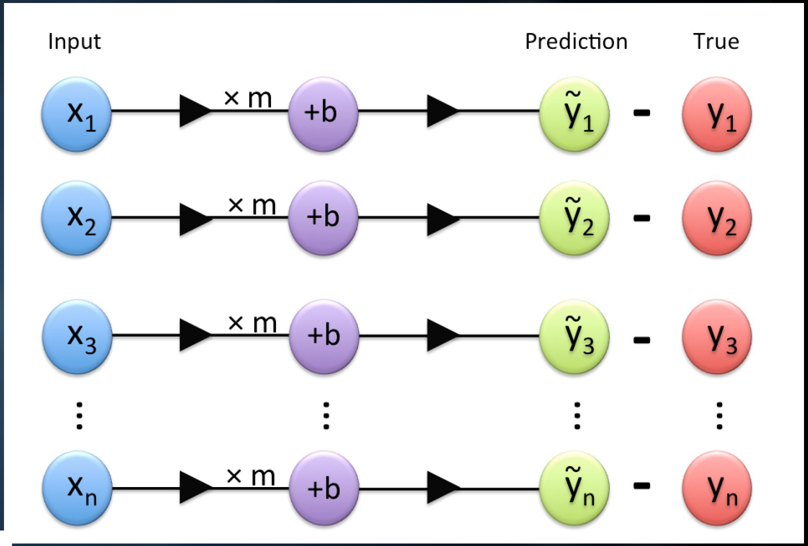
Ta-Da!!

- We've "learned" to approximate a large dataset via a *model* with parameters that can be adjusted
- This works for nonlinear fitting as well!



Same Thing, Different Views

Simple Neural Network:



Matrix Multiplication:

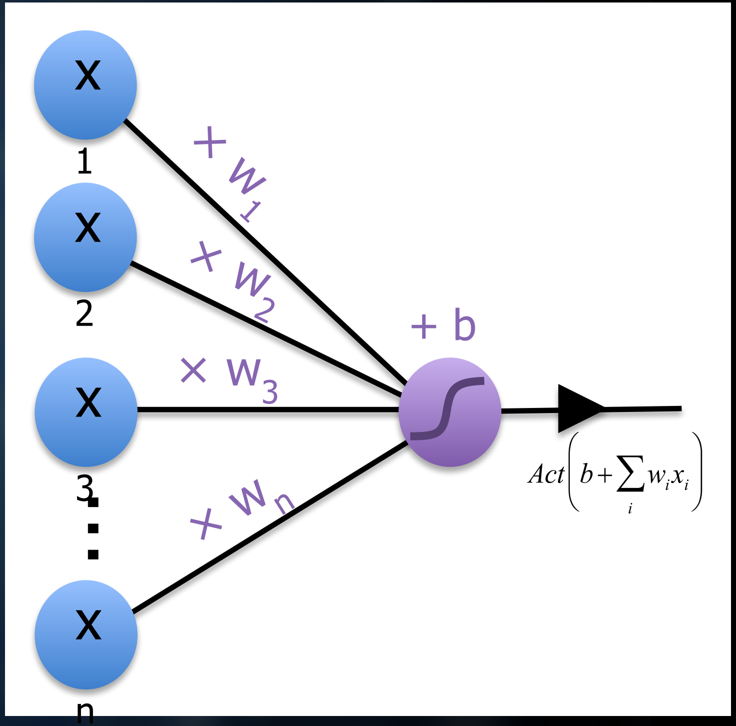
$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \\ \dots \\ \tilde{y}_n \end{pmatrix} = \begin{pmatrix} m & 0 & 0 & \dots & 0 \\ 0 & m & 0 & \dots & 0 \\ 0 & 0 & m & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{pmatrix} + \begin{pmatrix} b & 0 & 0 & \dots & 0 \\ 0 & b & 0 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & b \end{pmatrix}$$

GPUs are good at this!

(Artificial) Neural Networks?

