

Machine Learning & Deep Nets

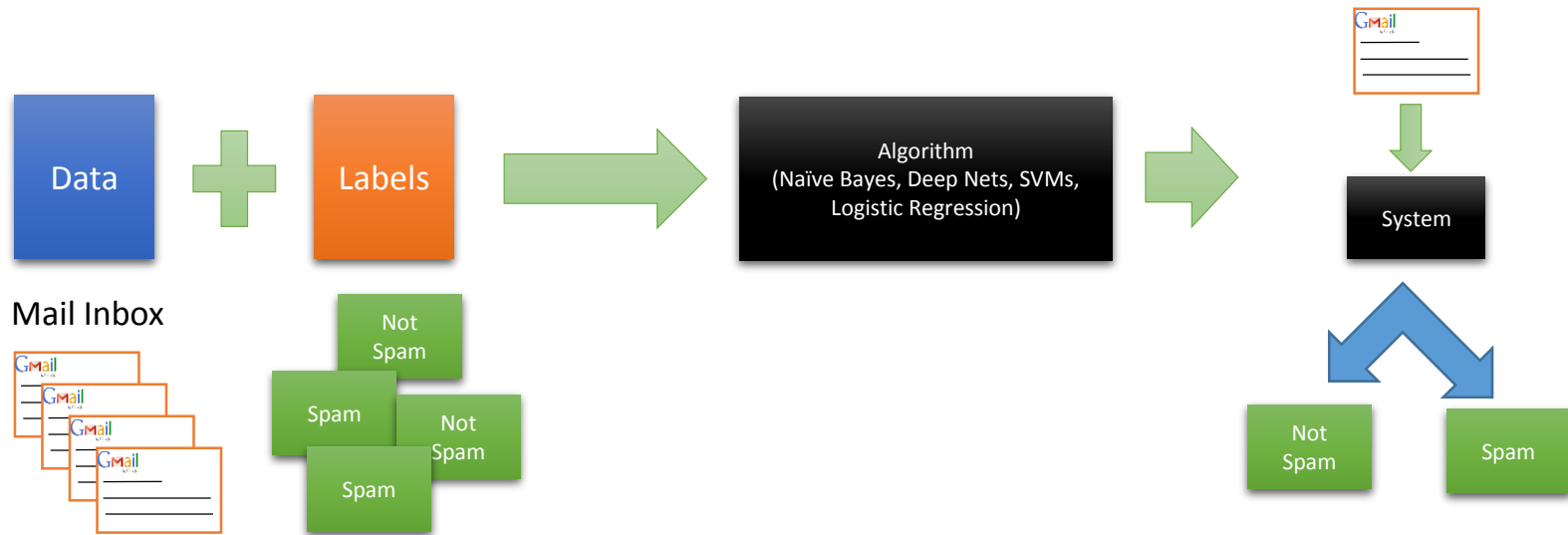
Leon F. Palafox

December 4th, 2014

Introduction

- What is Machine Learning?
 - Is a rebranding of Artificial Intelligence, since we don't really care about replicating intelligence.
 - Is a set of tools to analyze data to make predictions and get insights out of it.
 - Is a sub-branch of computer science and statistics.
- Areas of Machine Learning
 - Supervised Learning: Classification, Regression.
 - Unsupervised Learning (Knowledge Discovery): Clustering, Mixture Models.

Supervised Learning



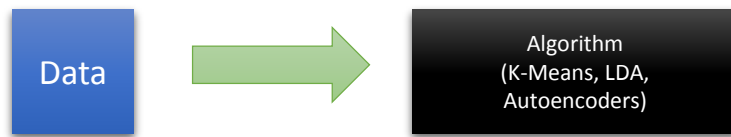
The set of elements that describe a single datum are called features, in this case, the features are the words in the e-mails.

Each category (spam, not spam) will have features that will characterize them.

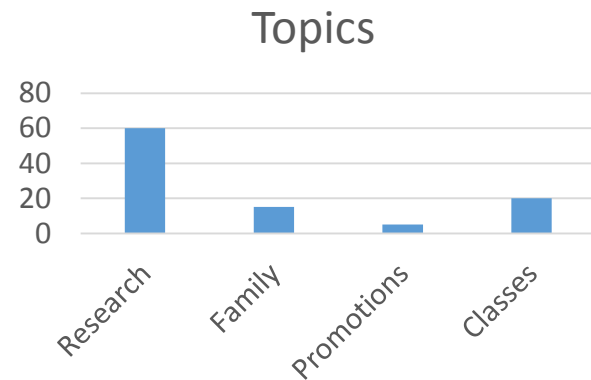
Spam: Offer, Viagra, medicine, Free, Conference in China

Not Spam: Hamilton, LPL, DTM, Mom, Dad

Unsupervised Learning



Mail Inbox



The set of elements that describe a single datum are called features, in this case, the features are the words in the e-mails.

