Deep Learning
(ML&DL)

Course Details

- Instructor
  - Fang Hui (方慧), Assistant Professor
  - email: fang.hui@mail.shufe.edu.cn

Course Details (cont.)

- Grading
  - Attendance and classroom performance (5%)
  - Individual Project (45%)
    - Research proposal (10%), deadline: April 26, 2019
    - Final Report (35)%
- Textbook
  - Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.
- Reference book:
  - Dive into Deep Learning, http://d2l.ai/

Research Proposal (no more than 2 pages)

- The major content of RP
  - Define a clear research and approach to answering it
  - Highlight its originality or significance
  - Explain how it adds to existing literature
- Structuring your proposal
  - Title
  - Introduction (research questions, objectives, significance/contribution, research plan)
  - References

Pre-requisites

- Proficiency in Python
  - All class assignments will be in Python
- Linear algebra
- Basic Probability and Statistics
- Machine Learning basics

Course Objectives

- Understanding deep learning – mainly neural networks
- Tips for training DL, regularization and optimization
- Familiarity with some of the DL structures: CNN, RNN, etc.
- Design and train networks for various tasks

Course Schedule

- Introduction to DL and DL basics
- Neural network basics and DNN
    - Regularization and optimization
    - Build networks and tips for training DNN
- Image processing and CNN
- NLP and RNN

Acknowledged course materials

- CMU 11-785 Introduction to Deep Learning http://deeplearning.cs.cmu.edu/
- Stanford CS231n: Convolutional Neural Networks for Visual Recognition http://cs231n.stanford.edu/index.html
- Stanford CS224d: Deep learning for natural language processing http://cs224d.stanford.edu/syllabus.html
- NTU ADL& MLDS https://www.csie.ntu.edu.tw/~yvchen/f106-adl/syllabus.html
- **Berkeley** https://courses.d2l.ai/berkeley-stat-157/syllabus.html

Lecture 1: Introduction to Deep Learning

课程：机器学习与深度学习

Overview

- What is Deep Learning?
- Deep Learning History
- Reasons for Exploring Deep Learning

Program for Solving Tasks

- Programs can do the things you ask them to do
- E.g. task: predicting positive or negative given a product review

" I love the restaurant"

program.py

+

if input contains "love", "like", etc.
output = positive

"The quality is bad"

program.py

-

if input contains "too much", "bad", etc.
output = negative

"This is not very expensive"

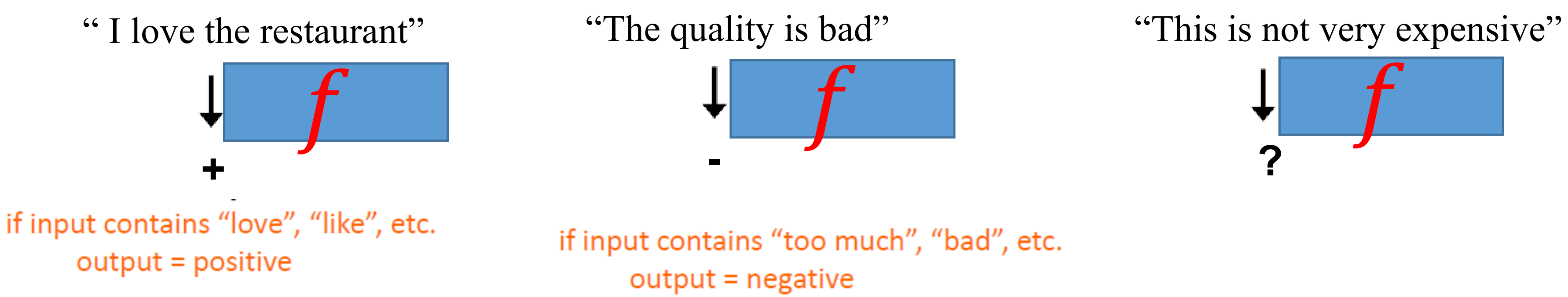
program.py

?

Some tasks are complex, and we do not know how to write a program (i.e. define the rules) to solve them.

Machine Learning ≈ Looking for a Function

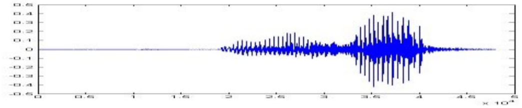
- Task: predicting positive or negative given a product review



Given a large amount of data, the machine learns what the function f should be.

Machine Learning ≈ Looking for a Function

- Speech Recognition

$f($  $) =$ "How are you"

- Image Recognition

$f($  $) =$ "Cat"

- Playing Go

$f($  $) =$ "5-5" (next move)

- Dialogue System

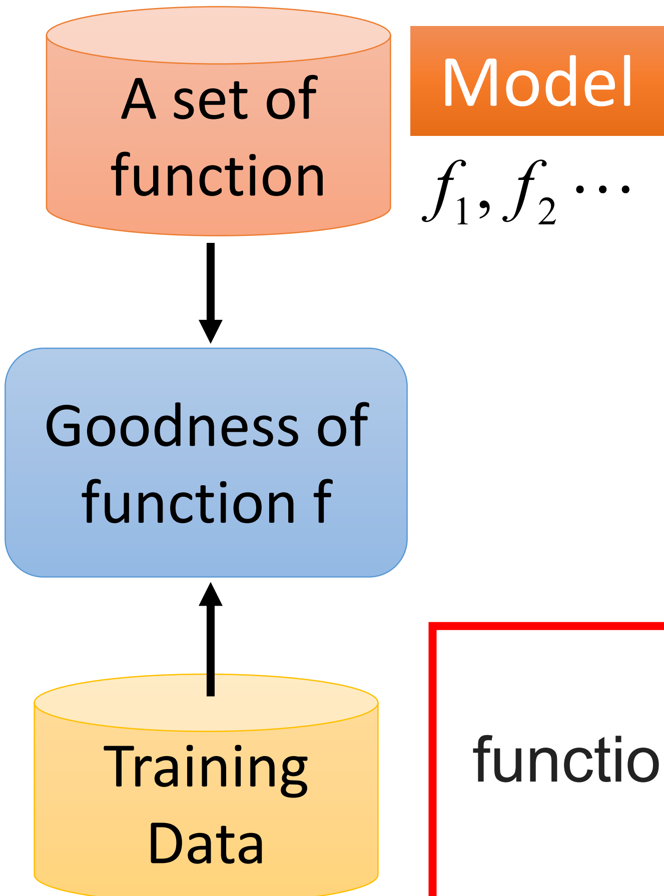
$f($ "Hi" $) =$ "Hello"

(what the user said) (system response)

Framework

Sentiment analysis

f("I love the restaurant") = "+" (positive)



$f1$("I love the restaurant") = "+"

$f2$("I love the restaurant") = "−"

Better!

$f1$("the quality is bad") = "−"

$f2$("the quality is bad") = "+"

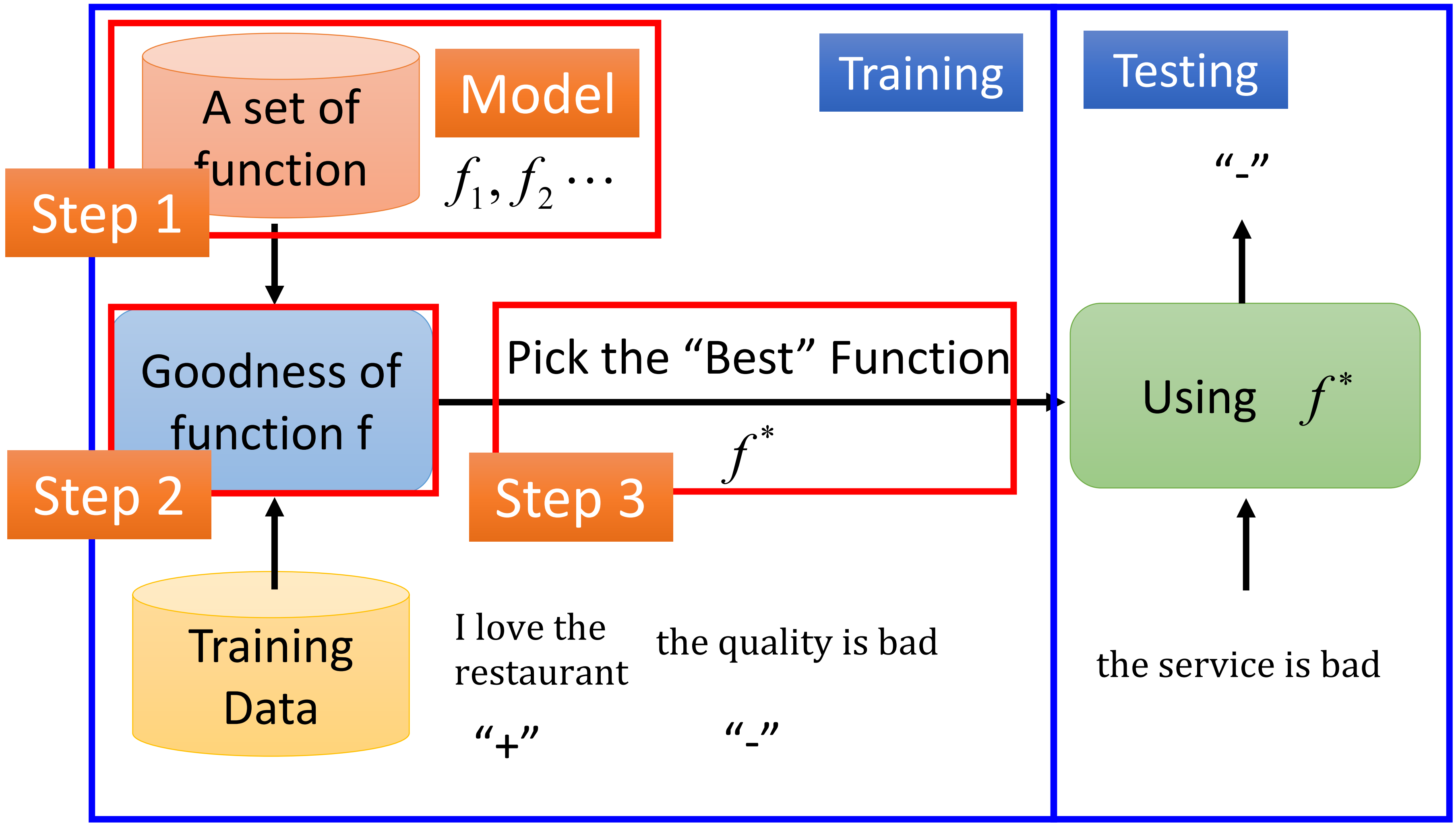
Supervised Learning

function input:	I love the restaurant	the quality is bad
function output:	"+"	"-"

