Machine Learning & Deep Nets

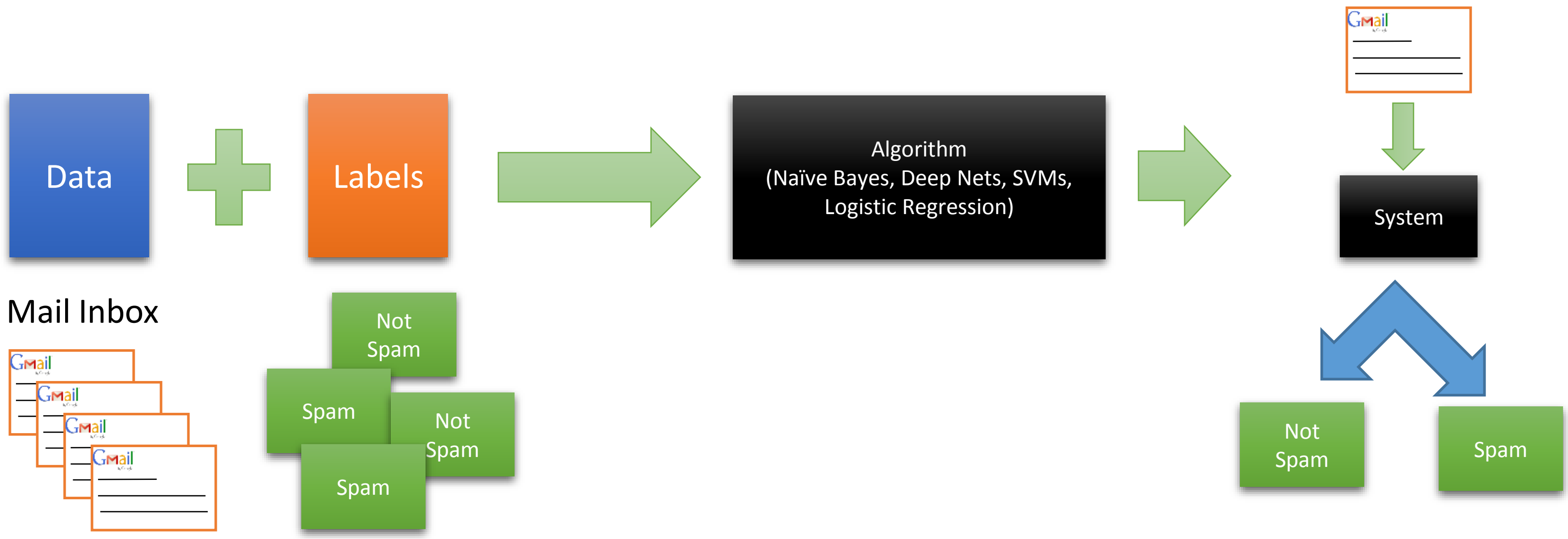
Leon F. Palafox

December 4^{th}, 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



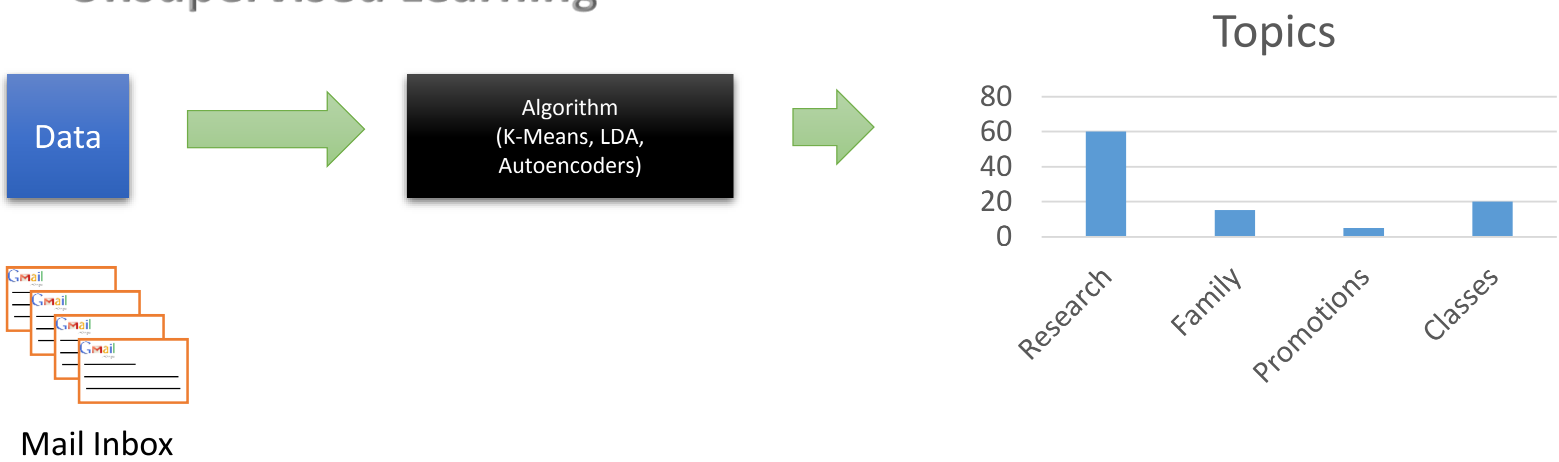
The set of elements that describe a single datum are called features, in this case, the features are the words in the e-mails.

