CS 60050
Machine Learning

Introduction

Some slides taken from course materials of Andrew Ng and P. Domingos

Course Logistics

- Instructor: Saptarshi Ghosh
  - Email: saptarshi@cse.iitkgp.ac.in
  - Office: CSE Department, room 207
- Class hours (attendance is compulsory)
  - Wednesday 11:00 – 11:55
  - Thursday 12:00 – 12:55
  - Friday 08:00 – 08:55
- Course Website:

http://cse.iitkgp.ac.in/~saptarshi/courses/ml2020s/

Evaluation

- Mid-semester exam (20%)
- End-semester exam (40%)
- Assignments (40%)
  - 4-5 in number
  - Programming in C / C++ / Java / Python

Text & Reference Materials

- R. Duda, et al., **Pattern Classification**, Wiley
- T. Mitchell, **Machine Learning**, McGraw-Hill
- C. Bishop, **Pattern Recognition and Machine Learning**, Springer
- Tan, Steinbach, Kumar, **Introduction to Data Mining**, Pearson

- Many resources available freely on Web

A Few Quotes

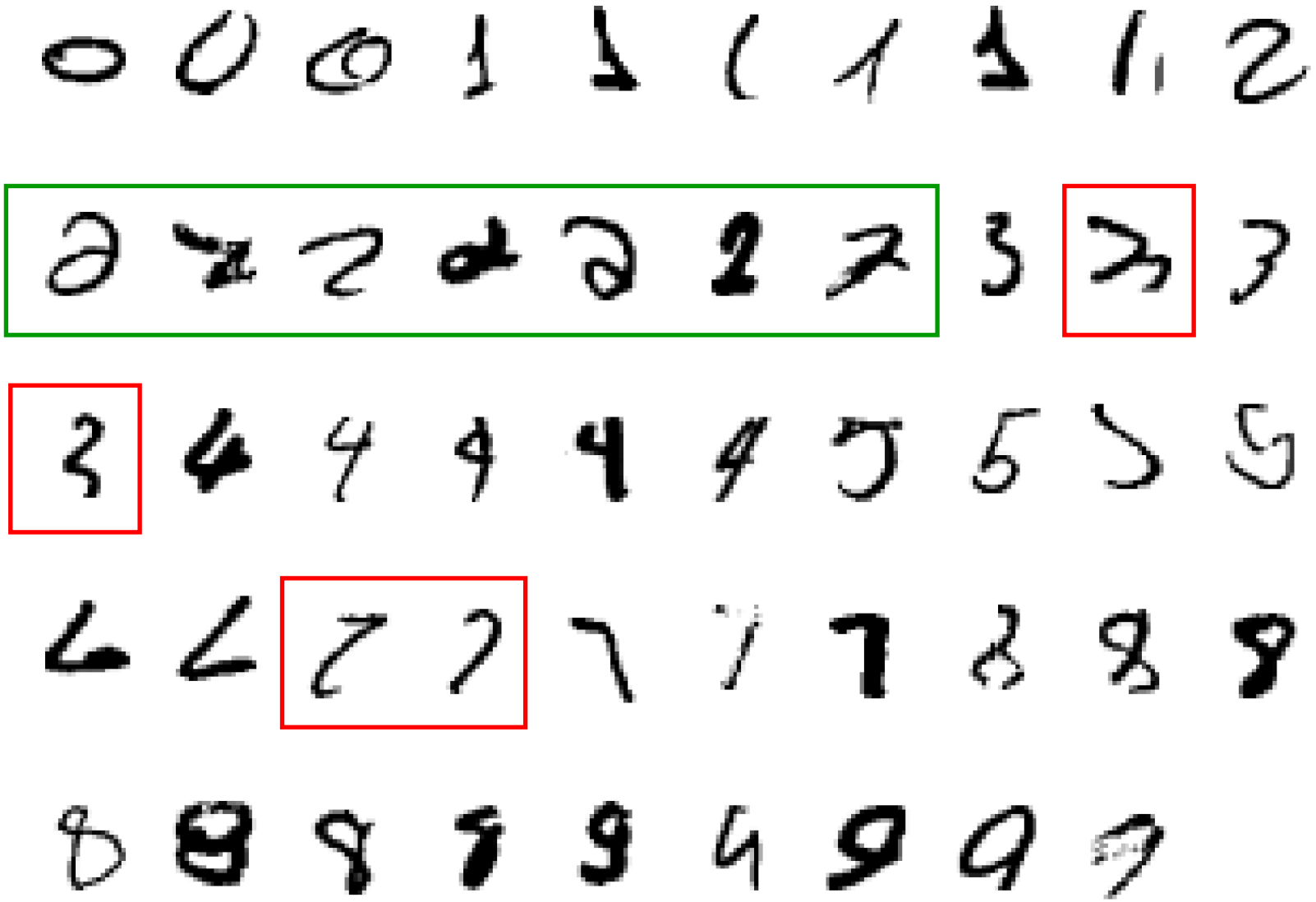
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet”
  (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing”
  (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)