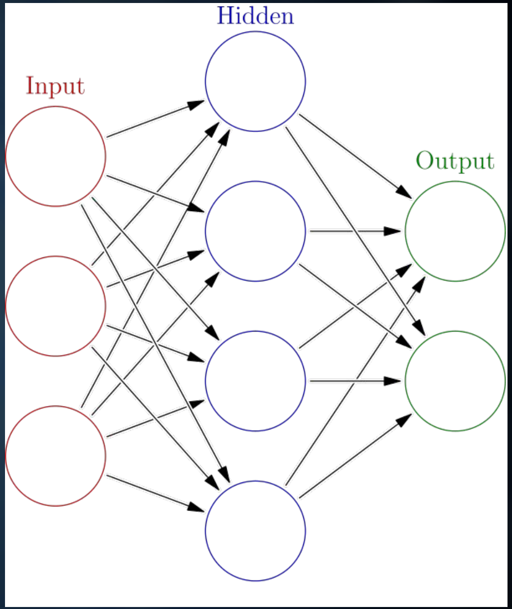
- (Loosely) Based on biological neurons

1. Sum weighted inputs (plus bias)
2. Apply (nonlinear) "activation" to sum to get output
   - Example activations: tanh, linear, rectified-linear



Early NN Work

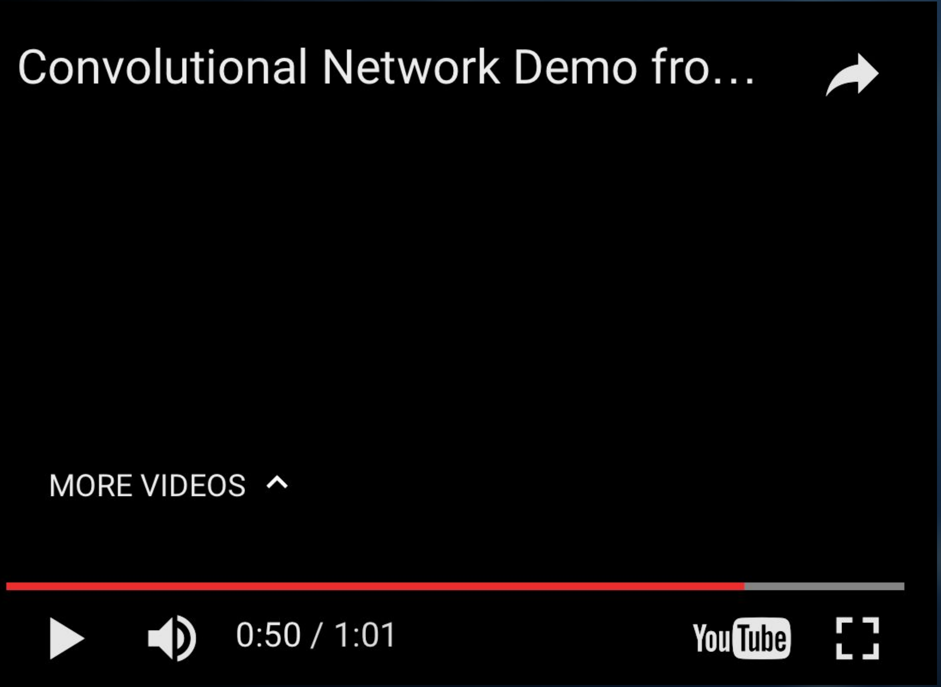
- **Multi-Layer Perceptron:**



- **If you can figure out how to handle it, can add multiple hidden layers**

‘Early’ Example

- **Image Recognition – LeCun 1993**

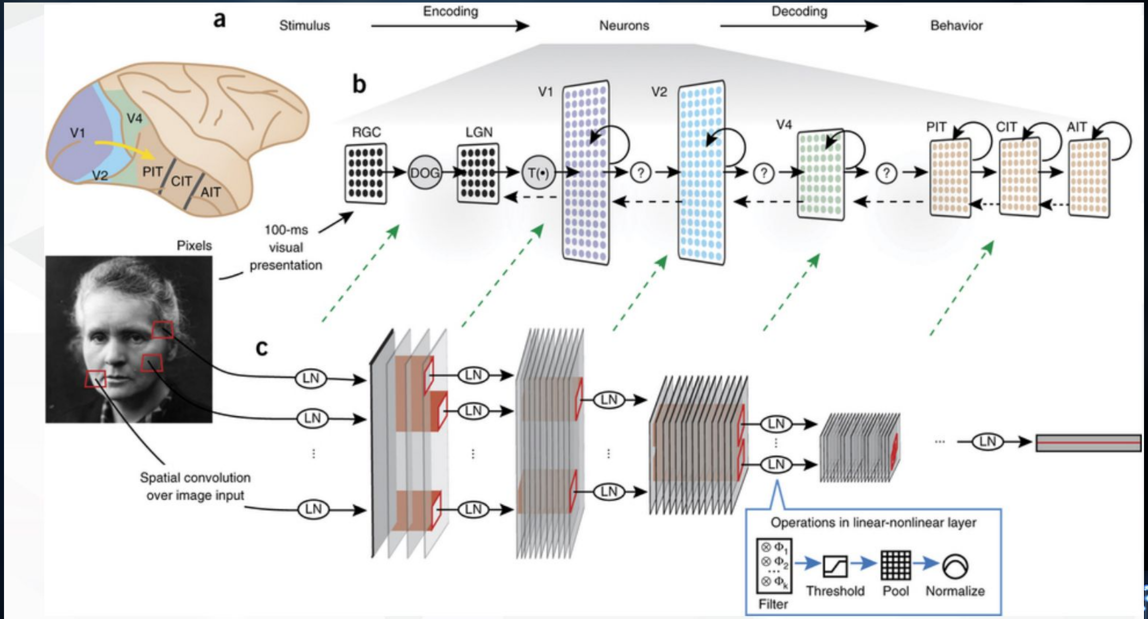


Lecun is now head of Facebook’s AI.

His advisor was Geoffrey Hinton.
Notable student of LeCun: Zaremba

Brain Science (Tie-In)

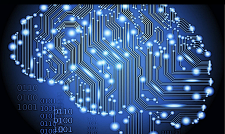
- **Hubel and Wiesel (1950s): Hooked electrodes to cat's visual cortex**
  - **Modern ConvNet:**



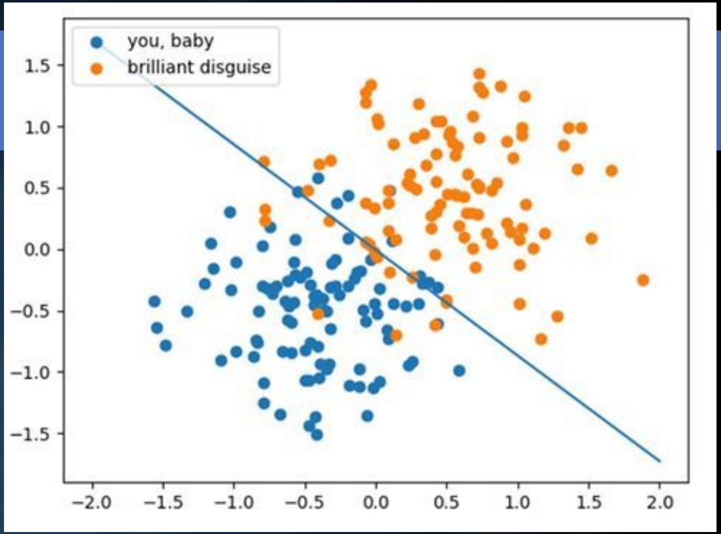
Yamins DLK, DiCarlo JJ. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 2016;19(3):356–365. Available from: http://www.nature.com/doifinder/10.1038/nn.4244

Deep Learning (/ NN's are now hot)

- **Term coined by Hinton, b/c NN's had fallen into disrepute**
- **Can now train networks that are many layers *deep***
- **Allows for hierarchy of abstraction**
- **What allowed NNs to go from "pariahs" to "stars"?**
  - **Tons of data to train with (thanks to Internet)**
  - **Improved theory**
    - Backpropagation, Batch Normalization
    - Avoiding vanishing/exploding gradients
    - More efficient architechtures
  - **Faster processors (to get through the data)**



Paradigms for ML



- **Tasks:**
  - **Regression (curve fitting)**
  - **Classification (example □)**
- **Contexts:**
  - **"Supervised Learning" where the 'right answer' is known (e.g. classification)**
  - **"Unsupervised Learning" – machine makes up its own answers (e.g. clustering)**