Each topic (clusters) will have features that will characterize them.

Research: Mars, Proposal, DTM, HiRISE, Machine Learning, Deep Nets, Bayesian

Family: Mom, House, Mexico

Promotions: Computer, PS4, Cheap, Amazon, Deal

Classes: Grades, Homework, Questions, Office Time

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough. Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson, a geneticist at Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

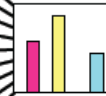
* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

ADAPTED FROM NCBI

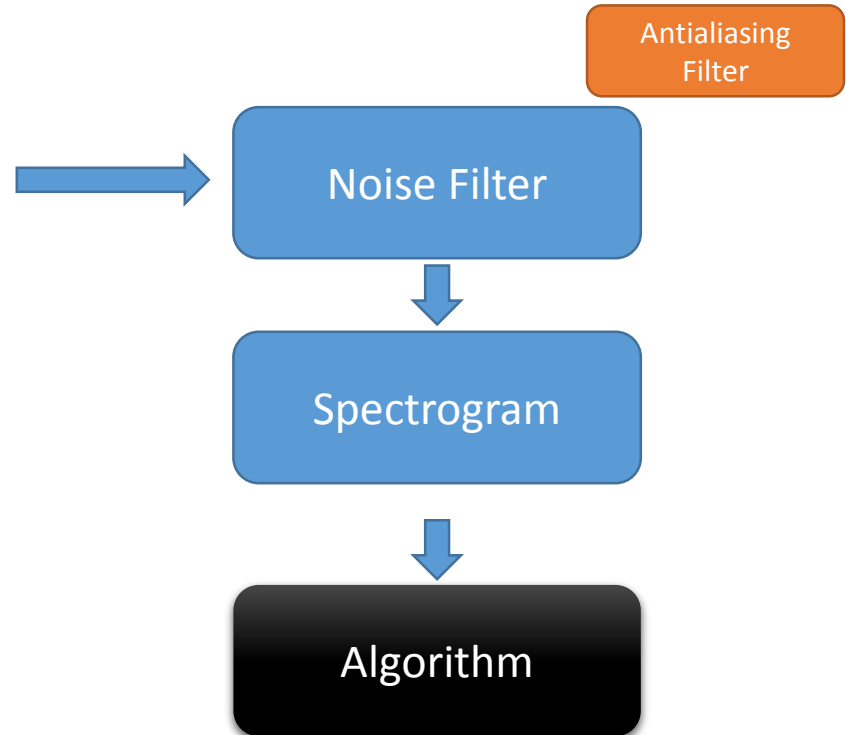
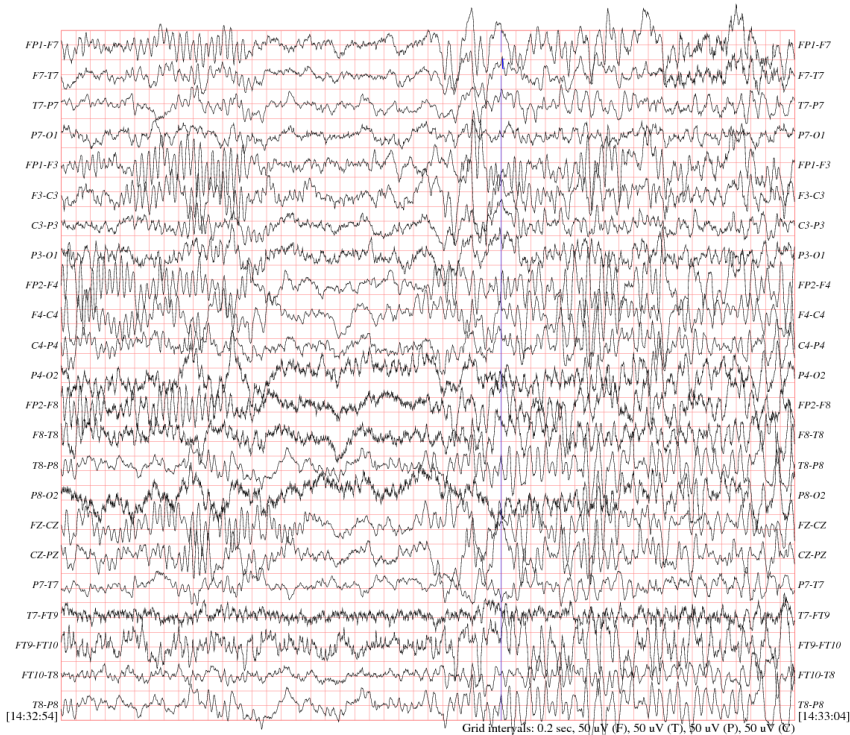
Topic proportions and assignments



