INTRODUCTION TO MACHINE LEARNING

INTRODUCTION TO MACHINE LEARNING

* Some contents are adapted from Dr. Hung Huang and Dr. Chengkai Li at UT Arlington

Sai Chandra, PhD
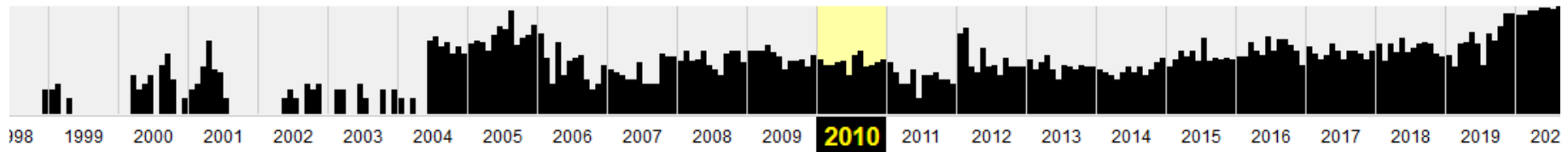Department of Computer Science @ UNLV

Begin with DATA and data mining

☐ Lots of data is being collected and warehoused

- ☐ Web data, e-commerce
- ☐ purchases at department/ grocery stores
- ☐ Bank/Credit Card transactions
- ☐ Social Network

How much data?

- Google processed 20 PB a day (2008)
- Wayback Machine has 3 PB + 100 TB/month (3/2009)
- Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)

998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 **2010** 2011 2012 2013 2014 2015 2016 2017 2018 2019 202

-> unlv.edu

KB: 2^10 bytes
MB: 2^20 bytes
GB: 2^30 bytes
TB: 2^40 bytes
PB: 2^50 bytes
EB: 2^60 bytes
ZB: 2^70 bytes
YB: 2^80 bytes

Ref: Ruoming Jin, PhD, Kent University

Type of Data

- Relational Data (Tables/Transaction/Legacy Data)

- Text Data (Web)

- Semi-structured Data (XML)

- Graph Data

  - Social Network, Semantic Web (RDF), …

- Streaming Data

What to do with these data?

- Aggregation and Statistics
  - Data warehouse and OLAP (Online analytical processing)
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling

Ref: Ruoming Jin, PhD, Kent University

DATA

SORTED

ARRANGED

PRESENTED VISUALLY

EXPLAINED WITH A STORY

AI with DATA

- Data Mining
- Machine Learning
- Pattern Recognition
- Data Science

Data mining?

- Lots of data is being collected and stored at enormous speeds (GB/hour)
  - Web data (web crawler)
  - Credit Card Transactions
  - Social Network Services
  - Wireless sensors
  - Genomic data
- Computers have become cheaper and powerful
- There is often "hidden" information in the data
- Traditional techniques infeasible for raw data

Data mining?

- Data Mining!!
  - KNOWLEDGE DISCOVERY FROM DATA
 - Extraction of interesting patterns or knowledge from huge amount of data

What's data mining?

□ Questions!

□ What is (not) data mining ?

- Look up phone number in phone directory
- Certain names are more prevalent in certain US locations (O'Brien, O'Rurke, O'Reilly… in Boston)
- Query a web search engine for information about "Amazon"
- Group together similar documents returned by search engine according to their context
- Certain words are prevalent in positive expression.

# Data Mining: Confluence of Multiple Disciplines

What is Machine Learning?