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



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

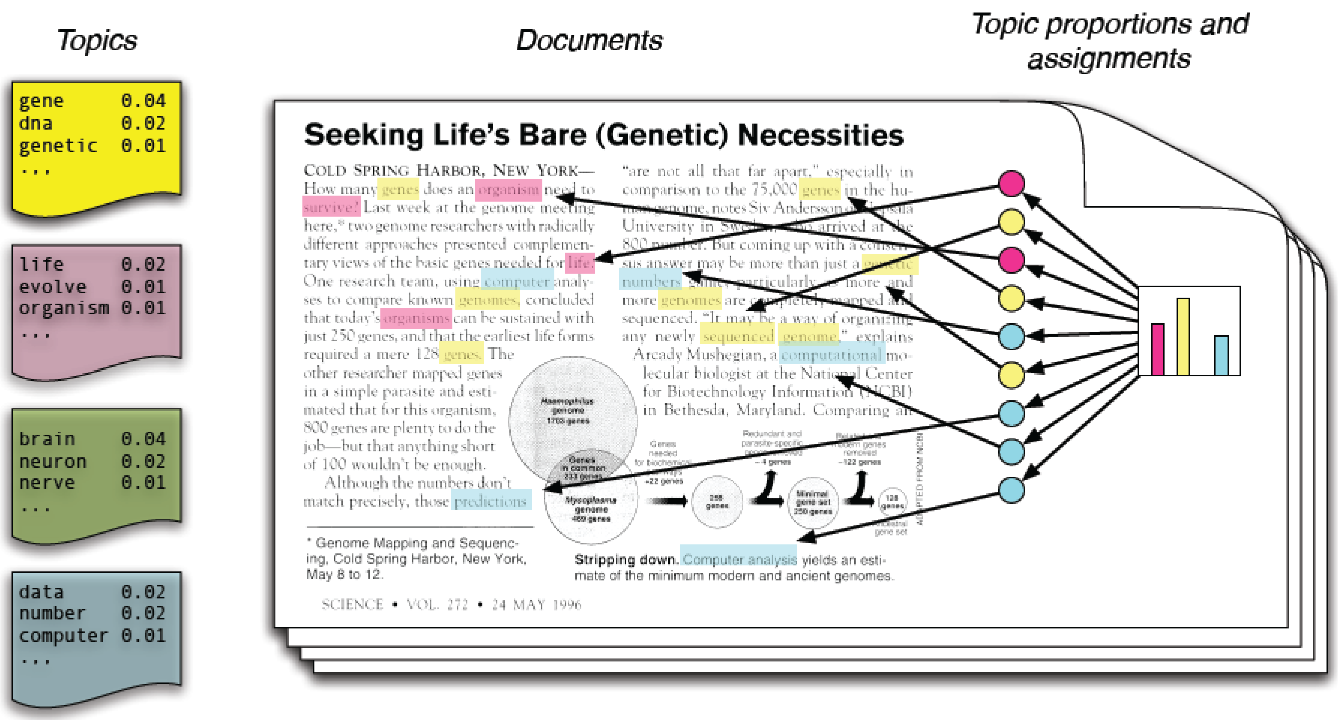
Each topic (clusters) will have features that will characterize them.

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

Family: Mom, House, Mexico

Promotions: Computer, PS4, Cheap, Amazon, Deal

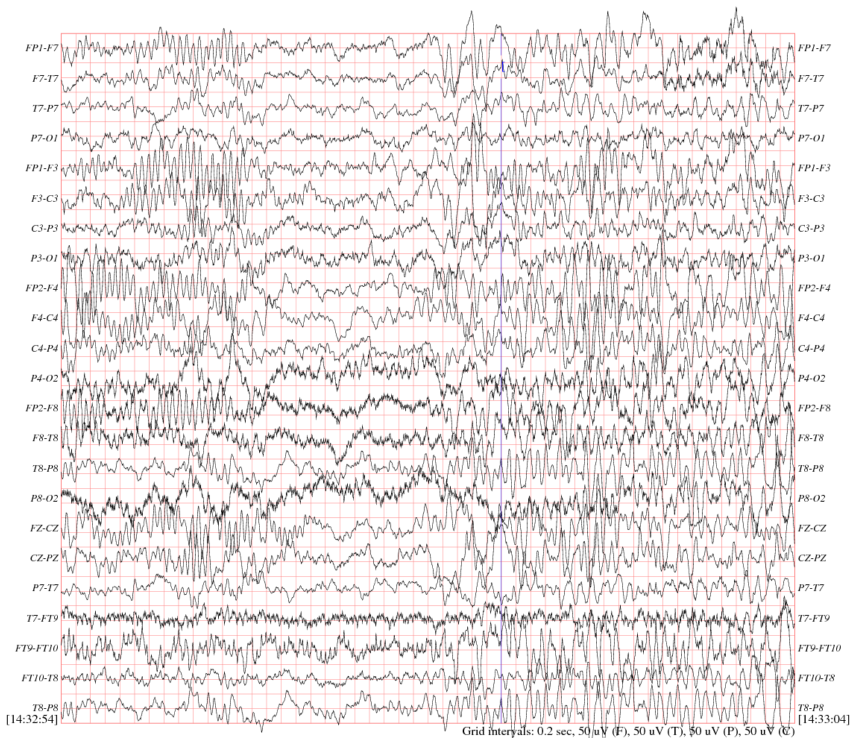
Classes: Grades, Homework, Questions, Office Time