Framework

Sentiment analysis:

f("I love the restaurant") = "+" (positive)



Three Steps for Machine Learning

Three Steps for Deep Learning

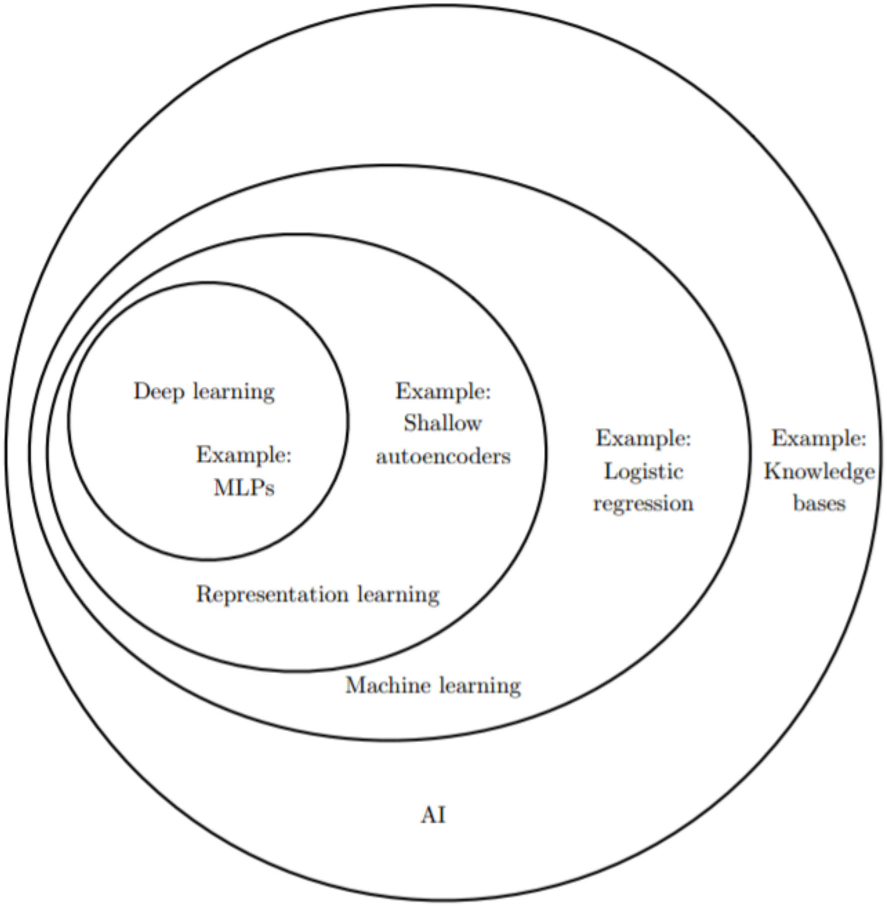
Step 1: define a Neural Network

Step 2: goodness of function

Step 3: pick the best function

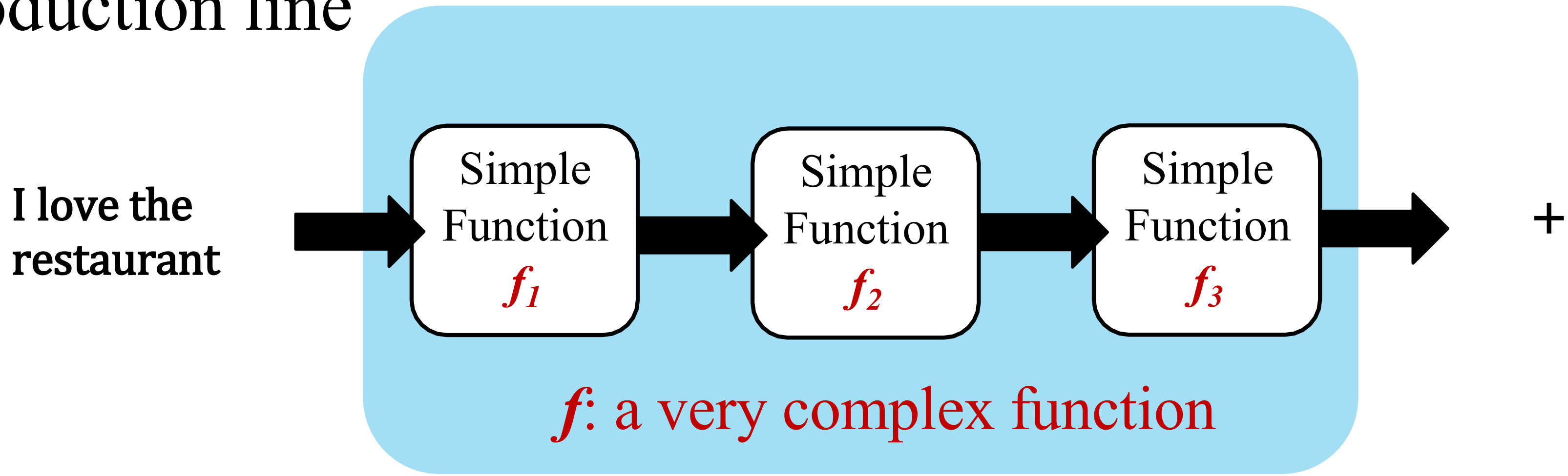
What is Deep Learning (DL)?