Turing Award, 1975
Nobel Prize in Economics, 1978
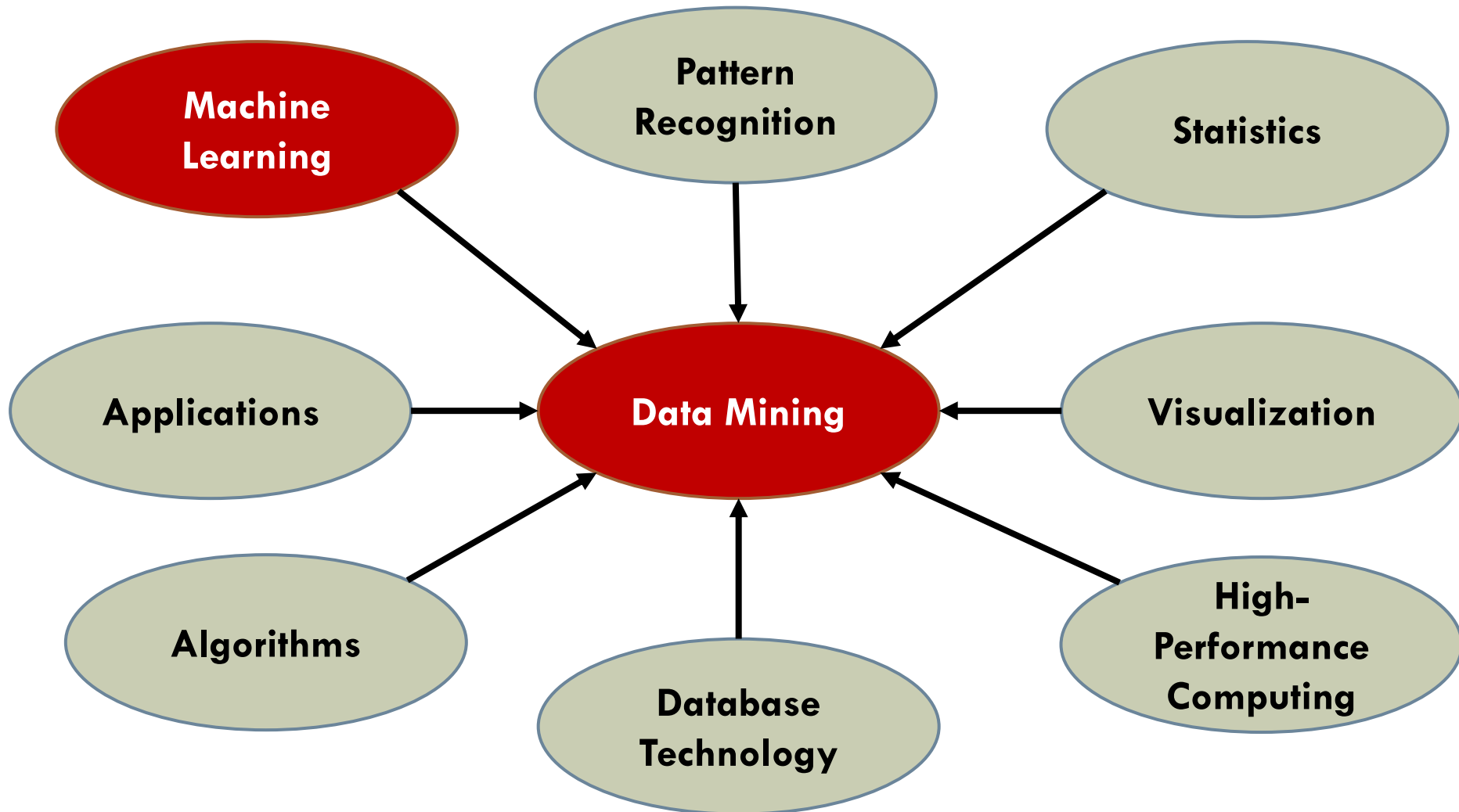
- Herbert A. Simon
  - "Learning is any **process** (**algorithm**) by which a **system** (**model**) improves **performance** (**accuracy**) from **experience** (**data**)"

- Data may include **hidden knowledge** that explains laws/rules/logics of a certain complex phenomenon

What is Machine Learning?

- Algorithms that train data and improve the performance by using the knowledge

- Why?
  - It is often too difficult to design a set of rules "by hand"
  - Machine learning is about automatically extracting relevant information from data and applying it to analyze new data

- Examples
  - Face Recognition
  - Speech recognition
  - Stock market prediction