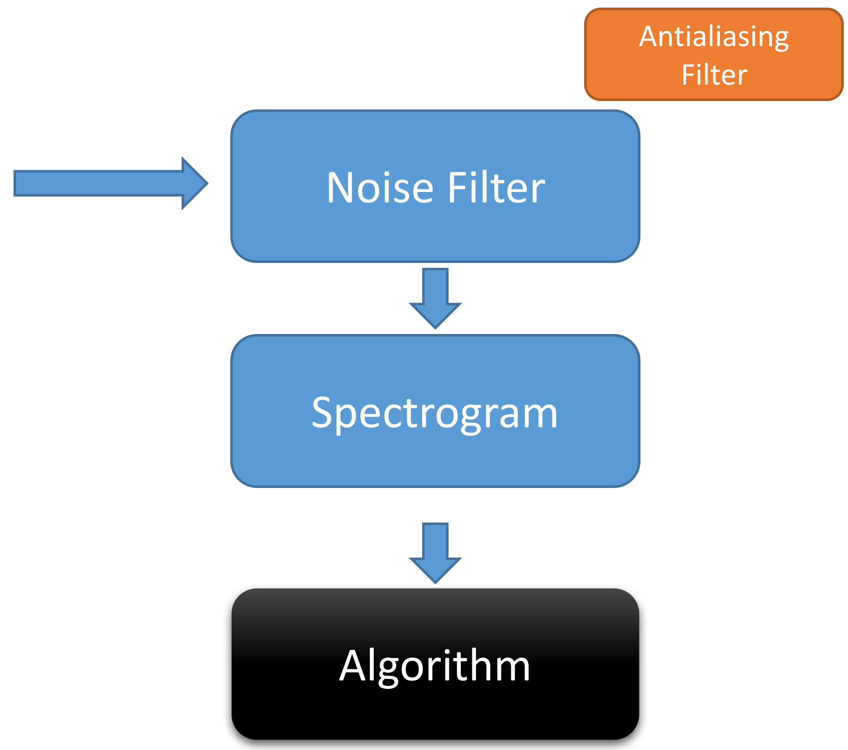


http://cs.stanford.edu/people/karpathy/nips2014/
http://sarah-palin.herokuapp.com/

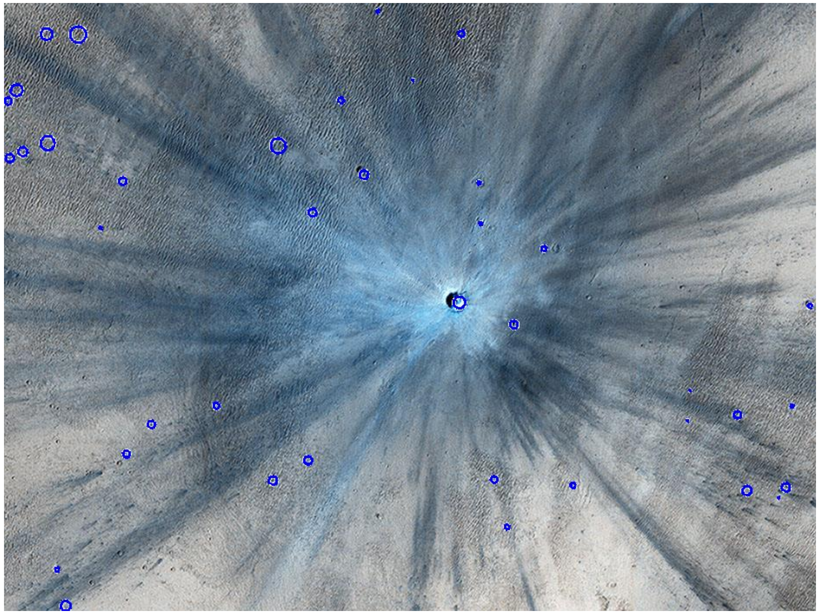
Preprocessing Data

- 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.