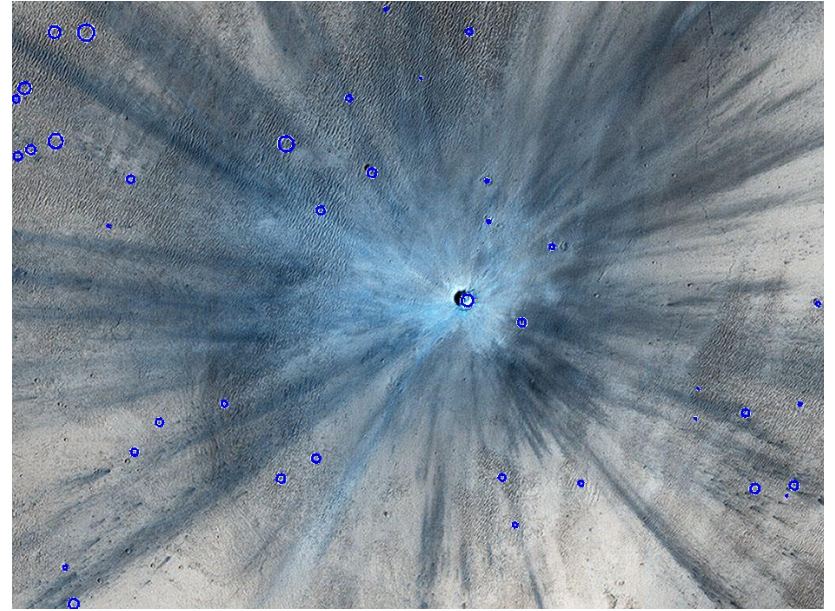
<http://cs.stanford.edu/people/karpathy/nips2014/>
<http://sarah-palin.herokuapp.com/>

Preprocessing Data

- It's a pain, but is needed



Preprocessing Data



Who uses Machine Learning

- Google: Spam Detection (Gmail), Ranking Algorithms (Google Search), Image Recognition (Google Image)
- Amazon: Recommendation Engines
- Facebook: Feed personalization, News personalization.
- Disney, NTT, Toyota, Ford, etc.

So what are Deep Nets?

- First we need to understand what are Neural Networks (NN).
 - NNs have gone through a heavy rebranding thorough the years.
 - In 1943, McCulloch and Pitts created the first model of an artificial neuron.
 - By 1958, Rosenblatt had come up with the Perceptron, the cornerstone of modern NN.
 - In 1986, Rumelhart started the connectionism euphoria.

