

# Machine Learning

Cornelia Caragea

Computer Science and Engineering  
University of North Texas

Acknowledgement: Sridhar Mahadevan and Andrew Ng

June 9, 2016

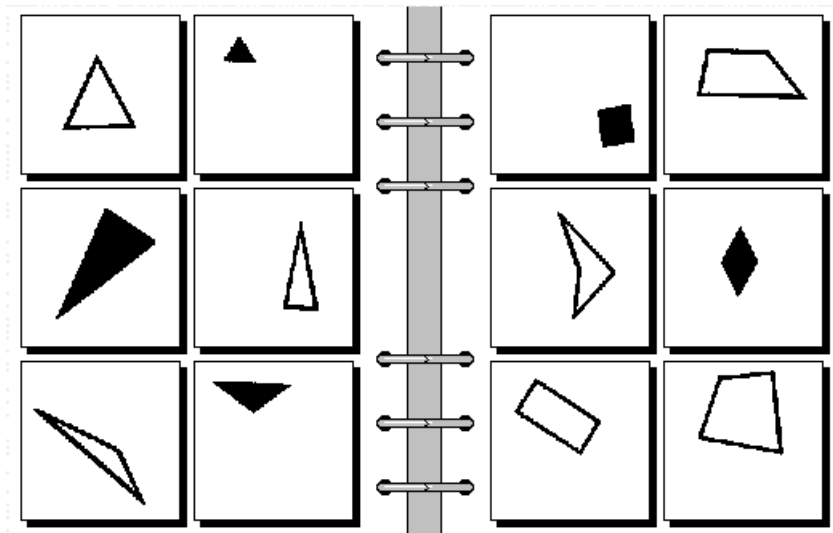
# Overview of Machine Learning

# What is "Learning"?

We "learn" many things:

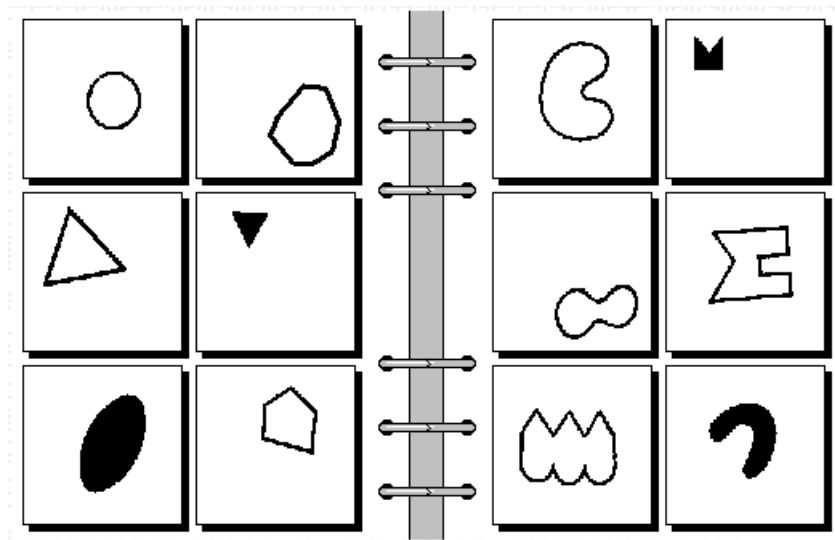
- **Motor skills:** walk, drive a bicycle, drive, play tennis or golf, play the piano.
- **Visual concepts:** man-made objects, faces, natural objects.
- **Language:** Speech recognition, read and write natural languages
- **Spatial knowledge:** Navigate between spatial locations, physical layout of a room.
- **Symbolic knowledge:** algebra, arithmetic, calculus.
- **Social rules:** how to interact with people, animals, machines....