- A subfield of machine learning



Stacked Functions Learned by Machine

- Production line

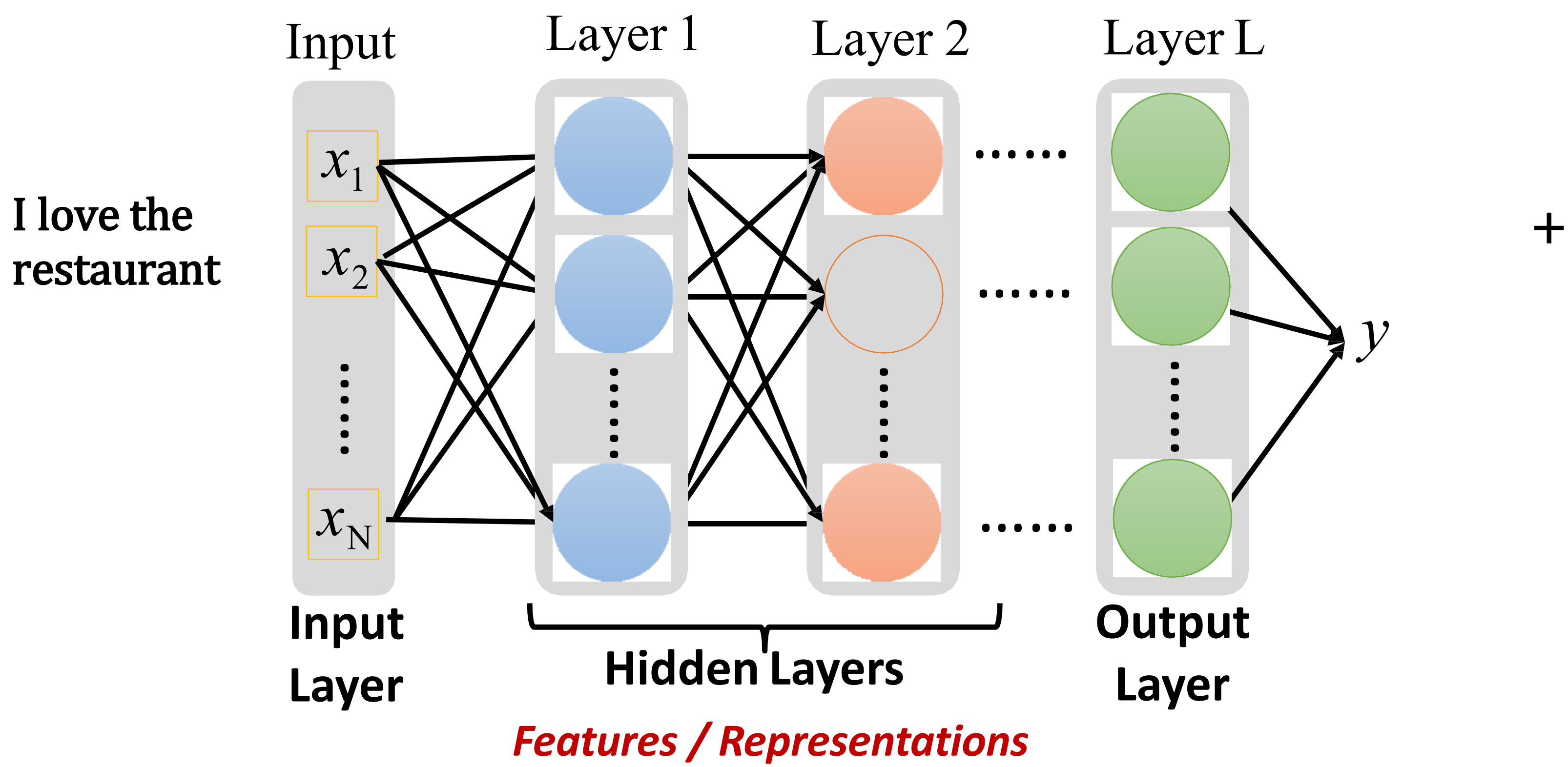


Deep Learning is as simple as linear model……

End to end training: what each function should do is learned automatically

DL usually refers to neural network based model

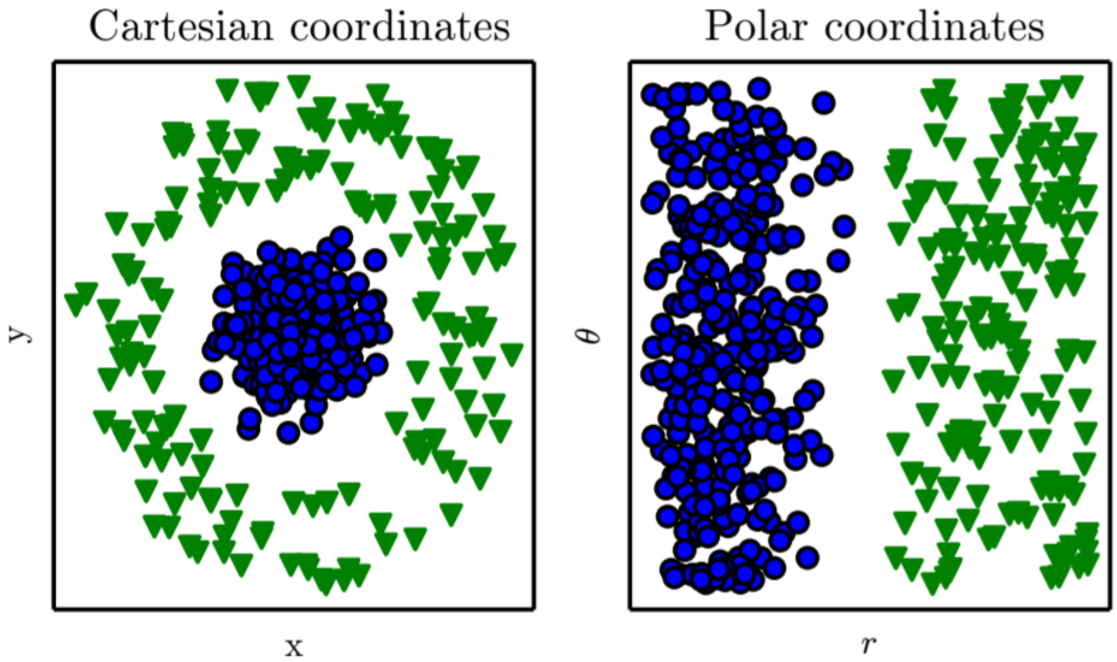
Stacked Functions Learned by Machine



Representation Learning attempts to learn good features/representation

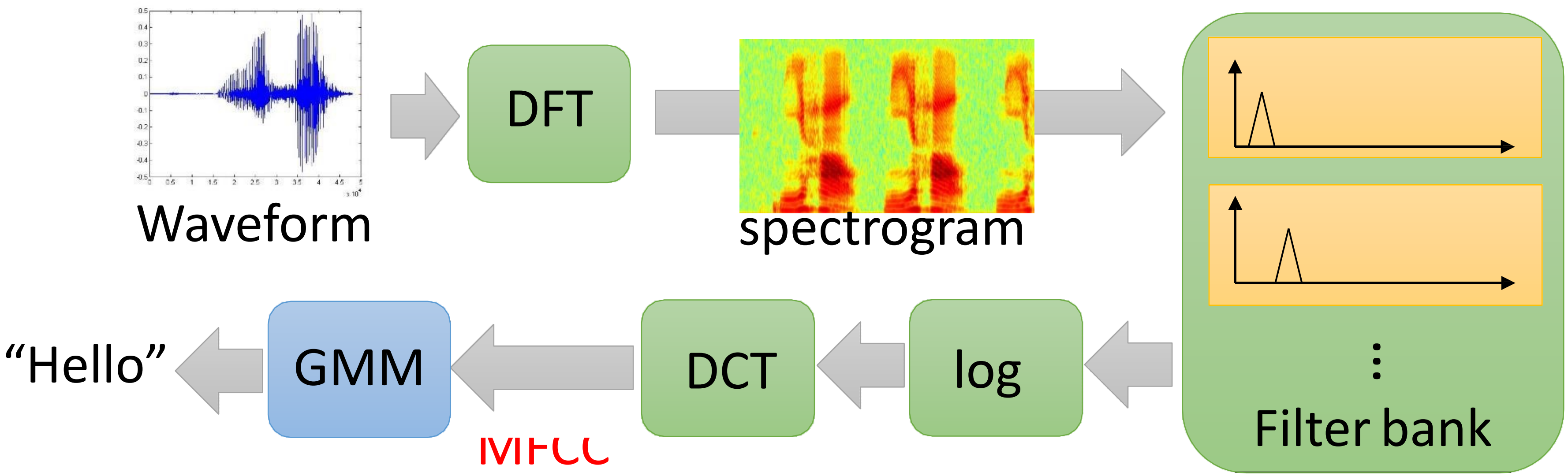
Deep learning algorithms attempt to learn multiple levels of representation and an output

Representations Matter



Deep v.s. Shallow – Speech Recognition

Shallow Model

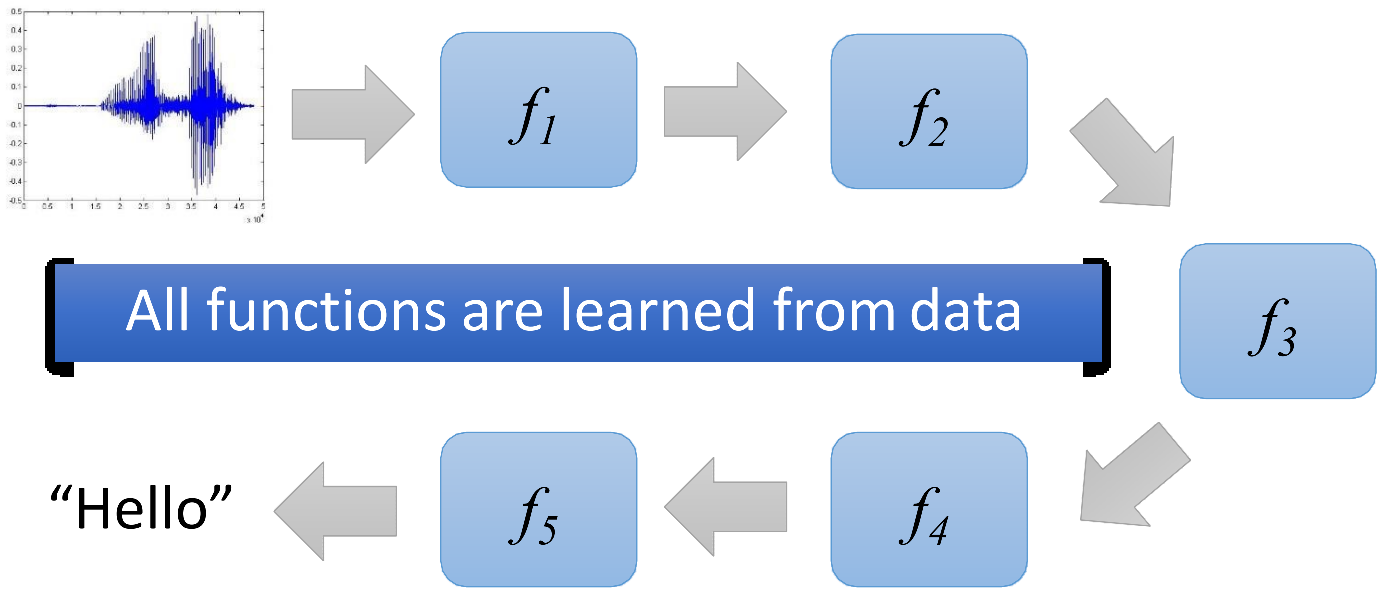


Each box is a simple function in the production line:

:hand-crafted :learned from data

Deep v.s. Shallow – Speech Recognition

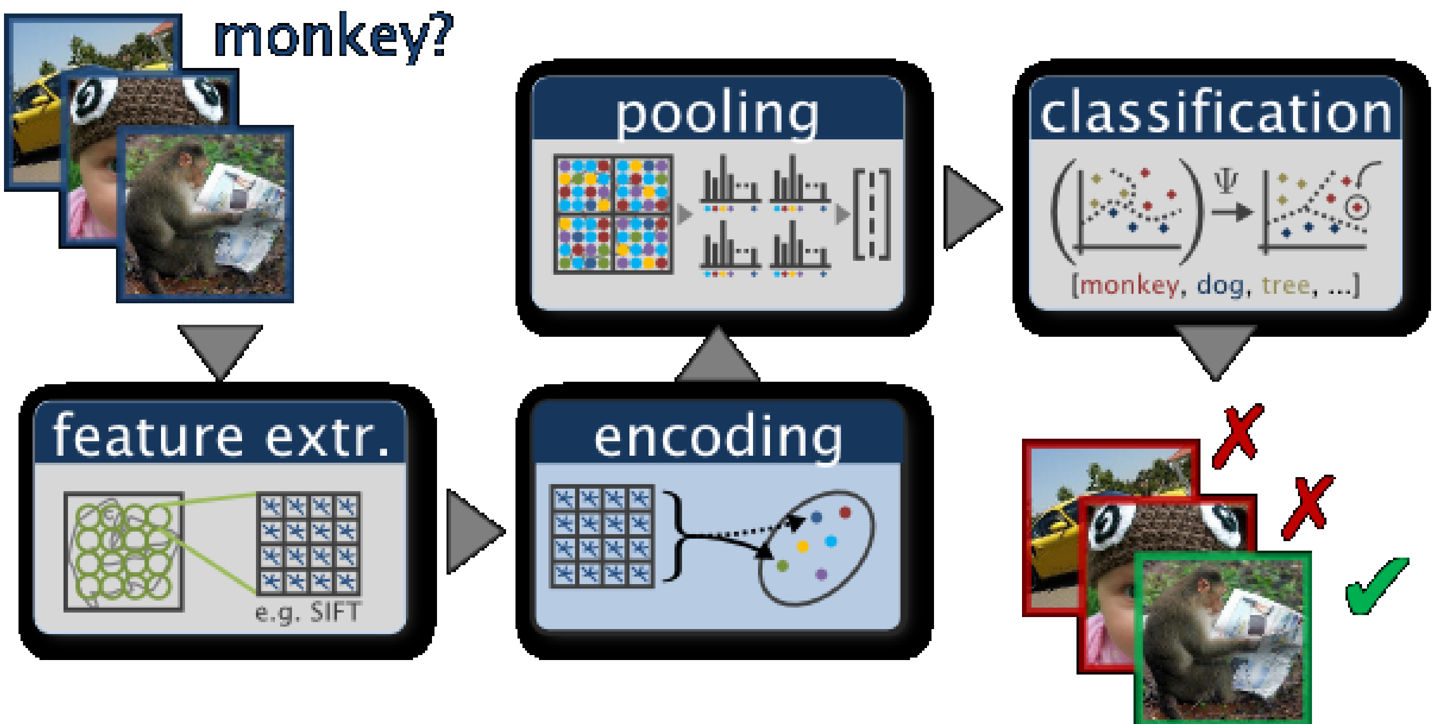
Deep Model



Less engineering labor, but machine learns more

Deep v.s. Shallow – Image Recognition

Shallow Model

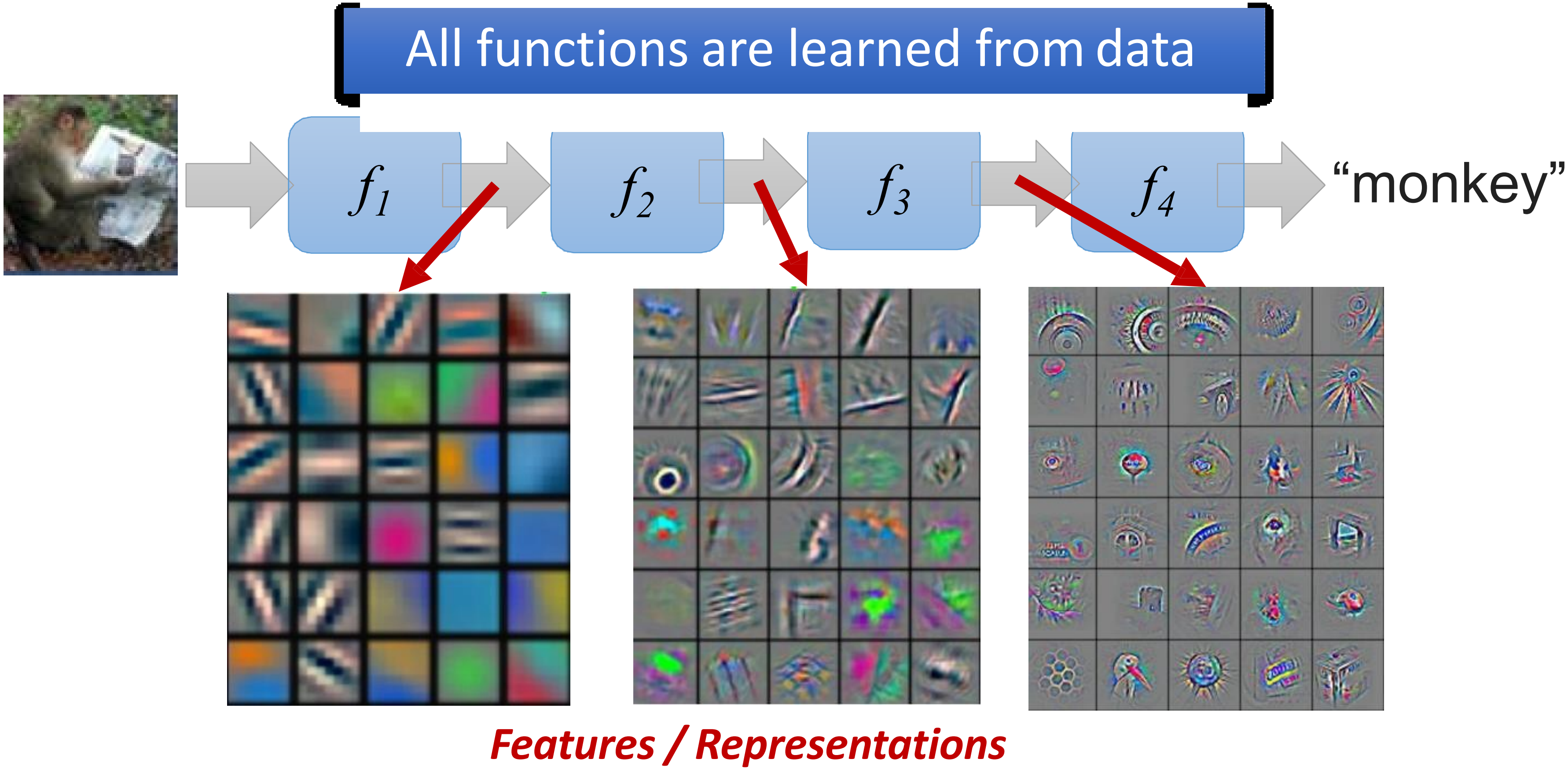


:hand-crafted :learned from data

http://www.robots.ox.ac.uk/~vgg/research/encoding_eval/

Deep v.s. Shallow – Image Recognition

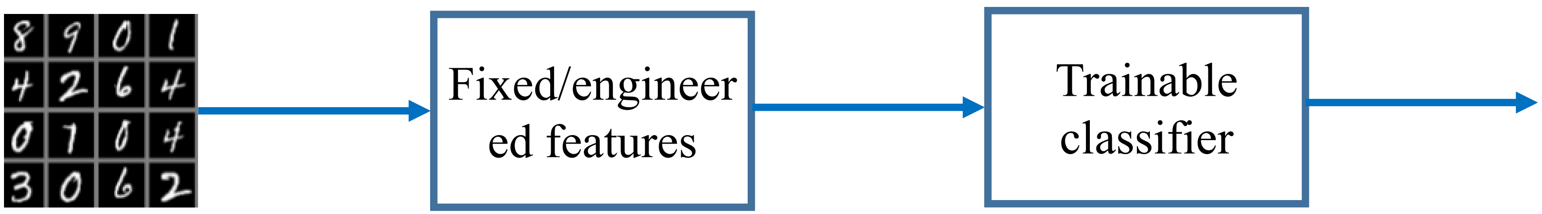
Deep Model



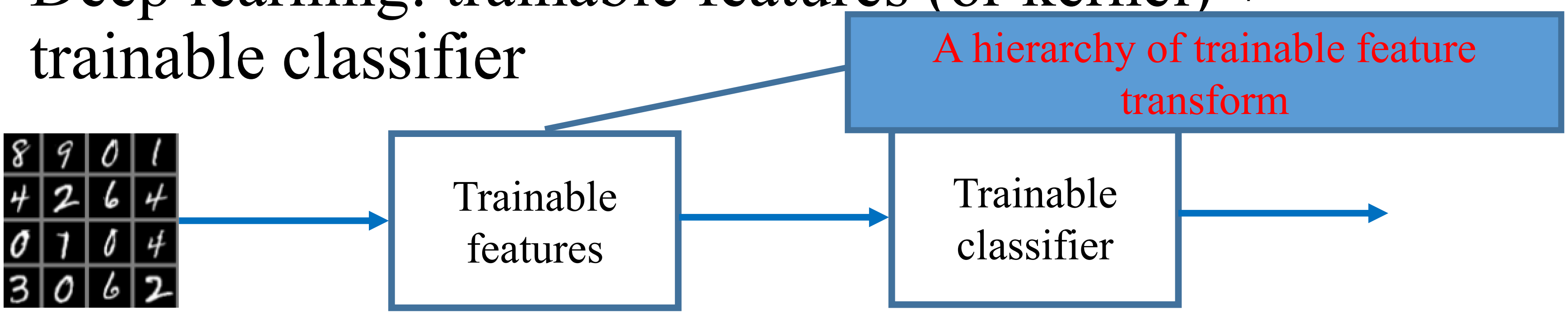
Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *European conference on computer vision*. Springer, Cham, 2014.