How These Things Learn

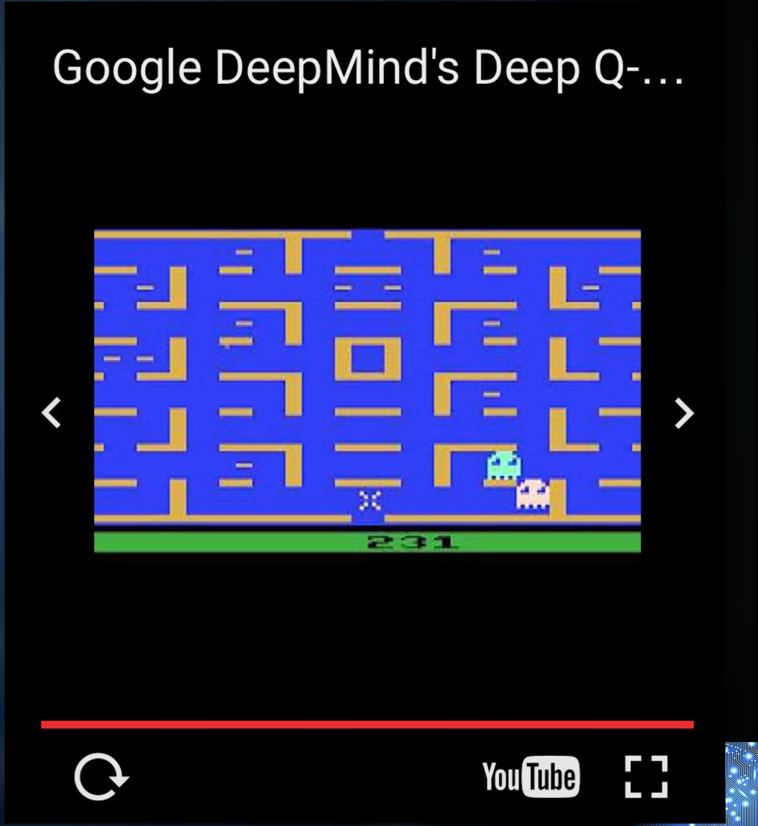
- **Training:** ***Repetition*** **and** ***Reinforcement***
- **Minimizing Loss, or Maximizing Reward**
- **Training vs. Learning?**

Aside: Not all ML is Neural Nets

- There's also:
  - Genetic Algorithms, Random Forests, Support Vector Machines, Gaussian Mixture Models, Hidden Markov Models, Bayesian Analysis, Independent Component Analysis, Non-negative Matrix Factorization, etc.
- However, NNs have been *highly* successful recently
  - More succesful than some of the others
- Can also be shown to be isomorphic to others in many cases of interest

Example: Games

- Maximize score, only given pixels
- Other examples:
  - Chess (Lai, Giraffe)
  - Go (DeepMind, AlphaGo)
  - Doom (CMU, link)
  - Flappy Bird (NEAT, link)
  - StarCraft (in progress)



Examples

- **Speech Recognition**
- **Language Translation**
- **Captioning Images** e.g., Google's: "Show & Tell":

- **Generating 'Fake' Images** **(search on GAN networks!)**
- **(Artistic) Style Transfer**
- **Self-driving cars/drones**
- **Me: Audio Signal Processing**
  - **Inspired by Zaremba & Sutskever...**

Learning to Execute

Zaremba & Sutskever, 2014
(arXiv:1410.4615)

- **Used "curriculum training"...**
  - **Curriculum Training** (Bengio et al., ICML '09) "Humans and animals learn much better when the examples are not randomly presented but organized in a meaningful order which illustrates gradually more concepts, and gradually more complex ones."
- **...to teach a network to read text as Python code and exectue it:**

Input:
```
j=8584
for x in range(8):
  j+=920
b=(1500+j)
print((b+7567))
```
Target: 25011.

- **Findings: ("improved curriculum")**
  - **Start easy & gradually get harder**
  - **But mix in a few hard ones early on too**
  - **Give it "equivalent" examples (doubling input & reversed input)**
  - **(Counterexamples?)**