# Bongard Problems





# Bongard Problems



# Bongard Problems

<http://www.foundalis.com/res/bps/bpidx.htm>

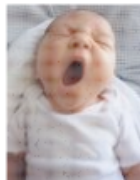
List of problems by incremental number

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Designers, & color-coding:	# of prob.
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40		
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	Mikhail M. Bongard	100
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	Douglas R. Hofstadter	56
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	Harry E. Foundalis	46
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	Joseph A. L. Insana	32
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	Peter Shanahan	5
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	Matthew J. Howells	5
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	Andreas Gunnarsson	16
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	Michael Ihde	1
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	Pablo Barenbaum	2
221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	Merse E. Gáspár	17
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260		
261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280		

# Object Recognition



# Name this activity



# The Challenge of Learning

- How is it possible that animals and humans are able to learn so much knowledge from a relatively small number of examples?
- Several possible explanations:
  - Most of what is learned is already built-in (The Blank Slate, Steve Pinker).
  - The brain is hardwired to learn specific classes of functions (e.g., language, faces, motor control).
  - Evolution has equipped the brain with some amazingly clever algorithms.
  - The brain is massively parallel ( $10^{12}$ ) neurons.

# Abstract Definition of "Learning"

Definition due to Arthur Samuel (1959):

*Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.*

Definition due to Herbert Simon (1980):

*"Learning" denotes changes in a system that are adaptive in that they enable the system to perform the same task or similar tasks drawn from the same population better over time.*

Definition due to Leslie Valiant (1986):

*"Learning" denotes knowledge acquisition in the absence of explicit programming.*

# Well-posed Learning Problem

Definition due to Tom Mitchell (1998):

*A computer program is said to learn from experience  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ .*

# Spam Filtering



Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task  $T$  in this setting?

- Classifying emails as spam or not spam
- Watching you label emails as spam or not spam
- The number (or fraction) of emails correctly classified as spam/not spam
- None of the above - this is not a machine learning problem



# Why Should Machines “Learn”?

- “Learning” can be viewed as a form of implicit programming.
- If the task changes over time, learning can make a machine adaptive.
- Learning may enable a machine to outperform human programming.

# Why Should Machines “Learn”?

- “Learning” can be viewed as a form of implicit programming.
- If the task changes over time, learning can make a machine adaptive.
- Learning may enable a machine to outperform human programming.
- We can now collect data on an unprecedented scale, but we need machine learning to make sense of the data!

# Why Study Machine Learning?

- “If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts”, Bill Gates quoted in NY Times, Monday March 3, 2004.

# Practical Applications of Machine Learning

- Spam filtering
- Speech/handwriting recognition
- Object detection/recognition
- Weather prediction
- Stock market analysis
- Search engines (e.g., Google)
- Ad placement on websites
- Adaptive website design
- Credit-card fraud detection
- Webpage clustering (e.g., Google News)
- Social Network Analysis
- Machine Translation (e.g., Google Translate)
- Recommendation systems (e.g., Netflix, Amazon)
- Classifying DNA sequences
- Automatic vehicle navigation
- Performance tuning of computer systems
- Predicting good compilation flags for programs
- ... and many more