Motivation for ML

- It is hard to write programs for certain tasks
  - Human face or handwriting recognition
  - Playing complex games like chess
  - Recommending movies that a person will like

- Why?
  - We do not ourselves know how to solve
  - Algorithm will be too complicated
  - Too many instances of the program needed (e.g., one for every user)

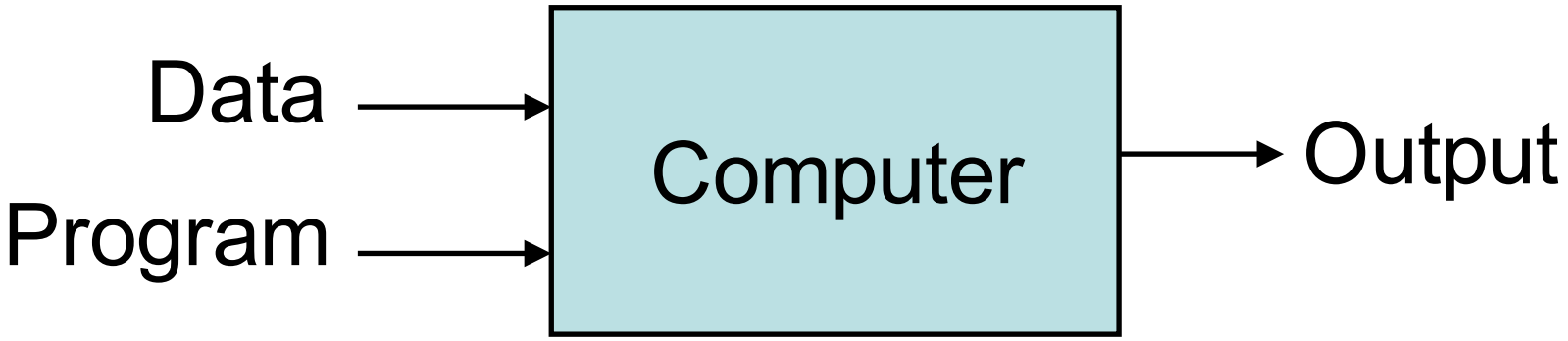
A classic example of a task that requires machine learning: It is very hard to say what makes a 2



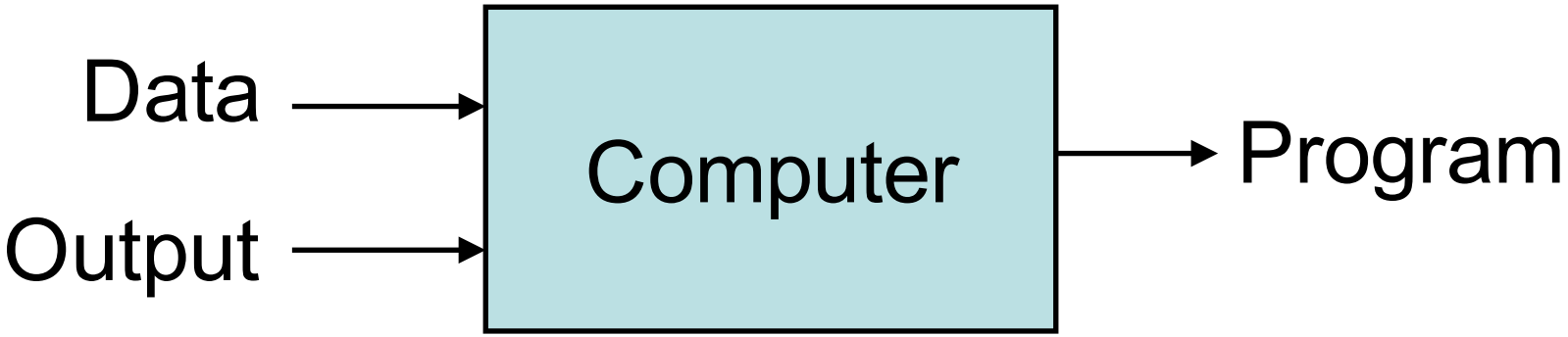
Motivation for ML

- Instead of writing a program by hand, collect lots of examples that specify the correct output for a given input.
- A machine learning algorithm takes these examples and produces a program that does the job.
- If done right, the program works for new cases as well as the ones we trained it on.