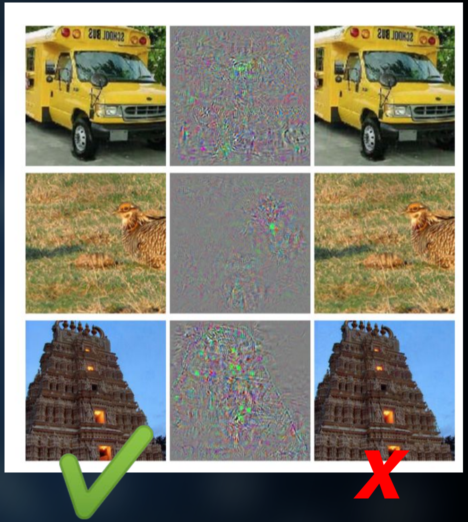
Human vs. Machine Learning

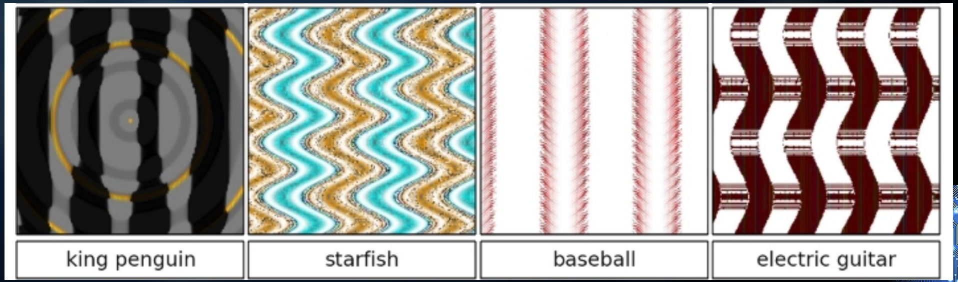
- **Human learners**
  - Have some innate structures / knowledge
  - Can learn based on minimal data ("never put your hand on the stove again!")
  - Can benefit from human model-coaching
  - Generalize poorly

- **Machine learners** (currently)
  - *Only* learn from the data
  - Require lots of data
  - Poor at Abstract Reasoning
    - Note: Facebook babI project
  - Generalize *very* poorly

Can Generalize Very Poorly

- **Can learn a variety of *specific* things, but results may not generalize in an intuitive way.**
  - Szegedy, et al; "we find that deep neural networks learn input-output mappings that are fairly discontinuous to a significant extend [sic]" https://arxiv.org/abs/1312.6199
  - Jeff Clune's Lab: "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images":





- **Still active research area**

¯_(ツ)_/¯

Machine Learning in Physics

- **ML Laser-tuning for making Bose-Einstein Condensate**

  Wigley et al., "Fast machine-learning online optimization of ultra-cold-atom experiments," *Nature Scientific Reports* 6, Article number: 25890 (2016)

  – "It did things a person wouldn't guess, such as changing one laser's power up and down, and compensating with another…It may be able to come up with complicated ways humans haven't thought of to get experiments colder and make measurements more precise."— Phys.Org

- **Solving quantum wavefunctions**

  Carleo & Troye, "Solving the quantum many-body problem with artificial neural networks," *Science* Vol. 355, Issue 6325, pp. 602-606 (10 Feb 2017)

  – "Carleo and Troyer harnessed the power of machine learning to develop a variational approach to the quantum many-body problem (see the Perspective by Hush). The method performed at least as well as state-of-the-art approaches, setting a benchmark for a prototypical two-dimensional problem."

- **Parameter estimation in particle accelerators**

  Baldi et al., "Parameterized neural networks for high-energy physics," *Eur. Phys. J.* C 76: 235 (2016)

Frontiers: A.I. in Education

Meet Jill Watson: Georgia Tech's first AI teaching assistant

Professor Ashok Goel recently gave a talk at TEDxSanFrancisco about the use of artificial intelligence to answer students' questions in the forums for his online Knowledge-Based Artificial Intelligence (KBAI) class. Dubbed Jill Watson, the AI teaching assistant was based on IBM's Watson platform, which is perhaps best known as the computer that beat two Jeopardy champions. Jill was developed specifically to handle the high number of forum posts by students enrolled in an online course that is a requirement for Georgia Tech's online master of science in computer science program.