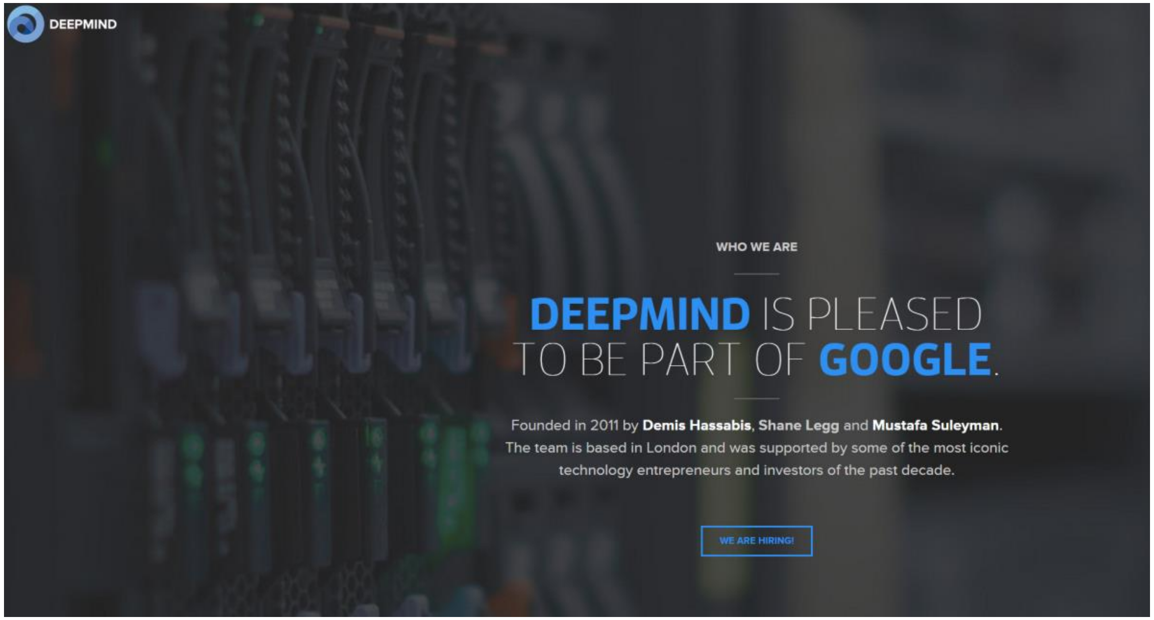
Background

- 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.

THE NEW YORKER

NOVEMBER 25, 2012

IS "DEEP LEARNING" A REVOLUTION IN ARTIFICIAL INTELLIGENCE?

BY GARY MARCUS

NEWS DESK

Tweet g+1

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.

日本リアルタイム
JAPANREALTIME

8:47 pm JST
Oct 1, 2014
JAPAN

NTT, Toyota Seek 'Deep Learning' Expertise

PREVIOUS
Big Win for Japanese Banks in Myanmar
NEXT
New Commemorative Coins Celebrate Shinkansen's 50th Anniversary

ARTICLE COMMENTS

DEEP LEARNING NTT PREFERRED NETWORKS TOYOTA MOTOR CORP

Email Print

By TAKASHI MOCHIZUKI CONNECT

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, as it struggles to keep pace with faster-growing

Scientists See Promise in Deep-Learning Programs

Hao Zhang/The New York Times

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

FACEBOOK TWITTER GOOGLE+ SAVE EMAIL SHARE PRINT REPRINTS

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.

点击查看本文中文版。

Connect With Us on Social Media
@nytimesscience on Twitter.
· Science Reporters and Editors on Twitter
Like the science desk on Facebook.

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.

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.

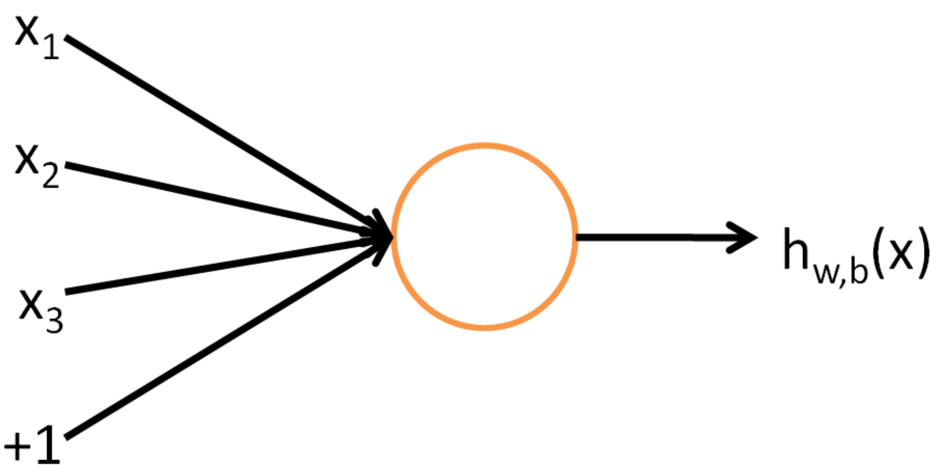
But what is new in recent months is the growing speed and accuracy of deep-learning programs, often called artificial neural networks or just "neural nets" for their resemblance to the neural connections in the brain.