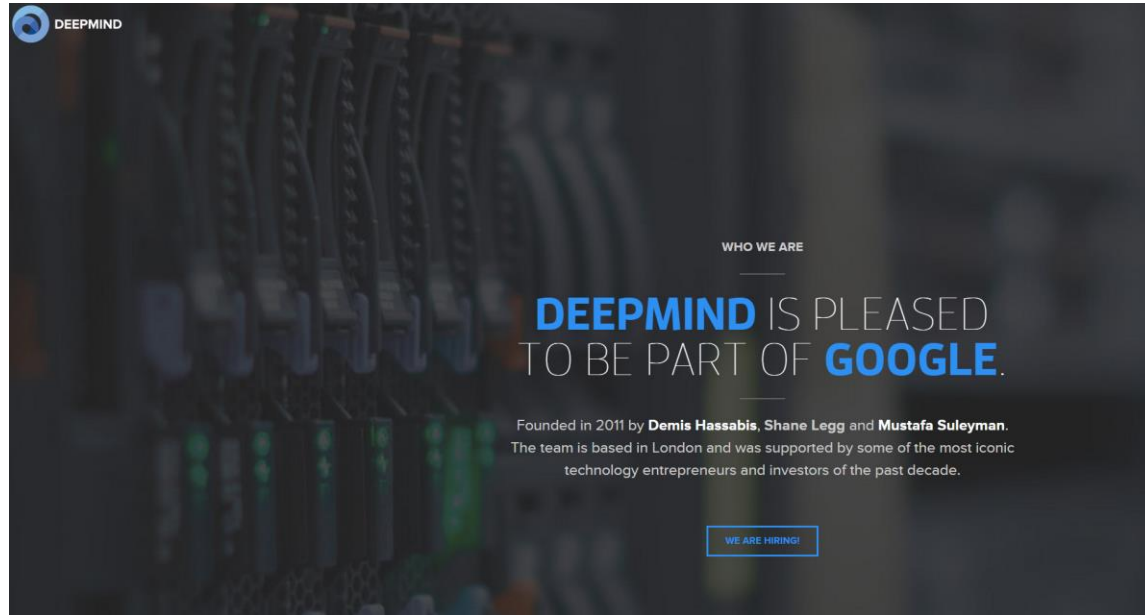
Background

- Processing power was still an issue and until 2006, common NNs were researched by only small clusters of people.
- Training was expensive, and the results only marginally better (or worse) than SVMs or Logistic Regression.
- In 2006, Hinton and Bengio made huge discoveries on how to train NNs and they rebranded them as Deep Nets.
- During this time, Convolutional Neural Networks (CNN) had been a great tool for image pattern recognition.

Motivation

- Deep Nets and CNNs, are by today standards the best algorithm for Image Pattern Recognition.
- The three Big Kahunas of NNs and Deep Nets, Geoffrey Hinton, Yann LeCun and Yoshua Bengio are working actively with Google, Facebook and University of Toronto, respectively.

Motivation



- In January Google bought DeepMind, a startup with no WebPage, no Product, a single NIPS (AI conference) Demo.
- They bought it for \$500 million.
- Facebook was deeply interested as well.

NEW YORKER
 READ SOMETHING THAT MEANS SOMETHING.
 \$1 A WEEK
 SUBSCRIBE RENEW GIVE A GIFT
 NON-U.S. ORDERS

Sign in | Link your subscription | TNY Store

THE NEW YORKER



NEWS CULTURE BOOKS & FICTION SCIENCE & TECH BUSINESS HUMOR MAGAZINE ARCHIVE SUBSCRIBE

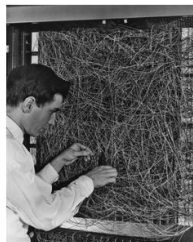
NOVEMBER 25, 2012

IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE?

BY GARY MARCUS



Can a new technique known as deep learning revolutionize artificial intelligence, as yesterday's front-page article at the *New York Times* suggests? There is good reason to be excited about deep learning, a sophisticated "machine learning" algorithm that far exceeds many of its predecessors in its abilities to recognize syllables and images. But there's also good reason to be skeptical. While the *Times* reports that "advances in an artificial intelligence technology that can recognize patterns offer the possibility of machines that perform human activities like seeing, listening and thinking," deep learning takes us, at best, only a small step toward the creation of truly intelligent machines. Deep learning is important work, with immediate practical applications. But it's not as breathtaking as the front-page story in the *New York Times* seems to suggest.



SIGN UP FOR NEWSLETTERS

E-mail address SIGN UP

ADVERTISEMENT

TAKING OFF 5,000 TIMES A DAY. TAKING ON EVERYTHING ELSE.
 America's most-awarded airline employees.

THE WALL STREET JOURNAL. JAPAN

SUBSCRIBE NOW >>

日本リアルタイム JAPAN REALTIME

FUKUSHIMA WATCH AUTOS ECONOMY & BUSINESS TECHNOLOGY POLITICS & POLICY LIFESTYLE & CULTURE

HOT TOPICS: BANK OF JAPAN SHINZO ABE DISPUTED TERRITORY NUCLEAR FUKUSHIMA DAICHI AGING POPULATION CHANGE LANGUAGE: ENGLISH

8:47 pm JST Oct 1, 2014 JAPAN

NIT, Toyota Seek 'Deep Learning' Expertise

ARTICLE COMMENTS

DEEP LEARNING NTT PREFERRED NETWORKS TOYOTA MOTOR CORP



By TAKASHI MOCHIZUKI (CONNECT)



More Enterprise SaaS Applications Than Any Other Cloud Services Provider

CLICK TO PLAY VIDEO

ORACLE

About Japan Real Time

Japan Real Time is a newsy, concise guide to what works, what doesn't and why in the one-time poster child for Asian development. It's a guide to how you can get the most out of...

The New York Times

Science

Search All NYTimes.com Go Capital One

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION ARTS STYLE TRAVEL JOBS REAL ESTATE AUTOS

ENVIRONMENT SPACE & COSMOS

Watch now

Watch why boiling an egg isn't as simple as boiling an egg.



Scientists See Promise in Deep-Learning Programs



A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

By JOHN MARKOFF
 Published: November 23, 2012

Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.

FACEBOOK

TWITTER

GOOGLE+

SAVE

EMAIL

SHARE

PRINT

REPRINTS

The advances have led to widespread enthusiasm among researchers who design software to perform human activities like seeing, listening and thinking. They offer the promise of machines that converse with humans and perform tasks like driving cars and working in factories, raising the specter of automated robots that could replace human workers.