Deep Learning vs. Traditional Learning

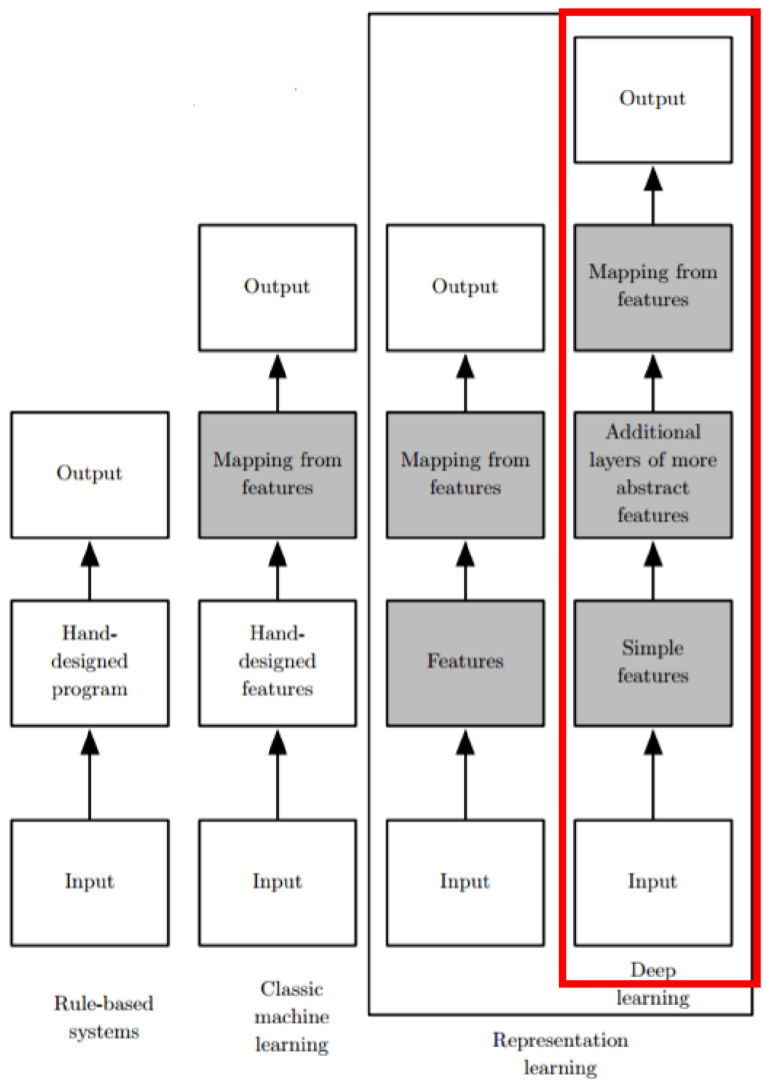
- Traditional learning: Fixed/engineered features (or fixed kernel) + trainable classifier



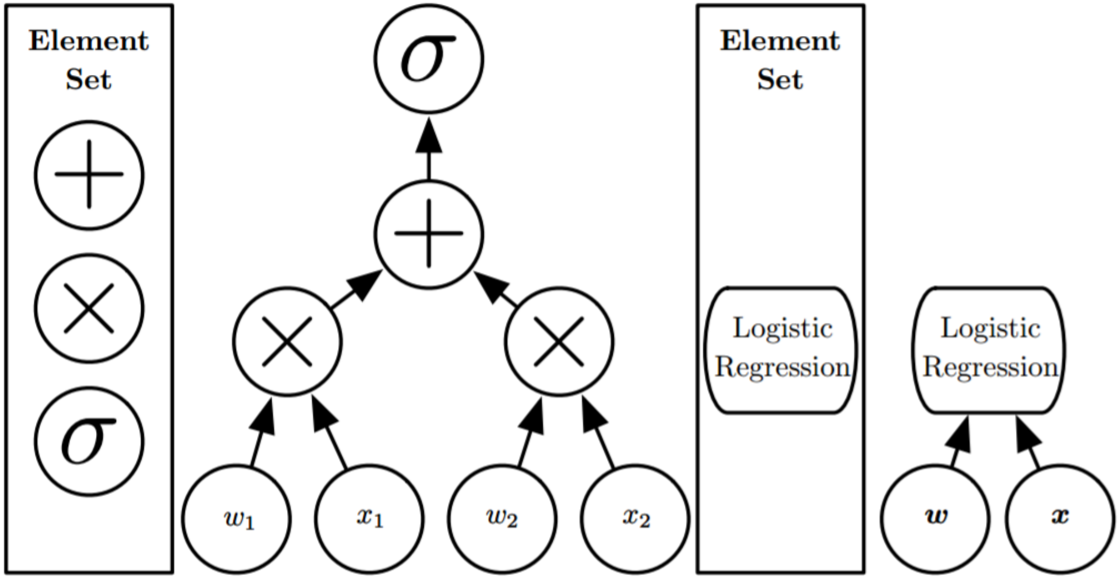
- Deep learning: trainable features (or kernel) + trainable classifier



Deep Learning vs. Others



Depth of Deep Learning



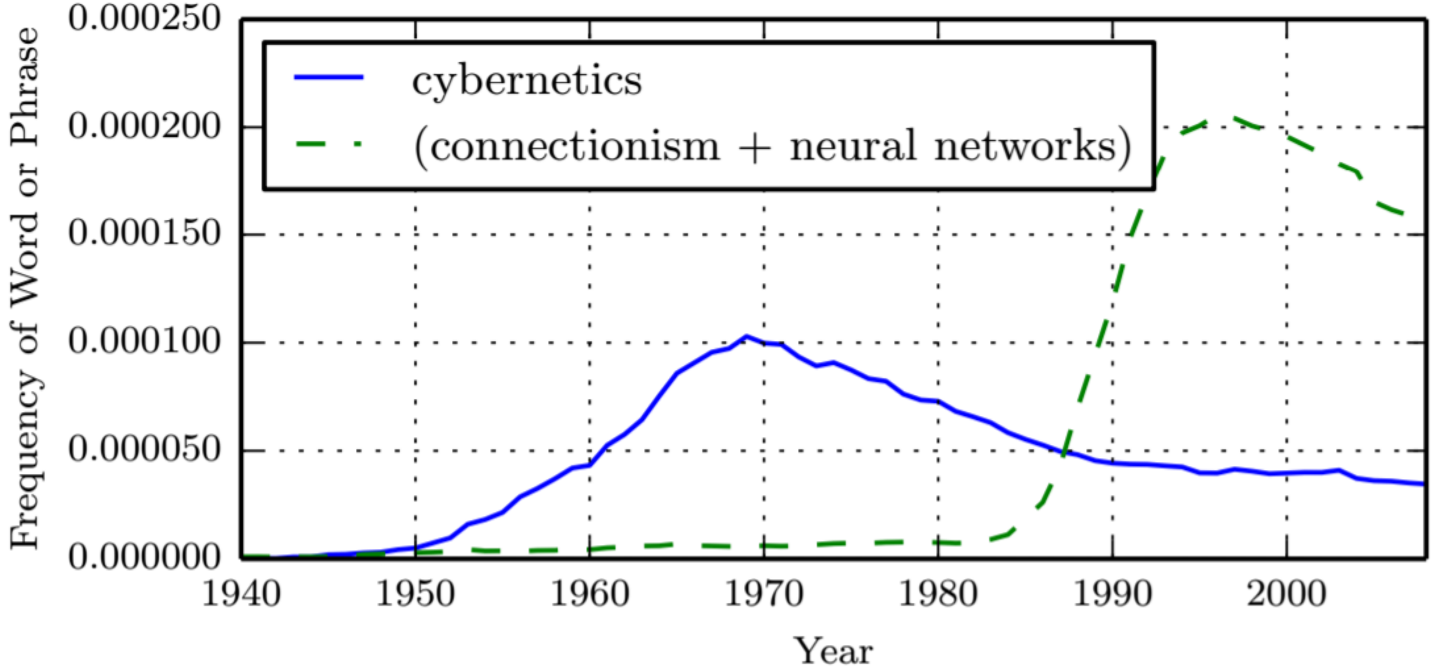
Which models are deep?

- 2-layer models No!
- Neural nets with 1 hidden layer No!
- SVMs and kernel methods No!
- Classification trees No!
- Graphical models Could be!