"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

MOST EMAILED RECOMMENDED FOR YOU
1. DEALBOOK Saks Flagship Store Is Appraised for Mortgage at $3.7 Billion
2. DEALBOOK U.S.-Backed Mortgages Put to Test in an Innovative Lawsuit
3. TODAY'S EDITORIALS Homeownership and Wealth Creation
4. ASK REAL ESTATE When Vermin Come to Visit
5. THE WORKING LIFE A Store Closes, but the Business Survives
6. DEALBOOK British Firm Starts Hedge Fund for Social Services
7. In Moscow, a Financial District in Name Only
8. On the Market in New York City

Perceptron

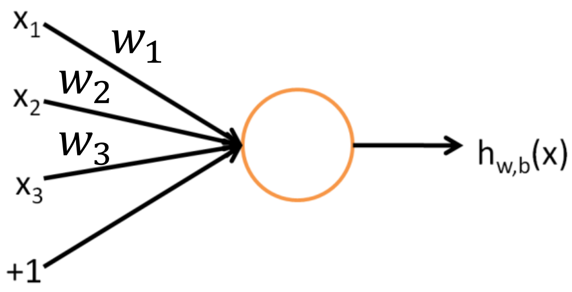
- Tries to mimic a real NN, since it has a nucleus that processes some inputs and give an output.



- $h_{w,b}(x)$ 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.

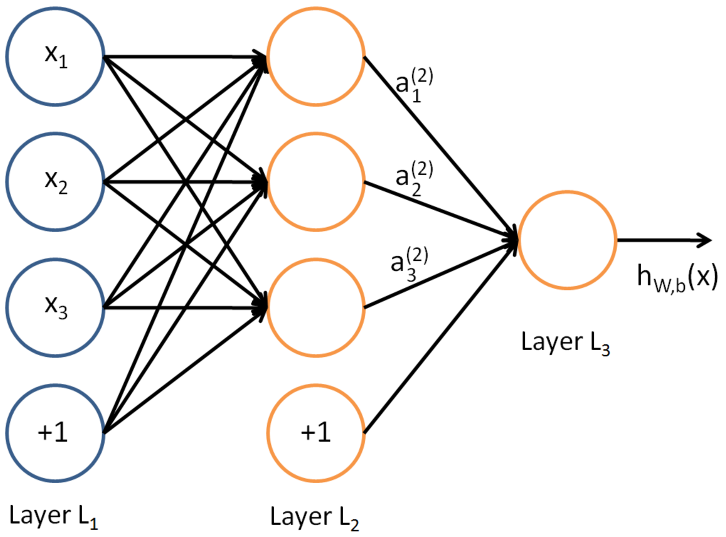
One of the most common activations functions is the sigmoid function:

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

This looks very familiar

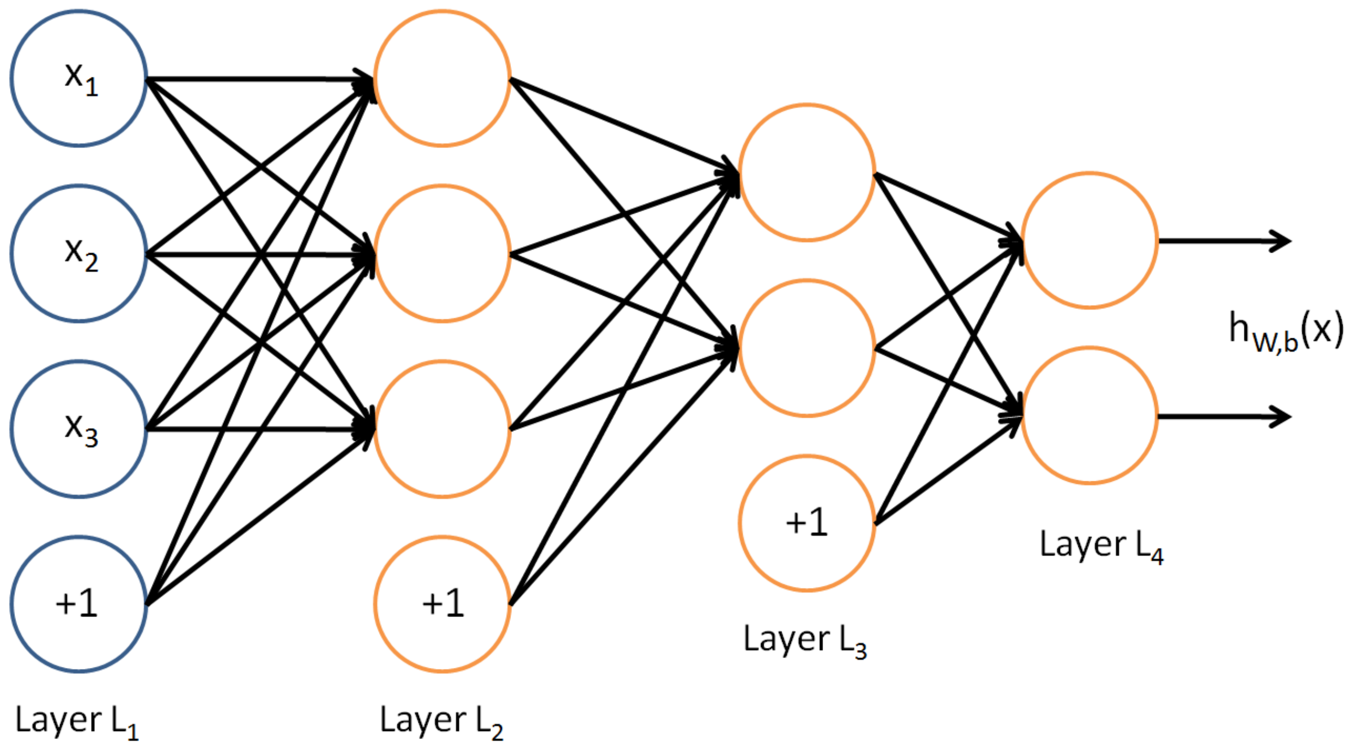
Neural Network

- 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



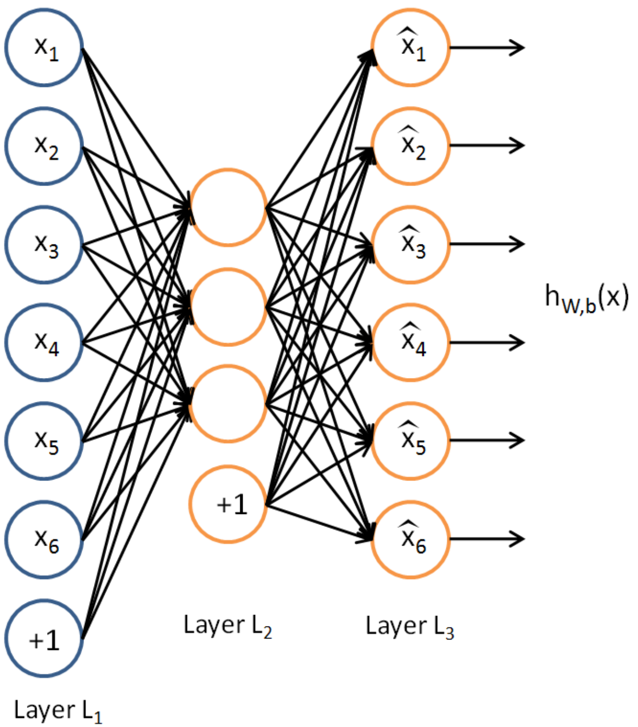
- We can add as many layers and outputs as we want, for example a two binary output allows us to classify in four classes.
- We also regularize NNs, since they can be also prone to overfitting.

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”