Connect With Us on Social Media @nytimescience on Twitter.

Science Reporters and Editors on Twitter

Like the science desk on Facebook.



Keith Penner
 A student team led by the computer scientist Geoffrey E. Hinton used

The technology, called deep learning, has already been put to use in services like Apple's Siri virtual personal assistant, which is based on Nuance Communications' speech recognition service, and in Google's Street View, which uses machine vision to identify specific addresses.

But what is new in recent months is the growing speed and accuracy of deep-learning programs, often called artificial neural networks or just "neural nets" for their resemblance to the neural connections in the brain.

"There has been a number of stunning new results with deep-learning methods," said Yann LeCun, a computer scientist at New York University who did pioneering

This cloud makes data make a difference.

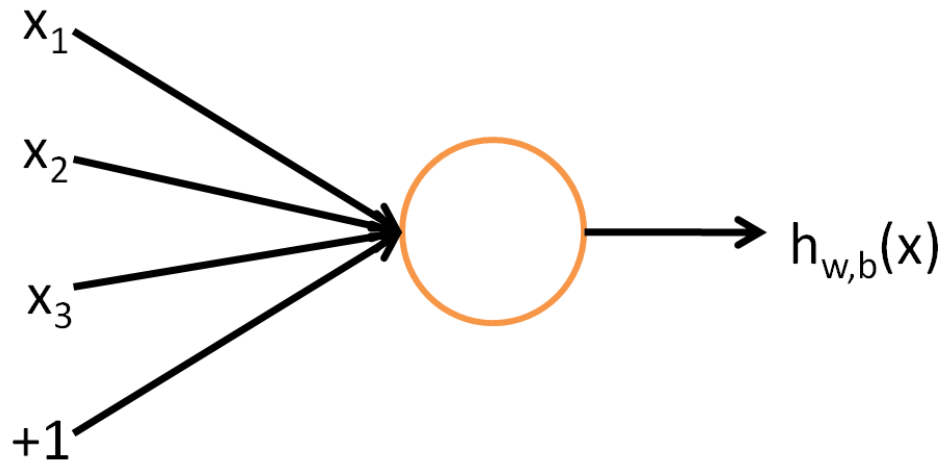


MOST EMAILED RECOMMENDED FOR YOU

- DEALBOOK: Saks Flagship Store Is Appraised for Mortgage at \$9.7 Billion
- DEALBOOK: U.S.-Backed Mortgages Put to Test in an Innovative Lawsuit
- TODAY'S EDITORIALS: Homeownership and Wealth Creation
- ASK REAL ESTATE: When Vermin Come to Visit
- THE WORKING LIFE: A Store Closes, but the Business Survives
- DEALBOOK: British Firm Starts Hedge Fund for Social Services
- In Moscow, a Financial District in Name Only
- On the Market in New York City

Perceptron

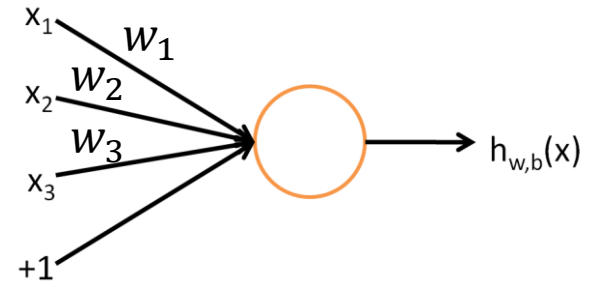
- Tries to mimic a real NN, since it has a nucleus that processes some inputs and give an output.



- $h_{w,b}(x)$ is a function of all the inputs, and is composed of two terms.

Perceptron

$$h_{w,b}(x) = f\left(\sum_{i=1}^3 W_i x_i + b\right)$$



f is called the activation function, and it works as a way to discretize the outputs of the perceptron.

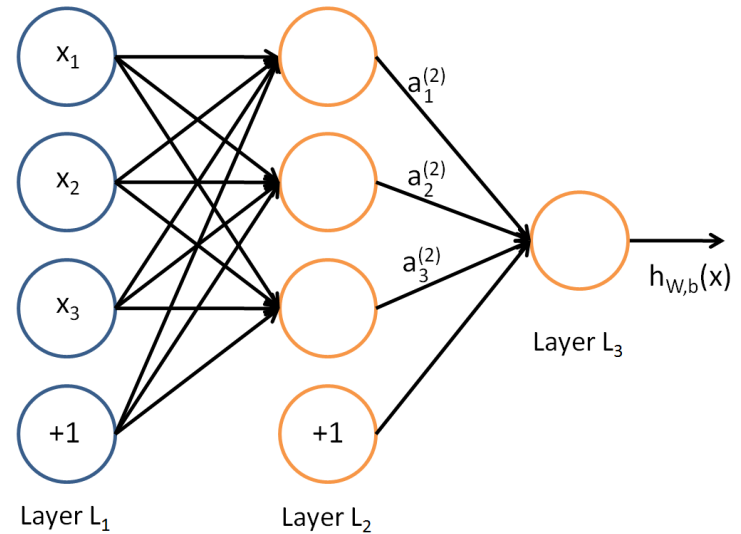
One of the most common activations functions is the sigmoid function:

$$f(z) = \frac{1}{1 + \exp(z)}$$

This looks very familiar

Neural Network

- Naturally, a NN is going to be a set of perceptrons interconnected within each other.



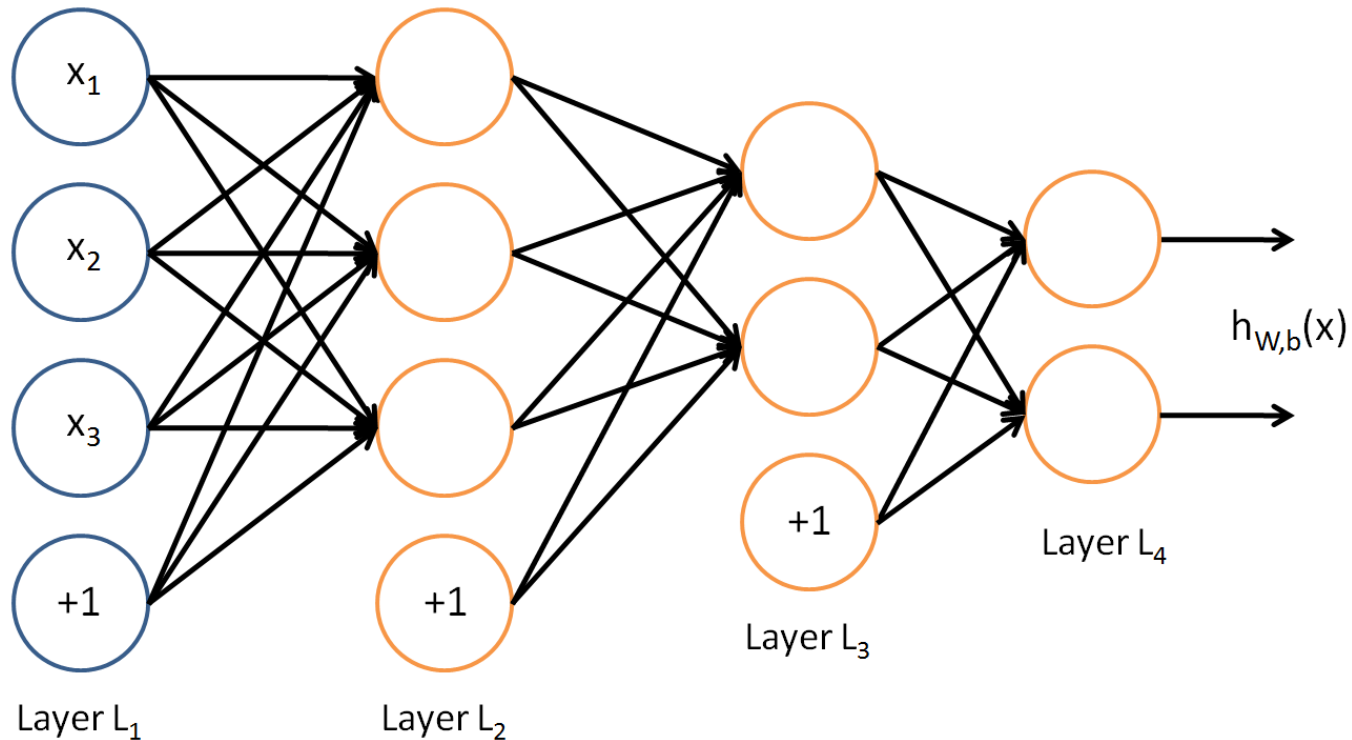
$$a_1^{(2)} = f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)})$$

$$a_2^{(2)} = f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)})$$

$$a_3^{(2)} = f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)})$$

$$h_{W,b}(x) = a_1^{(3)} = f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)})$$