Ramanathan and Chellappa, "Modeling Age Progression in Young Faces", 2006
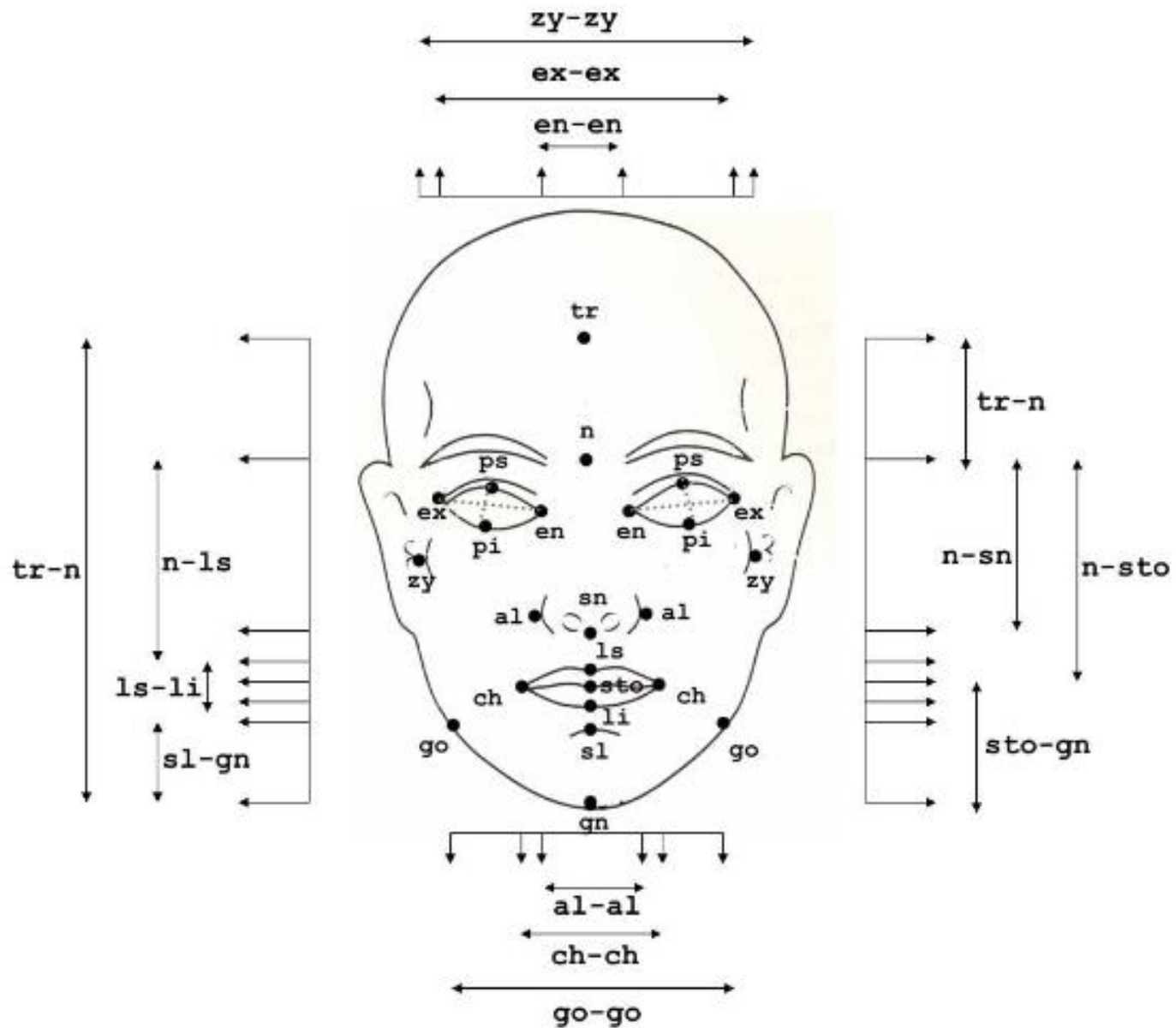
What is Machine Learning?

Face dataset

IMDb

460,723 images

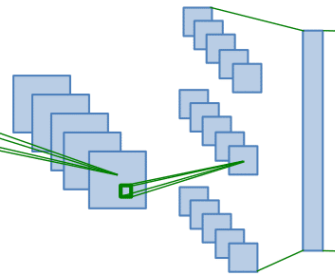Wikipedia

62,328 images

1. Input image   2. Face detection   3. Cropped face   4. Feature extraction   5. Prediction
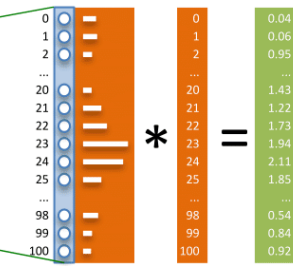
Mathias et al. detector   + 40% margin   VGG-16 architecture   Softmax expected value $\Sigma = 23.4$ years
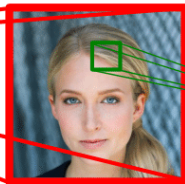
What is Machine Learning?