Deep Learning History

- Historical Waves
  - Cybernetics (1940s – 1960s)
  - Connectionism (1980s-1990s)
  - Deep learning (2006-)

Historical Waves

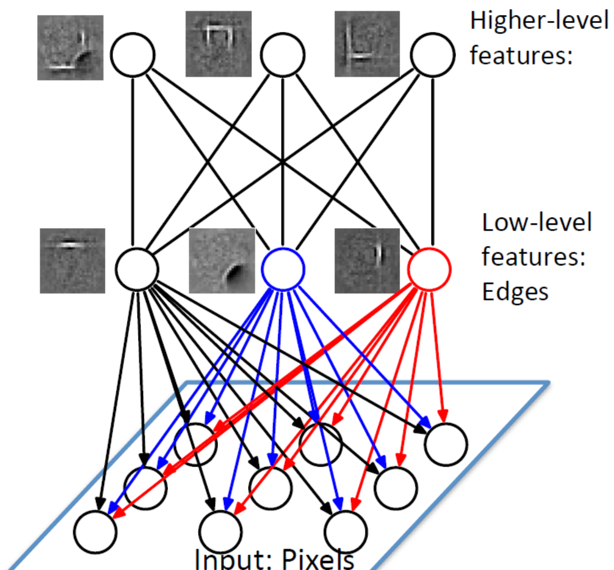


Ups and downs of Deep Learning

- 1960s: Perceptron (linear model, single layer nn)
- 1969: Perceptron has limitation
- 1980s: Multi-layer perceptron
  - Do not have significant difference from DNN today
- 1986: Backpropagation
  - Usually more than 3 hidden layers is not helpful
- 1989: 1 hidden layer is "good enough", why deep?
- 2006: RBM initialization (breakthrough)
- 2009: GPU
- 2011: Start to be popular in speech recognition
- 2012: win ILSVRC image competition
- 2015: "superhuman" results in Image and Speech Recognition

Important Breakthroughs

- Deep Belief Networks, 2006 (Unsupervised)
  - Hinton, G. E., Osindero, S. and Teh, Y., A fast learning algorithm for deep belief nets, Neural Computation, 2006.



Theoretical contributions:

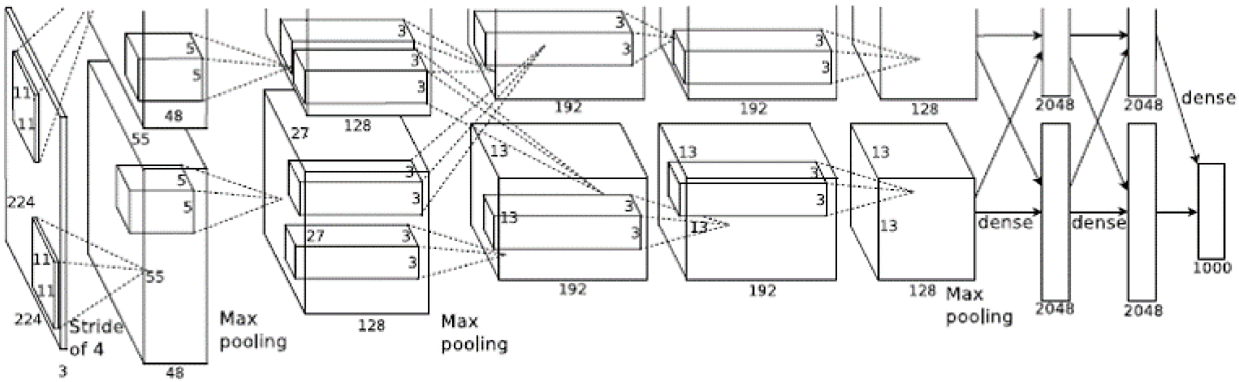
- Adding additional layers improves variation lower bound

Efficient Learning and Inference with multiple layers:

- Efficient greedy layer-by-layer learning algorithm
- Inferring the states of the hidden variables in the top most layer is easy

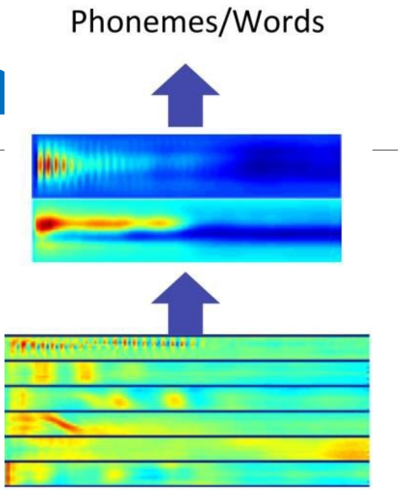
Important Breakthroughs

- Deep Convolutional Nets for Vision (Supervised)
  - Krizhevsky, A., Sutskever, I. and Hinton, G. E., ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012.



- Deep Nets for Speech (Supervised)
  - Hinton et. al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups, IEEE Signal Processing Magazine. 2012

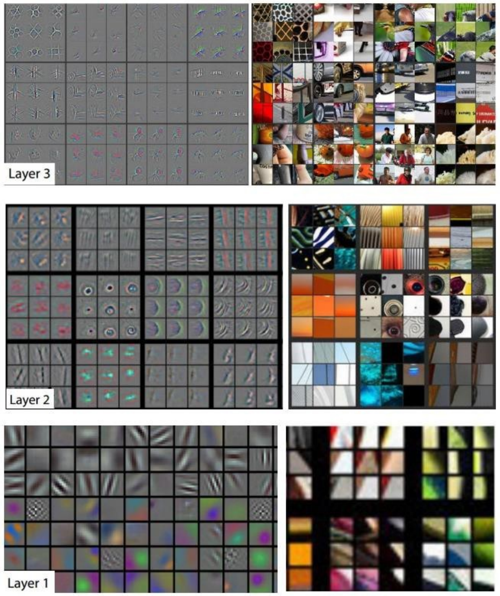
Deep Learning Breakthrough



First: Speech Recognition

Acoustic Model	WER on RT03S FSH	WER on Hub5 SWB
Traditional Features	27.4%	23.6%
Deep Learning	18.5% (-33%)	16.1% (-32%)

Second: Computer Vision



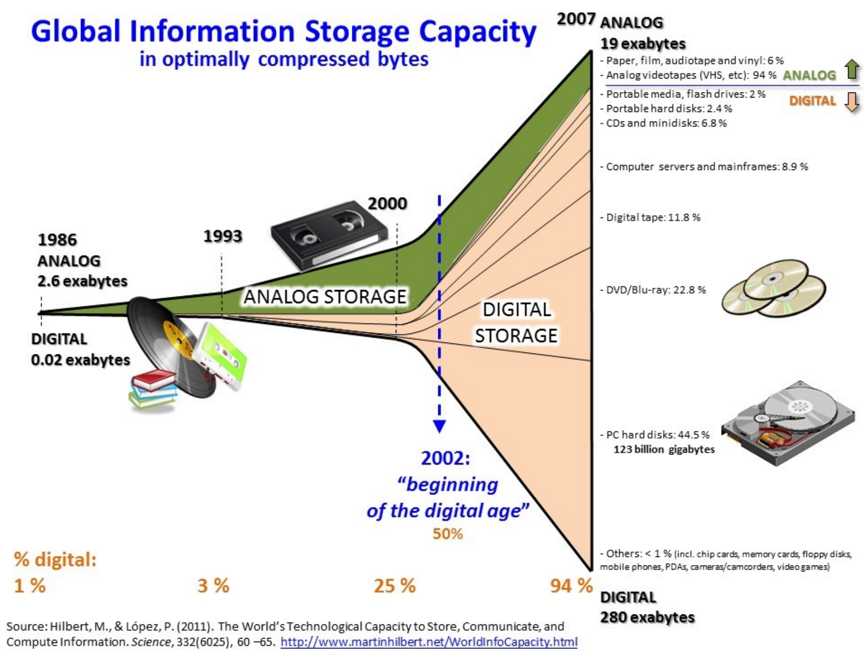
Ups and downs of Deep Learning

- 1960s: Perceptron (linear model, single layer nn)
- 1969: Perceptron has limitation
- 1980s: Multi-layer perceptron
  - Do not have significant difference from DNN today
- 1986: Backpropagation
  - Usually more than 3 hidden layers is not helpful
- 1989: 1 hidden layer is "good enough", why deep?
- 2006: RBM initialization (breakthrough)
- 2009: GPU
- 2011: Start to be popular in speech recognition
- 2012: win ILSVRC image competition
- 2015: "superhuman" results in Image and Speech Recognition

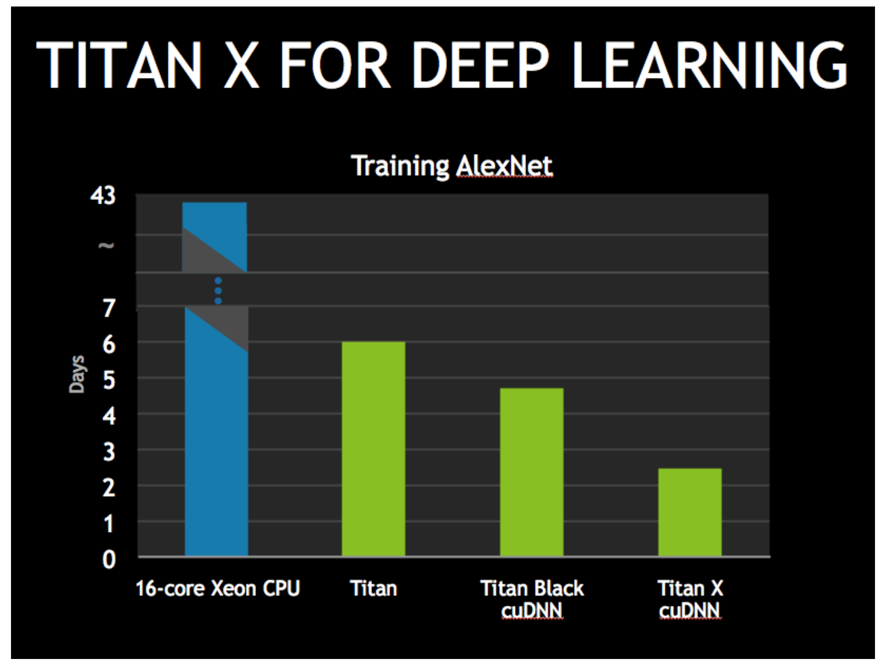
Why does deep learning show breakthrough in applications after 2010?