Neural Network



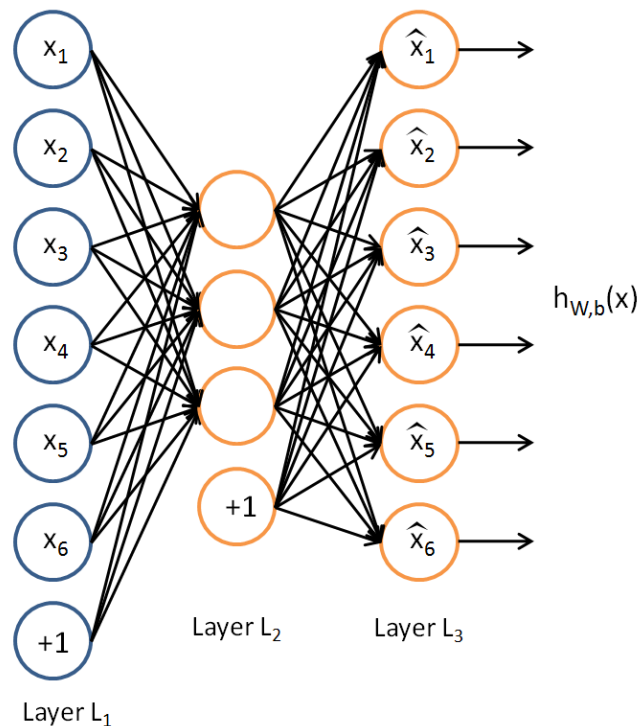
- We can add as many layers and outputs as we want, for example a two binary output allows us to classify in four classes.
- We also regularize NNs, since they can be also prone to overfitting.

Problems of NNs

- We need to answer two questions:
 - How many layers are enough to solve a problem?
 - How many hidden units should we use per layer?
- As you can imagine, training complexity increases as we increase hidden units.
 - This can be reduced by avoiding a full interconnection.
- The elephant in the room is called “Vanishing Gradient”

Autoencoders

- An autoencoder is a NN where the output and the input are the same.



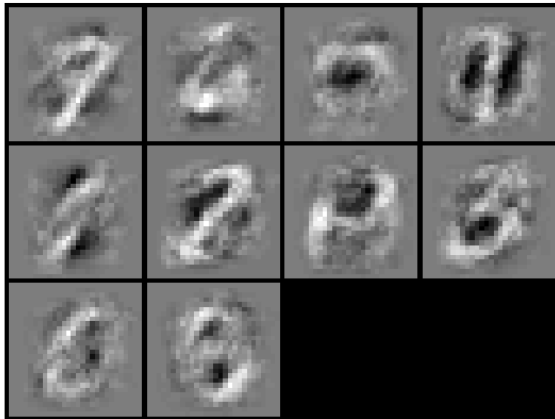
MNIST Dataset

- Dataset of handwritten digits
- Has a training set of 60,000 examples, and a test set of 10,000 examples.
- Each digit is an 28x28 image (784 pixels)
- Each digit has a label that identifies which digit it represents. (9 labels)



Autoencoders

- Why would I want both the input and the output to be the same.
- MNIST dataset as an example (28x28 input images)