Traditional Programming



Machine Learning



Human learning vs. ML

- Humans prefer to learn how to design the algorithm / program
- But for some very complex tasks, humans also learn from example inputs and outputs

Human learning vs. ML

- Some aspects common to both
- Desired: Generalization to new or unknown examples
  - E.g., teach a child to look out for vehicles while crossing a road
- Challenge: Learning a wrong model due to specific training examples used
  - E.g., teach a child to distinguish between ”safe” (cow, cat) and “dangerous” animals (lion, tiger)

Machine Learning definition

- Arthur Samuel (1959):

Field of study that gives computers the ability to learn without being explicitly programmed.

- Tom Mitchell (1998):

Well-posed Learning Problem: 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.

Some more examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns
  - Generating images or motion sequences
- Recognizing anomalies
  - Unusual sequences of credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant or unusual sound in your car engine.
- Prediction
  - Future stock prices or currency exchange rates

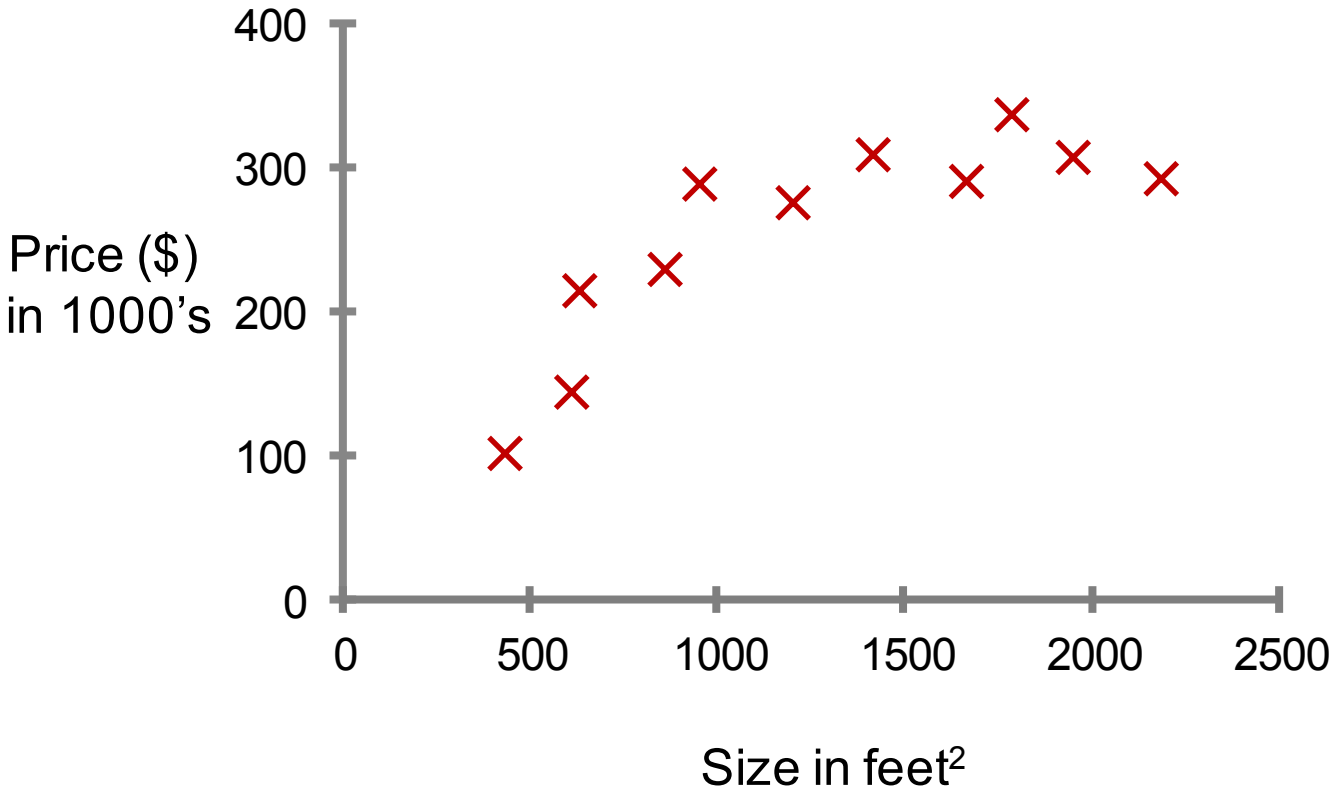
Some web-based examples of machine learning

- The web contains a lot of data. Tasks with very big datasets often use machine learning
- Spam filtering, fraud detection
  - The enemy adapts so we must adapt too.
- Recommendation systems
  - Need to adapt to millions of individuals. E.g., Youtube video, Netflix movies recommendation
- Information retrieval
  - Find documents or images with similar content.
- Data Visualization
  - Display a huge database in a revealing way

Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired / correct outputs or labels
- **Unsupervised learning**
  - Training data does not include desired outputs
- **Semi-supervised learning** (various forms)
  - Training data includes a few desired outputs
  - Training data has desired outputs, but for a different (related) task
- **Reinforcement learning**
  - Rewards from sequence of actions

Housing price prediction.