# IBM Jeopardy! Quiz Program

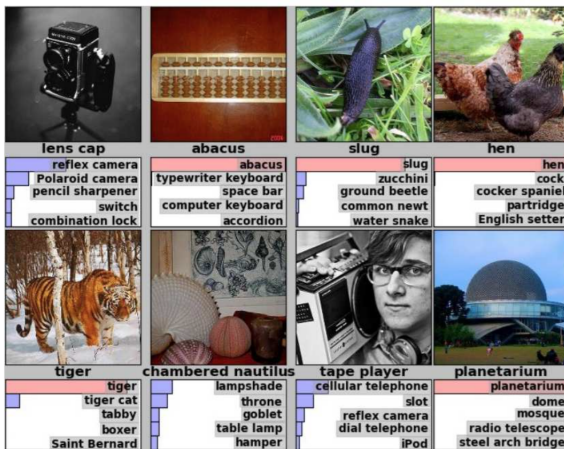


# Speech Recognition on Smart Phones

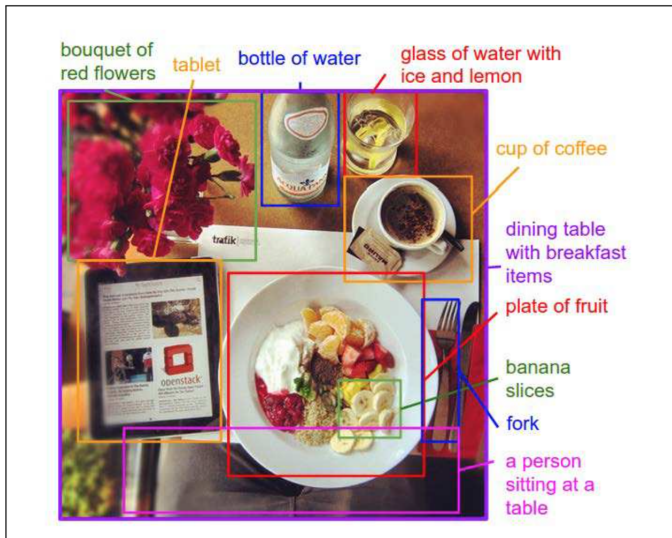


# ImageNet Vision Challenge

## Validation classification

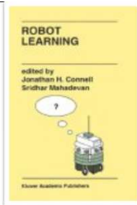


# Mapping Images to Text



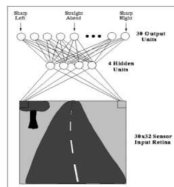


# Autonomous Driving



ALVINN learns  
from a human driver

## Learning to Drive



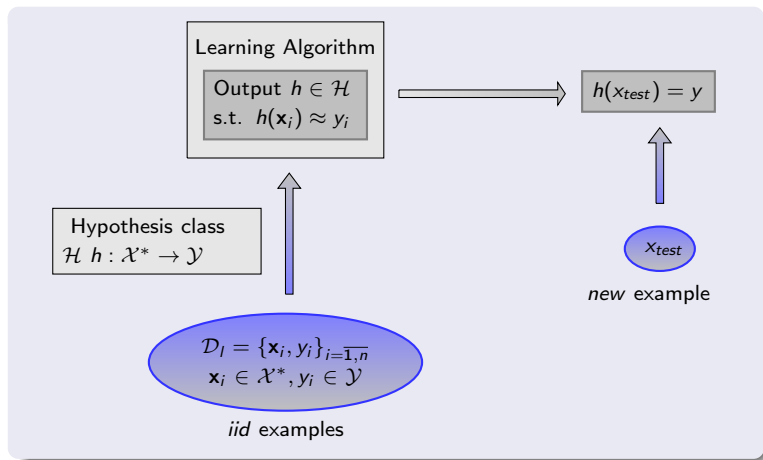
Neural  
Network

Can drive on actual highways at 65  
miles per hour!

# Three Fundamental Problems of Learning

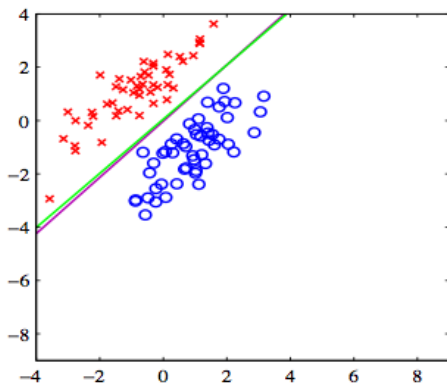
- **Classification:** Learning to predict discrete labels associated with given observations.
  - Binary classification: positive vs. negative examples
  - Multiclass classification: digit recognition
- **Regression:** Learning to predict continuous outputs associated with given observations
  - Example: how long does it take to bike to Northampton? How much does it cost to visit Florida? How much money can I make if do a PhD in CS?
- **Unsupervised learning:** Learning to group objects into categories, without any training labels.
  - Examples: density estimation, clustering

# Supervised Framework



Learning = Search in Hypothesis Class

# Linearly Separable Classification



- Spam vs. not spam
- Breast cancer (malignant, benign)

# Classification

- How would you write a program to distinguish a picture of **me** from a picture of **someone else**?
  - Provide examples pictures of **me** and pictures of **other people** and let a classifier learn to distinguish the two.
- How would you write a program to determine whether a sentence is **grammatical** or **not**?
  - Provide examples of **grammatical** and **ungrammatical** sentences and let a classifier learn to distinguish the two.
- How would you write a program to distinguish **cancerous** cells from **normal** cells?
  - Provide examples of **cancerous** and **normal** cells and let a classifier learn to distinguish the two.