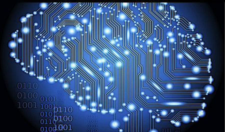
- More: Business Insider: "IBM's brilliant AI just helped teach a grad-level college course" (link)

- **Adaptive training systems for online / self-directed (AI-directed) study**
- **"8 Ways…" (next slide)**

by Tom Vander Ark, gettingsmart.com 11/26/2015

"8 Ways Machine Learning Will Improve Education"

1. **Content analytics that organize and optimize content modules:** Gooru, IBM Watson Content Analytics
2. **Learning analytics that track student knowledge and recommend next steps:**
   - Adaptive learning systems: DreamBox, ALEKS, Reasoning Mind, Knewton; Game-based learning: ST Math, Mangahigh
3. **Dynamic scheduling matches students that need help with teachers that have time:**
   - NewClassrooms uses learning analytics to schedule personalized math learning experiences.
4. **Grading systems that assess and score student responses to assessments and computer assignments at large scale, either automatically or via peer grading:**
   - Pearson's WriteToLearn and Turnitin's Lightside can score essays and detect plagiarism.
5. **Process intelligence tools analyze large amounts of structured and unstructured data, visualize workflows and identifying new opportunities:**
   - BrightBytes Clarity reviews research and best practices, creates evidence-based frameworks, & provides strength gap analysis.
   - Enterprise Resource Planning (ERP) systems like Jenzabar and IBM SPSS helps HigherEd institutions predict enrollment, improve financial aid, boost retention, and enhancing campus security.
6. **Matching teachers and schools:** MyEdMatch and TeacherMatch are eHarmony for schools.
7. **Predictive analytics and data mining to learn from expertise to:**
   - Map patterns of expert teachers
   - Improve learning, retention, and application.
8. **Lots of back office stuff:** EDULOG does school bus scheduling; Evolution, DietMaster.



Education: Course Offerings?

- Statistics is likely to become "hot"
  - Belmont CSM just hired a new statistician!
- "Data Science"
  - Computer Science
  - Mathematics
  - Hard Sciences (Physics!)
- Use of ML in classes (as projects)
  - As apps & architectures become "commodity," expect more/easier tools

Local: Belmont/Nashville

- **Digital Reasoning (Company, employs Belmont grads & former faculty)**
- **ASPIRE Research Co-op aspirecoop.github.io**

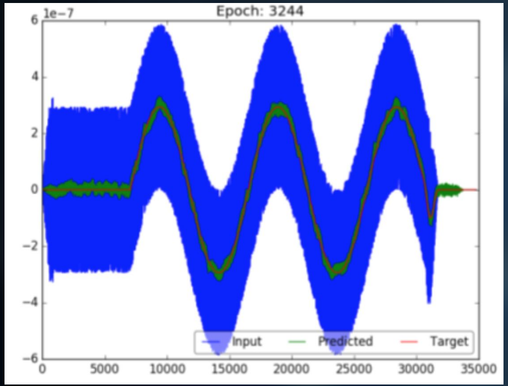
Next meeting
Mon Mar 20!
JAAC 1034

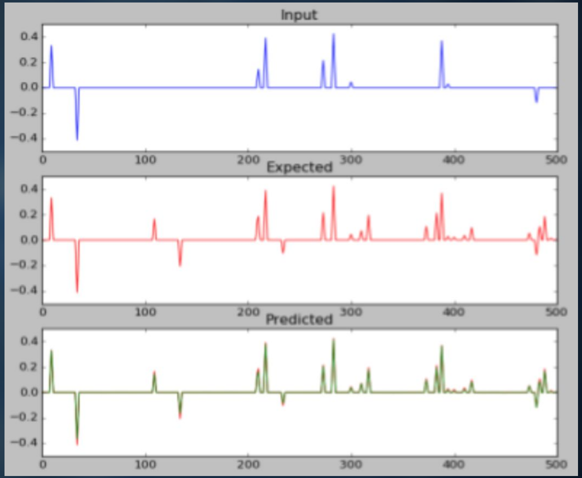
Further Study

- My ML blog http://drscotthawley.github.io
  - Includes long post of links to papers, tutorials & videos
- Textbooks:
  - neuralnetworksanddeeplearning.com (free, easy)
  - Artificial Intelligence: A Modern Approach (comprehensive)
  - Elements of Statistical Learning, Ch 1-4 & 7 (hardcore)
- Forbes: "The Future of Artificial Intelligence in Education"
- "Fun":
  - OpenAI Gym gym.openai.com
  - Kaggle competitions
- Find a tutorial and TRY IT! You might get hooked!

Extra: My Work: ML for Audio

- **Audio effect profiling (regression)**





- **Audio classification**
  - **99.7% on IDMT guitar effects database**
  - **80% on PhysioNet heart sounds diagnosis**
  - **Not great at binaural source localization:**