Reasons why Deep Learning works

Big Data



GPU



Reasons why Deep Learning works

Decade	Dataset	Memory	Floating Point Calculations per Second
1970	100 (Iris)	1 KB	100 KF (Intel 8080)
1980	1 K (House prices in Boston)	100 KB	1 MF (Intel 80186)
1990	10 K (optical character recognition)	10 MB	10 MF (Intel 80486)
2000	10 M (web pages)	100 MB	1 GF (Intel Core)
2010	10 G (advertising)	1 GB	1 TF (Nvidia C2050)
2020	1 T (social network)	100 GB	1 PF (Nvidia DGX-2)

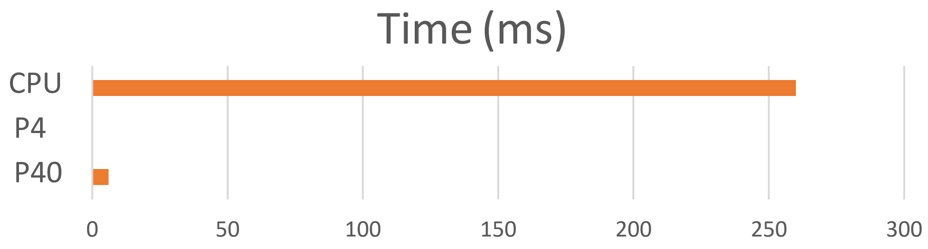
Why Speed Matters?

Training time

- Big data increases the training time
- Too long training time is not practical

Inference time

- Users are not patient to wait for the responses



GPU enables the real-world applications using the computational power

Recent Success with Deep Learning

- In many problems DL established the state of the art
  - Speech recognition
  - Translation
  - Image recognition
  - Caption generation
  - Signal enhancement
  - Chess
  - Recommender systems
  - …

Deep Learning vs. AI

- Deep learning is the fastest growing field in Artificial Intelligence (AI).

Deep Learning and the Employment Market

This guy didn't know about neural networks (a.k.a deep learning)

This guy learned about neural networks (a.k.a deep learning)

Source: http://deeplearning.cs.cmu.edu/

Deep Learning Trends

- Growing datasets
- Increasing model sizes
- Increasing accuracy, complexity and real-world impact
- New structures, DL theories

Do We Really Need Deep Architecture?

- Theoretician's dilemma
  - An NN with a single layer of enough hidden units can approximate any multivariate continuous function with arbitrary accuracy (Universality Theorem)
  - "Why would we need deep ones?"

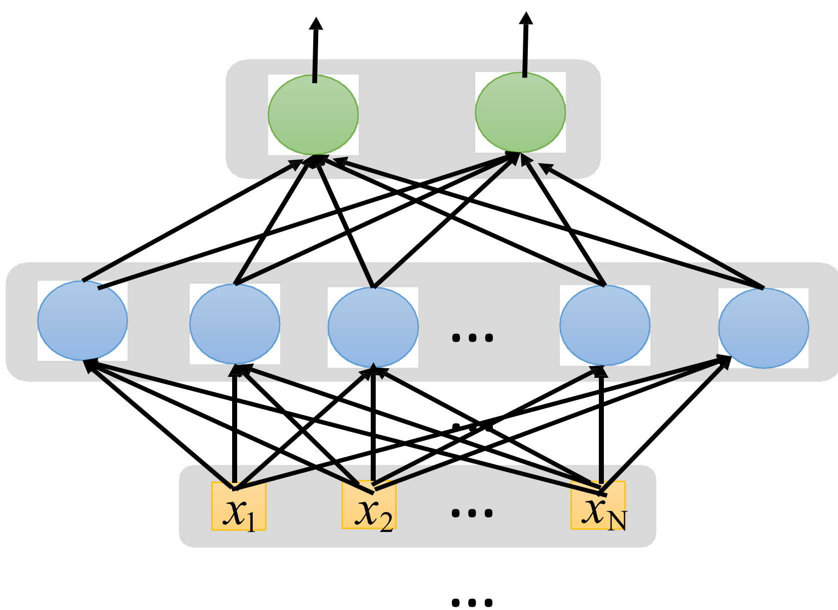
Reasons for Exploring Deep Learning

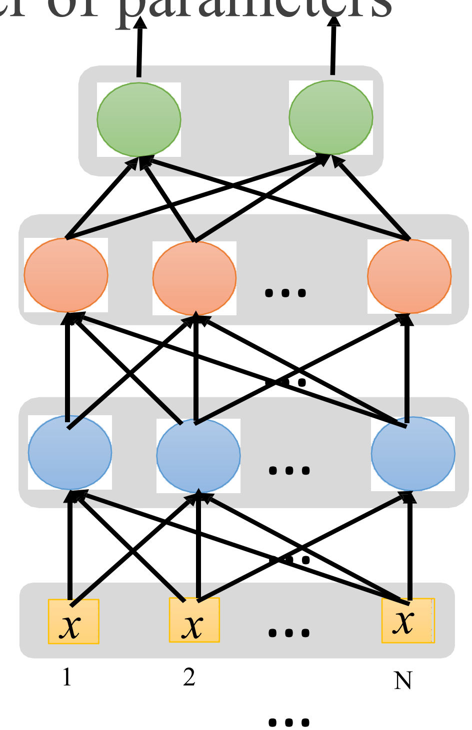
- Feature representation
  - Manually designed features are often over-specified, incomplete and take a long time to design and validate
  - **Learned features** are easy to adapt, fast to learn
- Deep machine are more efficient for representation certain classes of functions (AI tasks in vision, NLP, audition…)
  - Efficient parameterization
  - Trade space for time

Reasons for Exploring Deep Learning

- Effectiveness
  - Fat + Shallow v.s. Thin + Deep

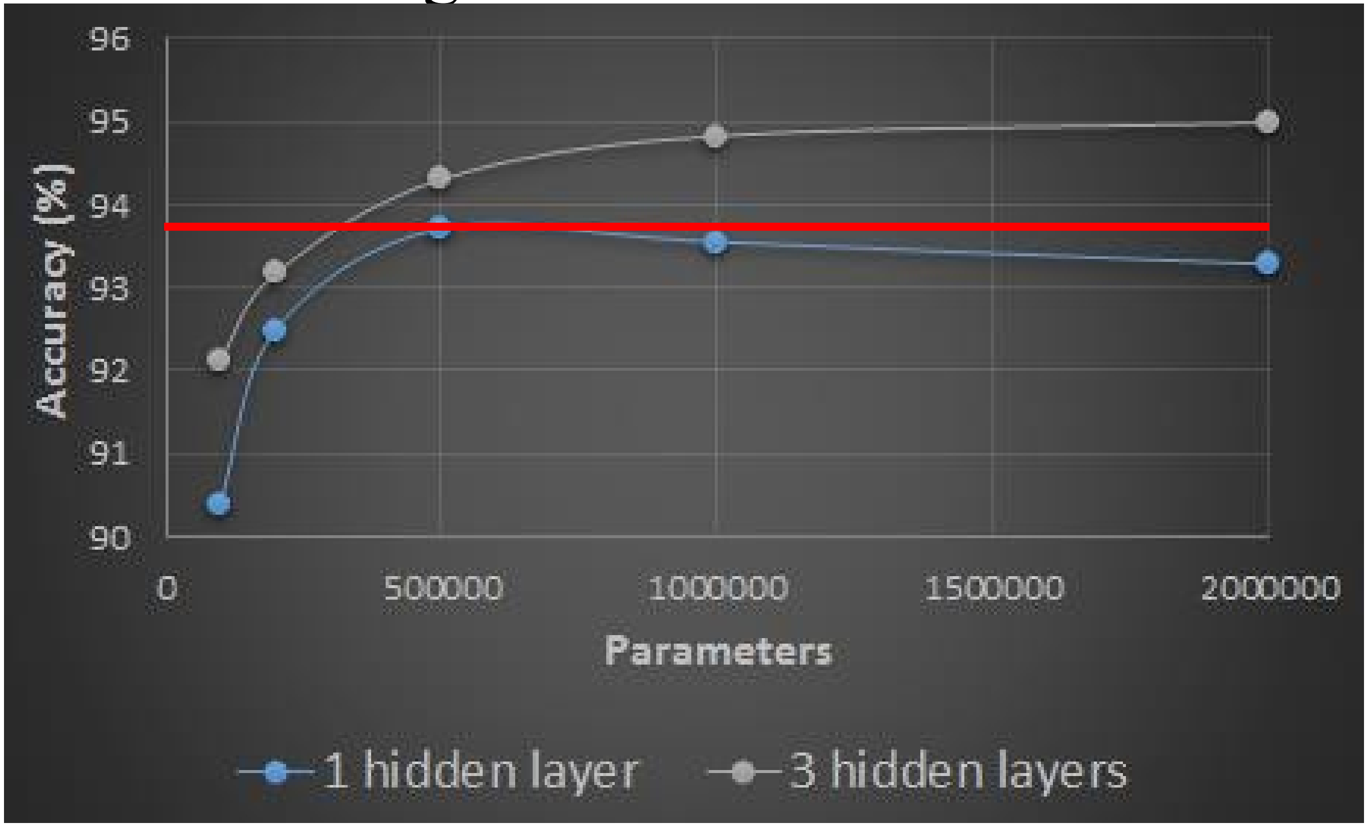
Two networks with the same number of parameters