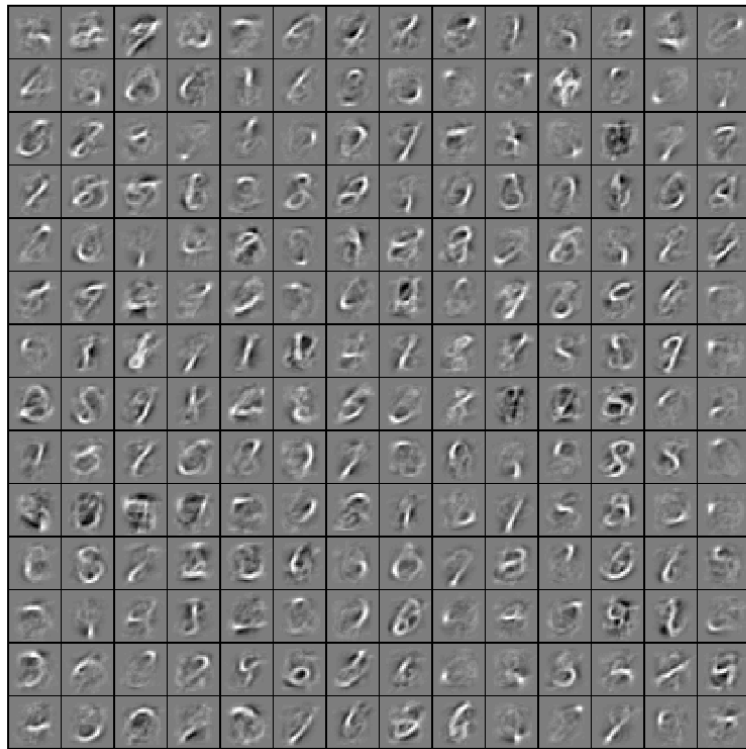


10 hidden units in Autoencoder

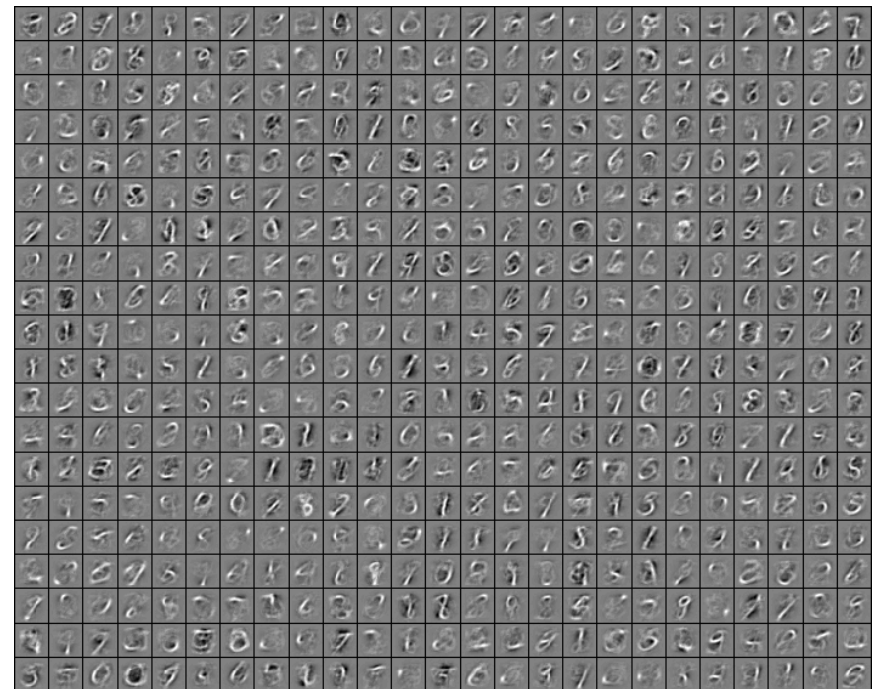


80 hidden units in Autoencoder

Autoencoders



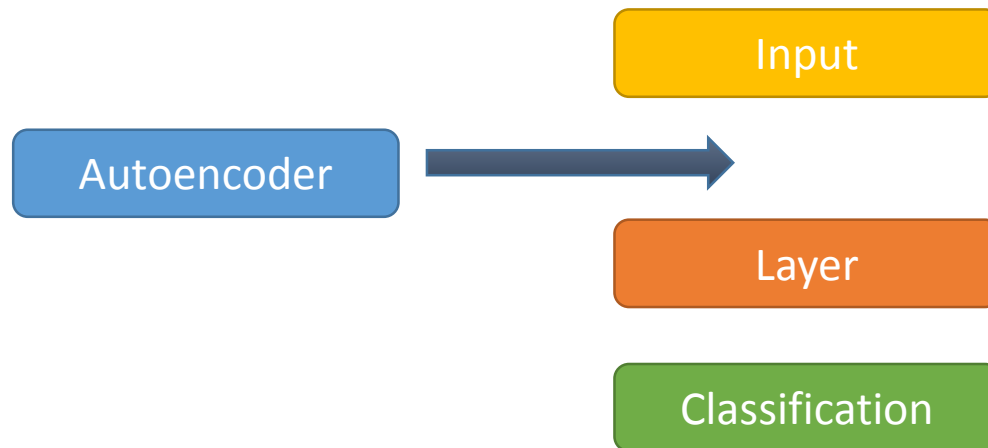
196 hidden units in Autoencoder



500 hidden units in Autoencoder

Autoencoders and Deep Nets

- We train an autoencoder, and plug it in a NN then train.



- This simple modification is one of the most important advancements in NN practice in the past 20 years.

Demo

- <http://www.clarifai.com/>

www.clarifai.com

clarifai

Demo News Contact Jobs

clarifai
Advanced Image Recognition

Recognize tens of thousands of categories, objects, and tags in any image.

Try our live demo »

Stay in the loop with news about Clarifai

Your email address I'm mainly interested in I'd like beta API access Keep me updated

Take our demo for a spin.
Send us an image, and we'll understand it using our model trained on 10,000 categories

Upload an image
Choose image
By uploading an image, you agree to our terms of use.

Or...

Enter an image URL
Image URL Image URL Go

Important notes

- We are still not entirely sure why it works:
 - Some people say is because using this as a random start saves us much hassle.
 - Some say that this artificially moves us to a better search space.
- Using the autoencoder as a preprocessing step, has been proven to help us save steps when it comes to preprocessing algorithms.
- The autoencoder can find circles, edges, etc by itself.