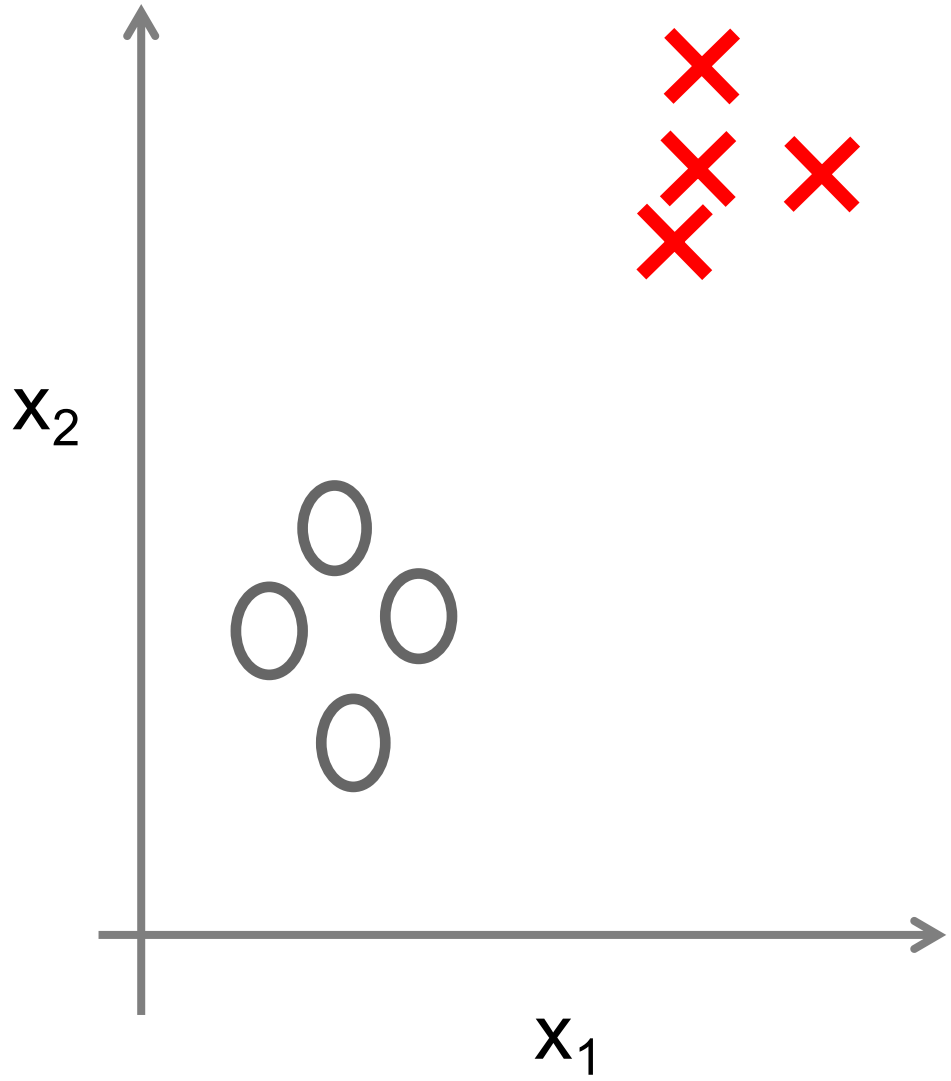


Supervised Learning

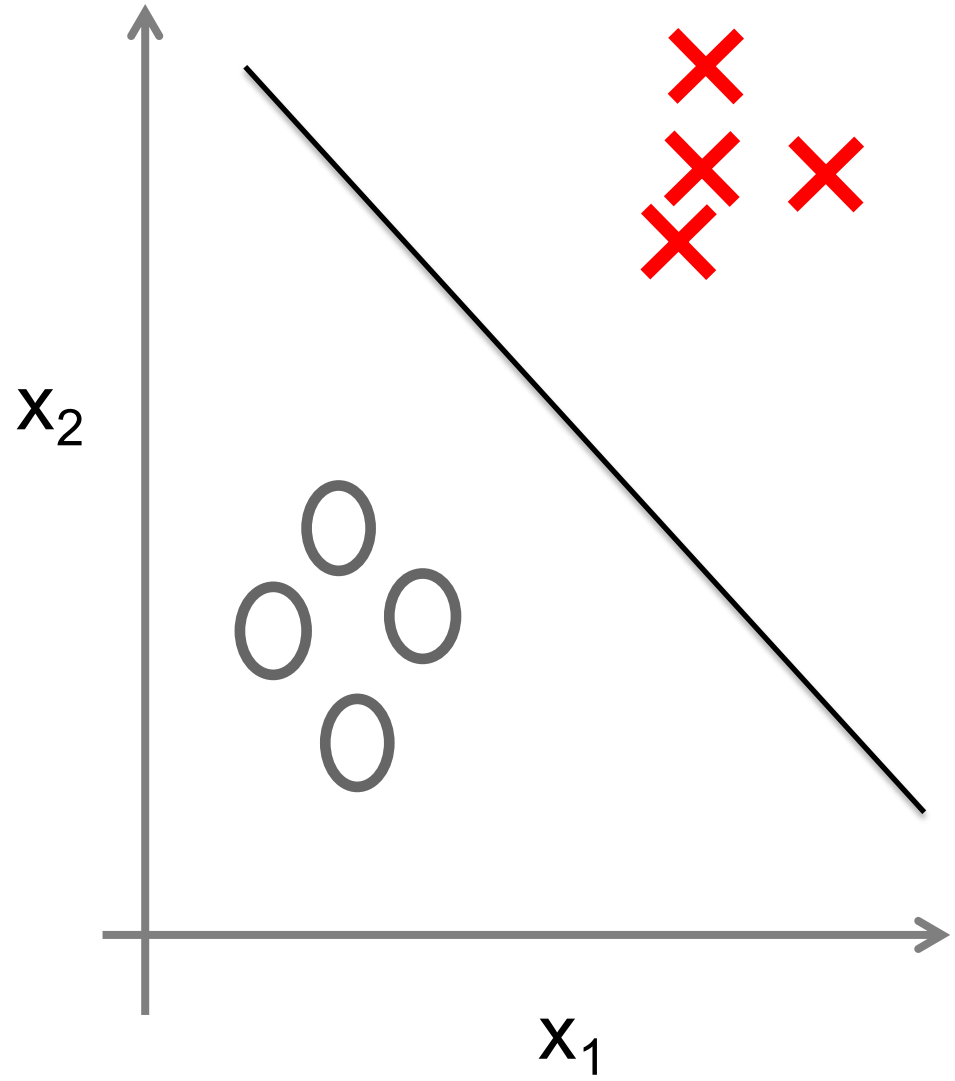
"right answers" given

Regression: Predict continuous valued output (price)