# Data ("weather" prediction)

## Example dataset:

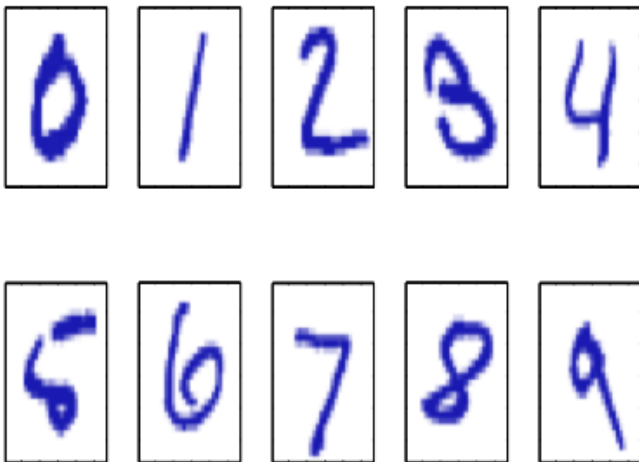
Class	Outlook	Temperature	Windy?
Play	Sunny	Low	Yes
No play	Sunny	High	Yes
No play	Sunny	High	No
Play	Overcast	Low	Yes
Play	Overcast	High	No
Play	Overcast	Low	No
No play	Rainy	Low	Yes
Play	Rainy	Low	No

## Three principle components:

1. Class label (aka "label", denoted  $y$ )
2. Features (aka "attributes")
3. Feature values (aka "attribute values", denoted  $x$ )  
 ⇒ Features can be binary, nominal or continuous

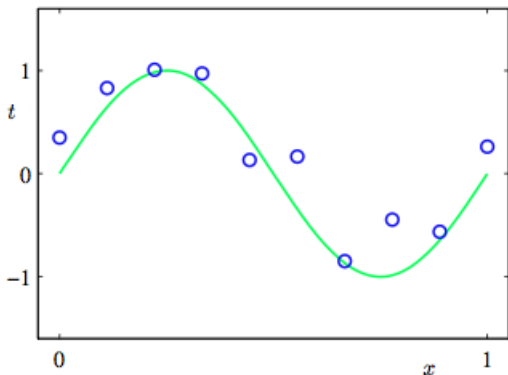
**A labeled dataset is a collection of  $(x, y)$  pairs**

# Digit Recognition



# Regression

Plot of a training data set of  $N = 10$  points, shown as blue circles, each comprising an observation of the input variable  $x$  along with the corresponding target variable  $t$ . The green curve shows the function  $\sin(2\pi x)$  used to generate the data. Our goal is to predict the value of  $t$  for some new value of  $x$ , without knowledge of the green curve.





## Question

You are running a company, and you want to develop learning algorithms to address each of two problems.

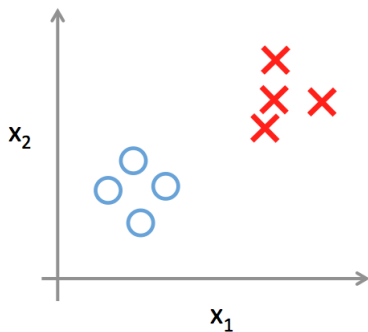
P1: You have a large inventory of identical items. Want to predict how many of them will sell over the next 3 months.

P2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked.

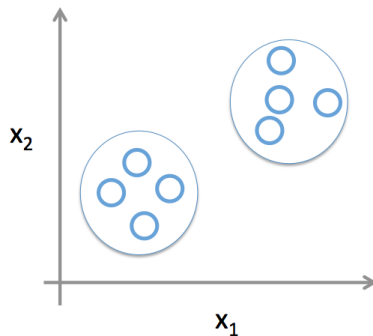
### Classification or regression problems?

- Treat both as classification problems.
- Treat P1 as a classification problem, P2 as a regression problem.
- Treat P1 as a regression problem, P2 as a classification problem.
- Treat both as regression problems.