Data

Training

Model

f(x)

New Data

Perform a task

☐ yes

☐ no

- Tasks
  - Classification
  - Regression
  - Clustering
  - Feature Selection
  - Reinforcement
  - Anomaly detection
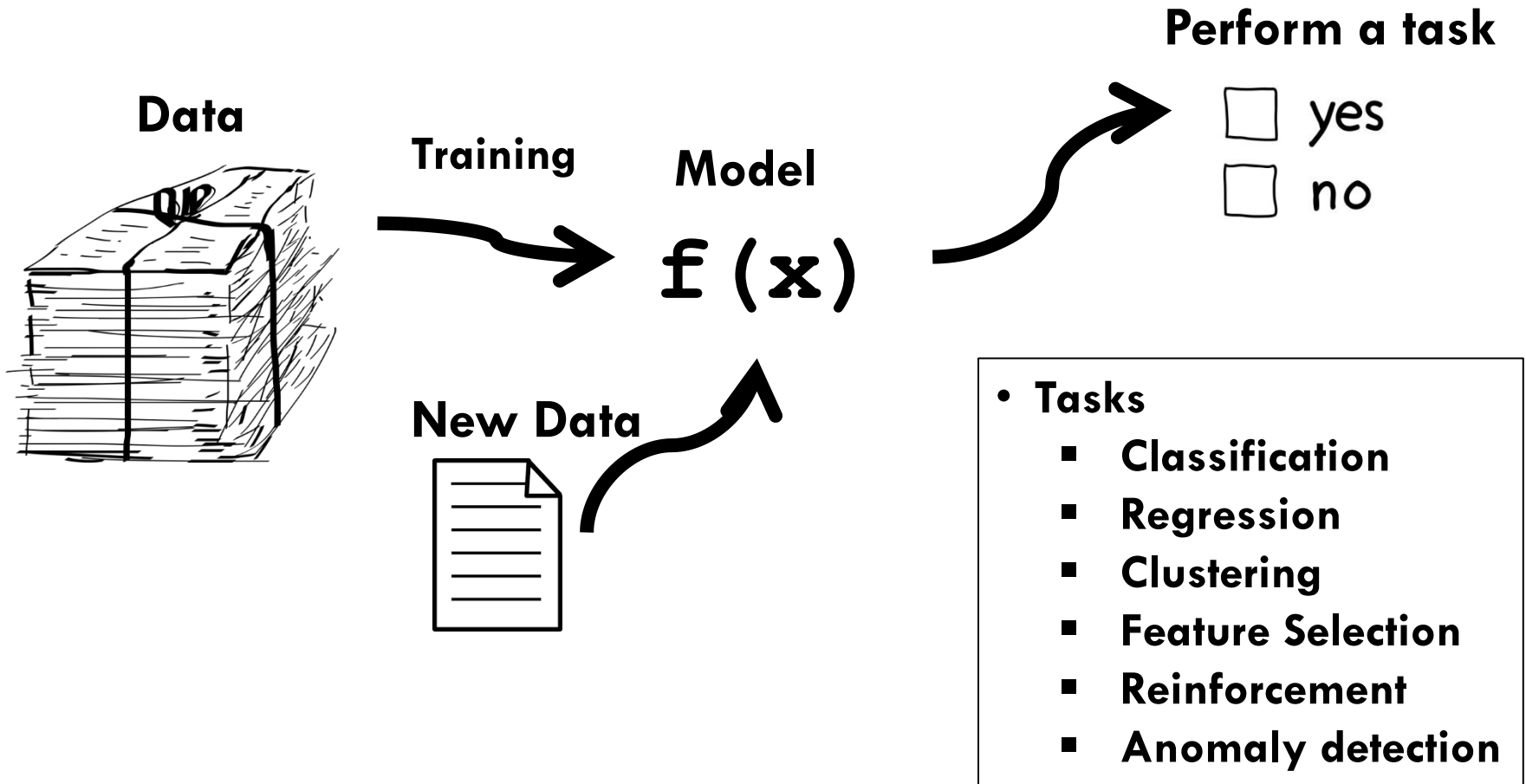
What is Machine Learning?