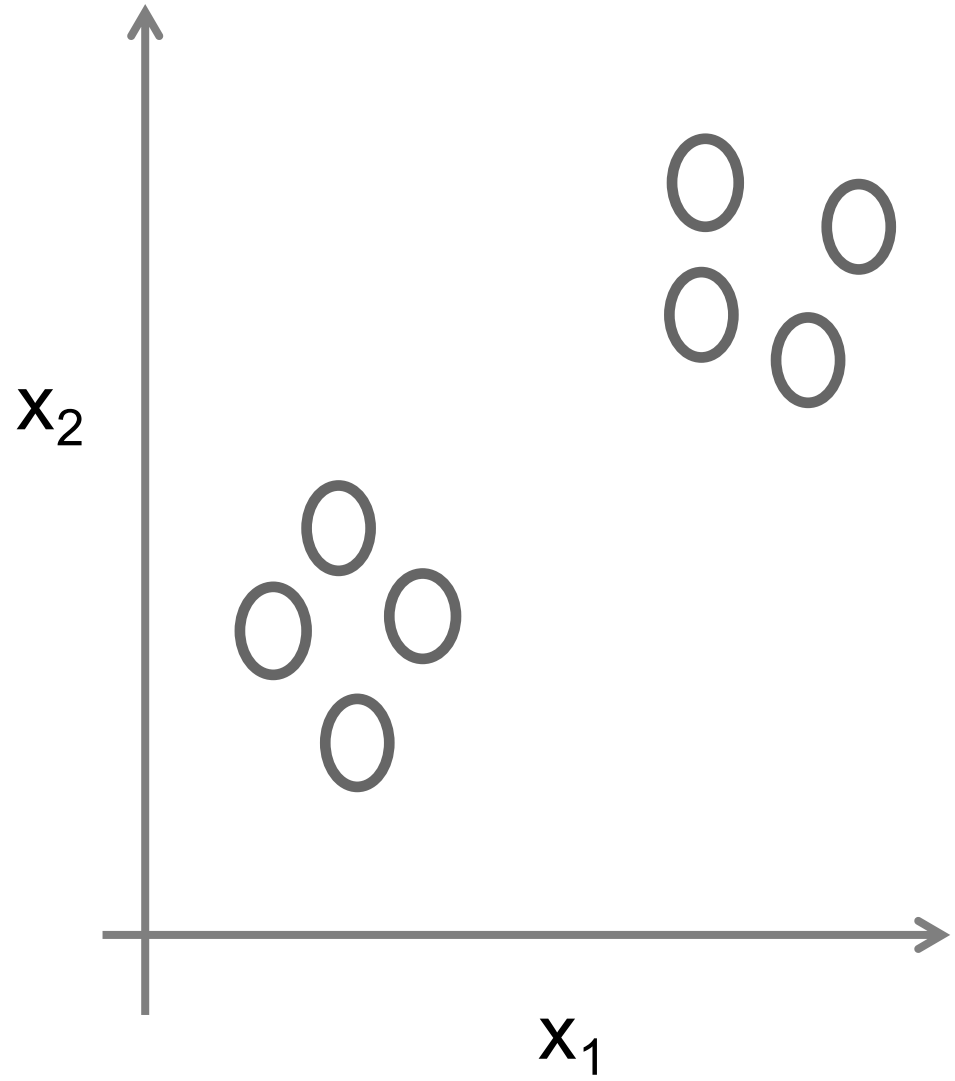
Supervised Learning: Classification



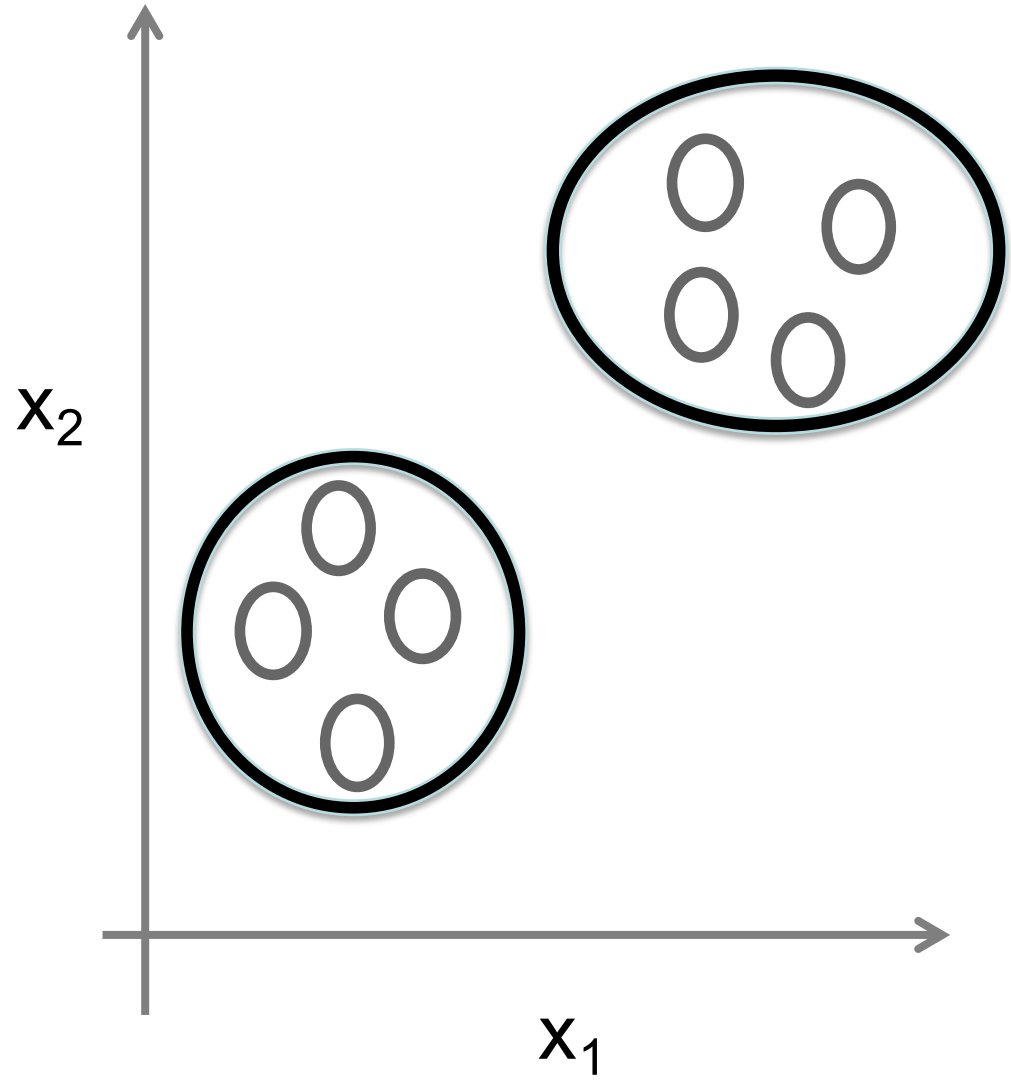
Supervised Learning: Classification



Unsupervised Learning: Clustering



Unsupervised Learning: Clustering



Google News

news.google.com

Web Images Videos Maps **News** Shopping Gmail more ▼ andrewyantakng@gmail.com | Web History | Settings ▼ | Sign out