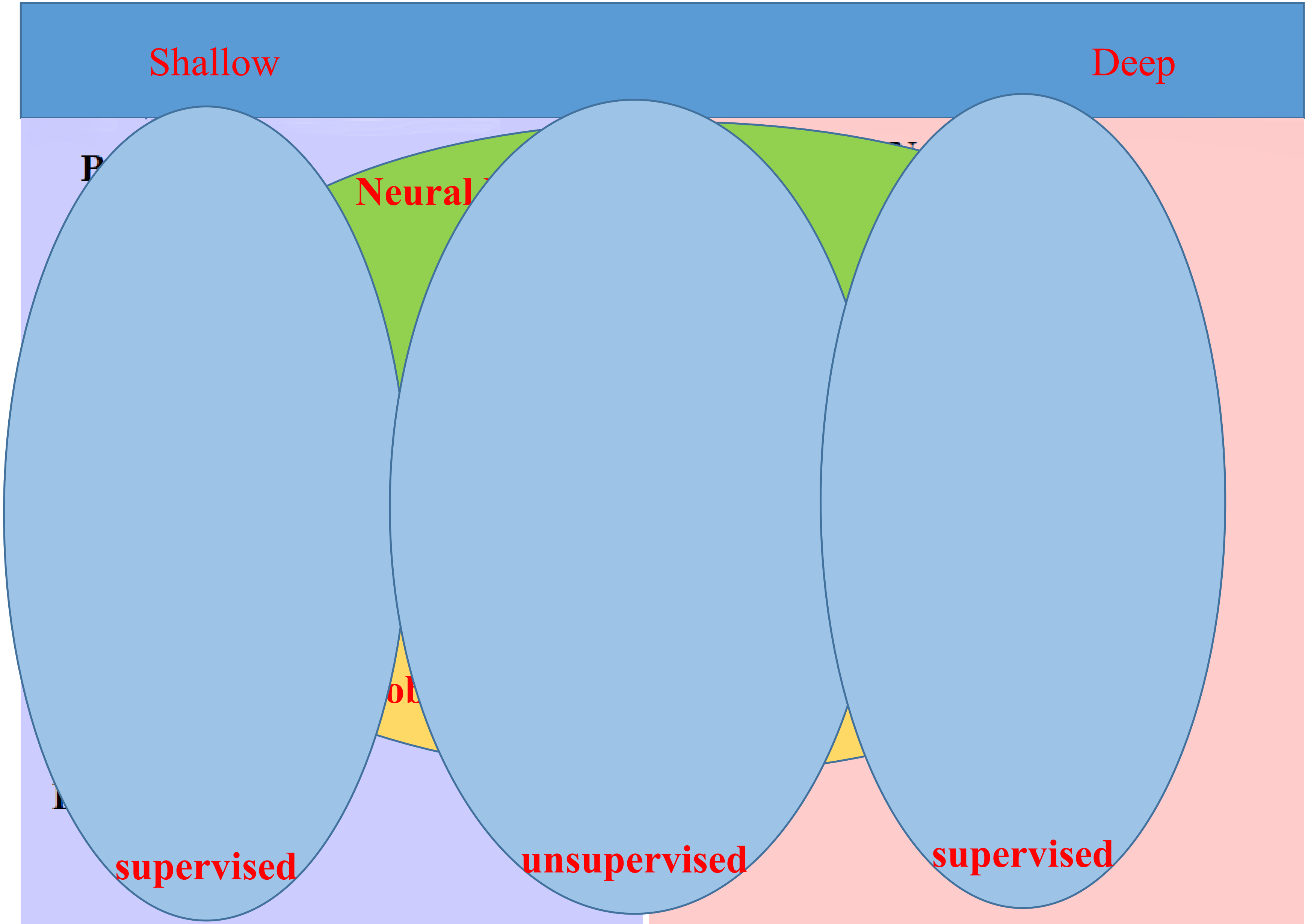


Reasons for Exploring Deep Learning

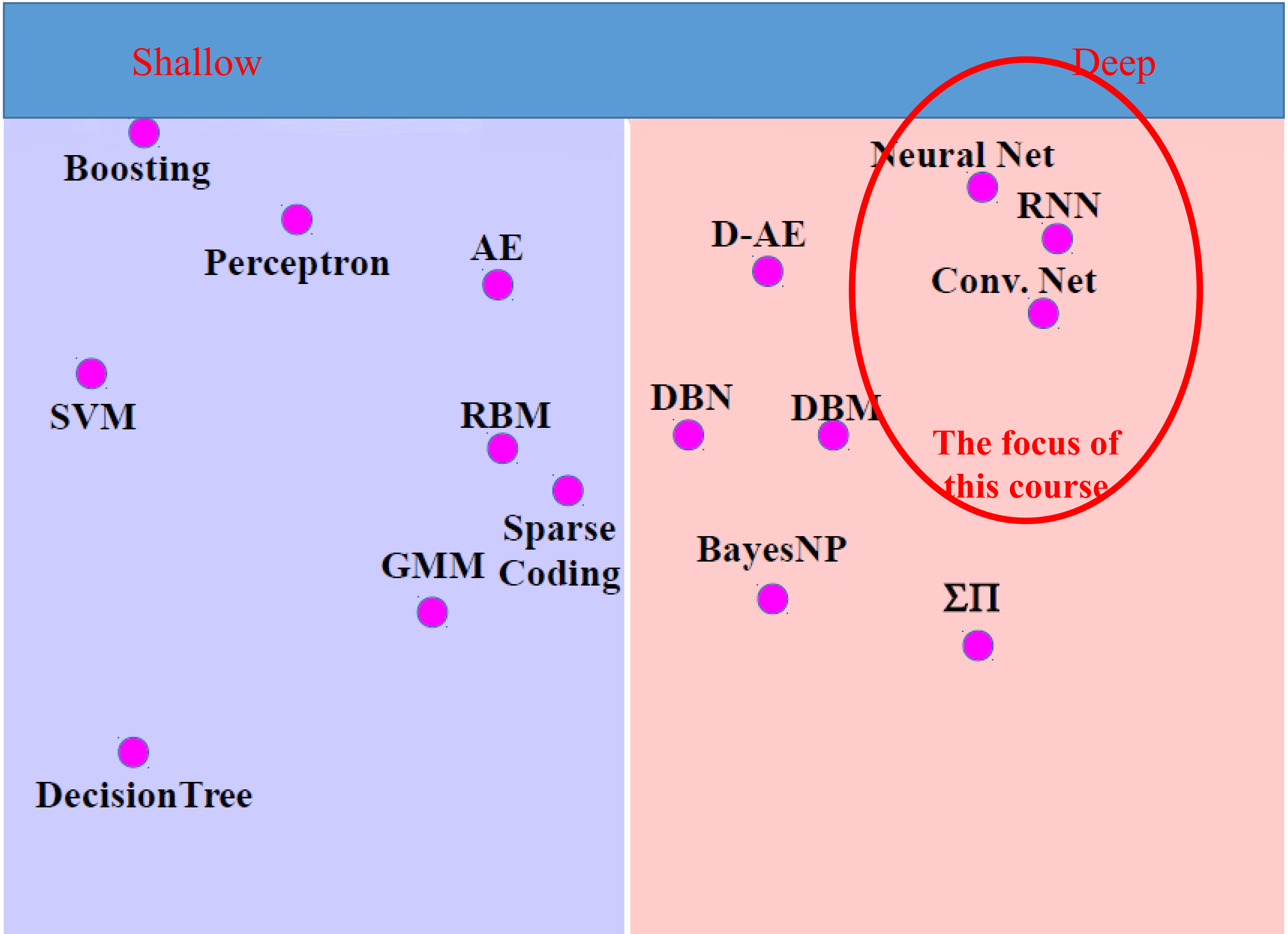
• Hand-Written Digit Classification



The deeper model uses less parameters to achieve the same performance



Source: Yann LeCun's deep learning course



Shallow
Deep
Boosting
Perceptron
AE
SVM
RBM
Sparse Coding
GMM
DecisionTree
Neural Net
RNN
D-AE
Conv. Net
DBN
DBM
The focus of this course
BayesNP
ΣΠ

Summary

- What is deep learning?
  - Deep learning algorithms attempt to learn multiple levels of representation and an output
- Deep learning history
  - not new, but it shows breakthrough in applications after 2010: big data and GPU
- Why deep learning?
  - v.s. shallow model, less engineering labor, but machine learns more
  - state-of-the-art in some areas
  - use less parameters to achieve the same performance

Reading Materials

- Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural networks* 61 (2015): 85-117.