# Unsupervised Learning



Supervised learning



Unsupervised learning

# Clustering News Articles

Google News

http://news.google.com/

CS 6780: Adv... (Projects) SHB ACM C... d Wellbeing Eminescu... FAREWELL YouTube ~ Lo... Channel Bernoulli di... encyclopedia Linear Algebra weka ~ Eclip... aka-src.jar) Romantic FM... line Player

Top Stories

- Mitt Romney
- Tropical cyclone
- Gabby Douglas
- Revolutionary Armed Forces of Colombia
- Driving under the influence
- Samsung Group
- Syria
- Sea ice
- Taliban
- Baltimore
- Dallas-Fort Worth Me...
- World
- U.S.
- Business
- Elections
- Technology
- Entertainment
- Sports
- Science
- Health
- Spotlight

Business »

**Japan Cuts Economic Assessment as BNP Says Contraction Looms**

Bloomberg - 15 minutes ago  
Japan's government downgraded its assessment of the world's third-biggest economy for the first time in 10 months as some analysts forecast that gross domestic product will shrink this quarter.

**Best Buy allows Schulze to pursue takeover bid**

CNET - 22 minutes ago  
Board allows founder and former CEO of the struggling electronics giant to conduct due diligence and form an investment group in takeover effort.

Written by Steven Must

More Business stories

Elections »

**GOP Penn. Senate candidate calls rape, unmarried pregnancy 'similar'**

New York Daily News - 1 hour ago  
Pennsylvania Senate hopeful Tom Smith sparked controversy Monday after he compared a pregnancy resulting from rape to "having a baby out-of-wedlock" - days after Rep. Todd Akin (R-Mo.

**Romney's tax secrecy is abuse of voters**

Syracuse.com (blog) - 17 hours ago  
By Readers' Page The main issue in this election is not the character of either Gov. Mitt Romney or President Barack Obama. It's the character of the American voter.

USA TODAY

More Elections stories

Technology »

**iPad Mini To Debut After iPhone 5**

InformationWeek - 4 hours ago  
Apple will reveal a smaller iPad at an event separate from the next-generation iPhone announcement. This is contrary to previous rumors and reports, which indicated that the new tablet and phone would be unveiled at the same event.

**Hardware Guru Bob Mansfield Sticking Around at Apple**

PC Magazine - 51 minutes ago  
By Damon Poeter Apple hardware guru Bob Mansfield has pulled a Michael Jordan, reversing his decision in June to retire from the company and instead will stay on to "work on future products," Apple announced Monday.

Written by Damon Poeter

**Connecting the Dots After Cyberattack on Saudi Aramco**

New York Times (blog) - 1 hour ago  
By NICOLE PERLROTH Publicly released details of a cyberattack on Saudi Aramco, the world's largest oil producer, appear to confirm reports that critical data on three-quarters of the company's PCs was replaced with the image of a

Artist Shmuel Shiva's Incredible Photorealistic Ballpoint Pen Drawings (PHOTOS)

Huffington Post

Benefits of Circumcision Are Said to Outweigh Risks

New York Times

Spotlight Video

PTI: Who Got The Better Deal?

ESPN - 1 hour ago  
Watch video

NFL Live OT: High Expectations Facing The 49ers

ESPN - 4 hours ago  
Watch video

NTV Women and power health

nivkenya - 5 hours ago  
Watch video

FRANCE 24 Tech 24 - 08/27/2012 TECH 24

France 24 - 9 hours ago  
Watch video

New Orleans Braces for Isaac

The Associated Press - 5 hours ago  
Watch video

Most popular

Gabby Douglas to Oprah: I was 'bullied,' called 'slave' during early gymnastics ...

New York Daily News - 2 hours ago

## Question

Of the following examples, which would you address using an unsupervised learning algorithm?

- Given email labeled as spam/not spam, learn a spam filter.
- Given a set of news articles found on the web, group them into set of articles about the same story.
- Given a database of customer data, automatically discover market segments and group customers into different market segments.
- Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

# Summary

- Machine learning is an exciting field of research with limitless practical application, as well as a deep scientific and intellectual challenge.
  - Broad applicability
    - Finance, robotics, vision, machine translation, medicine, etc.
  - Close connection between theory and practice
  - Open field, lots of room for new work
  - 12 IT skills that employers can't say no to (Machine Learning is #1)  
[http://www.computerworld.com/s/article/9026623/12\\_IT\\_skills\\_that\\_employers\\_can\\_t\\_say\\_no\\_to\\_](http://www.computerworld.com/s/article/9026623/12_IT_skills_that_employers_can_t_say_no_to_)
- "The beauty of machine learning? It never stops learning!"
  - <http://gigaom.com/2012/03/21/machine-learning-structure-data-2012/>