Google news

Search News Search the Web

Advanced news search

U.S. edition ▼ Add a section »

Top Stories

- Deepwater Horizon
- Fed meeting
- Foreign exchange market
- Lindsay Lohan
- IBM
- Tom Brady
- Toronto International Film Festival
- Paris Hilton
- Iran
- Hurricane Igor

- Starred
- San Francisco Bay Area
- World
- U.S.
- Business
- Sci/Tech
- More Top Stories
- Spotlight
- Health
- Sports
- Entertainment

All news

- Headlines
- Images

Top Stories

Christine O'Donnell »

White House official denies Tea Party-focused ad campaign

CNN International - Ed Henry - 1 hour ago

Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) -- A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...

Tea Party is misplacing the blame, former President Bill Clinton claims New York Daily News

GOP tea party backer defends Christine O'Donnell The Associated Press

Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon

all 726 news articles »

CNN Interna...

US Stocks Climb After Recession Called Over, Homebuilders Gain

MarketWatch - Kristina Peterson - 16 minutes ago

NEW YORK (MarketWatch) -- US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.

Longest recession since 1930s ended in June 2009, group says Los Angeles Times

Downturn Was Longest in Decades, Panel Confirms New York Times

Wall Street Journal - AFP - CNN - USA Today

all 276 news articles »

MyFox Phila...

Deepwater Horizon »

BP Oil Well, Site of National Catastrophe, Dies at One

Vanity Fair - Juli Weiner - 22 minutes ago

The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.

Video: Blown-out BP Well Finally Killed in Gulf The Associated Press

Weiss Doubts BP Would End Operations in Gulf of Mexico: Video Bloomberg

CNN International - Wall Street Journal (blog) - The Guardian - New York Times

all 2,292 news articles »

Reuters

Recent

Recession officially ended in June 2009
CNNMoney - Chris Isidore - 39 minutes ago

Hurricane Igor lashes Bermuda
USA Today - Gerry Broome - 5 minutes ago

'Explain what you want from us,' reads front-page editorial
msnbc.com - Olivia Torres - 10 minutes ago

Crisis response: Pakistan floods

San Francisco Bay Area - Edit

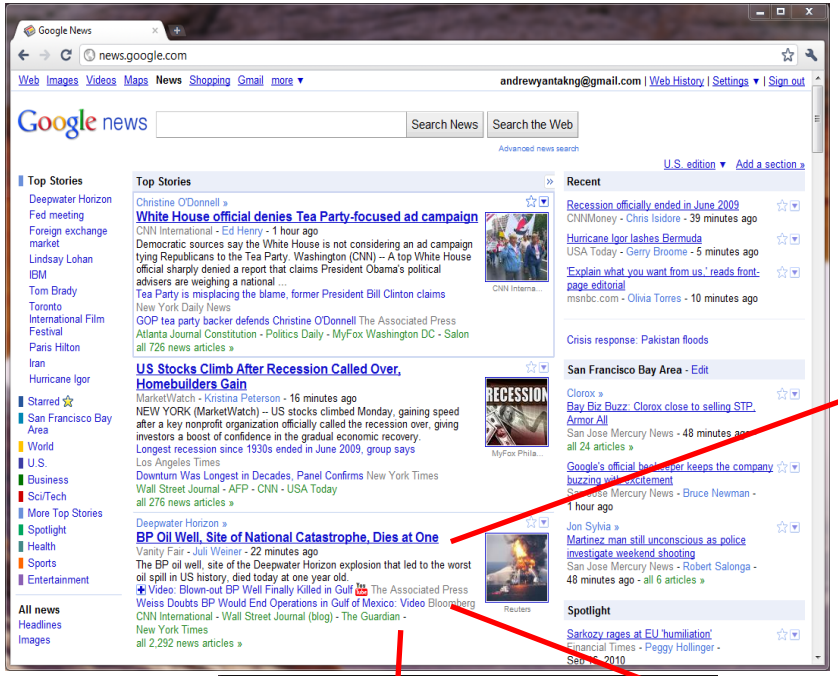
Clorox »
Bay Biz Buzz: Clorox close to selling STP, Armor All
San Jose Mercury News - 48 minutes ago - all 24 articles »