- Let a machine find certain patterns automatically to make a decision
  - Very related to Optimization
- However, machine learning is not "MAGIC"
- We should give a minimum guideline to a machine
  - Data
  - Model
- Machine will find the best setting of the model computationally

Types of Learning

- Supervised learning (Classification and Regression)
  - Given labeled data, classifying or predicting unlabeled new data
- Unsupervised learning (Clustering)
  - Given unlabeled data, inferring a function to describe hidden patterns
- Feature Selection/Feature Reduction
  - Selecting a subset of relevant features
- Semi-supervised learning
  - Given both labeled/unlabeled data, classifying or predicting unlabeled new data
- Ensemble Learning
- And many topics…

Foundation of Machine Learning

- Which Skills Are Most Valuable In Machine Learning? (Forbes)
  - Fundamentals of Statistics
  - Optimization (Mathematics, but don't be scared)
  - Building quantitative models
  - Understanding how models and data analysis actually apply to products and businesses
  - Knowing how to write high quality software
  - Working with large data sets

Ref: https://www.forbes.com/sites/quora/2017/09/29/which-skills-are-most-valuable-in-machine-learning/#79d6061b7c43

Examples: Object detection/regression

□ Google Vision API

  □ https://cloud.google.com/vision/

□ Age estimation from facial images

Input image							
Aligned face							
Apparent age	57	17	40	50	30	79	12
Predicted age	57.75	16.15	39.43	49.15	32.06	78.99	12.78

Examples: Classification

☐ Cat vs Dog from images

Sample of cats & dogs images from Kaggle Dataset
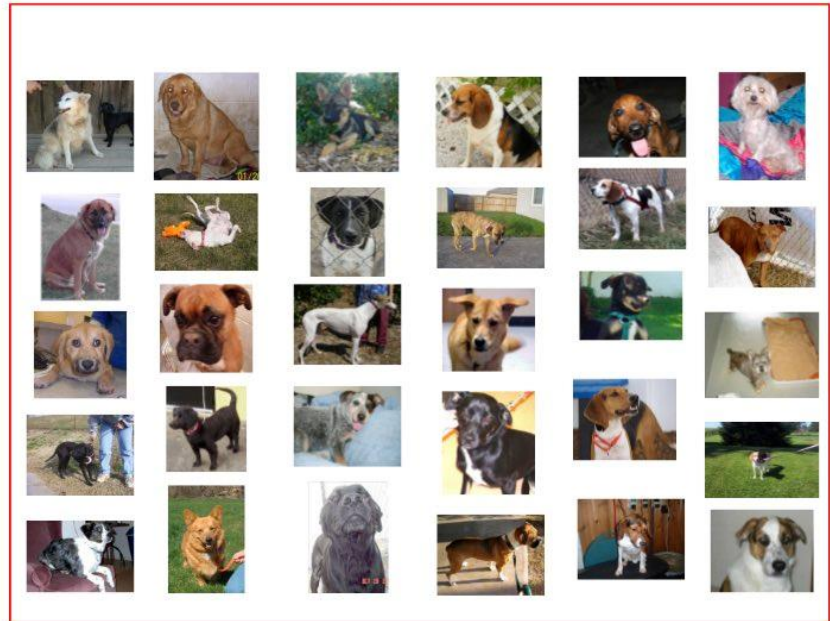
Examples: Classification

□ Vehicle Types from images

Microcar / Mini car	Subcompact car / Small car	Compact car / Medium car	Midi size car / Large car
Entry-level luxury car / Large car	Full-size car / Executive car	Mid-size luxury car / Executive car	Full-size luxury car
Convertible / Sport coupe	Grand tourer / Sport coupe	Sports car / Sport coupe	Supercar / Sport coupe
Roadster	Station wagon	Compact minivan / Multi Purpose vehicle (MPV)	Minivan / Large MPV
SUV - mini (off road) / Sport Utility Vehicle	SUV - compact (off road) / Sport Utility Vehicle	SUV - mid size (off road) / Sport Utility Vehicle	SUV - full size (off road) / Sport Utility Vehicle
Pickup truck - mini	Pickup truck - mid size	Pickup truck - full size	Pickup truck - full size / Heavy Duty