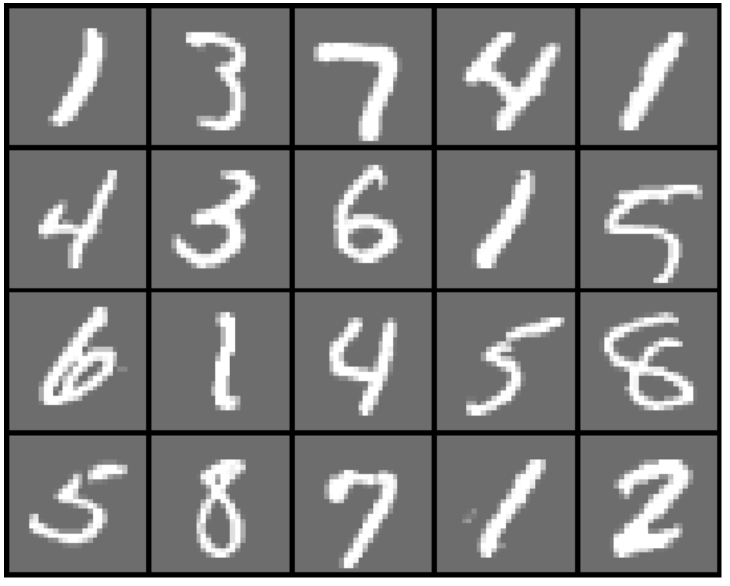
Autoencoders

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



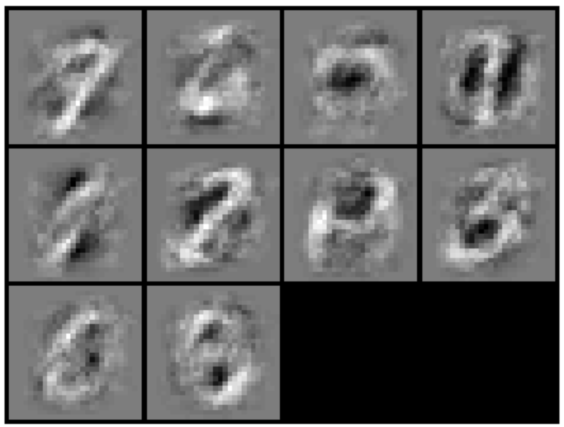
MNIST Dataset

- 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

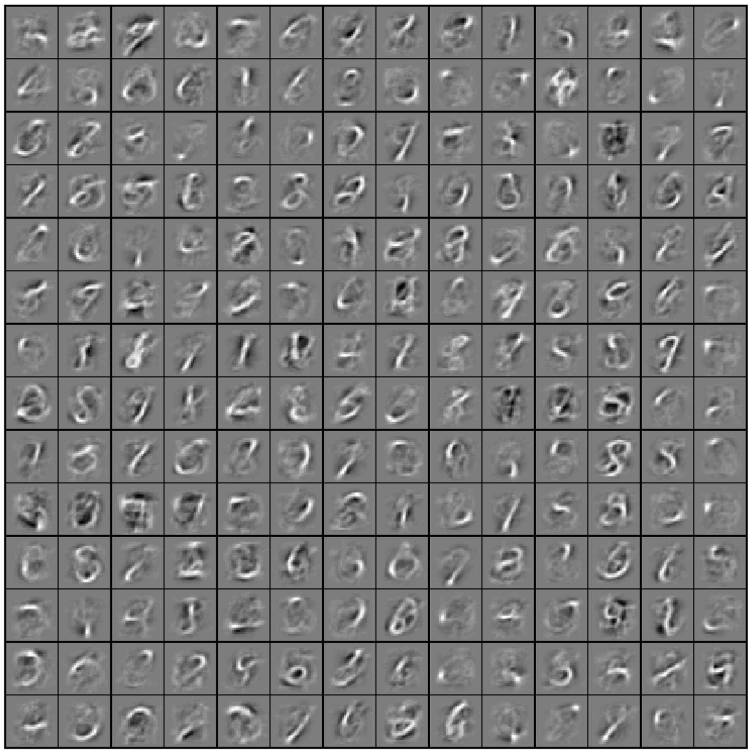
- Why would I want both the input and the output to be the same.
- MNIST dataset as an example (28x28 input images)