Google's official beekeeper keeps the company buzzing with excitement
San Jose Mercury News - Bruce Newman - 1 hour ago

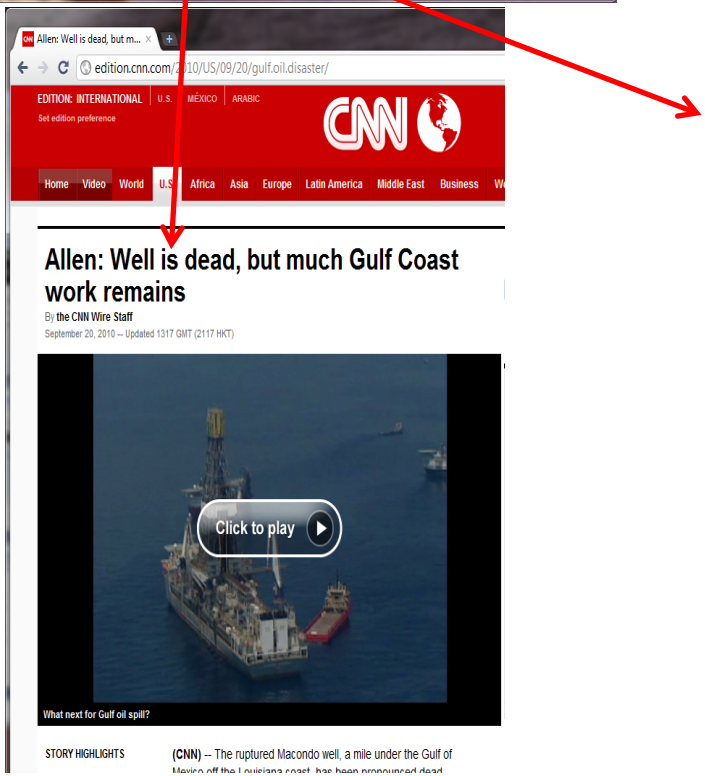
Jon Sylvia »
Martinez man still unconscious as police investigate weekend shooting
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Spotlight

Sarkozy rages at EU 'humiliation'
Financial Times - Peggy Hollinger - Sep 16, 2010


Google News
news.google.com
Google news
Search News
Search the Web
Top Stories
White House official denies Tea Party-focused ad campaign
US Stocks Climb After Recession Called Over, Homebuilders Gain
BP Oil Well, Site of National Catastrophe, Dies at One
Recent
Spotlight


BP Kills Macondo, But Its ...
blogs.wsj.com/source/2010/09/20/bp-kills-macondo-but-its-legacy-lives-on/
THE WALL STREET JOURNAL.
THE SOURCE
BP Kills Macondo, But Its Legacy Lives On
Article
Comments (2)
About The Source
Most Recent


Allen: Well is dead, but m...
edition.cnn.com/2010/US/09/20/gulf.oil.disaster/
EDITION: INTERNATIONAL
CNN
Allen: Well is dead, but much Gulf Coast work remains
By the CNN Wire Staff
September 20, 2010 -- Updated 1317 GMT (2117 HKT)
Click to play
What next for Gulf oil spill?
STORY HIGHLIGHTS
(CNN) -- The ruptured Macondo well, a mile under the Gulf of


BP oil spill cost hits nearl...
www.guardian.co.uk/environment/2010/sep/20/bp-oil-spill-dee
guardian.co.uk
News | Sport | Comment | Culture | Business | Money | Life & style
Business BP
BP oil spill cost hits nearly $10bn
BP has set up a $20bn compensation fund after the Deepwater Horizon disaster, which has so far paid out 19,000 claims totalling more than $240m
Julia Kollewe
guardian.co.uk, Monday 20 September 2010 08.33 BST
Article history
BP's costs for the Deepwater Horizon disaster have hit $10bn. Photograph: Ho/Reuters

What we will cover

- **Supervised learning**
  - Linear Regression (single and multiple variable)
  - Classification (logistic regression, decision tree, support vector machines, etc.)
  - Neural networks
- **Unsupervised learning**
  - Clustering
  - Dimensionality reduction
- **Ensemble Learning**: Bagging, Boosting
- **Recent advanced issues - fairness, interpretability in ML**
- **Theory of Generalization – the very basics**

ML in Practice

- Understanding domain, prior knowledge, and goals
- Data integration, cleaning, pre-processing, selection, etc.
- Feature selection – some features may be sensitive, protected, etc.
- Learning models
- Evaluating and interpreting results
- Consolidating and deploying discovered knowledge
- Loop