https://car-brand-names.com/types-of-cars/

Examples: Inference/Classification

Network Inference

Sky	Temp	Humid	Wind	Water
Sunny	Warm	Normal	Strong	Warm
Sunny	Warm	High	Strong	Warm
Rainy	Cold	High	Strong	Warm
Sunny	Warm	High	Strong	Cool

http://www.cs.umd.edu/hcil/science20/

Decision Trees
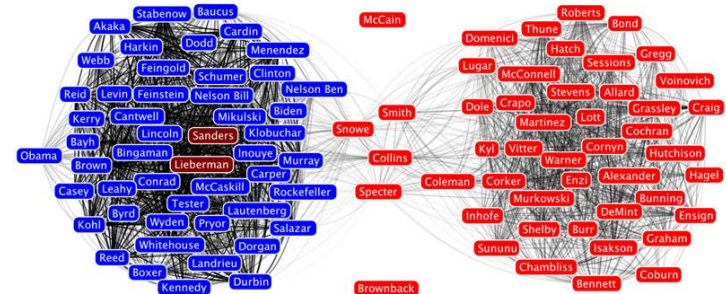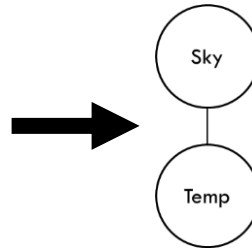
https://www.cs.cmu.edu/~bhiksha/courses/10-601/decisiontrees/
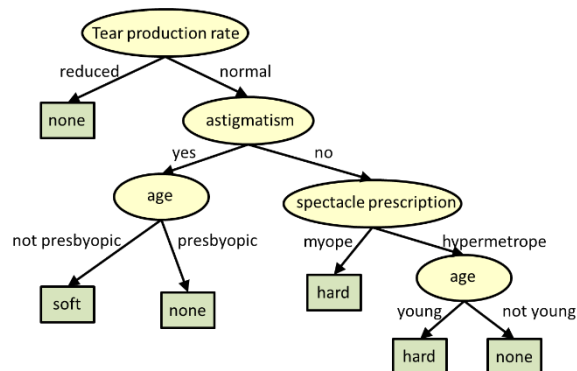
Examples: Classification

□ Handwritten Digit Recognition

$\longrightarrow$ 0, 1, …, 9

https://en.wikipedia.org/wiki/MNIST_database

Examples: Regression

□ Stock Market

Examples: Clustering

□ **Image segmentation**

   ■ Break up the image into similar regions

https://www.cc.gatech.edu/~hays/compvision2017/lectures/17.pdf

Discussion

- Let's discuss about machine learning examples
  - Problem?
  - Why should we consider machine learning for the problem?
  - What dataset do we need?

Discussion

- Algorithms are machine learning??
  - Shortest path algorithm (e.g., Dijkstra's algorithm)
  - Sorting algorithms

- Performance
  - Efficiency vs Accuracy

AI vs Data Mining vs Machine Learning

- There is considerable overlap among these, but some distinction can be made.

- Artificial Intelligence
  - Study of how to create intelligent agent. Not necessary to involve learning or induction.
- Machine Learning
  - Computer programs that learn some tasks from experience to improve performances.
- Data Mining
  - Study that has taken much of its inspiration and techniques from machine learning (and some, also, from statistics), but is put to different ends.

Relationships with…

- Big Data Analytics
- Data Mining
- Computer Vision
  - Image processing?
- Biomedical/Bioinformatics/Healthcare
- Text mining
- Natural Language Processing (NLP)

Journal/Conference

- Machine Learning
  - Conferences
    - ICML, NIPS, CVPR, ICCV, AAAI, IJCAI, ECML, ECCV, KDD, UAI, COLT
  - Journals
    - Journal of Machine Learning Research, Machine Learning, IEEE PAMI, IEEE TKDE

  - See "Google Scholar Metrics"

More details

- More details of Journals/Conferences
  - Scopes[1]
  - Paper types[2]
    - Original Papers, Discovery Notes, Application Notes, and Reviews (Survey)
  - Open-access journals
    - Otherwise via UNLV Online Library

[1] https://www.computer.org/web/tkde/about
[2] https://www.computer.org/web/tkde/author