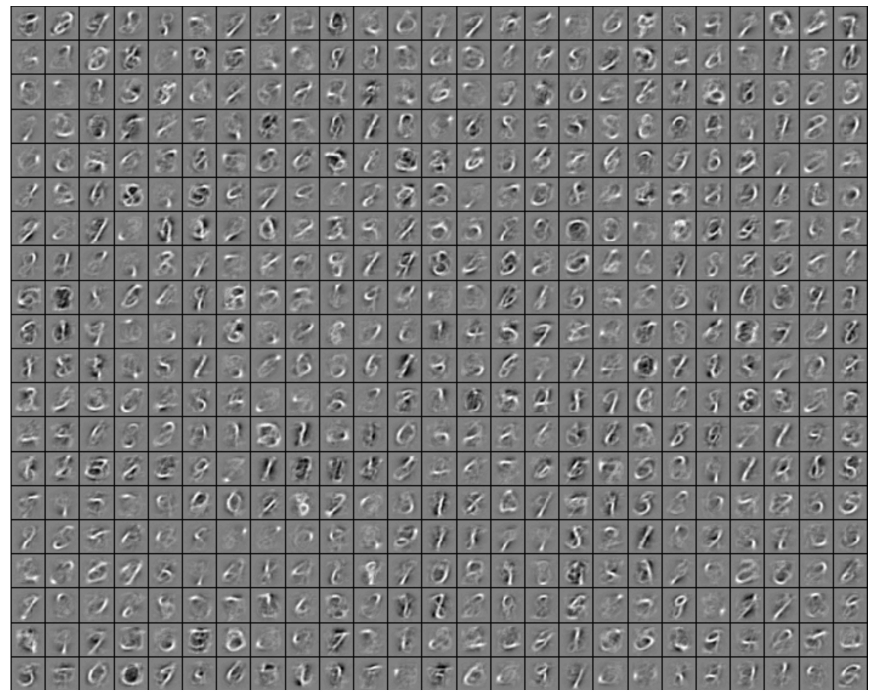
10 hidden units in Autoencoder

80 hidden units in Autoencoder

Autoencoders



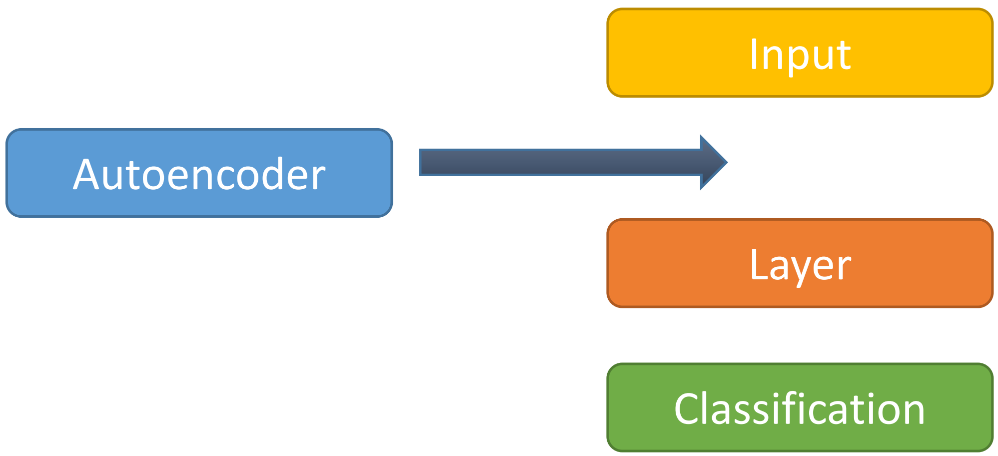
196 hidden units in Autoencoder



500 hidden units in Autoencoder

Autoencoders and Deep Nets

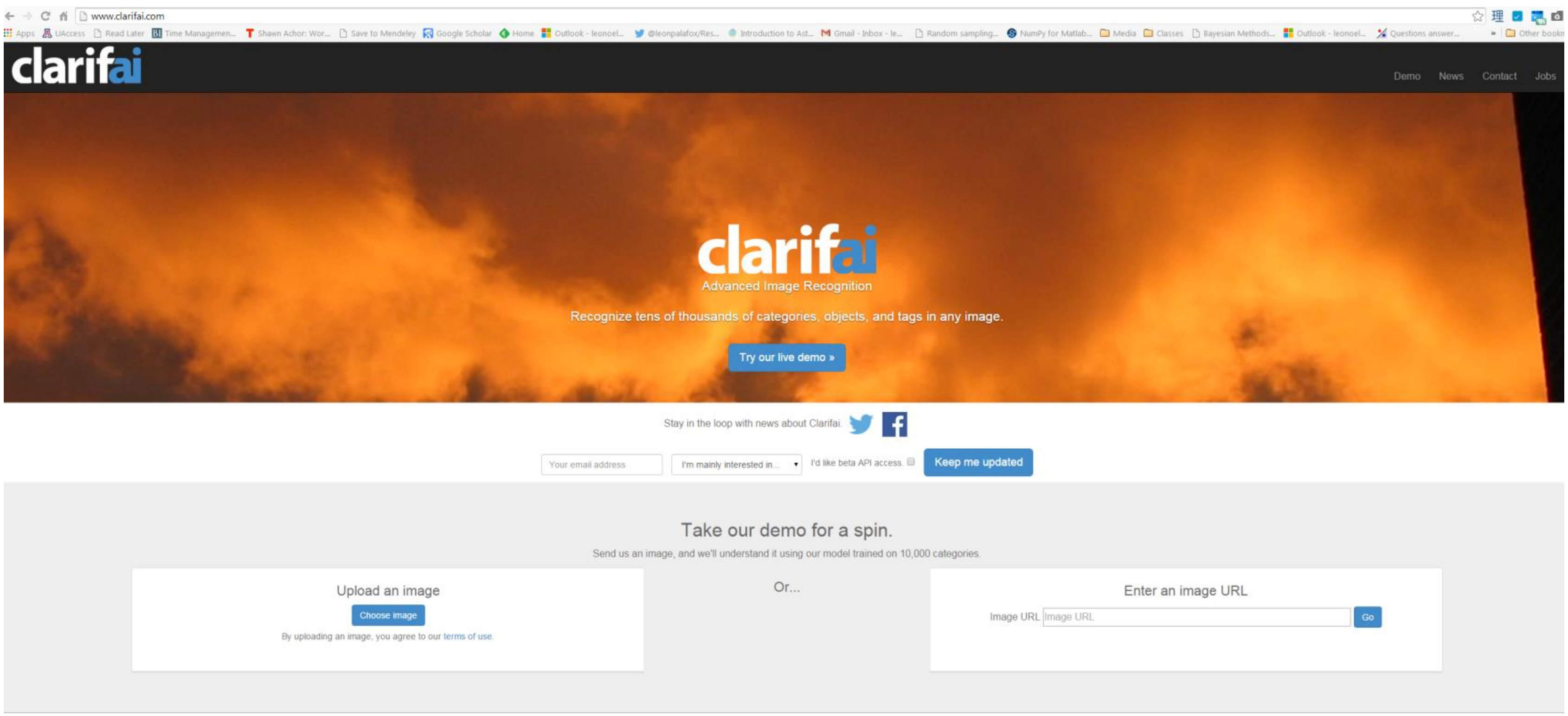
- We train an autoencoder, and plug it in a NN then train.



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

Demo

- http://www.clarifai.com/



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.